Preference Modeling in Data-Driven Product Design: Application in Visual Aesthetics

by

Yanxin Pan

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Design Science and Scientific Computing)
in The University of Michigan
2018

Doctoral Committee:

        Professor Richard Gonzalez, Co-Chair
        Professor Panos Y. Papalambros, Co-Chair
        Professor Honglak Lee
        Professor Ji Zhu

Email: yanxinp@umich.edu

ORCID: 0000-0001-9820-2732

For all the people

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Creating a form that is attractive to the intended market audience is one of the greatest challenges in product development given the subjective nature of preference and heterogeneous market segments with potentially different product preferences. Accordingly, product designers use a variety of qualitative and quantitative research tools to assess product preferences across market segments, such as design theme clinics, focus groups, customer surveys, and design reviews; however, these tools are still limited due to their dependence on subjective judgment, and being time and resource intensive. In this dissertation, we focus on a key research question: how can we understand and predict more reliably the preference for a future product in heterogeneous markets, so that this understanding can inform designers' decision-making?

We present a number of data-driven approaches to model product preference. Instead of depending on any subjective judgment from human, the proposed preference models investigate the mathematical patterns behind users choice and behavior. This allows a more objective translation of customers' perception and preference into analytical relations that can inform design decision-making. Moreover, these models are scalable in that they have the capacity to analyze large-scale data and model customer heterogeneity accurately across market segments. In particular, we use feature representation as an intermediate step in our preference model, so that we can not only increase the predictive accuracy of the model but also capture in-depth insight into customers' preference.

We tested our data-driven approaches with application in visual aesthetics preference. Our results show that the proposed approaches can obtain an objective measurement of aesthetic perception and preference for a given market segment. This measurement enables designers to reliably evaluate and predict the aesthetic appeal of their designs. We also quantify the relative importance of aesthetic attributes when both aesthetic attributes and functional attributes are considered by customers. This quantification has great utility in helping product designers and executives in design reviews and selection of designs. Moreover, we visualize the possible factors affecting customers' perception of product aesthetics and how these factors differ across different market segments. Those visualizations are incredibly important to designers as they relate physical design details to psychological customer reactions.

The main contribution of this dissertation is to present purely data-driven approaches that enable designers to quantify and interpret more reliably the product preference. Methodological contributions include using modern probabilistic approaches and feature learning algorithms to quantitatively model the design process involving product aesthetics. These novel approaches can not only increase the predictive accuracy but also capture insights to inform design decision-making.

CHAPTER I

Introduction

1.1 Introduction

Creating a form that is attractive to the intended market audience is one of the greatest challenges in product development. In the early design phase, designers translate the needs and desires (e.g. customers need an aesthetically appealing product) to actual design decisions. These decisions are usually defined using design attributes, which are the design properties that the people who will experience the product, namely, the criteria they will use to judge the product (e.g., luxuriousness, ease of use, etc). Design attributes may not be measurable, thus designers need to make a mapping from design attributes to the measurable design characteristics, which are the design properties that the designer can explicitly act upon by manipulating the design (e.g. color, length, shape). To create an attractive product, designers must understand how customers perceive the design of a product; however, this understanding can be difficult to gain due to the inherent subjectivity of preference and the heterogeneity of customers across market segments. Moreover, designers need to translate such understanding into the language of engineering so that it can inform design decision-making. In other words, designers need to understand which design characteristics can affect preference and how they affect it.

Designers may rely on their subjective judgment of product preference when mak-

ing related design decisions; however, this can bring a large risk in the design process in that such subjective insights may lead to the wrong preference reaction when implemented in the actual product. Accordingly, designers often use a variety of qualitative and quantitative methods to assess product preferences, such as design theme clinics, focus groups, customer surveys, and design reviews. While these methods can provide further insights into the rationale for product preference, they are still limited due to two main drawbacks. First, these methods require customers to translate their subjective perception into semantic or numerical assessment, and the designers must then translate the assessment back into an attractive design. Communication errors are likely to occur during this translation process as customers often cannot articulate accurately why they like or dislike a design (*Nisbett and Wilson*, 1977; *Silvera et al.*, 2002). Also, designers are often geographically or culturally distant from the potential customers. Second, these methods are not scalable because they are labor and resource intensive. Lack of scalability can be problematic especially for product domains where there are hundreds of market segments. These drawbacks present a research gap in translating customers' perception and preference into an attractive product in a more predictive and scalable way.

In this dissertation, we focus on a key research question: how can we understand and predict more reliably the aesthetic perception and preference for a future product in a heterogeneous market, so that this understanding can inform the designers' decision-making. To answer this question, we propose a number of data-driven approaches to model product preference. Instead of having a human analyst acting as the interpreter of data, the proposed preference models investigate the mathematical patterns behind users' responses. This allows a more objective translation of customers' preference into analytical relations that can inform design decision-making. Moreover, these models are scalable in that they have the capacity to analyze the datasets including a large number of customers and designs and model customer

heterogeneity accurately across market segments.

In particular, we test our proposed data-driven models with application in visual aesthetics preference. Product aesthetics have long been recognized as a critically important factor for the success of a product. An aesthetically appealing appearance not only evokes pleasant emotions but also communicates meaning, quality, and product integrity to the customers before they physically interact with the product. Customers exhibit an involuntary aesthetic reaction and infer the presence of other product attributes based on this aesthetics reaction. One can assume naturally that customers prefer beautiful designs to ugly ones, even in highly functional product domains. Moreover, customers are willing to pay more money for a more beautiful product.

1.2   Research Problem

Aesthetics is a very old concept with root in the Greek word *aisthesis*, whose original meaning can be translated as understanding through sensory perception (*Hekkert and Leder*, 2008). The definition of aesthetics has slightly changed over time. Nowadays, aesthetics refers to the pleasure attained from sensory perception (*Hekkert*, 2006). Sensory input traditionally can be visual, auditory, tactile (somatic), olfactory, and gustatory; today we also recognize other sensory input such as motion, temperature or pain. Product aesthetics is defined here as the process of how products communicate meaning and evoke emotion through the senses (*Khalid and Helander*, 2006). The aesthetic reaction can be quick, often representing a first reaction to the product (*Reber et al.*, 2004). In addition, an aesthetic reaction often encompasses an overall assessment of the product. Customers are attracted by the aesthetic appeal of the designed artifact and they tend to relate the function of a product to its aesthetic appeal (*Norman*, 2005).

The process, which includes aesthetic preferences in design decision making, is
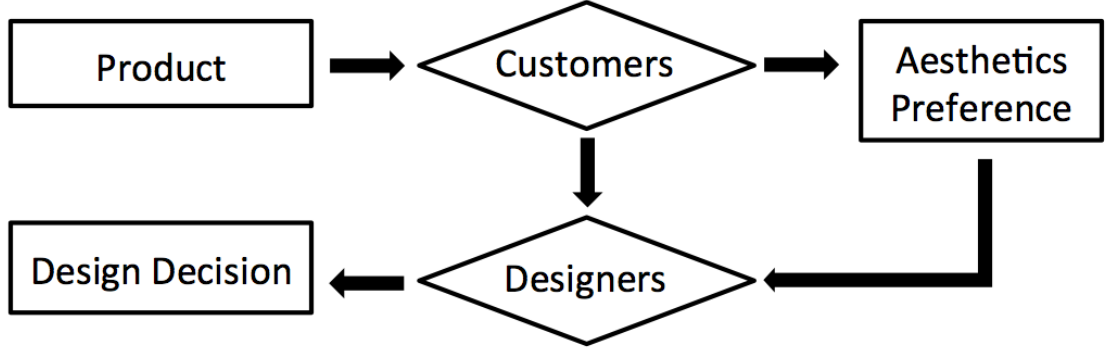
Figure 1.1: A symbolic model for the design process including aesthetic preferences in design decisions

very complex. To study this process, we illustrate the process using a symbolic model, which is an abstract description of the real world giving an approximate representation of more complex functions of physical systems, in figure 1.1. In this aesthetics preference model, customers are the perceivers of the product aesthetics, who receive the sensory information from the product, resulting in an aesthetics perception or preference (adapted from (*Leder et al.*, 2004; *Hekkert*, 2006)). Designers observe the users choices or behaviors, then infer the underlying aesthetics perception and preference, followed by interpreting the possible factors affecting aesthetics preference, finally, designers refer to this interpretation when making design decisions. Moreover, we can use quantitative model to represent this symbolic model by mathematical relations.

In this dissertation, the general research problem is to model the design process including aesthetics preferences quantitatively. The goal of this quantitative model is to extract useful information about product aesthetics more objectively in order to inform design decision-making more reliably. While product aesthetics involve many senses, in this dissertation we focus only on modeling aesthetics related to visual sensory perception. From here on, when referring to aesthetics we mean specifically visual aesthetics. This choice is made in part because visual input is dominant in product

aesthetics (*Goldstein and Brockmole*, 2016; *Hekkert*, 2006). While the methodology we present may be applicable to other sensory input, this generalization is beyond the scope of the dissertation.

Within the general research problem above we extract and address four specific research questions:

1. Does the product achieve the desired aesthetic design attributes for a given market segment?

2. How important are the aesthetic design attributes when compared with functional attributes and price?

3. What are the possible factors affecting customers' perception of product aesthetic appeal?

4. How do these factors differ across different market segments?

1.3   Aesthetics Preference Models

To address the research questions we develop data-driven aesthetics preference models that can map the visual information to the customers' aesthetic perception and preference. We then investigate the interpretation of this mapping in order to understand what are the factors that account for the observed emotional responses. Finally, we examine how we can apply the findings to support the decision-making process of practicing designers.

A model-based quantitative process for including aesthetic preferences in design decision making typically consists of five steps: design and user representation, preference function model, aesthetics labeling, interpretation, and design decision as shown in Figure 1.2. The first step, serving as input to the preference function model, is a mathematical representation of the design. This input is different from that in a
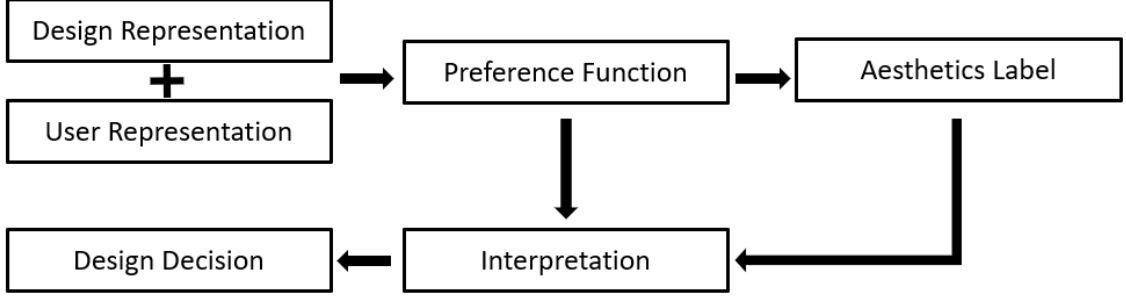
Figure 1.2: A model-based process to account for aesthetic preferences in design decisions

design optimization model, which classifies the input into design variables that are quantities specifying different states of a system by assuming different values and design parameters that are quantities that are given one specific value in any particular model statement. The input design representation includes all measurable design quantities that believed to be predictive to product aesthetics. This representation can include design variables (as shown in Chap II), design parameters (as shown in Chap V), design characteristics (as shown in Chap II and III), and measurable design attributes (as shown in Chap III). In addition, design representation can also be non-parametric, for example, the design images (as shown in Chap IV and V). A representation of the customers may also be included as part of the model input especially when heterogeneous markets are considered (as shown in Chap II and V). We denote the design representation as $\mathbf{X}_d$ and the customer representation as $\mathbf{X}_c$. The second step is the creation of the preference function which can predict the customer preference for the third step of the model, the aesthetics preference. This preference function, denoted as f, can implicitly capture the aesthetics preference by predicting subsequent user behaviors or user choices. Specifically, we want to estimate the preference function, so that $y = f(\mathbf{X}_d, \mathbf{X}_c)$, where aesthetics labels, denoted as $\mathbf{y}$, are the quantities resulting from users potential aesthetics preferences, such as users choices (e.g. Chap III), crowdsourced responses (e.g. Chap IV), and users behavior

(e.g. Chap V). The fourth step is interpretation, where the preference relation is quantitatively analyzed through mathematical tools and visualization. For example, we can determine the optimal design by maximizing the preference as shown in Chap II, investigate the influence of visual sensory by control experiments as shown in Chap III, approximately inverse the preference function as shown in Chap IV, and visualize the salient design regions as shown in Chap V. Those tools play the same role as *designers* in the symbolic model. Finally, design decisions are made based on the results of this analysis.

There are three challenges in aesthetics preference modeling. The first challenge is to determine an appropriate mathematical representation of the design and its customers in the model. The explicit mathematical representations used in quantitative preference model are usually a set of mathematical elements, which are supposed to contain influential information in aesthetics phenomena. Unlike other types of preference, aesthetics preference is dominantly evoked by visual inputs. The visual sensory information is best preserved in the form of an image or video, which are not the conventional modalities of input data in a quantitative model. As a result, we need to transform this visual information into an explicit representation that can be used in a quantitative model.

The second challenge is how to handle the complexity of the nonlinear mapping between the design representation and the aesthetics labels. In the product design community, linear logit functions have been widely used in preference modeling, because the logit function is easy to compose and to interpret, and it has a solid theoretical foundation for model diagnostics under known assumptions; however, the assumptions behind the logit function, such as linearity and independence, may not be appropriate in most real design situations. As a result, conclusions from these models maybe inaccurate or inappropriate. The predictive accuracy of linear logit models may be low especially in scenarios where the preference relationship is highly

nonlinear.

Sophisticated nonlinear models have been developed to model this process with high accuracy, such as content-based collaborative filtering (*Pazzani and Billsus*, 2007), kernel methods (*Ren et al.*, 2013), and Bayesian approach (*Srivastava and Schrater*, 2012); however, these bring the third challenge for aesthetic preference modeling. With the nonlinear models, predictive performance is significantly improved at the cost of reduced interpretability. In contrast, with the linear models, interpretability is often possible but predictive power is relatively poor due to assumptions that typically do not hold, namely, linearity (*Bodenhofer and Klawonn*, 2004), feature independence (*Holt*, 1986; *Torrance et al.*, 1995), homogeneity (*Birol et al.*, 2006; *Feick and Higie*, 1992), and complex noise distributions (*Althaus*, 2003). Subsequently, we must face the trade-off between interpretability and predictive accuracy or develop new approaches to interpret nonlinear functions.

1.4   Related Work

1.4.1   Aesthetics Research

Research in aesthetics can be generally classified into two categories: descriptive study and quantitative study. In a descriptive study, design researchers or practicing designers play the role of the *preference function* and interpretation in the aesthetics preference model. They link the products and customers with the aesthetics preference through their observation and then translate the observation into design principles. The descriptive study of aesthetics dates back to centuries ago, for example, the use of the golden ratio. In recent decades, much of descriptive study research focused on finding properties of objects and simple patterns that determine aesthetic preference.

Researchers have proposed three classes of product properties: psychophysical,

organizational and meaningful properties (see (*Hekkert*, 1995) for an overview). The psychophysical properties refer to the product properties that can be quantified, such as size and color. These properties affect aesthetics preference relationally and contextually. For example, it has been demonstrated that hues are preferred in the order blue, green or red, and yellow (*McManus et al.*, 1981). The organizational properties focus on how our visual system organizes information by analyzing edges, contours, blobs, and basic geometrical shapes. Gestalt psychology is a well-known organizational property, which studies the relationship between elements and the whole. There are several principles in Gestalt psychology; for example, the *principle of similarity* demonstrates that elements that look similar in shape, color, or size, are likely seen as belonging together. Meaningful properties are the subjective properties that are determined based on one's knowledge, culture, and previous experiences. A pioneering study in this category is 'familiarity breeds liking' (*Zajonc*, 1968), which suggests that people may prefer a familiar design solution. Descriptive research in aesthetics has provided some valuable principles. While these principles have been successfully employed in practice, there is still a need for more detailed studies in aesthetics due to the heterogeneity of aesthetics perception. This need is further exacerbated for customer-centered products.

Quantitative research in aesthetics allows designers to investigate aesthetics from a much closer distance. Rather than having the human as the interpreter, quantitative studies in aesthetics aim at investigating the mathematical relationship between design, customer, and aesthetic preference. A common approach is to first decompose a complex form into design characteristics. For example, the form of an automobile can be broken into headlight, grill, etc. This decomposition process is also called atomization (*Durgee*, 1988). After atomizing the form, how design characteristics affect aesthetics preference can be determined either separately or jointly. (*Orsborn et al.*, 2009) shows an example of quantifying aesthetic preference in a utility function

via atomization. While designers can mathematically represent the form of design through atomization, this approach is still limited as the visual sensory information is not well preserved. Moreover, the three challenges discussed previously remain to be solved.

1.4.2   Feature Learning

Feature learning is a research field that has been applied to improve the representation of input data in various applications. Feature learning methods transform the original data representation into a feature representation, which can be effectively exploited in the predictive modeling task. The feature representation consists of features which are functions of the design variables in the original design representation, for example, a nonlinear combination of the geometric variables, the mean value of a set of pixel values, or the design variable itself.

There is a long history of using features to represent designs in the design community. These design features may be manually defined by designers based on an obvious interpretation or classification of the features, such as a set of parametric handles to manipulate vehicle silhouettes (*Petiot and Dagher*, 2011; *Poirson et al.*, 2013; *Reid et al.*, 2013; *Ren et al.*, 2013). Using finite shape grammars to form more complex representations is another type of design representation (*McCormack et al.*, 2004; *Orsborn et al.*, 2006; *Pugliese and Cagan*, 2002). Hybrid approaches that learn the set of handles have been studied, for example, autoencoders for 3D object manipulation to affect attribute ratings (*Yumer et al.*, 2015), and representations that combine hand-crafted and implicitly learned representations to capture design freedom and brand recognition (*Burnap et al.*, 2016a).

In addition to the design-specific features, there is a vast amount of work from computer vision researchers on features automatically learned from the data. These features are more general for a variety of tasks and have achieved great success such as

object recognition(*Krizhevsky et al.*, 2012), speech recognition (*Hinton et al.*, 2012), and natural language process (*Collobert and Weston*, 2008). In this context features are more abstract data-derived mathematical constructs and may not have immediate or obvious interpretation. The feature learning methods can be divided into two categories: supervised feature learning and unsupervised feature learning. Supervised feature learning means learning features from labeled data, such as neural networks (*Krizhevsky et al.*, 2012) and supervised dictionary learning (*Mairal et al.*, 2009). These feature learning functions may be generalized to the task outside of the training data set (*Gatys et al.*, 2015; *Karayev et al.*, 2013). Unsupervised feature learning means learning features from unlabeled data, such as sparse coding (*Lee et al.*, 2006), deep autoencoders (*Kingma and Welling*, 2013), and deep Boltzmann machine(*Salakhutdinov and Hinton*, 2009). These feature learning methods are promising for use in aesthetics preference modeling.

1.4.3   Preference Learning

Preference is also an important topic in the machine learning field. The research related to this topic is often referred to as preference learning. In general, the goal of preference learning is to construct ("learn") a predictive preference model from observed preference data. Specifically, there is a training set, which is a set of samples whose preferences are known. A preference learning task is to learn a function from the training set that predicts preferences for a new set of samples. Preference learning is often viewed as a supervised learning task, whose training samples are pairs consisting of an input object (also called independent variable in statistics) and an associated output label (also called dependent variable in statistics).

There is an overlap between preference learning and aesthetics preference modeling, as both of approaches aim at constructing the functions linking the input product representation and its preference label using data. Despite this commonality, there
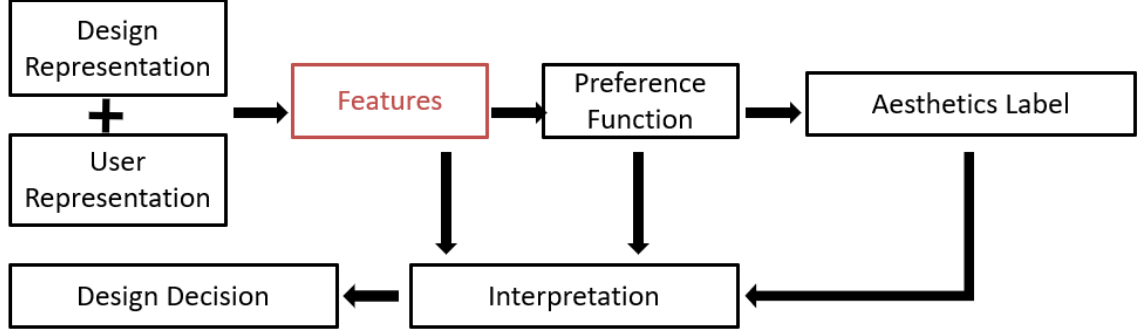
Figure 1.3: The proposed aesthetics preference model

are several differences distinguishing preference learning from aesthetics preference modeling. Although a predictive preference function is an important component of an aesthetics preference model, the primary goal of the model is to inform design decision making. To achieve this goal, the model must be interpretable to enhance the designers' understanding of the targeted users. Preference learning primarily emphasizes the use of the preference function, which is usually highly nonlinear, to predict future user choices, and understanding of why the choices are made is of secondary interest. As a result, preference learning may have high predictive accuracy, but low interpretability. This difference prevent designers from directly employing preference learning model to inform design decision making.

1.4.4 Proposed Approach

In this dissertation, we propose to use the feature learning method as an intermediate step between the original design representation and aesthetics labeling as shown in Figure 1.3. The idea of modeling preference via feature learning is motivated by the success of using feature learning to improve prediction accuracy in various applications (*Girshick et al.*, 2014; *Guyon and Elisseeff*, 2003; *Li et al.*, 2010; *Mittelman et al.*, 2013). Moreover, research in the design community has shown that consumers prefer to perceive a product through the high-level design attributes such as compact

design, which cannot be directly controlled by the designers, instead of the original design characteristics or variables such as "the length is 10 cm", which can be directly controlled by designers. Feature learning may help designers to discover rules for how to change the perceived design attributes by changing the design characteristics or variables accordingly.

Aesthetic preference modeling can benefit from using feature representation in several ways. First, feature learning can transform multimodal data into a unified representation so that the preference model can relate information in different data modalities together to infer preference. Second, previous research has shown that transforming data variables into more easily human-memory "chunked" perceptual features justifies the linear models commonly used in the design community (*Livingstone and Hubel*, 1987). Classic classification models, such as the l2 regularized logit model, may have superior prediction accuracy with features as the model variables rather than with the original variables as shown in (*Burnap et al.*, 2014). Third, instead of directly interpreting the highly nonlinear aesthetic preference function, it may be easier to interpret a preference mapping with the feature representation as an intermediate stage, because the interpretation process can be decomposed into two steps: interpreting the features and interpreting the less complex preference function. These issues are explored in more detail in the later chapters of the dissertation.

1.5   Dissertation Contribution

The main contribution of this dissertation is to demonstrate that we can quantify product preference using a purely data-driven approach in a way that has value for the practicing designer. We aim to provide designers a method to more objectively measure the product preference. In addition, we aim to develop a number of quantitative models that interpret how portions of the design space affect product perception and preference in both homogeneous and heterogeneous markets. Moreover, these

models are scalable to hundreds or thousands of markets, an important consideration for enterprises engaged in product design across globally dispersed markets.

Methodological contributions include introducing feature representation as an intermediate step in preference modeling. Dissertation results show that there is an increase of preference prediction accuracy when using feature learning methods, as compared with the original data representation. These results also suggest features indeed represent better the customers underlying design preferences, thus offering deeper insights to inform design decisions. Furthermore, the dissertation addresses how to deal with multimodal data forms, and demonstrates training models using large-scale multimodal data including 2D images, numerical labels, and crowdsourced response data.

1.6   Dissertation Overview

The rest of the dissertation is organized as follows: Chapter 2 presents a product preference model which uses feature representation as an intermediate step between design representation and preference function. This model is applied to predict automobile purchase decisions. The results show that the use of features offers improvement in prediction accuracy, and the interpretation and visualization of these feature representations can be used to support data-driven design decisions. Moreover, a theoretical error bound is given to guarantee the model fitness. Chapter 3 presents a novel approach to quantify more objectively aesthetic attributes as well as their relative importance vs. other attributes. Specifically, this approach provides designers quantitative evidence to evaluate whether a design concept achieves the desired aesthetic design attributes; it further investigates the relative importance of aesthetic attributes when compared with functional attributes and price. These findings offer deeper insights on how customers make trade-off between product aesthetics and function. Chapter 4 presents a data-driven method building on features from a

deep convolutional neural network. This data-driven method can predict aesthetic attribute values for given design images. Importantly, this model can identify visual attention regions that affect customers' aesthetics perception. Chapter 5 presents a scalable deep learning approach that predicts how customers across different market segments perceive aesthetic designs, and provides a visualization to aid product design. Chapter 6 gives a summary of results, contributions, and future directions.

CHAPTER II

Improving Design Preference Prediction Accuracy with Feature Learning

2.1 Introduction

In this chapter we explore how we can introduce features to improve preference prediction accuracy while being able to extract insights for design decisions. Much research has been devoted to develop design preference models that predict customer design choices. A common approach is to: (i) collect a large database of previous purchases that includes customer data, e.g., age, gender, income, and purchased product design data, e.g., number of cylinders, length, curb weight — for an automobile; and (ii) statistically infer a design preference model that links customer and product variables, using conjoint analysis or discrete choice analysis such as logit, mixed logit, and nested logit models (*Berkovec and Rust*, 1985; *McFadden and Train*, 2000).

However, a customer may not purchase a vehicle solely due to interactions between these two sets of variables, e.g., a 50-year old male prefers 6-cylinder engines. Instead, a customer may purchase a product for more 'meaningful' design attributes that are functions of the original variables, such as environmental sustainability or sportiness (*Reid et al.*, 2012; *Norman*, 2005). These meaningful intermediate functions of the original variables, both of the customer and of the design, are hereafter

termed *features*. We posit that using customer and product features, instead of just the original customer and product variables, may increase the prediction accuracy of the design preference model.

Our goal then is to find features that improve this preference prediction accuracy. To this end, one common approach is to ask design and marketing domain experts to choose these features intuitively, such as a design's social context (*He et al.*, 2014) and visual design interactions (*Sylcott et al.*, 2013). For example, eco-friendly vehicles may be a function of miles per gallon (MPG) and emissions, whereas environmentally active customers may be a function of age, income, and geographic region. An alternative explored in this chapter is to find features 'automatically' using feature learning methods studied in computer science and statistics. As shown in Figure 2.1, feature learning methods create an intermediate step between the original data and the design preference model by forming a more efficient "feature representation" of the original data. Certain well-known methods such as principal component analysis may be viewed similarly, but more recent feature learning methods have shown impressive results in 1D waveform prediction (*Hinton et al.*, 2012) and 2D image object recognition (*Krizhevsky et al.*, 2012).

We conduct an experiment on automobile purchasing preferences to assess whether three feature learning methods increase design preference prediction accuracy: (1) principle component analysis, (2) low-rank + sparse matrix decomposition, and (3) exponential family sparse restricted Boltzmann machines (*Salakhutdinov et al.*, 2007). We cast preference prediction as a binary classification task by asking the question, "given customer $\mathbf{x}$, do they purchase vehicle p or vehicle q." Our data set is comprised of 1,161,056 data points generated from 5582 real passenger vehicle purchases in the United States during model year 2006 (MY2006).

The first contribution of this work is an increase of preference prediction accuracy by 2%-7% just using simple "single-layer" feature learning methods, as compared
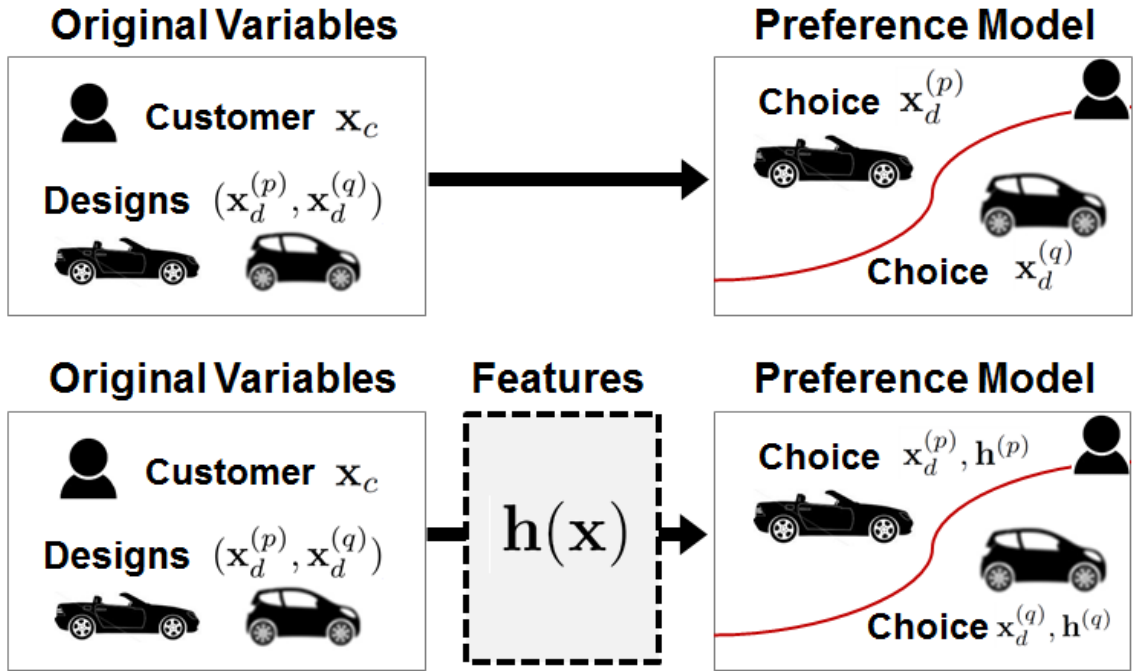
Figure 2.1: The concept of feature learning as an intermediate mapping between variables and a preference model. The diagram on top depicts conventional design preference modeling (e.g., conjoint analysis) where an inferred preference model discriminates between alternative design choices for a given customer. The diagram on bottom depicts the use of features as an intermediate modeling task.

with the original data representation. These results suggest features indeed better represent the customer's underlying design preferences, thus offering deeper insight to inform decisions during the design process. Moreover, this finding is complementary to recent work in crowdsourced data gathering (*Burnap et al.*, 2015; *Panchal*, 2015) and nonlinear preference modeling (*Chapelle and Harchaoui*, 2004; *Evgeniou et al.*, 2007)) since they do not affect the preference model or data set itself.

The second contribution of this work is to show how features may be used in the design process. We show that feature interpretation and feature visualization offer designers additional tools for augmenting design decisions. First, we interpret the most influential pairings of vehicle features and customer features to the preference task, and contrast this with the same analysis using the original variable representation. Second, we visualize the theoretically optimal vehicle for a given customer within the learned feature representation, and show how this optimal vehicle, which does not exist, may be used to suggest design improvements upon current models of vehicles that do exist in the market.

Methodological contributions include being the first to use recent feature learning methods on heterogeneous design and marketing data. Recent feature learning research has focused on homogeneous data, in which all variables are real-valued numbers such as pixel values for image recognition (*Krizhevsky et al.*, 2012; *Lee et al.*, 2011); in contrast, we explicitly model the heterogeneous distribution of the input variables, for example 'age' being a real-valued variable and 'General Motors' being a categorical variable. Subsequently, we give a number of theoretical extensions: First, we use exponential family generalizations for the sparse restricted Boltzmann machines, enabling explicit modeling of statistical distributions for heterogeneous data. Second, we derive theoretical bounds on the reconstruction error of the low-rank + sparse matrix decomposition feature learning method.

The remainder of the chapter is structured as follows: Section 2.2 discusses efforts

to increase prediction accuracy by the design community, as well as feature learning advances in the machine learning community. Section 2.3 sets up the preference prediction task as a binary classification problem. Section 2.4 details three feature learning methods and their extension to suit heterogeneous design and market data. Section 2.6 details the experimental setup of the preference prediction task, followed by results showing improvement of preference prediction accuracy. Section 2.5 proves the theoretical bounds on the reconstruction error of the low-rank + sparse matrix decomposition feature learning methods. Section 2.7 details how features may be used to inform design decisions through feature interpretation and feature visualization. Section 2.8 summarizes the chapter.

2.2 Background and Related Work

Design preference modeling has been investigated in design for market systems, where quantitative engineering and marketing models are linked to improve enterprise-wide decision making (*Wassenaar and Chen*, 2003; *Lewis et al.*, 2006; *Michalek et al.*, 2005). In such frameworks, the design preference model is used to aggregate input across multiple stakeholders, with special importance on the eventual customer within the targeted market segment (*Chen et al.*, 2013).

These design preference models have been shown to be especially useful for the design of passenger vehicles, as demonstrated across a variety of applications such as engine design (*Wassenaar et al.*, 2005), vehicle packaging (*Kumar et al.*, 2007), brand recognition (*Burnap et al.*, 2016a), and vehicle styling (*Orsborn et al.*, 2009; *Reid et al.*, 2012; *Sylcott et al.*, 2013). Connecting many of these research efforts is the desire for improved prediction accuracy of the underlying design preference model. With increased prediction accuracy, measured using "held out" portions of the data, greater confidence may be placed in the fidelity of the resulting design conclusions.

Efforts to improve prediction accuracy involve: (i) developing more complex sta-

tistical models to capture the heterogeneous and stochastic nature of customer preferences; examples yuinclude mixed and nested logit models (*McFadden and Train*, 2000; *Berkovec and Rust*, 1985), consideration sets (*Morrow et al.*, 2014), and kernel-based methods (*Chapelle and Harchaoui*, 2004; *Evgeniou et al.*, 2007; *Ren et al.*, 2013); and (ii) creating adaptive questionnaires to obtain stated information more efficiently using a variety of active learning methods (*Toubia et al.*, 2003; *Abernethy et al.*, 2008).

This work is different from (i) above in that the set of features learned is agnostic of the particular preference model used. One can just as easily switch out the l^2 logit design preference model used in this paper for another model, whether it be mixed logit or a kernel machine. This work is also different from (ii) in that we are working with a set of revealed data on actual vehicle purchases, rather than eliciting this data through a survey. Accordingly, this work is among recent efforts towards data-driven approaches in design (*Tuarob and Tucker*, 2015), including design analytics (*Van Horn and Lewis*, 2015) and design informatics (*Dym et al.*, 2005), in that we are directly using data to augment existing modeling techniques and ultimately suggest actionable design decisions.

2.2.1 Feature learning

Feature learning methods capture statistical dependencies implicit in the original variables by "encoding" the original variables in a new feature representation. This representation keeps the number of data the same while changing the length of each data point from M variables to K features. The idea is to minimize an objective function defining the reconstruction error between the original variables and their new feature representation. If this representation is more meaningful for the discriminative design preference prediction task, we can use the same supervised model (e.g., logit model) as before to achieve higher predictive performance. More details are given in

Section 2.4.

The first feature learning method we examined is principal component analysis (PCA). While not conventionally referred to as a feature learning method, PCA is chosen for its ubiquitous use and its qualitative difference from the other two methods. In particular, PCA makes the strong assumption that the data is Gaussian noise distributed around a linear subspace of the original variables, with the goal of learning the eigenvectors spanning this subspace (*Friedman et al.*, 2001). The features in our case are the coefficients of the original variables when projected onto this subspace or, equivalently, the inner product with the learned eigenvectors.

The second feature learning method is low-rank + sparse matrix decomposition (LSD). This method is chosen as it defines the features implicitly withing the preference model. In particular, LSD decomposes the "part-worth" coefficients contained in the design preference model (e.g., conjoint analysis or discrete choice analysis) into a low-rank matrix plus a sparse matrix. This additive decomposition is motivated by results from the marketing literature suggesting certain purchase consideration are linearly additive (*Gonzalez and Wu*, 1999), and thus well captured by decomposed matrices (*Evgeniou et al.*, 2005). An additional motivation for a linear decomposition model is the desire for interpretability (*Hauser and Rao*, 2004). Predictive consumer marketing oftentimes uses these learned coefficients to work hand-in-hand with engineering design to generate competitive products or services (*Papalambros and Wilde*, 2000). Such advantages are bolstered by separation of factors captured by matrix decomposition, as separation may lead to better capture of heterogeneity among market segments (*Lenk et al.*, 1996). Readers are referred to (*Netzer et al.*, 2008) for further in-depth discussion.

The third feature learning method is the exponential family sparse restricted Boltzmann machine (RBM) (*Smolensky*, 1986; *Lee et al.*, 2008). This method is chosen as it explicitly represents the features, in contrast with the LSD. The method is a

special case of a Boltzmann machine, an undirected graph model in which the energy associated within an energy state space defines the probability of finding the system in that state (*Smolensky*, 1986). In the RBM, each state is determined by both visible and hidden nodes, where each node corresponds to a random variable. The visible nodes are the original variables, while the hidden nodes are the feature representation. The "restricted" portion of the RBM refers to the restriction on visible-visible connections and hidden-hidden connections, later detailed and depicted in in Section 2.4 and Figure 2.4, respectively.

All three feature learning methods are considered "simple" in that they are single-layer models. The aforementioned results in 1D waveform speech recognition and 2D image object recognition have been achieved using hierarchical models, built by stacking multiple single-layer models. We chose single-layer feature learning methods here as an initial effort and to explore parameter settings more easily; as earlier noted, there is limited work on feature learning methods for heterogeneous data (e.g., categorical variables) and most advances are currently only on homogeneous data (e.g., real-valued 2D image pixels).

2.3   Preference Prediction as Binary Classification

We cast the task of predicting a customer's design preferences as a binary classification problem: Given customer j, represented by a vector of heterogeneous customer variables $\mathbf{x}_c^{(j)}$, as well as two passenger vehicle designs p and q, each represented by a vector of heterogeneous vehicle design variables $\mathbf{x}_d^{(p)}$ and $\mathbf{x}_d^{(q)}$, which passenger vehicle will the customer purchase? We use a real data set of customers and their passenger vehicle purchase decisions as detailed below (*Maritz Research Inc.*, 2007).

Table 2.1: Customer variables $\mathbf{x}_c$ and their variable types

Customer Variable	Type	Customer Variable	Type
Age	Real	U.S. State Cost of Living	Real
Number of House Members	Real	Gender	Categorical
Number of Small Children	Real	Income Bracket	Categorical
Number of Med. Children	Real	House Region	Categorical
Number of Large Children	Real	Education Level	Categorical
Number of Children	Real	U.S. State	Categorical
U.S. State Average Income	Real		

2.3.1 Customer and vehicle purchase data from 2006

The data used in this work combines the Maritz vehicle purchase survey from 2006 (*Maritz Research Inc.*, 2007), the Chrome vehicle variable database (*Chrome Systems Inc.*, 2008), and the 2006 estimated U.S. state income and living cost data from the U.S. Census Bureau (*United States Census Bureau*, 2006) to create a data set with both customer and passenger vehicle variables. These combined data result in a matrix of purchase records, with each row corresponding to a separate customer and purchased vehicle pair, and each column corresponding to a variable describing the customer (e.g., age, gender, income) or the purchased vehicle (e.g., # cylinders, length, curbweight).

From this original data set, we focus only on the customer group who bought passenger vehicles of size classes between mini-compact and large vehicles, thus excluding data for station wagons, trucks, minivans, and utility vehicles. In addition, purchase data for customers who did not consider other vehicles before their purchases were removed, as well data for customers who purchased vehicles for another party.

The resulting database contained 209 unique passenger vehicle models bought by 5582 unique customers. The full list of customer variables and passenger vehicle variables can be found in Tables 2.1 and 2.2. The variables in these tables are grouped into three unit types: Real, binary, and categorical, based on the nature of the variables.

Table 2.2: Design variables $\mathbf{x}_d$ and their variable types

Design Variable	Type	Design Variable	Type
Invoice	Real	AWD/4WD	Binary
MSRP	Real	Automatic Transmission	Binary
Curbweight	Real	Turbocharger	Binary
Horsepower	Real	Supercharger	Binary
MPG (Combined)	Real	Hybrid	Binary
Length	Real	Luxury	Binary
Width	Real	Vehicle Class	Categorical
Height	Real	Manufacturer	Categorical
Wheelbase	Real	Passenger Capacity	Categorical
Final Drive	Real	Engine Size	Categorical
Diesel	Binary		

2.3.2 Choice set training, validation, and testing split

We converted the data set of 5582 passenger vehicle purchases into a binary choice set by generating all pairwise comparisons between the purchased vehicle and the other 208 vehicles in the data set for all 5582 customers. This resulted in $N = 1,161,056$ data points, where each datum indexed by n consisted of a triplet (j, p, q) of a customer indexed by j and two passenger vehicles indexed by p and q, as well as a corresponding indicator variable $y^{(n)} \in \{0, 1\}$ describing which of the two vehicles was purchased.

This full data were then randomly shuffled, and split into training, validation, and testing sets. As previous studies have shown the impact on prediction performance given different generations of choice sets (*Shocker et al.*, 1991), we created 10 random shufflings and subsequent data splits of our data set, and run the design preference prediction experimental procedure of Section 2.6 on each one independently. This work is therefore complementary to studies on developing appropriate choice set generation schemes such as (*Wang and Chen*, 2015). Full details into the data processing procedure are given in Section 2.6.

2.3.3 Bilinear design preference utility

We adopt the conventions of utility theory for the measure of customer preference over a given product (*Von Neumann and Morgenstern*, 2007). Formally, each data point consists of a pairwise comparison between vehicles p and q for customer j, with corresponding customer variables $\mathbf{x}_c^{(j)}$ for $j \in \{1, \ldots, 5582\}$ and original variables of the two vehicle designs, $\mathbf{x}_d^{(p)}$ and $\mathbf{x}_d^{(q)}$ for $p, q \in \{1, \ldots, 209\}$. We assume a bilinear utility model for customer j and vehicle p:

$$U_{jp} = \left[\mathrm{vec}\left(\mathbf{x}_c^{(j)} \otimes \mathbf{x}_d^{(p)} \right)^T, \left(\mathbf{x}_d^{(p)} \right)^T \right] \omega, \tag{2.1}$$

where $\otimes$ is an outer product for vectors, $\mathrm{vec}\,(\cdot)$ is vectorization of a matrix, $[\cdot, \cdot]$ is concatenation of vectors, and ω is the part-worth vector.

2.3.4 Design preference model

The preference model refers to the assumed relationship between the bilinear utility model described in Section 2.3.3 and a label indicating which of the two vehicles the customer actually purchased. While the choice of preference model is not the focus of this paper, we pilot-tested popularly used models including l^1 and l^2 logit model, naïve Bayes, l^1 and l^2 linear as well as kernelized support vector machine, and random forests.

Based on these pilot results, we chose the l^2 logit model due to its widespread use in the design and marketing communities (*Netzer et al.*, 2008; *Fuge*, 2015); in particular, we used the primal form of the logit model. Equation (2.2) captures how the logit model describes the probabilistic relationship between customer j's preference for either vehicle p or vehicle q as a function of their associated utilities given by Equation (2.1). Note that ϵ are Gumbel-distributed random variables accounting for noise over the underlying utility of the customer j's preference for either vehicle p or vehicle q.

$$P^{(n)} = P_{(j,p,q)} = P\left(U_{jp} + \epsilon_{jp} > U_{jq} + \epsilon_{jq}\right) = \frac{e^{U_{jp}}}{e^{U_{jp}} + e^{U_{jq}}} \tag{2.2}$$

Parameter Estimation

We estimate the parameters of the logit model in Eq. (2.2) using conventional convex loss function minimization using the log-loss regularized with the l^2 norm.

$$\min_{\omega,\alpha} \frac{1}{N} \sum_{n=1}^{N} (y^{(n)} \log P^{(n)} + (1 - y^{(n)}) \log(1 - P^{(n)})) + \alpha \left\| \omega \right\|^2 \tag{2.3}$$

where $y^{(n)} = y_{(jpq)}$ is 1 if customer j chose vehicle p to purchase, and 0 if vehicle q was purchased; and α is the l^2 regularization hyperparameter. The optimization algorithm used to minimize this regularized loss function was stochastic gradient descent, with details of hyperparameter settings given in Section 2.6.

2.4   Feature Learning

We present three qualitatively different feature learning methods as introduced in Section 2.2: (1) principal component analysis, (2) low-rank + sparse matrix decomposition, and (3) exponential family sparse restricted Boltzmann machine. Furthermore, we discuss their extensions to better suit the market data described in Section 2.3, as well as derivation of theoretical guarantees.

2.4.1   Principal Component Analysis

Principal component analysis (PCA) maps the original data representation $\mathbf{x} = [x_1, x_2, \ldots, x_M]^T \in \mathbb{R}^{M \times 1}$ to a new feature representation $\mathbf{h} = [h_1, h_2, \ldots, h_K]^T \in \mathbb{R}^{K \times 1}$, $K \leq M$, with an orthogonal transformation $\mathbf{W} \in \mathbb{R}^{M \times K}$. Assume that the original data representation $\mathbf{x}$ has zero empirical mean (otherwise we simply subtract the empirical mean from $\mathbf{x}$). The mapping is given by:
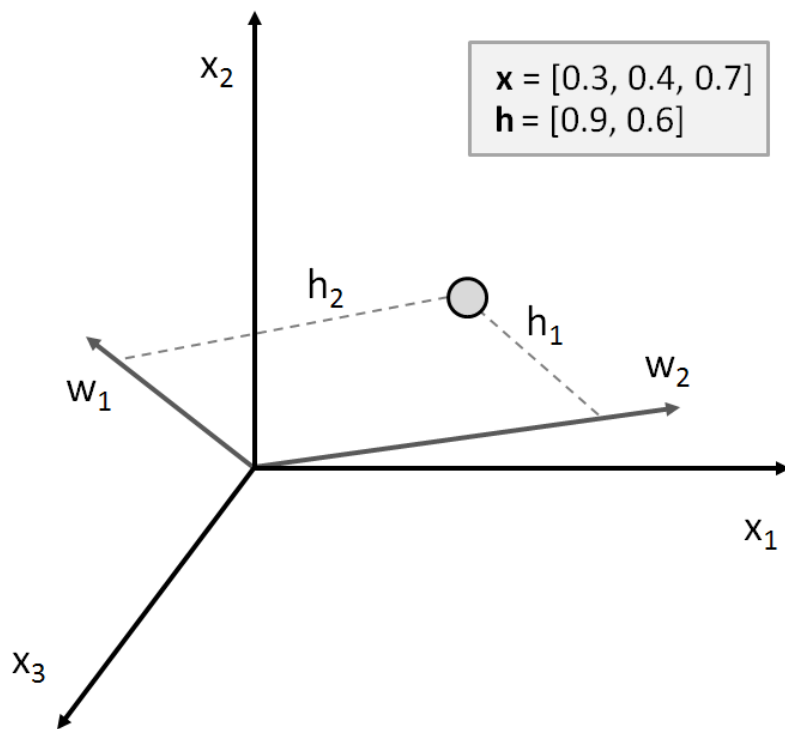
Figure 2.2: The concept of principle component analysis shown using an example with a data point represented by three original variables $\mathbf{x}$ projected to a two dimensional subspace spanned by $\mathbf{w}$ to obtain features $\mathbf{h}$.

$$\mathbf{h} = \mathbf{x}^T \mathbf{W} \tag{2.4}$$

The PCA representation has the following properties: (1) h_1 has the largest variance, and the variance of h_i is not smaller than the variance of h_j for all $j < i$; (2) the columns of $\mathbf{W}$ are orthogonal unit vectors; and (3) $\mathbf{h}$ and $\mathbf{W}$ minimize the reconstruction error ϵ:

$$\epsilon = ||\mathbf{x} - \mathbf{h}||^2 \tag{2.5}$$

When the q columns of $\mathbf{W}$ consist of the first q eigenvectors of $\mathbf{x}^T \mathbf{x}$, the above properties are all satisfied, and the PCA feature representation can be calculated by Equation (2.4). Since PCA is a projection onto a subspace, the features $\mathbf{h}$ in this case are not "higher order" functions of the original variables, but rather a linear mapping from original variables to a strictly smaller number of linear coefficients over the eigenvectors.

2.4.2 Low-Rank + Sparse Matrix Decomposition

The utility model U_{rp} given in Equation (2.1) can be rewritten into matrix form, in which Ω is a matrix reshaped from the "part-worth" coefficients vector ω:

$$U_{rp} = [\left(\mathbf{x}_c^{(j)}\right)^T, 1]\Omega \mathbf{x}_d^p \tag{2.6}$$

The decomposition of the original part-worth coefficients into a low-rank matrix and a sparse matrix may better represent customer purchase decisions than the large coefficient matrix of all pairwise interactions given in Equation (2.1) and as detailed in Section 2.2. Accordingly, we decompose Ω into a low-rank matrix $\mathbf{L}$ of rank r superimposed with a sparse matrix $\mathbf{S}$, i.e. $\Omega = \mathbf{L} + \mathbf{S}$. This problem may be solved in the general case exactly with the following optimization problem:

$$\min_{\mathbf{L},\mathbf{S}} l(\mathbf{L}, \mathbf{S}; \mathbf{X}_c, \mathbf{X}_d, \mathbf{y}) \tag{2.7}$$

$$s.t. \quad rank(\mathbf{L}) \leq r$$

$$\mathbf{S} \in \mathcal{C}$$

where $\mathbf{X}_u$ and $\mathbf{X}_c$ are the full set of customer and vehicle data, $\mathbf{y}$ is the vector of whether customer j chose vehicle p or vehicle q, $l(\cdot)$ is the log-loss without the l^2 norm,

$$l(\mathbf{L}, \mathbf{S}; \mathbf{X}_c, \mathbf{X}_d, \mathbf{y})$$

$$= \frac{1}{N} \sum_{n=1}^{N} (y^{(n)} \log P^{(n)} + (1 - y^{(n)}) \log(1 - P^{(n)})) \tag{2.8}$$

and $\mathcal{C}$ is a convex set corresponding to the sparse matrix $\mathbf{S}$. As this problem is intractable (NP-hard), we instead learn this decomposition of matrices using an approximation obtained via regularized loss function minimization:

$$\min_{\mathbf{L},\mathbf{S}} l(\mathbf{L}, \mathbf{S}; \mathbf{X}_c, \mathbf{X}_d, \mathbf{y}) + \lambda_1 ||\mathbf{L}||_* + \lambda_2 ||\mathbf{S}||_1 \tag{2.9}$$

where $||\cdot||_*$ is the nuclear norm to promote low-rank structure, and $||\cdot||_1$ is the l_1-norm.

In particular, while a number of low-rank regularizations may be used to solve Eq. (2.9), e.g., trace norm and log-determinant norm (*Fazel*, 2002). We choose the nuclear norm as it may be applied to any general matrix, while the trace norm and log-determinant regularization are limited to positive semidefinite matrices. Moreover, the nuclear norm is often considered optimal as $||\mathbf{L}||_*$ is the convex envelop of $Rank(\mathbf{L})$, implying that $||\mathbf{L}||_*$ is the largest convex function smaller than $Rank(\mathbf{L})$
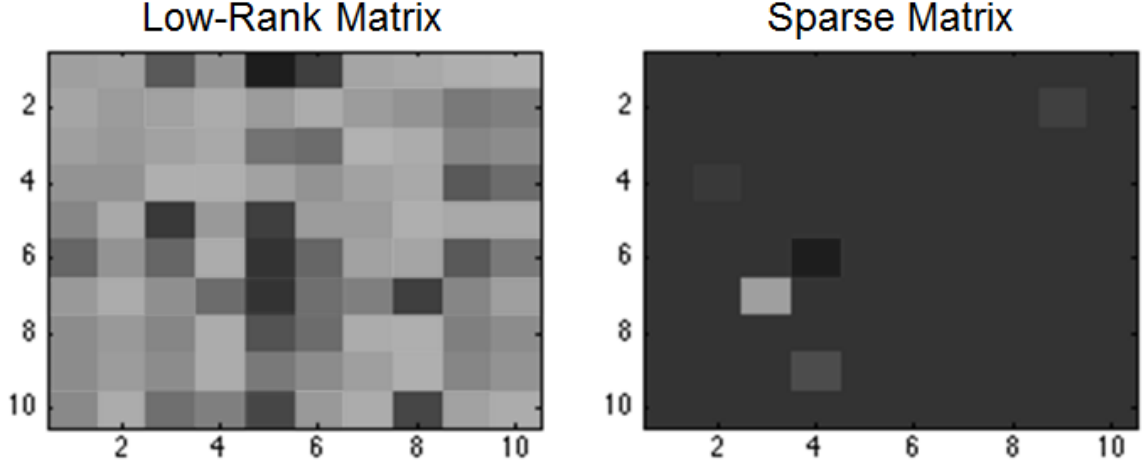
Figure 2.3: The concept of low-rank + sparse matrix decomposition using an example "part-worth coefficients" matrix of size 10 x 10 decomposed into two 10 x 10 matrices with low rank or sparse structure. Lighter colors represent larger values of elements in each decomposed matrix.

(*Fazel*, 2002).

Definition II.1. For matrix $\mathbf{L}$,the nuclear norm is defined as,

$$||\mathbf{L}||_* := \sum_{i=1}^{\min(dim(L))} s_i(\mathbf{L})$$

where $s_i(\mathbf{L})$ is a singular value of $\mathbf{L}$.

2.4.2.1 Parameter Estimation

The non-differentiability of the convex low-rank + sparse approximation given in Eq. (2.9) necessitates optimizations techniques such as augmented Lagrangian (*Tomioka et al.*, 2010), semi-definite programming (*Liu and Yan*, 2014), and proximal methods (*Parikh and Boyd*, 2013). Due to theoretical guarantees on convergence, we choose to train our model using proximal methods which are defined as follows.

Definition II.2. Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a closed proper convex function. The

---

Algorithm 1 Low-Rank + Sparse Matrix Decomposition

Input: Data $\mathbf{X}_c$, $\mathbf{X}_d$, $\mathbf{y}$
Initialize $\mathbf{L}^0 = \mathbf{0}, \mathbf{S}^0 = \mathbf{0}$
repeat
   $\mathbf{L}^{t+1} = prox_f(\mathbf{L}^t - \eta_t \nabla_{\mathbf{L}^t} l(\mathbf{L}, \mathbf{S}; \mathbf{X}_c, \mathbf{X}_d, \mathbf{y}))$
   $\mathbf{S}^{t+1} = prox_S(\mathbf{S}^t - \eta_t \nabla_{\mathbf{S}^t} l(\mathbf{L}, \mathbf{S}; \mathbf{X}_c, \mathbf{X}_d, \mathbf{y}))$
until $\mathbf{L}^t$, $\mathbf{S}_i^t$ are converged

---

proximal operator of f is defined as

$$\mathbf{prox}_f(\mathbf{v}) = \arg\min_{\mathbf{x}} \left(f(\mathbf{x}) + \frac{1}{2}||\mathbf{v} - \mathbf{x}||_2^2 \right)$$

With these preliminaries, we now detail the proximal gradient algorithm used to solve Eq. 2.9 using low-rank and l^1 proximal operators. Denote $f(\cdot) = \lambda_1 || \cdot ||_*$, and its proximal operator as $prox_f$. Similarly denote the proximal operator for the l^1 regularization term by $prox_S$, $i = 1, \ldots n$. Details of calculating $prox_f$ and $prox_S$ is given in the separate Section 2.5 below to maintain continuity of the exposition here..

With this notation, the proximal optimization algorithm to solve Equation (2.9) is given by Algorithm 1. Moreover, this algorithm is guaranteed to converge with constant step size as given by the following lemma (*Parikh and Boyd*, 2013).

Lemma II.3. *Convergence Property*

When ∇l is Lipschitz continuous with constant ρ, this method can be shown to converge with rate $O(\frac{1}{k})$ when a fixed step size $\eta_t = \eta \in (0, 1/\rho]$ is used. If ρ is not known, the step sizes η_t can be found by a line search; that is, their values are chosen in each step.

2.4.2.2 Error Bound on Low-Rank + Sparse Estimation

We additionally prove a variational bound that guarantees this parameter estimation method converges to a unique solution with bounded error as given by the following theorem.

Theorem II.4. *Error Bound on Low-Rank+Sparse Estimation*

$$|\triangle l| \leq \lambda_1 \min \left(dim(L_0) \right) ||\mathbf{L}^* - \mathbf{L}_0||_2$$

where $\mathbf{L}^*$ *is the optima of problem (2.9) and* $\mathbf{L}_0$ *is the matrix minimizing the loss function* $l(\cdot)$.

The proof of this theorem is given in Section 2.5.

2.4.3 Restricted Boltzmann machine

The restricted Boltzmann machine (RBM) is an energy-based model in which an energy state is defined by a layer of M visible nodes corresponding to the original variables $\mathbf{x}$ and a layer of K features denoted as $\mathbf{h}$. The energy for a given pair of original variables and features determines the probability associated with finding the system in that state; like nature, systems tend to states that minimize their energy and thus maximize their probability. Accordingly, maximizing the likelihood of the observed data $\mathbf{x}^{(1)} \ldots \mathbf{x}^{(N)} \in \mathbb{R}^M$ and its corresponding feature representation $\mathbf{h}^{(1)} \ldots \mathbf{h}^{(N)} \in \mathbb{R}^K$ is a matter of finding the set of parameters that minimize the energy for all observed data.

While traditionally this likelihood consists of binary variables and binary features, as described in Table 2.1 and Table 2.2, our passenger vehicle purchase data set consists of M_G Gaussian variables, M_B binary variables, and M_C categorical variables. We accordingly define three corresponding energy functions E_G, E_B, and E_C, in which each energy function connects the original variables and features via a weight matrix $\mathbf{W}$, as well as biases for each original variable and feature, $\mathbf{a}$ and $\mathbf{b}$ respectively.

Real-valued random variables (e.g., vehicle curb weight) are modeled using the Gaussian density. The energy function for Gaussian inputs and binary hidden nodes

is:

$$
\begin{aligned}
E_G(\mathbf{x}, \mathbf{h}; \theta) = & -\sum_{m=1}^{M_G}\sum_{k=1}^{K} h_k w_{km} x_m \\
& + \frac{1}{2}\sum_{m=1}^{M_G}(x_m - b_m)^2 - \sum_{k=1}^{K} a_k h_k
\end{aligned}
\tag{2.10}
$$

where the variance term is clamped to unity under the assumption that the input data are standardized.

Binary random variables (e.g., gender) are modeled using the Bernoulli density. The energy function for Bernoulli nodes in both the input layer and hidden layer is:

$$
\begin{aligned}
E_B(\mathbf{x}, \mathbf{h}; \theta) = & -\sum_{m=1}^{M_B}\sum_{k=1}^{K} h_k w_{km} x_m \\
& - \sum_{m=1}^{M_B} x_m b_m - \sum_{k=1}^{K} a_k h_k
\end{aligned}
\tag{2.11}
$$

Categorical random variables (e.g., vehicle manufacturer) are modeled using the categorical density. The energy function for categorical inputs with Z_m classes for m-th categorical input variable (e.g., Toyota, General Motors, etc.) is given by:

$$
\begin{aligned}
E_C(\mathbf{x}, \mathbf{h}; \theta) = & -\sum_{m=1}^{K_m}\sum_{k=1}^{K}\sum_{z=1}^{Z_m} h_k w_{kmz}\delta_{mz} x_{mz} \\
& - \sum_{m=1}^{M_C}\sum_{z=1}^{Z_m} \delta_{mz} x_{mz} b_{mz} - \sum_{k=1}^{K} a_k h_k
\end{aligned}
\tag{2.12}
$$

where $\delta_{mz} = 1$ if $x_{mz} = 1$ and 0 otherwise.

Given these energy functions for the heterogeneous original variables, the probability of a state with energy $E(\mathbf{x}, \mathbf{h}; \theta) = E_G(\mathbf{x}, \mathbf{h}; \theta) + E_B(\mathbf{x}, \mathbf{h}; \theta) + E_C(\mathbf{x}, \mathbf{h}; \theta)$, in which $\theta = \{\mathbf{W}, \mathbf{a}, \mathbf{b}\}$ are the energy function weights and bias parameters, is defined by the Boltzmann distribution.

$$
P(\mathbf{x}, \mathbf{h}) = \frac{e^{-E(\mathbf{x}, \mathbf{h}; \theta)}}{\sum_{\mathbf{x}}\sum_{\mathbf{h}} e^{-E(\mathbf{x}, \mathbf{h}; \theta)}}
\tag{2.13}
$$

34

$\mathbf{X} = [x_1, x_2]$ $\mathbf{h} = [h_1, h_2, h_3, h_4]$
(variables)   (features)
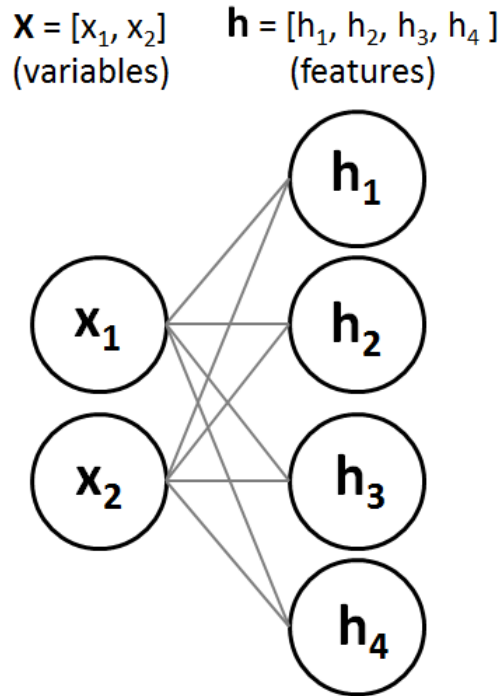
Figure 2.4: The concept of the exponential family sparse restricted Boltzmann machine. The original data are represented by nodes in the visible layer by $[x_1, x_2]$, while the feature representation of the same data is represented by nodes in the hidden layer $[h_1, h_2, h_3, h_4]$. Undirected edges are restricted to being only between the original layer and the hidden layer, thus enforcing conditional independence between nodes in the same layer.

The "restriction" on the RBM is to disallow visible-visible and hidden-hidden node connections. This restriction results in conditional independence of each individual hidden unit h given the vector of inputs $\mathbf{x}$, and each visible unit x given the vector of hidden units $\mathbf{h}$.

$$P\left(\mathbf{h}|\mathbf{x}\right) = \prod_{n=1}^{N} P\left(h_n|\mathbf{x}\right) \tag{2.14}$$

$$P\left(\mathbf{x}|\mathbf{h}\right) = \prod_{k=1}^{K} P\left(x_k|\mathbf{h}\right) \tag{2.15}$$

The conditional density for a single binary hidden unit given the combined K_G Gaussian, K_B binary, and K_C categorical input variables is then:

$$\sigma\big(a_n + \sum_{k=1}^{K_G} w_{nk}x_k + \sum_{k=1}^{K_B} w_{nk}x_k + \sum_{k=1}^{K_C}\sum_{d=1}^{D_k} w_{nk}\delta_{kd}x_k\big) \tag{2.16}$$

where $\sigma(s) = \frac{1}{1+\exp(-s)}$ is a sigmoid function.

For an input data point $\mathbf{x}^{(n)}$, its corresponding feature representation $\mathbf{h}^{(n)}$ is given by sampling the "activations" of the hidden nodes.

$$[P(h_1 = 1|\mathbf{x}, \theta), \, ..., \, P(h_N = 1|\mathbf{x}, \theta)] \tag{2.17}$$

Parameter Estimation

To train the model, we optimize the weight and bias parameters $\theta = \{\mathbf{W}, \mathbf{b}, \mathbf{a}\}$ by minimizing the negative log-likelihood of the data $\{\mathbf{x}^{(1)} \ldots \mathbf{x}^{(N)}\}$ using gradient

descent. The gradient of the log-likelihood is:

$$
\begin{aligned}
\frac{\partial}{\partial \theta} \sum_{n=1}^{N} \log P\left(\mathbf{x}^{(n)}\right) &= \frac{\partial}{\partial \theta} \sum_{n=1}^{N} \log \sum_{\mathbf{h}} P\left(\mathbf{x}^{(n)}, \mathbf{h}\right) \\
&= \frac{\partial}{\partial \theta} \sum_{n=1}^{N} \log \sum_{\mathbf{h}} \frac{e^{-E\left(\mathbf{x}^{(n)}, \mathbf{h}\right)}}{\sum_{\mathbf{x}, \mathbf{h}} e^{-E\left(\mathbf{x}^{(n)}, \mathbf{h}\right)}} \\
&= \sum_{n=1}^{N} \mathbb{E}_{\mathbf{h} \mid \mathbf{x}^{(n)}}\left[\frac{\partial}{\partial \theta} E\left(\mathbf{x}^{(n)}, \mathbf{h}\right)\right] \\
&\quad - \mathbb{E}_{\mathbf{h}, \mathbf{x}}\left[\frac{\partial}{\partial \theta} E(\mathbf{x}, \mathbf{h})\right]
\end{aligned}
\tag{2.18}
$$

The gradient is the difference of two expectations, the first of which is easy to compute since it is "clamped" at the input datum $\mathbf{x}$, but the second of which requires the joint density over the entire $\mathbf{x}$ space for the model.

In practice, this second expectation is approximated using the contrastive divergence algorithm by Gibbs, sampling the hidden nodes given the visible nodes, then the visible nodes given the hidden nodes, and iterating a sufficient number of steps for the approximation (*Hinton*, 2002). During training, we induce sparsity of the hidden layer by setting a target activation β_k, fixed to 0.1, for each hidden unit h_k (*Lee et al.*, 2008). The overall objective to be minimized is then the negative log-likelihood from Equation (2.18) and a penalty on the deviation of the hidden layer from the target activation. Since the hidden layer is made up of sigmoid densities, the overall objective function is:

$$
\begin{aligned}
&\sum_{n=1}^{N} \log \sum_{\mathbf{h}} P\left(\mathbf{x}^{(n)}, \mathbf{h}\right) \\
&+ \lambda_3 \sum_{k=1}^{K}\left(\beta_k^{(n)} \log h_k + \left(1 - \beta_k^{(n)}\right) \log \left(1 - h_k\right)\right),
\end{aligned}
\tag{2.19}
$$

where λ_3 is the hyperparameter trading off the sparsity penalty with the log-likelihood.

2.5 Proof of Low-Rank Matrix Estimation Guarantee

Though the low-rank matrix is estimated jointly with the decomposed matrices as well as the loss function, an accurate estimation of the low-rank matrix can still be achieved as guaranteed by the bound as in this section. =

We subsequently provide a variational bound of the divergence of the estimated likelihood from the true likelihood.

To simplify the notation in our proof, we redefine the following notation:

$$(m, n) = dim(\Omega)$$

Before our proof, however, we state the following relevant prepositions.

Proposition 1. *The proximal operator for the nuclear norm $f = \lambda_1 || \cdot ||_*$ is the singular value shrinkage operator $\mathcal{D}_{\lambda_1}$.*

Consider the singular value decomposition (SVD) of a matrix $\mathbf{X} \in \mathbb{R}^{m \times n}$ with rank r.

$$\mathbf{X} = \mathbf{U}\Sigma\mathbf{V}^T$$

$$\mathcal{D}_{\lambda_1}(\mathbf{X}) = \mathbf{U}\mathcal{S}_{\lambda_1}(\Sigma)\mathbf{V}^T \tag{2.20}$$

where the soft-thresholding operator $\mathcal{S}_{\lambda_1}(\Sigma) = diag(\{max(s_i - \lambda_1, 0)\}_{i=1,...,\min(m,n)})$. Moreover, $\mathcal{S}_{\lambda_2}(\cdot)$ is also the proximal operator for the l^1 norm.

The matrix decomposition structure of our model builds on the separable sum property(*Parikh and Boyd*, 2013):

Proposition 2. *Separable Sum Property*

If f is separable across two variables x and y, i.e., $f(\mathbf{x}, \mathbf{y}) = f_1(\mathbf{x}) + f_2(\mathbf{y})$, then,

$$prox_f(\mathbf{x}, \mathbf{y}) = (prox_{f_1}(\mathbf{x}), prox_{f_2}(\mathbf{y})) \tag{2.21}$$

Our proof proceeds as follows. Let us denote the optima of problem (2.9) as $\mathbf{L}^*$, the gradient of the loss function $l(\cdot)$ w.r.t $\mathbf{L}^*$ as $\nabla_{\mathbf{L}^*}l$, and the matrix minimizing the loss function $l(\cdot)$ as $\mathbf{L}_0$.

We next prove the following theorems: Theorem II.5 provides a tight bound on $\nabla_{\mathbf{L}^*}l$. Corollary II.6 bounds the estimation error for the learned matrix $\mathbf{L}^*$. Theorem II.7 follows by bounding the divergence of likelihood from the true data distribution where $l(\cdot)$ is a likelihood function.

First, we make the weak assumption that the optimization problem given in Equation (2.9) is strictly convex, since a necessary and sufficient condition is that the saddle points for $l(\cdot)$ and the regularization terms are not overlapping.

Theorem II.5. *Loss Function Gradient Bound.*

$$||\nabla_{\mathbf{L}^*}l||_2 \leq \lambda_1 \min(m, n)$$

Proof. Under the strictly convex assumption, the stationary point (i.e., the optima $\mathbf{L}^*$ for the optimization problem (2.9)) is unique. By Lemma II.3, iterations of the proximal gradient optimization method $\mathbf{L}_k$ converge to this optima $\mathbf{L}^*$. According to the fixed point equation for $\mathbf{L}$ (Algorithm 1), we have,

$$\mathbf{L}^* = prox_f(\mathbf{L}^* - \eta\nabla_{\mathbf{L}^*}l) \tag{2.22}$$

Denote $\mathbf{L}^* - \eta\nabla_{\mathbf{L}^*}l$ as $\mathbf{M}$, representing the argument of the proximal operator at the optimal low-rank estimation. The singular value decomposition (SVD) for $\mathbf{L}^*$, $\mathbf{M}$,

and $prox_f(\mathbf{M})$ yields,

$$\mathbf{L}^* = \mathbf{U}\Sigma\mathbf{V}^T \tag{2.23}$$

$$\mathbf{M} = \mathbf{U}_M\Sigma^M\mathbf{V}_M^T \tag{2.24}$$

$$prox_f(\mathbf{M}) = \mathbf{U}_{prox}\Sigma^{prox}\mathbf{V}_{prox}^T \tag{2.25}$$

where $\mathbf{U}, \mathbf{U}_{prox} \in \mathbb{R}^{m\times r}$; $\mathbf{V}^T, \mathbf{V}_{prox}^T \in \mathbb{R}^{r\times n}$; $\Sigma, \Sigma^{prox} \in \mathbb{R}^{r\times r}$ with $\Sigma = diag(\{s_i\}_{i=1,\ldots,r})$, $\Sigma^{prox} = diag(\{s_i^{prox}\}_{i=1,\ldots,r})$. $\mathbf{U}_M \in \mathbb{R}^{m\times m}$, $\mathbf{V}_M^T \in \mathbb{R}^{n\times n}$ and Σ^M is a $m \times n$ rectangular diagonal matrix.

Without loss of generality, assume that $s_1 > s_2 > \cdots > s_r > 0$, i.e., these singular values are distinct and positive, thus ensuring column orderings are unique. Thus, we may assert that $\mathbf{U} = \mathbf{U}_{prox}$, $\mathbf{V} = \mathbf{V}_{prox}$ and $\Sigma = \Sigma^{prox}$ due to the uniqueness of SVD for distinct singular values in $\mathbf{L}^* = prox_f(\mathbf{M})$.

According to Proposition 1,

$$prox_f(\mathbf{M}) = \mathbf{U}_M \max\left(\Sigma^M - \lambda_1\eta\mathbf{I}, \mathbf{0}\right)\mathbf{V}_M^T \tag{2.26}$$

Note that the dimensionality of $prox_f(\mathbf{M})$ is less than that of the value of M. To bridge the gap between them, we define diagonal sub-matrices Σ_+^M and Σ_-^M. (In other words, we partition Σ^M into two sub-matrices Σ_+^M and Σ_-^M.) For all singular values s_i^M of $\mathbf{M}$, $i = 1, 2, \ldots, \min(m, n)$, if $s_i^M - \lambda_1\eta \geq 0$, then s_i^M is a diagonal element of the sub-matrix Σ_+^M, otherwise, s_i^M is a diagonal element of the sub-matrix Σ_-^M. Hence, $max(\Sigma_+^M - \lambda_1\eta\mathbf{I}, \mathbf{0}) = \Sigma_+^M - \lambda_1\eta\mathbf{I}$ and $max(\Sigma_-^M - \lambda_1\eta\mathbf{I}, \mathbf{0}) = \mathbf{0}$.

$$prox_f(\mathbf{M}) = \mathbf{U}_M^+\left(\Sigma_+^M - \lambda_1\eta\mathbf{I}\right)(\mathbf{V}_M^+)^T$$

where $\mathbf{U}_M^+$ ($\mathbf{V}_M^+$) are left-singular (right-singular) vectors corresponding to Σ_+^M, $\mathbf{U}_M^-$ and $\mathbf{V}_M^-$ are also defined respectively. Again, due to the uniqueness of SVD, we have $\mathbf{U}_M^+ = \mathbf{U}$ and $\mathbf{V}_M^+ = \mathbf{V}$

We now rewrite the SVD formula for $prox(\mathbf{M})$ and $\mathbf{M}$ as,

$$prox(\mathbf{M}) = \mathbf{U}\left(\Sigma_+^M - \lambda_1 \eta \mathbf{I}\right)\mathbf{V}^T \tag{2.27}$$

$$\begin{aligned}
\mathbf{M} &= \mathbf{U}_M \Sigma^M \mathbf{V}_M^T \\
&= \mathbf{U}_M^+ \Sigma_+^M (\mathbf{V}_M^+)^T + \mathbf{U}_M^- \Sigma_-^M (\mathbf{V}_M^-)^T \\
&= \mathbf{U}\Sigma_+^M \mathbf{V}^T + \mathbf{U}_M^- \Sigma_-^M (\mathbf{V}_M^-)^T
\end{aligned} \tag{2.28}$$

By definition of $\mathbf{M}$,

$$\mathbf{M} = \mathbf{U}\Sigma\mathbf{V}^T - \eta\nabla_{\mathbf{L}^*}l \tag{2.29}$$

Equation (2.22) and (2.27) indicates that,

$$\mathbf{U}\Sigma\mathbf{V}^T = \mathbf{U}(\Sigma_+^M - \lambda_1 \eta \mathbf{I})\mathbf{V}^T \tag{2.30}$$

$$\Sigma = \Sigma_+^M - \lambda_1 \eta \mathbf{I} \tag{2.31}$$

By Equation (2.28), (2.29), and (2.31), we have

$$\begin{aligned}
-\nabla_{\mathbf{L}^*}l &= \mathbf{U}\left(\Sigma_+^M - \Sigma\right)\mathbf{V}^T + \mathbf{U}_M^- \Sigma_-^M (\mathbf{V}_M^-)^T \\
&= \lambda_1 \mathbf{U}\mathbf{V}^T + \frac{1}{\eta}\mathbf{U}_M^- \Sigma_-^M (\mathbf{V}_M^-)^T
\end{aligned} \tag{2.32}$$

Note that every diagonal element s^M_{-i} in Σ^M_- satisfies $0^M_{-i} \le \lambda_1 \eta$. Hence,

$$||\nabla_{\mathbf{L}^*} l||_2 \le \lambda_1 ||\mathbf{U}\mathbf{V}^T||_2 + \frac{1}{\eta}||\mathbf{U}^-_M \Sigma^M_- (\mathbf{V}^-_M)^T||_2$$

$$\le \lambda_1 \sum_i^r ||U_{:i} V^T_{:i}||_2 + \sum_{j=1}^{\min(m,n)-r} \frac{s^M_{-i}}{\eta}||[\mathbf{U}^-_M]_{:j} ([\mathbf{V}^-_M]_{:j})^T||_2 \qquad (2.33)$$

$$\le \lambda_1 \min(m,n) \qquad (2.34)$$

where $\mathbf{U}_{:i}$ or $\mathbf{V}_{:i}$ is the i-th column in matrix $\mathbf{U}$ or $\mathbf{V}$, and $[\mathbf{U}^-_M]_{:j}$ or $[\mathbf{V}^-_M]_{:j}$ is the j-th column in matrix $\mathbf{U}^-_M$ or $\mathbf{V}^-_M$. $\qquad \square$

Summarizing the proof of Theorem II.5, the gradient of the loss function at the estimated low-rank matrix is bounded by a unit ball within the original problem space that has radius of the low-rank regularization parameter λ_1. The relaxation of the bound partially comes from the second term in inequality (2.33). This implies that the bound is tighter if the rank of L^* is increased.

Based on the gradient bound given in Theorem II.5, we now bound the estimation error of the learned low-rank matrix $\mathbf{L}^*$. Although the value of the bound is not explicit in this proof, in some cases we are able to explicitly calculate its value.

Corollary II.6. *Learned Low-Rank Matrix Estimation Error. The error $||\mathbf{L}^* - \mathbf{L}_0||_2$ is bounded by the diameter of minimum-sized ball that include the following set*

$$\{\mathbf{L} : ||\nabla_{\mathbf{L}} l||_2 \le \lambda_1 \min(m,n)\}$$

Proof. The proof directly follows from Theorem II.5 and the fact that $\nabla_{\mathbf{L}_0} l = \mathbf{0}$. $\qquad \square$

Since the loss function $l(\cdot)$ is convex, the Euclidean norm of its gradient $\nabla_{\mathbf{L}} l$ is non-decreasing as the Euclidean distance $||\mathbf{L} - \mathbf{L}_0||_2$ is increasing.

When the loss function is sharp around its minima, then $\{\mathbf{L} : ||\nabla_{\mathbf{L}} l||_2 \leq \lambda_1 \min(m, n)\}$ is a small region which implies that $\mathbf{L}^*$ is a good estimation of $\mathbf{L}_0$.

We next bound the likelihood divergence when the loss function $l(\cdot)$ is a likelihood function. To do this, we use Theorem II.5 and Corollary II.6 to construct a variational bound.

Theorem II.7. *Variational Bound on Estimated Likelihood*

$$|\triangle l| \leq \lambda_1 \min(m, n) \, ||\mathbf{L}^* - \mathbf{L}_0||_2$$

Proof. By the Lagrangian mean value theorem, there exists $\mathbf{L}_1 \in \{\mathbf{L} : \mathbf{L}_{ij} \in [\mathbf{L}_{ij}^*, \mathbf{L}_0^{ij}]\}$ such that,

$$l(\mathbf{L}^*; \mathbf{X}, \mathbf{S}) - l(\mathbf{L}_0; \mathbf{X}, \mathbf{S}) = \langle \nabla_{\mathbf{L}_1} l, (\mathbf{L}^* - \mathbf{L}_0) \rangle$$
$$\leq ||\nabla_{\mathbf{L}_1} l||_2 ||\mathbf{L}^* - \mathbf{L}_0|| \tag{2.35}$$

where $\langle \mathbf{A}, \mathbf{B} \rangle$ denotes inner product of $vec(\mathbf{A})$ and $vec(\mathbf{B})$, in which $vec(\cdot)$ is the matrix vectorization operator.

Because of the convexity of $l(\cdot)$, $||\nabla_{\mathbf{L}_1} l||_2 \leq ||\nabla_{\mathbf{L}^*} l||_2$. By Theorem II.5,

$$|\triangle l| = l(\mathbf{L}^*; \mathbf{X}, \mathbf{S}) - l(\mathbf{L}_0; \mathbf{X}, \mathbf{S}) \tag{2.36}$$
$$\leq \lambda_1 \min(m, n) \, ||\mathbf{L}^* - \mathbf{L}_0||_2 \tag{2.37}$$

$\square$

Summarizing the proof of Theorem II.7, the variational bound of the estimated likelihood depends on both the bound of gradient of the likelihood function $l(\cdot)$ given in Theorem II.5 and the property of the likelihood function in the neighborhood of

its optima $\mathbf{L}_0$ as described by Corollary II.6.

2.6   Experiment

In this section we present a computational experiment to assess how preference prediction accuracy changes when using the same preference model on three different representations of the same data set. The preference model used, as discussed in Section 2.3.4, is the l^2 logit, while the three representations are the original variables, low-rank + sparse features, and RBM features. The same experimental procedure was run on each of these three representations, where the first representation acts as a baseline for prediction accuracy, and the next two representations demonstrate the relative gain in preference prediction accuracy when using features.

In addition, we performed an analysis of how the hyperparameters affected design preference prediction accuracy for the hyperparameters used in the PCA, LSD, and RBM feature learning methods. For PCA, the hyperparameter was the dimensionality K of the subspace spanned by the eigenvectors of the PCA method. For LSD, the hyperparameters were the rank penalty λ_1, which affects the rank of the low-rank matrix L, and the sparsity penalty λ_2, which influences the number of non-zero elements in the sparse matrix S, both found in Equation (2.9). For RBM, the hyperparameters were the sparsity penalty λ_3, which controls the number of features activated for a given input datum, and the overcompleteness factor γ, which defines by what factor the dimensionality of the feature space is larger than the dimensionality of the original variable space, both of which are found in Equation (2.19).

The detailed experiment flow is summarized below and illustrated in Figure 2.6:

1. The raw choice data set of pairs of customers and purchased designs, described in Section 2.3.1, was randomly split 10 times into 70% training, 10% validation, and 20% test sets. This was done in the beginning to ensure no customers in
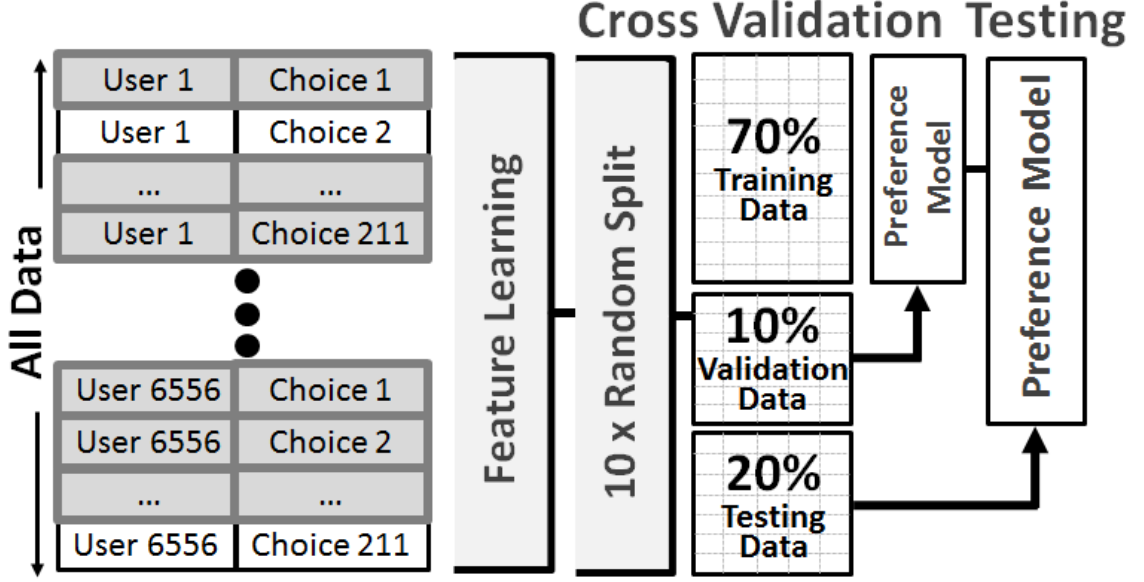
Figure 2.5: Data processing, training, validation, and testing flow.

the training sets ever existed in the validation or test sets.

2. Choice sets were generated for each training, validation, and test sets for all 10 randomly shuffled splits as described in Section 2.3.2. This process created a training data set of 832,000 data points, a validation data set of 104,000 data points, and a testing data set of 225,056 data points, for each of the 10 shuffled splits.

3. Feature learning was conducted on the training sets of customer variables and vehicle variables for a vector of 5 different values of K for PCA features, a grid of 25 different pairs of low-rank penalty λ_1 and sparsity penalty λ_2 for the LSD features, and a grid of 56 different pairs of sparsity λ_3 and overcompleteness γ hyperparameters for RBM features. For PCA features, these hyperparameters were $K \in \{30, 50, 70, 100, 150\}$. For LSD features, these hyperparameters were $\lambda_1 \in \{0.005, 0.01, 0.05, 0.1, 0.5\}$ and $\lambda_2 \in \{0.005, 0.01, 0.05, 0.1, 0.5\}$. For RBM, these hyperparameters were $\lambda_3 \in \{4.0, 5.0, 6.0, 7.0, 8.0, 9.0, 10.0\}$ and $\gamma \in \{0.25, 0.5, 0.75, 1.0, 1.5, 2.0, 2.5, 3.0\}$. These hyperparameter settings were

Table 2.3: Averaged preference prediction accuracy on held-out test data using the logit model with the original variables or the three feature representations. Average and standard deviation were calculated from 10 random training and testing splits common to each method, while test parameters for each method were selected via cross validation on the training set.

Design Preference Model	Feature Representation	Prediction Accuracy (Std. dev.) (p-value) N=10,000	Prediction Accuracy (Std. dev.) (p-value) N=1,161,056
Logit Model	Original Variables (No Features)	69.98% (1.82%) (N/A)	75.29% (0.98%) (N/A)
Logit Model	Priciple Component Analysis	61.69% (1.24%) (1.081e-7)	62.03% (0.89%) (8.22e-10)
Logit Model	Low-Rank +Sparse Matrix Decomposition	76.59% (0.89%) (3.276e-8)	77.58% (0.81%) (4.286e-8)
Logit Model	Exponential Family Sparse RBM	74.99% (0.64%) (2.3e-5)	75.15% (0.81%) (0.136)

selected by pilot testing large ranges of parameter settings to find relevant regions for upper and lower hyperparameter bounds, with numbers of hyperparameters selected based on computational constraints.

4. Each of the validation and testing data sets were encoded using the feature learning methods learned for each of the 5 PCA hyperparameters K, 25 (λ_1, λ_2) LSD hyperparameter pairs, and 56 (λ_3, γ) RBM hyperparameter pairs.

5. The encoded feature data was combined with the original variable data in order to separate linear term effects of the original variables with higher order effects from the features. While this introduces a degree of information redundancy between features and original variables, the regularization term in Equation 2.3 mitigates effects of collinearity. Each datum consists of the features concatenated with the original variables, then input into the bilinear utility model. Specifically, for some customer features $\mathbf{h}_u$ and customer variables $\mathbf{x}_u$, we used

$\mathbf{h}_{u'}^T := [\mathbf{x}_u^T, \mathbf{h}_u^T]$ to define the new representation of the customer; likewise, for some vehicle features $\mathbf{h}_c$ and vehicle variables $\mathbf{x}_c$, we used $\mathbf{h}_{c'}^T := [\mathbf{x}_c^T, \mathbf{h}_c^T]$ to define the new representation of the customer. Combined with Equation (2.1), a single data point used for training is the difference in utilities between vehicle p and vehicle q for a given customer r.

$$\left[\mathbf{h}_{u'}^{(r)} \otimes \left(\mathbf{h}_{c'}^{(p)} - \mathbf{h}_{c'}^{(q)} \right), \mathbf{h}_{c'}^{(p)} - \mathbf{h}_{c'}^{(q)} \right] \tag{2.38}$$

Note that the dimensionality of each datum could range above 100,000 dimensions for the largest values of γ.

6. For each of these training sets, 6 logit models were trained in parallel over minibatches of the training data, corresponding to 6 different settings of the l^2 regularization parameter $\alpha = 0.00001, 0.0001, 0.001, 0.01, 0.1, 1.0$. These logit models were optimized using stochastic gradient descent, with learning rates inversely related to the number of training examples seen (*Bottou*, 2010).

7. Each logit model was then scored according to its respective held-out validation data set. The hyperparameter settings $(\alpha_{BASELINE})$ for the original variables, (K_{PCA}, α_{PCA}) for PCA feature learning, (λ_1, λ_2) for LSD feature learning, and $(\lambda_3, \gamma, \alpha_{RBM})$ for RBM feature learning with the best validation accuracy were saved. For each of these four sets of best hyperparameters, Step 3 was repeated to obtain the set of corresponding features on each of the 10 random shuffled training plus validation sets.

8. Logit models corresponding to the baseline, PCA features, LSD features, and RBM features were retrained for each of the 10 randomly shuffled and combined training and validation. The prediction accuracy for each of these 10 logit models was assessed on the corresponding "held out" test sets in order to give

average and standard deviations of the design preference predictive accuracy for the baseline, PCA features, LSD features, and RBM features.

2.6.1 Results

Table 2.3 shows the averaged test set prediction accuracy of the logit model using the original variables, PCA features, LSD features, and RBM features. Prediction accuracy averaged over 10 random training and held-out testing data splits are given, both for the partial data $N = 10,000$ and the full data $N = 1,161,056$ cases. Furthermore, we include the standard deviation of the prediction accuracies and a 2-sided t-test relative to the baseline accuracy for each feature representation.

The logit model trained with LSD features achieved the highest predictive accuracy on both the partial and full data sets, at 76.59% and 77.58%, respectively. This gives evidence that using features can improve design preference prediction accuracy as the logit model using the original variables achieved an averaged accuracy of 69.98% and 75.29%, respectively. The improvement in design preference prediction accuracy is greatest for the partial data case, as evidenced by both the LSD and RBM, yet the improvement with the full data case shows that the LSD feature learning method is still able to improve prediction accuracy within the capacity of the logit model. The RBM results for the full data case do not show significant improvement in prediction accuracy. Finally, we note a relative loss in design preference prediction accuracy when using PCA as a feature learning method, both for the partial and full data sets, suggesting the heavy assumptions built into PCA are overly restrictive.

The parameter settings for the LSD feature learning method give additional insight to the preference prediction task. In particular, the optimal settings of λ_1 and λ_2 obtained through cross validation on the 10 random training sets was ranged from $r = 29$ to $r = 31$. This significantly reduced rank of the part-worth coefficient matrix given in Eq. (2.1) suggests that the vast majority of interactions between customer

variables and design variables given in Table 2.1 and Table 2.2 do not significantly contribute to overall design preferences. This insight allows us to introspect into important feature pairings on a per-customer basis to inform design decisions.

We have shown that even "simple" single-layer feature learning can significantly increase predictive accuracy for design preference modeling. This finding signifies that features more effectively capture the design preferences than the original variables, as features form functions of the original variables more representative of the customer's underlying preference task. This offers designers opportunity for new insights if these features can be successfully interpreted and translated to actionable design decisions; however, given the relatively recent advances in feature learning methods, interpretation and visualization of features remains an open challenge–see Section 2.7 for further discussion.

Further increases to prediction accuracy might be achieved by stacking multiple feature learning layers, often referred to as "deep learning". Such techniques have recently shown impressive results by breaking previous records in image recognition by large margins (*Krizhevsky et al.*, 2012). Another possible direction for increasing prediction accuracy may be in developing novel architectures that explicitly capture the conditional statistical structure between customers and designs. These efforts may be further aided through better understanding of the limitations of using feature learning methods for design and marketing research. For example, the large number of parameters associated with feature learning methods results in greater computational cost when performing model selection; in addition to the cross-validation techniques used in this paper, model selection metrics such as BIC and AIC may give further insight along these lines.

2.7 Using Features for Design

Using features can support the design process in at least two directions: (1) Features interpretation can offer deeper insights into customer preferences than the original variables, and (2) feature visualization can lead to a market segmentation with better clustering than with the original variables. These two directions are still open challenges given the relative nascence of feature learning methods. Further investigation is necessary to realize the above design opportunities and to justify the computational cost and implementation challenges associated with feature learning methods.

The interpretation and visualization methods may be used with conventional linear discrete choice modeling (e.g., logit models). However, deeper insights are possible through interpreting and visualizing features, assuming that features capture more effectively the underlying design preference prediction task of the customer as shown through improved prediction accuracy on held-out data. Since we are capturing "functions" of the original data, we are more likely to interpret and visualize feature pairings such as "eco-friendly" vehicle and "environmentally conscious" customer; such pairing may ultimately lead to actionable design decisions.

2.7.1 Feature Interpretation of Design Preferences

Similar to PCA, LSD provides an approach to interpret the learned features by looking at the linear combinations of original variables. The major difference between features learned using PCA versus LSD is their different linear combinations; in particular, features learned by LSD are more representative as they contain information from both the data distribution and the preference task, while PCA features only contain information from the data distribution.

As introduced in section 2.4.2, the weight matrix Ω is decomposed into a low-rank matrix $\mathbf{L}$ and a sparse matrix $\mathbf{S}$, i.e. $\Omega = \mathbf{L} + \mathbf{S}$. The nonzero elements in the sparse matrix $\mathbf{S}$ may be interpreted as the weight of the product of its corresponding original

design variables and customer variables. As for the low-rank matrix $\mathbf{L}$, features can be extracted by linearly combining the original variable according to the singular value decomposition (SVD) for $\mathbf{L}$. The singular value decomposition is a factorization of the $(m+1) \times n$ matrix $\mathbf{L}$ in the form $\mathbf{L} = \mathbf{U\Sigma V}$, where $\mathbf{U}$ is a $(m+1) \times (m+1)$ unitary matrix, Σ is an $m \times n$ rectangular diagonal matrix with non-negative real numbers $\sigma_1, \sigma_2, \ldots, \sigma_{\min(m+1,n)}$ on the diagonal, and $\mathbf{V}$ is a $(n) \times (n)$ unitary matrix. Rewriting Equation (2.6):

$$
\begin{aligned}
U_{rp} &= \left[\left(\mathbf{x}_c^{(j)} \right)^T, 1 \right] \mathbf{L} \mathbf{x}_d^p + \left[\left(\mathbf{x}_c^{(j)} \right)^T, 1 \right] \mathbf{S} \mathbf{x}_d^p \\
&= \left[\left(\mathbf{x}_c^{(j)} \right)^T, 1 \right] \mathbf{U\Sigma V} \mathbf{x}_d^p + \left[\left(\mathbf{x}_c^{(j)} \right)^T, 1 \right] \mathbf{S} \mathbf{x}_d^p \\
&= \sum_{i=1}^{min(m+1,n)} \sigma_i \left[\left(\mathbf{x}_c^{(j)} \right)^T, 1 \right] \mathbf{u}_i \mathbf{v}_i \mathbf{x}_d^p + \left[\left(\mathbf{x}_c^{(j)} \right)^T, 1 \right] \mathbf{S} \mathbf{x}_d^p
\end{aligned}
\tag{2.39}
$$

where $\mathbf{u}_i$ is the i-th column of matrix U, and $\mathbf{v}_i$ is the i-th row of matrix V. The i-th user feature $\left[\left(\mathbf{x}_c^{(j)} \right)^T, 1 \right] \mathbf{u}_i$ is a linear combination of original user variables; the i-th design feature $\mathbf{v}_i \mathbf{x}_d^p$ is a linear combination of original design variables; and σ_i represents the importance of this pair of features for the customer's design preferences.

Interpreting these features in the vehicle preference case study, we found that the most influential feature pairing (i.e., largest σ_i) corresponds to preference trends at the population level: Low price but luxury vehicles are preferred, and Japanese vehicles receive the highest preference while GM vehicles receive the lowest preference. The second most influential feature pairing represents a rich customer group, with preferred vehicle groups being both expensive and luxurious. The third most influential feature pairing represents an elder user group, with their preferred vehicles as large but with low net horsepower.
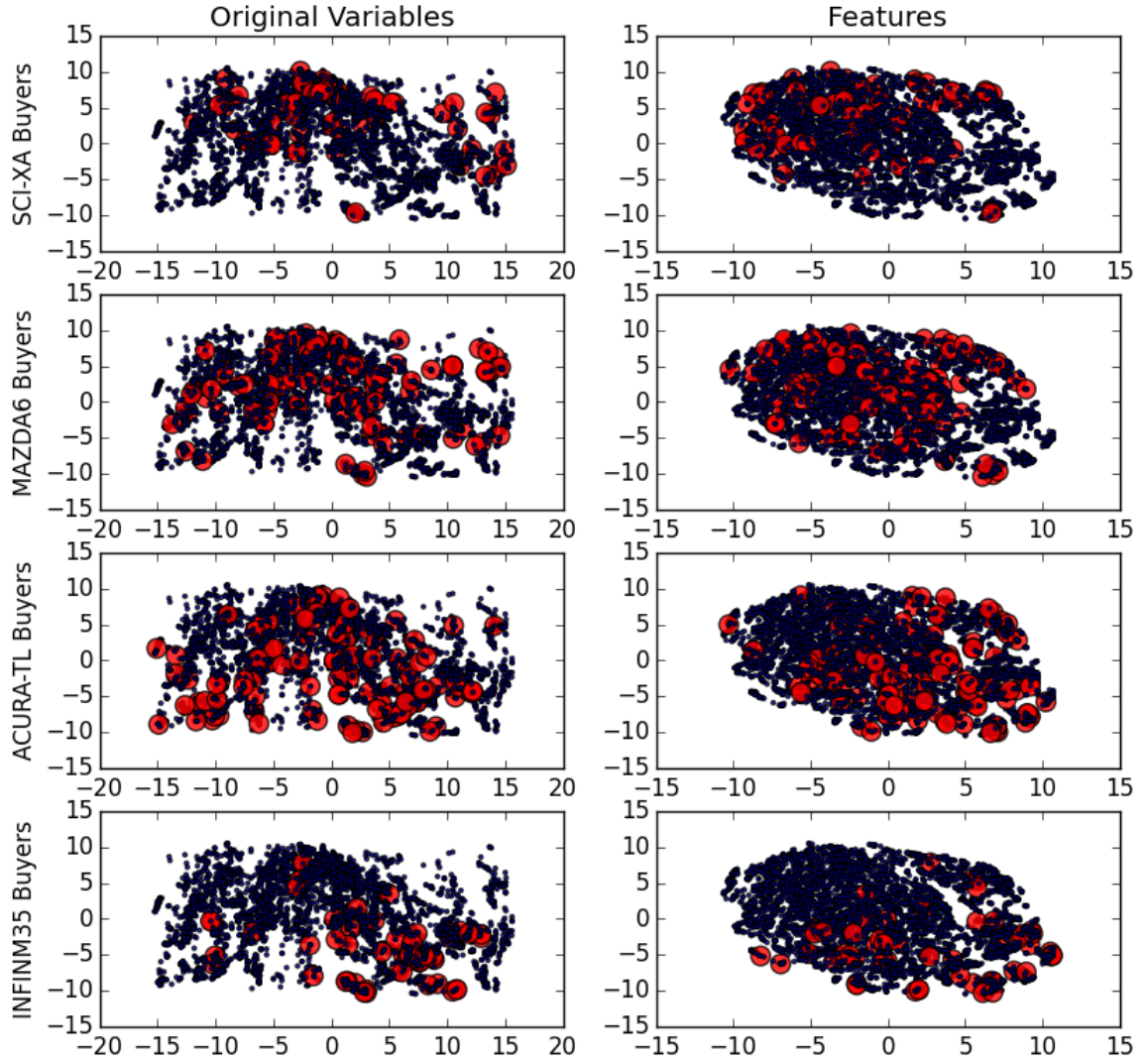
Figure 2.6: Optimal vehicle distribution visualization. Every point represents the optimal vehicle for one consumer. In the left column, the optimal vehicle is inferred using the utility model with original variables. In the right column, LSD features are used to infer the optimal vehicle. In the first row, the optimal vehicles from SCI-XA customers are marked in big red points. Similarly, the optimal vehicles from MAZDA6, ACURA-TL and INFINM35 customers are marked in big red points respectively.

2.7.2 Features Visualization of Design Preferences

We now visualize features to understand what insights for design decision making. Specifically, we make early-stage inroad to visual market segmentation performed in an estimated feature space, thus clustering customers in a representation that better captures their underlying design preference decisions.

We begin by looking at the utility model U_{rp} given in Equation (2.1) and note that the inner product between Ω and the variables $\mathbf{x}_u^{(r)}$ representing customer r may be interpreted as customer r's optimal vehicle, denoted $\mathbf{x}_{opt}^{(r)}$:

$$\mathbf{x}_{opt}^{(r)} = \left(\mathbf{x}_u^{(r)}\right)^T \Omega_{out} + \mathbf{1}^T \Omega_{main} \tag{2.40}$$

where Ω_{out} is the matrix reshaped from the coefficients of Ω corresponding to the outer product given in Equation (2.1), Ω_{main} is the matrix reshaped from the remaining coefficients, and $\mathbf{1}$ is a vector consisting of 1's with the same dimension as $\mathbf{x}_u^{(r)}$. We rewrite the utility model U_{rp} given in Equation (2.1) in terms of the optimal vehicle $\mathbf{x}_{opt}^{(r)}$:

$$U_{rp} = \left(\mathbf{x}_{opt}^{(r)}\right)^T \mathbf{x}_d^p \tag{2.41}$$

According to the geometric meaning of inner product, the smaller the angle between $\mathbf{x}_d^p$ and $\mathbf{x}_{opt}^{(r)}$ is, the larger will be the utility U_{rp}. In this way, we have an interpretable method of improving upon the actual purchased vehicle design in the form of an 'optimal' vehicle vector. This optimal vehicle vector could be useful for a manufacturer developing a next-generation design from a current design, particularly as the manufacturer would target a specific market segment.

We now provide a visual demonstration of using an optimal vehicle derived from feature learning to suggest a design improvement direction. First, we calculate the optimal vehicle using Equation (2.40) for every customer in the data set. Then, we

visualize these optimal vehicle points by reducing their dimension using t-distributed stochastic neighbor embedding (t-SNE), an advanced nonlinear dimension reduction technique that embeds similar objects into nearby points (*van der Maaten*, 2008). Finally, optimal vehicles from targeted market segments are marked in red.

Figure 2.6 shows the optimal vehicles for the SCI-XA, MAZDA6, ACURA-TL and INFINM35 customer groups using red points respectively. We observe that the optimal vehicle moves from the left-top corner to the right-bottom corner as the purchased vehicles become more luxurious using the LSD features, while the optimal vehicles in the original variable representation show overlap, especially for MAZDA6 and ACURA-TL customers. In other words, we are visualizing what has been shown quantitatively through increased preference prediction accuracy; namely, that optimal vehicles represented using LSD features as opposed to the original variables result in a larger separation of various market segments' optimal vehicles.

The contribution of this demonstration is not the particular introspection on the chosen example with MAZDA6 and ACURA-TL customers. Instead, this demonstration is significant as it suggests it is possible to perform feature-based market segmentation purely using visual analysis. Such visual analysis is likely to be more useful to practicing designers and marketers, as it abstracts away the underlying mathematical mechanics of feature learning.

2.8   Summary

Following our premise from Chapter 1, feature learning is a promising method to improve design preference prediction accuracy without changing the design preference model or the data set. This improvement is obtained by transforming the original variables to a feature space acting as an intermediate step as shown in Figure 2.1. Thus, feature learning complements advances in both data gathering and design preference modeling.

We presented three feature learning methods–principal component analysis, low-rank plus sparse matrix decomposition, and sparse exponential family restricted Boltzmann machines–and applied them to a design preference data set consisting of customer and passenger vehicle variables with heterogeneous unit types, e.g., gender, age, # cylinders.

We then conducted an experiment to measure design preference prediction accuracy involving 1,161,056 data points generated from a real purchase dataset of 5582 customers. The experiment showed that feature learning methods improve preference prediction accuracy by 2-7% for a small and full dataset, respectively. This finding is significant, as it shows that features offer a better representation of the customer's underlying design preferences than the original variables. Moreover, the finding shows that feature learning methods may be successfully applied to design and marketing data sets made up of variables with heterogeneous data types; this is a new result as feature learning methods have primarily been applied on homogeneous data sets made up of variables of the same distribution.

Feature interpretation and visualization offer a promise for using features to support product decisions during the design process. Specifically, interpreting features can give designers deeper insights of the more influential pairings of vehicle features and customer features, while visualization of the feature space can offer deeper insights when performing market segmentation. These new findings suggest opportunities to develop feature learning algorithms that are not only more representative of the customer preference task as measured by prediction accuracy but also easier to interpret and visualize by a domain expert. Methods allowing easier interpretation of features would be valuable when translating the results of more sophisticated feature learning and preference prediction models into actionable design decisions.

CHAPTER III

Quantification of Visual Aesthetics

3.1 Introduction

In this chapter we examine at how we can quantify visual aesthetics. In particular we examine multimodal inputs to modeling that include both words and images.

Aesthetics has been long recognized as a primary factor affecting the success of product design, thus a key task in the design process is to create a design concept that is aesthetically attractive to the target markets. This is one of the most challenging tasks for designers, as it can be hard for them to understand how customers perceive the visual design of a product. Several attempts have been made to assist in objectively measuring customer aesthetics perception and aesthetics preference. Survey is a common approach to quantify aesthetics. Customer responses are collected via rating tasks with a semantic differential such as basic vs. luxury. However, a rating based method is subject to scale difference. For example, a 7 out of 10 luxury score may mean an ultimate luxury product to some respondents, while it may mean an entry luxury product to others. This issue likely adds noise to subsequent design decision-making. In addition, the interview-based approach is also limited, as respondents cannot articulate why they think a design is aesthetically appealing or not (*Nisbett and Wilson*, 1977; *Silvera et al.*, 2002). Designers may use engineering tools such as eye-tracking to obtain sophisticated responses; however, this approach is

not scalable due to being time and resource intensive. Some designers deploy a mathematical model relating the design representation and aesthetics attributes (*Orsborn et al.*, 2009). The design representation consists of design characteristics believed influential for product aesthetics, such as geometric characteristics. The functional form, such as a linear logit form, is explicitly assumed beforehand and followed by estimating the part-worth coefficients of the assumed functional form. These methods heavily rely on the predetermined design representation and functional form. Results are likely plausible if the design representation cannot sufficiently represent all influential visual clues or if the functional form is inappropriate.

Designing an aesthetically attractive form is never an isolated task. The functionality of a product should be jointly considered when modeling the aesthetics preference. Conjoint analysis is widely used for quantifying the relative importance of product features and measuring preference at the individual level (*Green and Srinivasan*, 1990). It would be very useful if conjoint analysis could be used in the scenario where both functional attributes and aesthetics attributes are presented. However, this is a very challenging task, because design concepts are represented by a written description in conjoint analysis. The written description of aesthetics attributes is probably imprecise and unrealistic, hence it is unlikely for respondents to make a reliable decision in conjoint analysis.

In this chapter, we aim to quantify the aesthetics attributes of design and the relative importance of aesthetics attributes when both aesthetics and functions are considered by potential customers. We take a novel approach to avoid the issue of scale difference and any assumption of the functional form. Specifically, rather than rating tasks, we ask respondents to rank the images of several design concepts according to a given aesthetic attribute. This task is more intuitive for human evaluation (*Hubel and Wiesel*, 1962; *Burnap et al.*, 2016a). These rankings are aggregated into the value of the aesthetic attribute through a modified version of PageRank, where the

functional form is implicitly captured in a Markov Chain. According to the value of aesthetics attributes, we select representative design images for each aesthetics attributes and their levels; accordingly, we can use design images to better represent aesthetics attributes in the conjoint analysis. The relative importance of aesthetics can be estimated subsequently.

We test the proposed approach on an SUV aesthetics study. This study covers 373 SUV models designed on the U.S. market. 3,302 respondents participated in the ranking survey and 900 respondents participated in the conjoint analysis. Our results show that we can assess the aesthetics attributes and aesthetics preference over functional attributes of SUV. We can obtain reasonable assessment for aesthetics attributes and their relative importance. A control experiment also demonstrates that aesthetics attributes achieve much higher importance when represented by design images.

The main contribution of this work is to provide designers a novel approach to objectively quantify the aesthetics as well as its relative importance. This approach circumvents the limitations of previous approaches, both in terms of the issue of scale difference in subjective perception of aesthetics and in terms of inappropriate assumptions of functional forms. More importantly, the results offer deep insights to inform design decisions during the design process.

3.2 Related Work

3.2.1 Aesthetics Measurement

Theories about aesthetics date back to Plato, namely, that "Beauty is in the eye of the beholder."(*Dukerich et al.*, 2002) Plato pointed out that the primary challenge in measuring aesthetics is subjectivity. Many attempts have been made to develop engineering design tools to translate this subjective concept in an engineering language

so that designers can improve the aesthetics of product. Examples includes golden ratio, emotional design, and craftsmanship. In addition, researchers also investigate how to quantitatively measure aesthetics. Pioneer works include Birkhoffs mathematical formula for aesthetics (*Birkhoff*, 1933), which defines aesthetics as the ratio of order over complexity. In recent years, design researchers use a utility function to relate the details of form to the aesthetics. Eye-tracking has also been successfully applied to measuring aesthetics, for example, measuring the aesthetics appeal of website through the fixation time (*Du and MacDonald*, 2014; *Reid et al.*, 2010). Our proposed method differs from previous work in that (1) we avoid making any assumption on the functional form for aesthetics; (2) this method is scale to hundreds and thousands of designs that are not feasible for eye-tracking.

3.2.2 Trade-offs Between Aesthetics and Functions

Form means the aesthetics or appearance of a product in the design community. The relationship between form and function is frequently discussed. "Form follows function" is a mainstream principle that originated in modernist architecture. This principle suggests that the shape of an object should primarily relate to its intended function or purpose (*Norman*, 2005). Another point of view is that narrowly applying "form follows function" may preclude product differentiation. The reason is that it is likely that products may be reducible to a single optimal form with the highest functional performance (*Meikle*, 2010). The "MAYA" (Most Advanced Yet Acceptable) principle proposed by designer Loewy suggests that product designs are bounded by the functional constraints of math, materials, and logic, but their acceptance is constrained by social expectations (*Hekkert et al.*, 2003).

Advanced engineering tools have been developed to assist designers in making decisions related to the trade-off between form and function. A common idea of these tools is to use the feedback from potential customers in the decision-making process.

The feedback can be revealed data such as purchase record (as shown in Chap II) or stated data such as responses to surveys (*Orsborn et al.*, 2009). As noted several times before, the most popular tool to model consumers' preference through a set of choice data is conjoint analysis. The proposed approach here to quantify the relative importance of aesthetics attributes is also based on conjoint analysis. The approach here differs from previous work in that aesthetic attribute are represented by design images rather than being described by words. Inclusion of images in a survey has been investigated in some application domains, such as housing preference (*Jansen et al.*, 2009). The effect of including images differs substantially across different application domains. Conclusions from these works cannot be safely generalized into the product design domain. In the work presented here, we investigate the effect of including images in conjoint analysis only in the product design scenario.

3.3 Methodology

3.3.1 Overview

We introduce the general approach for quantifying aesthetics and its relative importance, then detail each stage in the following sections. The overall approach consists of three stages:

1. Stage 1: *Aesthetics data collection.* The first stage is to collect customer aesthetic perceptions for the predefined aesthetics attributes. Respondents are asked to rank several randomly selected design images along with one randomly selected semantic differential such as 'Sporty' vs. 'Conservative'. Those rankings are further analyzed in Stage 2. Detailed experimental settings are discussed in section 3.4.1

2. Stage 2: Quantify aesthetics attributes. The second stage is to quantify aesthetics attributes for each design images using the rankings collected in the

first stage. Specifically, the rankings are converted into the value of aesthetics attributes via a modified Pagerank method, which is described in section 3.3.2. Each design image receives a score for each aesthetics attribute.

3. Stage 3: *Quantify the relative importance.* The third stage is to quantify the relative importance of aesthetics via conjoint analysis with images. Those design images with an extreme value of a given aesthetic attribute are selected as the representative images for the corresponding level of that attribute. The product profiles in our conjoint analysis consist of design images representing the aesthetics attributes as well as textual description of functional attributes such as "transmission". The relative importance is quantified using the choice data in conjoint analysis as described in section 3.3.3

3.3.2 Quantifying Aesthetics Attributes Using a Modified Pagerank Algorithm

The method that we use to quantify aesthetics attributes modifies the Pagerank algorithm, which was originally developed to determine the relative importance of web pages. We extend this algorithm so that it can quantify the relative aesthetics appeal of designs. The underlying assumption of the proposed algorithm is that more aesthetically appealing designs are likely to be ranked higher than other designs. Another assumption is that if a design is ranked higher than an aesthetically appealing design, then that design should be more aesthetically appealing.

We next formalize the rankings, designs, and scores.

Assume that there are N design images. In *Stage 1*, we ask respondents to rank several randomly selected design images. No matter how many design images are ranked at one time, we can always convert the original ranking into several binary rankings. For example, the respondents rank three design images (say design image A, B, and C) at one time. We can convert the original ranking (A > B > C) into 3

binary rankings (A > B, A > C and B > C). To simplify the notation, the rankings in the rest of this chapter refer to binary rankings.

Assume that attribute value for $i-th$ design is s_i. To extract these attribute values, rather than simply aggregate these rankings, we weigh the rankings in a recursive formula:

$$s_i = \sum_{j=1,2,\ldots,N,j\neq i} s_j \frac{\#\{i > j\}}{\#\{i' > j, i' = 1, 2, \ldots, N\}} \tag{3.1}$$

where $\#\{i > j\}$ denotes the number of rankings that design i is ranked higher than design j, and $\#\{i' > j, i' = 1, 2, \ldots, N\}$ denotes the number of rankings that design j is ranked lower than any design i.

Intuitively, if design i is frequently ranked higher than design j, i.e., $\#\{i > j\}$ is large, then it means the comparison results $\{i > j\}$ is reliable; therefore, it should be given more weight. If design j is frequently ranked lower than the competing designs, i.e., # {i'¿j, i' = 1,2,..., N} is large, then it means design j should have a relatively lower value; therefore, it should be given less weight.

Denote the weight matrix as $\mathbf{P}$, whose element P_{ij} is:

$$P_{ij} = \frac{\#\{i > j\}}{\#\{i' > j, i' = 1, 2, \ldots, N\}} \tag{3.2}$$

And the score vector $S = (s_1, s_2, \ldots, s_N)$

Rewrite the recursive formula in matrix notation:

$$\mathbf{S} = \mathbf{SP} \tag{3.3}$$

Now we interpret the formula in matrix notation from the perspective of a Markov Chain. Markov Chain is a stochastic model describing a sequence of possible events in which the probability of each event depends only on the state attained in the

previous event (*Gagniuc*, 2017). The proposed recursive formula can be regarded as a Markov Chain with the states as designs, and with transition probability matrix $\mathbf{P}$. Intuitively, we can describe this Markov Chain as an agent jumping from the current design to another with a higher attribute value according to certain probability $\mathbf{P}$. According to Markov Chain theory, the vector S in equation (3.3) is referred as the stationary distribution of this Markov Chain. The attribute value of a design is the probability of choosing that design after a large number of comparisons.

However, only when a Markov Chain satisfies specific properties, its stationary distribution $\mathbf{S}$ exists and is unique. To achieve uniqueness in the stationary distribution, we make two modifications to convert the raw transition matrix P to a stochastic, irreducible, and aperiodic matrix (*Brin and Page*, 2012).

First, the rows in P containing only 0's are replaced with $\frac{1}{D}\mathbf{e}^T$, where $\mathbf{e}^T$ is a column vector consisting of 1's, and T denotes the transpose operator. This modification results in a stochastic matrix denoted $\mathbf{P}_s$ as given:

$$\mathbf{P}_s = \mathbf{P} + \mathbf{Q}(\frac{1}{D}\mathbf{e}^T) \tag{3.4}$$

where $\mathbf{Q}_i = 1$ if $\mathbf{P}_i = \mathbf{0}$ and $\mathbf{Q}_i = 0$ otherwise.

To convert $\mathbf{P}_s$ into an irreducible and aperiodic matrix $\mathbf{P}_g$, we add a dumping factor $\gamma \in (0, 1)$:

$$\mathbf{P}_g = \gamma\mathbf{P}_s + (1 - \gamma)(\frac{1}{D}\mathbf{e}\mathbf{e}^T) \tag{3.5}$$

With these modifications, there exists a unique stationary distribution S for $\mathbf{P}_g$ and this vector S will be the scores of aesthetics attributes for designs. The modified recursive formula is:

$$\mathbf{S} = \mathbf{S}\mathbf{P}_g \tag{3.6}$$

Second, the aesthetics attribute values can be computed algebraically or through

63

an iterative method. According to Markove Chain theory, the score vector S in equation 3.6 is the eigenvector of matrix $\mathbf{P}_g$. The corresponding eigenvalue is one. Since the matrix $\mathbf{P}_g$ is stochastic, irreducible, and aperiodic, there exists a unique eigenvector with eigenvalue one. Computing the score vector through eigenvalue decomposition is straightforward; however, the computational complexity of the composition is about $\mathcal{O}(n^3)$. Hence when there are many design images, computing the scores algebraically is expensive. In this case, the iterative method, which is also referred as the power method, is an efficient solution. Recalled that the score vector is the stationary distribution that remains unchanged in the Markov Chain as time progresses, we can calculate S using the iterative method described below:

Starting with an arbitary vector S_0. the operator $\mathbf{P}_g$ is applied in succession. i.e.,

$$\mathbf{S}_{t+1} = \mathbf{S}_t \mathbf{P}_g \tag{3.7}$$

until there is little change between iterations. i.e.,

$$|\mathbf{S}_{t+1} - \mathbf{S}_t| < \epsilon \tag{3.8}$$

where ϵ is a predefined stopping criteria.

3.3.3 Quantify Aesthetics Preference Using Conjoint Analysis with Images

Conjoint analysis asks respondents to state preferences for several product profiles and then chooses the most preferred product profile. Each profile is described by several attribute-levels (e.g. attribute: drivetrain, level: all-wheel-drive). Figure 3.1 shows an example of a conjoint task with written description for functional attributes.

In this section, we introduce how we design the conjoint task so that we can mea-

If these were your only options, which would you choose?

	Profile 1	Profile 2	Profile 3	Profile 4	
Drivetrain	All-wheel-drive	Front-wheel-drive	Front-wheel-drive	All-wheel-drive	NONE: I wouldn't choose any of these.
Fuel efficiency	30 mpg	40 mpg	40 mpg	20 mpg	
Engine type	6 cylinder gas engine	8 cylinder gas engine	6 cylinder hybrid	4 cylinder gas engine	
Wheel size	18 inch	18 inch	16 inch	20inch	
Sunroof	No sunroof	Panoramic roof	Pop-up roof	Panoramic roof	
Price	$25,000	$35,000	$45,000	$35,000	
	SELECT	SELECT	SELECT	SELECT	SELECT

Figure 3.1: An example of conjoint task with written description of functional attributes

sure the joint aesthetics and function preference. To include aesthetics attributes in conjoint tasks, one can simply add a written description of aesthetics attributes in the product profile as shown in Figure 3.2. However, there are several drawbacks: First, styling attributes may be difficult to describe precisely in a few words. Second, a written description of styling may be inadequate for respondents to understand the difference among options. Third, styling attribute is a subjective concept; thus respondents are likely to have different interpretation and perception of a given written description. For example, some respondents may consider the styling "luxurious" as the luxurious level of the top premium cars such as Rolls-Royce, while other respondents may consider the styling "luxurious" as the level of mid-class cars such as BMW. Our approach is to use design images as the description of aesthetics attributes. Respondents will see the product profile consisting of written descriptions of functional attributes and images visualizing the aesthetics attributes. Figure 3.3 shows an example of such a conjoint task.

There are several benefits of using images to describe the aesthetics attributes. First, though styling attributes may be difficult to describe in a few words, such attributes may be easily visualized by a single image. Second, images may stimulate respondents' awareness of styling attributes, so that respondents may better understand and appreciate different options and thus may make better choices. Third,

If these were your only options, which would you choose?

	Profile 1	Profile 2	Profile 3	Profile 4	
Aesthetics	Luxurious	Basic	Luxurious	Basic	NONE: I wouldn't choose any of these.
	Traditional	Traditional	Innovative	Innovative	
	Sporty	Conservative	Conservative	Sporty	
Drivetrain	All-wheel-drive	Front-wheel-drive	Front-wheel-drive	All-wheel-drive	
Fuel efficiency	30 mpg	40 mpg	40 mpg	20 mpg	
Engine type	6 cylinder gas engine	8 cylinder gas engine	6 cylinder hybrid	4 cylinder gas engine	
Wheel size	18 inch	18 inch	16 inch	20inch	
Sunroof	No sunroof	Panoramic roof	Pop-up roof	Panoramic roof	
Price	$25,000	$35,000	$45,000	$35,000	
	SELECT	SELECT	SELECT	SELECT	SELECT

Figure 3.2: An example of conjoint task with written description of functional attributes and aesthetics attributes

If these were your only options, which would you choose?

	Profile 1	Profile 2	Profile 3	Profile 4	
Aesthetics					NONE: I wouldn't choose any of these.
Drivetrain	All-wheel-drive	Front-wheel-drive	Front-wheel-drive	All-wheel-drive	
Fuel efficiency	30 mpg	40 mpg	40 mpg	20 mpg	
Engine type	6 cylinder gas engine	8 cylinder gas engine	6 cylinder hybrid	4 cylinder gas engine	
Wheel size	18 inch	18 inch	16 inch	20inch	
Sunroof	No sunroof	Panoramic roof	Pop-up roof	Panoramic roof	
Price	$25,000	$35,000	$45,000	$35,000	
	SELECT	SELECT	SELECT	SELECT	SELECT

Figure 3.3: An example of conjoint task with written description of functional attributes and image description of aesthetics attributes

respondents' perception of styling attributes may be more homogeneous when the attributes are represented by images as it is less open to respondents' interpretation when describing the styling attributes by written text. More importantly, including the vehicle images may enhance the realism of the conjoint tasks as such tasks are more similar to the scenario when respondents are making a *real* purchase decision.

We choose these design images according to the aesthetics value that we obtained in Stage 2. For a given aesthetics attribute level, we choose a group of candidate design images whose value of that aesthetics attributes-level is high. We also try to select the candidate design images that differ substantially from others in the group.

For example, we try to include candidate design images from different brands. In this way, we can reduce the possible confounding effect from unexpected or uncontrolled factors. Respondents will be shown only one randomly selected design image for the corresponding attribute-level. In the meantime, the written description of functional attributes remains the same.

Conjoint survey results in a set of choice data. Using this data, we can establish a discrete choice model to describe respondents' preference quantitatively, then conduct further analysis. The discrete choice model is built on utility theory. Utility is a ubiquitous concept in economics as an abstract measurement of the degree of goal-attainment or want-satisfaction provided by a product or service. We cannot directly measure how much utility a person may gain from a product; however, we can make inferences about utility based on the persons behavior if we presume that people act rationally, which means that people choose the product with higher utility. In random utility models we assume that the utility u_{ij} provided to individual i by product j is composed of a deterministic component v_{ij}, which can be calculated based on observed characteristics, and a stochastic error component ϵ_{ij}, which is unobserved,

$$u_{ij} = v_{ij} + \epsilon_{ij} \tag{3.9}$$

The total value of product j for respondent i is v_{ij}. In conjoint analysis, we assume that this value can be discomposed. A product or service can be viewed as a bundle of its attribute levels. If we understand how consumers value different attribute level of a product, we can derive the value of various products that are created based on some combinations of these features. The total value equals the sum of sub-values of its attribute levels.

$$v_{ij} = \sum_{k=1}^{K} \sum_{l=1}^{L_k} \beta_{ikl} z_{ijkl} \tag{3.10}$$

where z_{ijkl} is dummy variables, k is the index of attributes, and l is the index of at-

tribute level. β_{ikl} is called as part-worth, which measures how much the corresponding attribute-level influences the customers' decision. The discrete choice model in conjoint analysis follows the logit choice rule. The probability of respondent i choosing product j is P_{ij}

$$P_{ij} = \frac{\exp(v_{ij})}{\sum_k \exp(v_{ik})} \tag{3.11}$$

Such a discrete choice model can provide many insights on customers' preference. In this work, we focus on characterizing the relative importance of each attribute. Specifically, we consider how much difference each attribute could make in the total utility of a product. The relative importance of attribute k for customer i is denoted as $\mathbf{I}_{ik}$ and is defined below:

$$\mathbf{I}_{ik} = \frac{\max_l(\beta_{ikl}) - \min_l(\beta_{ikl})}{\sum_k \left(\max_l(\beta_{ikl}) - \min_l(\beta_{ikl})\right)} \tag{3.12}$$

3.4 Experiments

In this section, we describe three experiments. Experiment I details how to crowd-source ranking responses as described in Stage 1. Experiment II details how to quantify aeshtehtics attributes using the ranking responses collected in Experiment I. Experiment III shows how to design a conjoint analysis to quantfiy the joint preference using the aesthetics attributes obtained in Experiment II. In these experiments, we focused on Sport Utility Vehicle (SUV). We tried to quantify 3 pairs of semantic differentials: sporty vs. conservative, luxurious vs. basic, innovative vs. traditional.

3.4.1 Experiment I: Crowdsourced Ranking Responses

The ranking responses were collected through a crowdsourcing web application. In this web application, respondents were asked to rank four randomly selected SUV images from the same viewpoint along one randomly selected semantic differential such
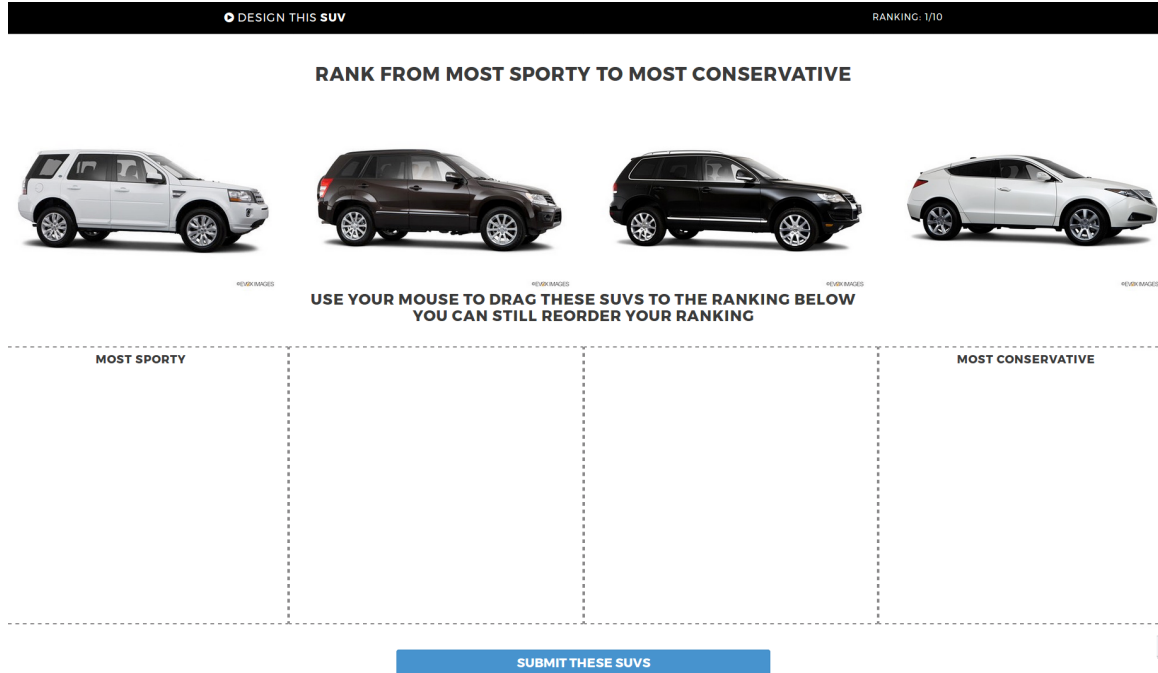
Figure 3.4: A snapshot of the ranking page in the crowdsourcing web application.

as 'Sporty' vs. 'Conservative'. Each respondent was asked to complete 10 rankings with different images but the same semantic differential. This experiment included images of 373 SUVs which were designed on the U.S. market from 2010 to 2014. We collected the responses from 3,302 respondents who had bought SUV in the past 5 years. Figure 3.4 shows a snapshot of the ranking page in the crowdsourcing web application. A more detailed description of this experiment can be found in Chapter V. The statistics of customer demographic data show that this experiment recruited respondents from a broad distribution, suggesting it well covered the potential customer space.

3.4.2 Experiment II: Quantify Aesthetics Attributes

In this experiment, we first coverted the rankings collected in Experiment I into binary rankings. Then, for each aesthetics attribute, we calculated the value of the given attribute of SUV images using the method proposed in section 3.3.2. In our

experiment, we set the dumping facor γ in Equation 3.5 at 0.85.

3.4.3 Experimental III: Quantify the joint aesthetics and function preference

In this experiment, we demonstrated how to assess customer's joint aesthetics and functions preference. In addition, we investigated how this preference changed when including the same product functional and styling attributes but representing the aesthetics attributes in different ways, i.e., written text vs. vehicle images in conjoint analysis.

We recruited 900 respondents who had bought SUV in the past five years. Respondents were randomly divided into two groups. One group took the conjoint survey with the textual description of both functional attributes as well as styling attributes. The other group took the conjoint survey with the same textual description for functional attributes but with the aesthetics attribute represented by a SUV image. The represented SUV images were selected according to value of the given aesthetics attribute, which was obtained in experiment II. A total of 8 functional attributes as well as *price* are included in this experiment. Those functional attributes are believed to be the most important functional attributes according to internal confidential study at General Motors.

3.5 Results and Discussion

3.5.1 Product Aesthetics Measurement

Figure 3.5 shows the aesthetics values that we obtained using the method proposed in section 3.3.2 for all 373 SUVs. As a sanity check, figure 3.6 lists the top 10 SUVs for each aesthetics attribute. The brand image of these SUVs match well with their aesthetics values. Moreover, the plots of sorted aesthetics values indicate

that most SUVs have an aesthetics value in the medium range, while a few of them achieve a distinguished high or low aesthetics value. This pattern may agree with a general design objective that the product should be aesthetically attractive to its own target segment, instead of being distinct in every dimension. For example, Cadillac Escalade targets at the luxury SUV market; thus, it manages to look luxurious but not necessarily sporty or innovative.

3.5.2 Relative Importance of Aesthetics

Figure 3.7 shows the relative importance of aesthetics and function attributes in the conjoint analysis whose aesthetics attribute is represented by design images. As shown in this figure, aesthetics attribute is the second most important attribute. This was also suggested in qualitative studies such as JD Power Initial Quality Study (*Tews*, 2016), J.D. Power APEAL Study (*Dobrian*, 2016). However, the aesthetics attribute is only the fourth most important attribute when the aesthetics attribute is represented by textual description as shown in Figure 3.8. This difference indicates that images have an impact on respondents' preferences.

Though there are various benefits of representing aesthetics attribute using images as discussed in section 3.3.3 and demonstrated through our SUV study, previous research show that inclusion of images in a questionnaire may also have drawbacks. The primary drawback are accidental and non-systematically varied details in images (for example, lighting condition, shadows, and background color) (*Jansen et al.*, 2009). To overcome this drawback, we use the design images taken in the same photograph studio to control the lighting condition etc.. In addition, we reduce the influence from accidental varied details by having a group of candidate design images per attribute level. We intentionally choose the candidate images that differ from others within one group, especially different brand, which has long been recognized as an influential factor for vehicle preference. More importantly, studies have shown that

many respondents can recognize the brand of the vehicle even though the image is completely debranded. To further investigate the possible influence from brand, we included the brand as an attribute in our utility model using the same choice data as experiment III. Figure 3.9 shows that brand is the second most important attribute. While the relative importance of aesthetics is reduced, aesthetics is still the third most important attribute. This indicates that even though brand is an influential factor for customers' preference, aesthetics still influence customers' preference. However, the effect of brand and aesthetics cannot be fully disentangled in the market segment where little product differentiation exists. For example, Jeep dominates the market of sporty, basic, and traditional SUV. Both Jeep brand as well as the sporty, basic, and traditional aesthetics level received a negative partworth, hence, brand may be the confounding factor for the negative partworth of this aesthetics level.

3.6   Summary

Aesthetics is a critical important factor for the success of product design. It is challenging for designers to predict whether the design concept is aesthetically appealing for its potential customers. More importantly, designers need to understand how customers make trade-offs between design aesthetics and functionality of the product so that the design concept can be adjusted accordingly.

We introduced a new approach to measure the product aesthetics and the relative importance of aesthetics. Specifically, we built on the Pagerank method and developed a Markov Chain-based ranking aggregation method. This methods converts a partial rankings of design images collected from respondents into a numerical assessment of given aesthetics attribute for each design image. Then, we selected design images to represent aesthetics attributes in a conjoint analysis. Functional attributes are also included in this conjoint analysis. In this way, we enable quantification of the relative importance of aesthetics attributes as well as the joint preference of aesthetics and

functions.

Experiments were conducted to test this approach. We included design images for 373 Sport Utility Vehicles as well as over 4,000 respondents. Our results showed that this research approach is indeed able to obtain reasonable aesthetics attribute values. Further, we showed the aesthetics attributes are the second most important attribute for SUVs.

Future research could address ways to further increase the realism and reduce the possible confounding effect of the conjoint analysis for product aesthetics. Instead of using images, artistic works may be a compromise between increasing the realism and reducing the confounding factors. In addition, virtual reality and augmented reality techniques *Berg and Vance* (2017) may have the potential to make respondents feel like they are making a real purchase decision. Another promising direction is to include the design images as the input of preference function. In this way, the visual information of a design can be well preserved in the images data rather than being lost due to atomization in conjoint analysis where a complex form can only be represented with lesser features. To include design images in the preference function requires efforts on developing mathematical tools to relate the design images with customer aesthetics preference. Pioneering works in this direction include (*Pan et al.*, 2017), which manage to predict the aesthetics preferences and interpret it, but haven't been applied to model the joint preference over aesthetics and functions.
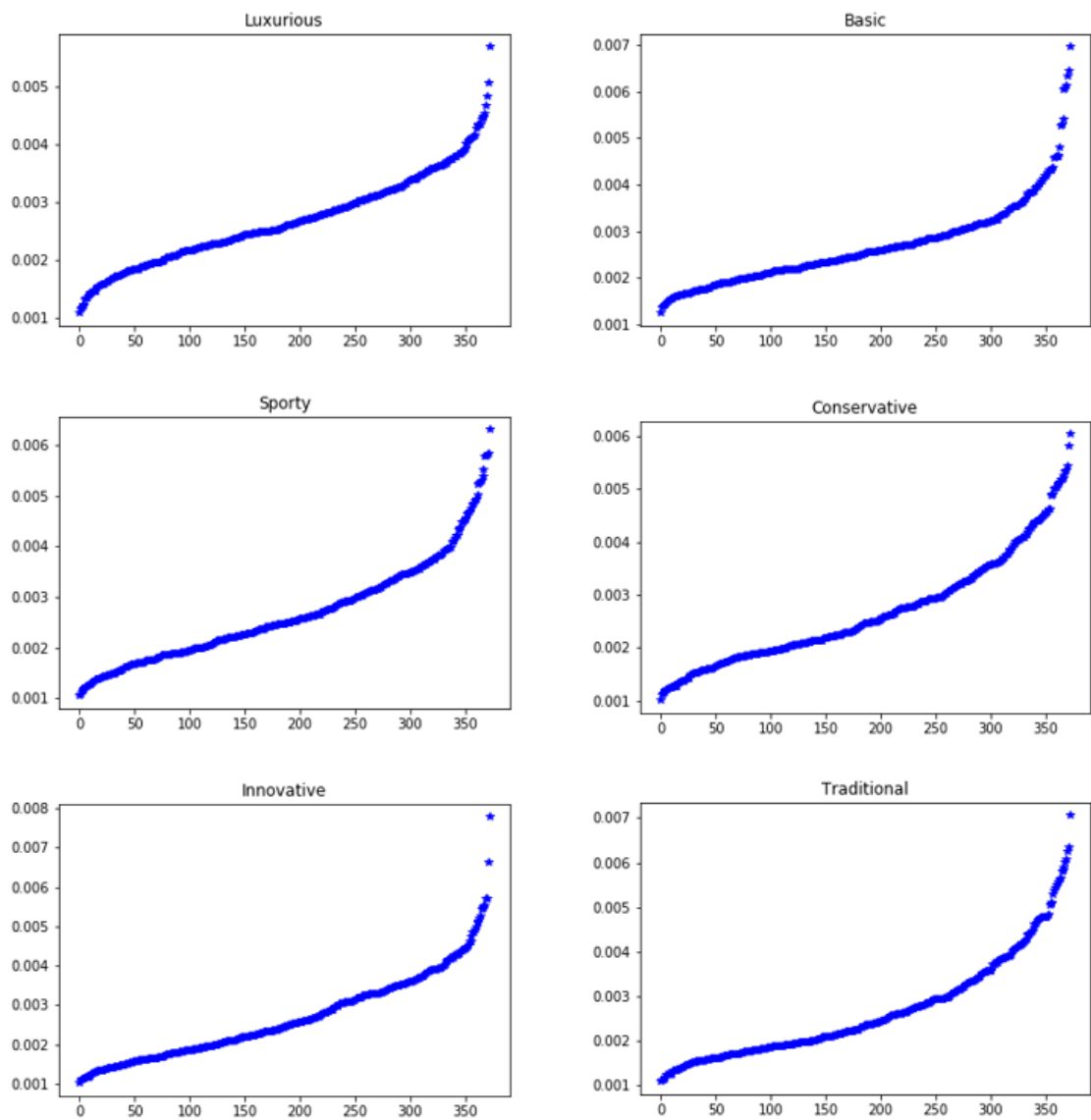
Figure 3.5: Sorted Aesthetics Values

	Attribute-level: Luxurious		
	Year	Make	Model
1	2010	Cadillac	Escalade
2	2013	Land Rover	Range Rover Evoque
3	2012	Land Rover	Range Rover Evoque Coupe
4	2012	Cadillac	Escalade ESV
5	2012	Mercedes-Benz	R-class
6	2014	Cadillac	SRX
7	2011	Cadillac	Escalade
8	2011	Mercedes-Benz	M-class hybrid
9	2011	Cadillac	SRX
10	2013	Cadillac	Escalade hybrid

	Attribute-level: Basic		
	Year	Make	Model
1	2011	Jeep	Wrangler
2	2010	Jeep	Wrangler
3	2012	Jeep	Wrangler
4	2013	Jeep	Wrangler
5	2014	Jeep	Wrangler
6	2010	Jeep	Patriot
7	2010	Jeep	Wrangler Unlimited
8	2012	Jeep	Liberty
9	2014	Jeep	Liberty
10	2012	Subaru	Forester

	Attribute-level: Sporty		
	Year	Make	Model
1	2013	BMW	X6
2	2010	Acura	ZDX
3	2011	Infiniti	FX
4	2012	Land Rover	Range Rover Evoque Coupe
5	2014	Nissan	Juke
6	2011	BMW	X6
7	2013	Land Rover	Range Rover Evoque
8	2012	BMW	X6
9	2011	Acura	ZDX
10	2011	Infiniti	FX

	Attribute-level: Conservative		
	Year	Make	Model
1	2012	GMC	Yukon XL
2	2010	Ford	Flex
3	2012	Ford	Expedition EL
4	2014	Chevrolet	Suburban
5	2014	Ford	Flex
6	2010	Ford	Expedition
7	2014	GMC	Yukon XL
8	2013	Chevrolet	Suburban
9	2012	Ford	Flex
10	2010	Ford	Expedition EL

	Attribute-level: Innovative		
	Year	Make	Model
1	2013	Land Rover	Range Rover Evoque Coupe
2	2010	Infiniti	FX
3	2014	Nissan	Juke
4	2012	Porsche	Cayenne Hybrid
5	2013	Infiniti	FX
6	2011	Nissan	Juke
7	2011	Acura	ZDX
8	2014	Infiniti	QX70
9	2012	Acura	ZDX
10	2012	Nissan	Juke

	Attribute-level: Traditional		
	Year	Make	Model
1	2014	Jeep	Wrangler
2	2013	Jeep	Wrangler
3	2011	Jeep	Liberty
4	2012	Jeep	Wrangler
5	2013	Jeep	Patriot
6	2010	Jeep	Wrangler
7	2011	Jeep	Wrangler
8	2013	Ford	Expedition EL
9	2010	Jeep	Patriot
10	2011	Jeep	Wrangler Unlimited

Figure 3.6: Top 10 SUVs for each aesthetics attribute levels

Attribute Importance - With Images

- Price
- Aesthetics attribute
- Functional attribute 1
- Functional attribute 2
- Functional attribute 3
- Functional attribute 4
- Functional attribute 5
- Functional attribute 6
- Functional attribute 7
- Functional attribute 8
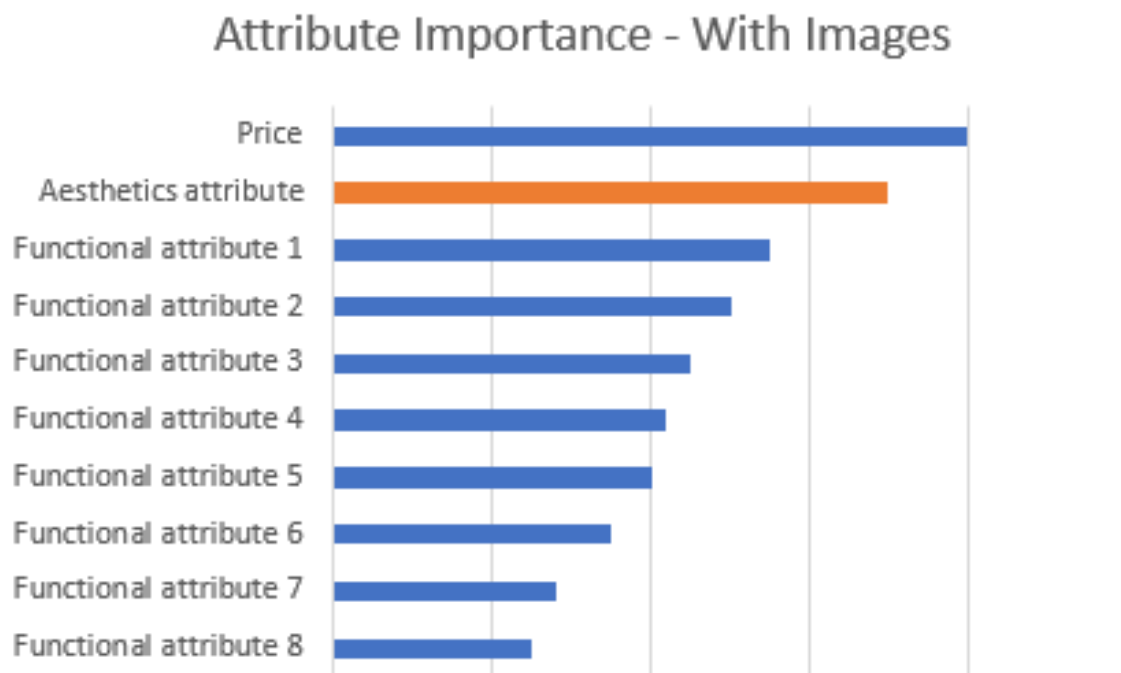
Figure 3.7: The relative importance of attribute in conjoint analysis with images. The value of relative importance is hidden due to the Intelligence Properties Protection for General Motors
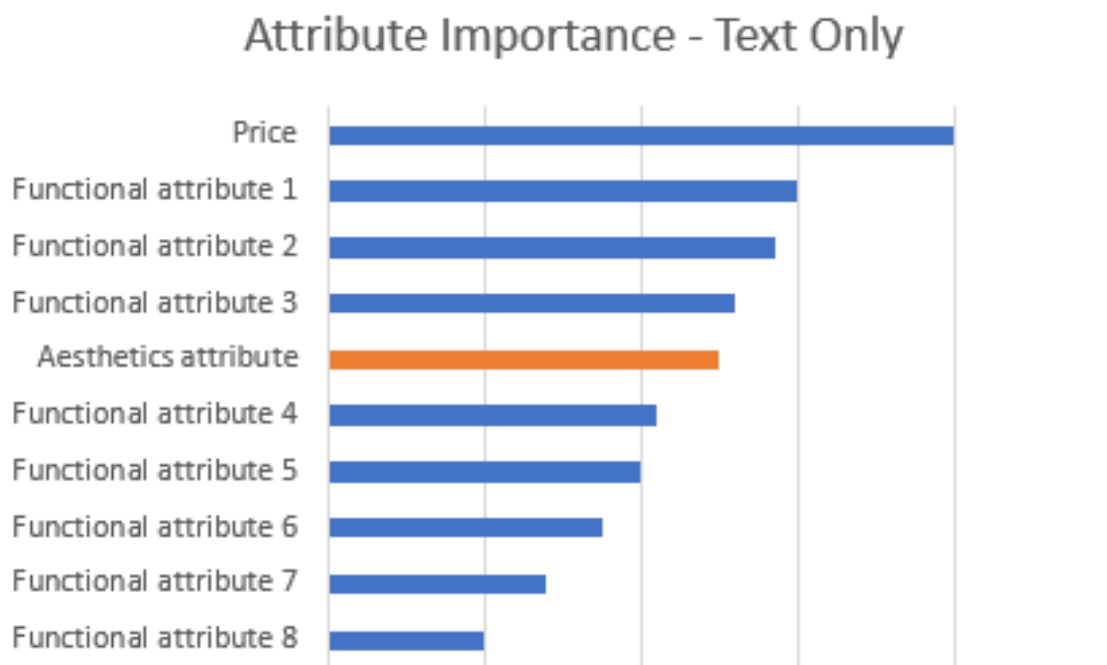
Figure 3.8: The relative importance of attribute in conjoint analysis with textual description. The value of relative importance is hidden due to the Intelligence Properties Protection for General Motors
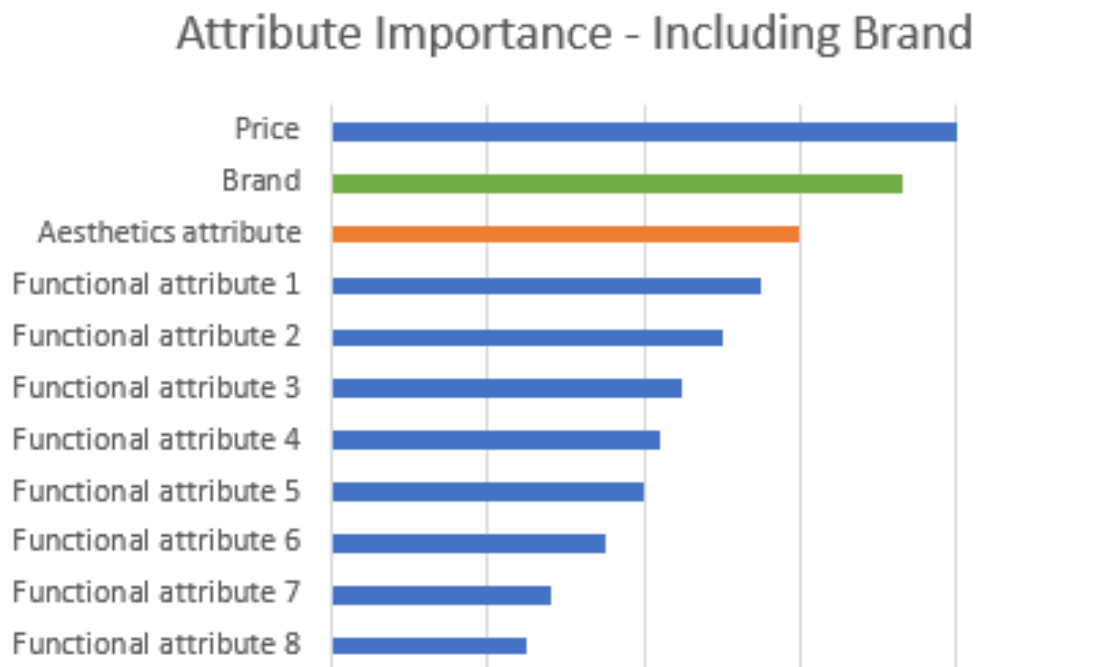
Figure 3.9: The relative importance of attribute in conjoint analysis when including brand. The value of relative importance is hidden due to the Intelligence Properties Protection for General Motors

CHAPTER IV

Identifying Design Regions of Visual Attraction

4.1 Introduction

In this chapter, we investigate on how we can predict visual aesthetics attribute and what are the possible factors affecting customers' perception of product aesthetic appeal.The aesthetic appeal of designed artifacts has been long recognized as significantly affecting customer preferences; examples include golden section proportions, Gestalt Psychology, form versus function, emotional design and craftsmanship. Both practicing designers and design researchers have focused on this important topic, see, e.g., (*Coates*, 2003; *Norman*, 2005; *Orsborn et al.*, 2009; *Reid et al.*, 2010, 2013). This notion is particularly true for increasingly commoditized products such as automobiles, as standardization across product components and manufacturing processes are pushing product differentiation to moreso to perceptual attributes such as aesthetic styling and corresponding visual attraction (*Bloch*, 1995; *Moulson and Sproles*, 2000).

To better understand the factors affecting visual attraction, we extend previous research on both descriptive and predictive aspects of aesthetic appeal. Descriptive studies of aesthetic appeal have examined the saliency of design features and their propensity to draw perceptual attention (*Crilly et al.*, 2004). Berlynes theory of appeal, for example, aggregates sensory information and models aesthetic appeal by balancing novelty and arousal and trading off meaning and recognition (*Berlyne*,

79
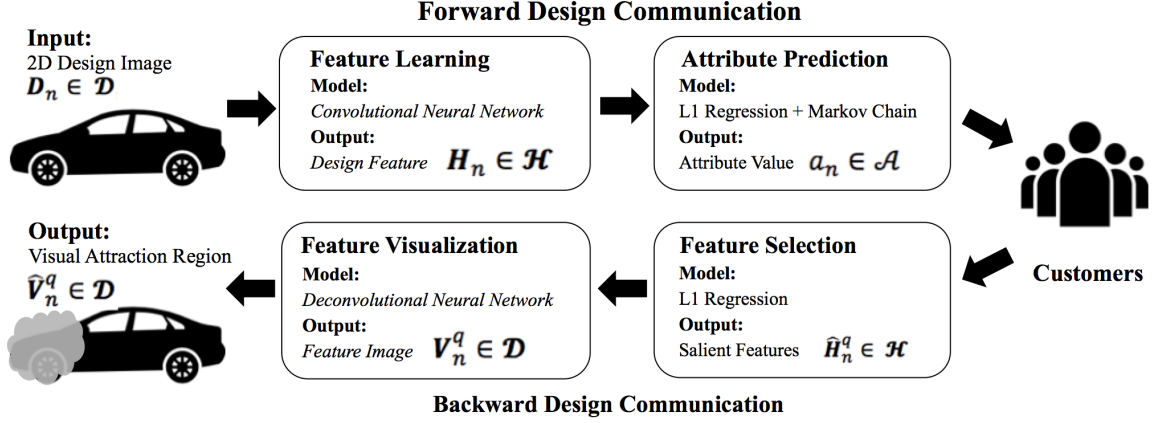
Figure 4.1: Overview of design process using the proposed quantitative communication model. The goal is to predict a region of visual attraction, denoted in grey given a particular design.

1971). The designers intent is often focused on actively drawing visual attention to salient regions of a design in a communication between designers and customers (*Crilly et al.*, 2004; *Monö et al.*, 1997). This communication may begin with a pleasant initial impression of the customers due to the attractive appearance of the design and cement that impression by expressing attributes important to them (*Norman*, 2005).

Predictive studies of factors that affect aesthetic appeal model which design features evoke particular visual design attributes. Linear models of forward communication such as conjoint analysis (*Ben-Akiva et al.*, 1999) and Kansei engineering (*Nagamachi*, 1995) have been used to capture and predict design attributes as functions of design features. These models may use design features implicitly learned (*Orsborn et al.*, 2009; *Ren et al.*, 2013), hand-crafted features (*Kelly et al.*, 2011; *Orsborn et al.*, 2009; *Petiot and Dagher*, 2011; *Reid et al.*, 2010), or learned through dimensionality reduction (*Yumer et al.*, 2015). Another approach is to use eye-tracking methods where the subjects gaze and fixation time to a given design stimuli are measured and correlated to behavioural information such as consumer choice (*Du and MacDonald*, 2014; *Marshall et al.*, 2014; *Reid et al.*, 2013).

To quantiatively capture these descriptive and predictive factors of aesthetic ap-

peal and visual attention, we adopt the framework of design as a communication between designers and customers. This framework suggests that design communication occurs from designer to customer, hereafter referred to as forward communication. We extend this framework to include communcation from customer to designer, or a backward communication direction of customer response. This forward- backward design communication concept shown in Figure 4.1 draws heavily from previous literature butthe formalism introduced in this paper is novel. In the forward communication direction, the high dimensionality of realistic design representations and complexity of the nonlinear mapping between this representation and an attribute value creates a challenging statistical estimation problem (*Burnap et al.*, 2015). Sophisticated nonlinear models have been developed to model this process with high accuracy, such as kernel methods (*Ren et al.*, 2013) and feature learning (*Burnap et al.*, 2016c). With nonlinear models, predictive performance of the underlying physics is significantly improved at the cost of reduced interpretability. With linear models, interpretability is often possible but predictive power is relatively poor due to assumptions that typically do not hold, such as linearity, feature independence, homogeneity, and complex noise distributions.

In the backward communication direction, inverting a nonlinear function to model backward communication poses significant challenges. The backward process is often quantified using experiment-based approaches such as eye-tracking and stated responses (*Duchowski*, 2002; *Chang et al.*, 2013; *Du and MacDonald*, 2014; *Marshall et al.*, 2014). These approaches work well in analysing overall aesthetic performance, but do not typically provide information about each aesthetic attribute separately. Moreover, these backward approaches do not currently use information from the forward communication.

Motivated by a collaboration with practicing automotive designers, our research goal is to capture this forward and backward communication by identifying regions of

a design that draw visual attention. We introduce a data-driven method to simultaneously quantify both the forward and the backward communication. This method does not require humans to directly provide attention data, instead this method predicts the attention region in the given design from humans feedback on its attribute values in four stages: (i) feature learning, (ii) attribute prediction, (iii) feature selection, and (iv) feature visualization. The resulting mathematical model has three goals: (i) assess aesthetic attributes based on the design representation (in our application study these are pixel-based 2D images); (ii) invert the nonlinear model to predict corresponding attention region; and (iii) leverage useful information from both communication directions. The modeling tools employed consist of a convolutional neural network, L1 regression, a crowdsourced ranking Markov chain, and a deconvolutional neural network. The four data sources we use for modeling are summarized in Table 4.1. We conducted an experiment involving four steps: (i) learn design features of 2D car images through a convolutional neural network trained by ImageNet (*Deng et al.*, 2009) and Flickr (*Karayev et al.*, 2013) data sets; (ii) use L1 regression to model the relation between the design features and the design attribute values determined by a crowdsourced ranking Markov chain; (iii) determine salient features according to the L1 regression model; and (iv) determine visual attraction regions by visualizing the selected salient features using a deconvolutional neural network. The L1 regression was chosen to introduce sparcity thus reducing the complexity of the number of design features needed to relate to the design attributes. The major contribution of this research is the the extension of previous quantification of forward only design-customer communication to a combined forward-backward communication using a purely data-driven approach and multiple large-scale data resources. The purely data-driven approach can also be used alongside existing methods such as eye tracking and dimensionality reduction of 2D and 3D designs.

Table 4.1: Description of the four data sets used in this work

Dataset	ImageNet	Flickr	Vehicle Images	Design Attribute Ranking
Num. of Data	15,000,000+	80,000	110	5,054
Annotation	Object Name	Style Tags	Vehicle Make and Model	Design Attribute
Source	Open Source	Open Source	Search Engine	Crowdsourcing

4.2 Related Work

We build on literature from visual attention studies from art and product design, and data features for representing 2D images from biology, computer vision, and the design community.

4.2.1 Visual attention in design

Perhaps the earliest experimentally recorded investigation of design attention was conducted to analyse regions of eye-gaze fixation of 55 artistic pictures by 200 participants (*Buswell*, 1935). Such eye-tracking approaches have been successful in optimizing the layout of product placement, advertisements, and labelling objects in supermarkets. Readers are referred to (*Duchowski*, 2002) for a comprehensive and crossdisciplinary review of eye-tracking research. Recently, these methods have been applied to design research, including vehicle face attribute assessment with Kansei engineering (*Chang et al.*, 2013), design representation comparison (*Reid et al.*, 2013), relations with vehicle face component size changes (*Du and MacDonald*, 2014), and technical diagram assessment (*Ruckpaul et al.*, 2015).

4.2.2 Data features for design representation

We review data features from several perspectives: biology, computer vision, and design. A feature is a general term for a function of the underlying design variables, used to represent the design at a particular level of fidelity. For example, complex 3D

meshes underlying a realistic design concept can also be represented by a set of control points. This set of control points may be a more efficient the feature representation of the realistic design concept, as it is able to preserve the important design information in the space with lower dimensions (*Ren et al.*, 2013; *Yumer et al.*, 2015).

At a neurophysiological level, visual attention can be modeled in a bottom-up fashion according to perception pathways (*Hubel and Wiesel*, 1962). Such pathways are analogous to the forward direction of our model, from the 2D design image space to the design attribute space, see Figure 4.1. Similarities have been shown between neural network data features and Gabor features (*Marĉelja*, 1980) known to model visual cortex V1 and V2 cell receptive fields (*Lee et al.*, 2008).

There is vast amount of foundational and ongoing work from computer vision researchers on hand-crafted image features and implicitly-learned (i.e., learned purely from data) image features. Hand-crafted features tend to outperform implicitly-learned features due to the reduction in the uncertainty of the true data-generating mechanism. For example, features learned for face recognition take advantage of facial symmetry and facts such as two eyes are separated by a nose and mouth. One the other hand, implicitly-learned features in so-called feature extraction tend to be more general for a variety of tasks. Such features include HOG features (*Dalal and Triggs*, 2005), and features learned in convolutional neural network (*Krizhevsky et al.*, 2012).

There are data features specific to design, for example, in investigating how design attributes vary according to corresponding variability in a design representation. These design representations may be hand-crafted, such as a set of parametric handles to manipulate vehicle silhouettes (*Petiot and Dagher*, 2011; *Poirson et al.*, 2013; *Reid et al.*, 2010). These design representations have also been created implicitly using finite shape grammars that together form more complex representations (*McCormack et al.*, 2004; *Orsborn et al.*, 2006; *Pugliese and Cagan*, 2002). Hybrid approaches that learn the set of handles have been studied, for example, autoencoders for 3D object
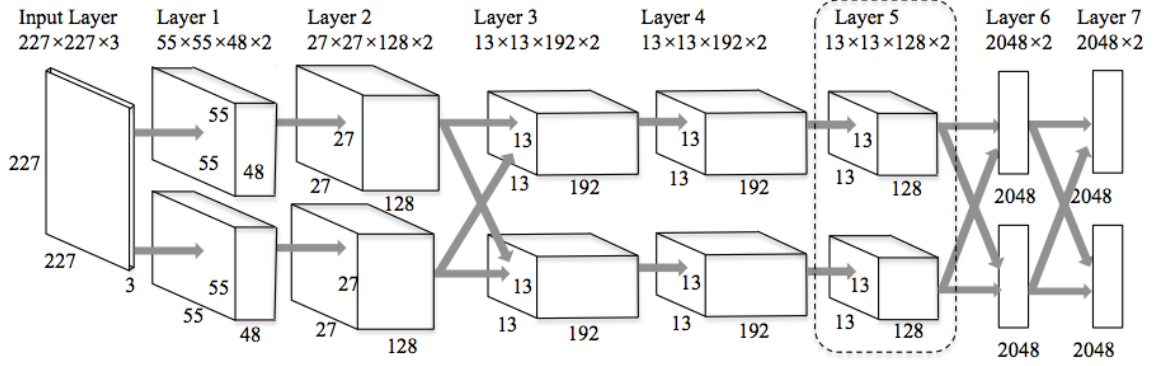
Figure 4.2: AlexNet convolutional neural network structure.

manipulation to affect attribute ratings (*Yumer et al.*, 2015), and representations that combine hand-crafted and implicitly learned representations to capture design freedom and brand recognition (*Burnap et al.*, 2016a).

4.3 Method

We model how customers perceive aesthetic design attributes and build a mapping from the design image space $\mathcal{D}$ to the attribute space $\mathcal{A}$ through an intermediate step in the design feature space $\mathcal{H}$, and then inverting this design-attribute mapping to predict visual attraction regions $\mathcal{V}$ in the original design image space $\mathcal{D}$. The four modeling steps–feature learning, attribute prediction, feature selection, and feature visualization–are detailed below as well as in Figures 4.24.4.

4.3.1 Feature learning using deep convolutional neural network

A deep convolutional neural network is a hierarchical model consisting of multiple layers (which could be conceptualized as layers of neurons following the organization of neurons in the human cortex), with each layer extracting higher-level data features from the previous layer. The output of each layer is a collection of features of the input image. Recent research has successfully applied these deep convolutional neural

85

network features to aesthetic related tasks such as style recognition (*Karayev et al.*, 2013)] and artistic image generation (*Gatys et al.*, 2015).

The features learned from a deep convolutional neural network depend on its structure and training data. Here, we learn design features using the structure of AlexNet (*Krizhevsky et al.*, 2012), detailed in Figure 4.2, due to its record-beating performance on the ImageNet 2012 classification benchmark (*Deng et al.*, 2009). Originally, AlexNet was trained on the ImageNet dataset, which consists of over 15 million images with over 22,000 class labels of the objects in image (e.g., dog breeds and strawberries). In addition, further fine-tuning was obtained by using additional images from the Flickr dataset (*Karayev et al.*, 2013), which itself consists of 80,000 images with more-specific style labels (e.g., 'melancholy,' 'ethereal') to modify higher layers to be more specific to desired aesthetic concepts. (*Karayev et al.*, 2013) show that mid-level features (layer 5 and layer 6) in AlexNet outperforms hand-tuned features in style recognition tasks and achieves the same level of prediction accuracy as participants in Amazon Mechanical Turk in a photographer group membership prediction task.

Accordingly, for any design image $\mathbf{D}_n \in \mathcal{D}$, using the deep convolutional neural network with structure and training procedure described above, we choose the feature outputs in layer 5 (see Figure 4.2) as the feature representation of the input design image, as these design features contain both the 2D image-specific information (usually contained in lower layers) and design attribute-specific information (usually contained in higher layers); we denote this feature representation as $\mathbf{H}_n \in \mathcal{H}$.

Design Attributes

a_n

$\beta^1 = 0$ β^2 β^{m-1} $\beta^m = 0$

Design Features H_n h_n^1 h_n^2 $\cdots$ h_n^{m-1} h_n^m

Salient Features $\hat{H}_n^2 = (0, h_n^2, 0, ..., 0)$ $\hat{H}_n^{m-1} = (0, ..., 0, ... h_n^{m-1}, 0)$
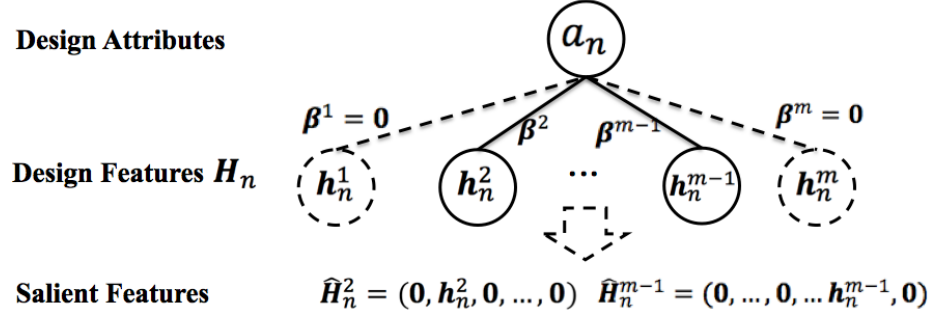
Figure 4.3: L1 Regression.

4.3.2 Design attribute prediction using crowdsourced Markov chain and L1 regression

4.3.2.1 Crowdsourced Markov chain

To obtain design attribute values for each 2D vehicle design image (e.g., the Toyota Prius may be 0.07 aggressive and 0.86 youthful), we assumed a ranked list of all 2D vehicle designs for each attribute. To obtain these ranked lists, we crowdsourced evaluations in the form of partial ranked lists partial ranking into a full ranking by an aggregation model based on Markov chain theory. Specifically, we assumed that the full ranking corresponds to the probability mass of individual designs of the stationary distribution of an ergodic Markov chain. We obtain the stationary distribution by using a modified version of the PageRank algorithm (*Brin and Page*, 2012); see Chap III for more implementation details.

4.3.2.2 L1 regression

Previous research has shown that transforming highly nonlinear design variable relationships into more easily human-memory "chunked" perceptual attributes justifies the linear models commonly used in the design community (*Hubel and Wiesel*, 1962). Accordingly, we model the relation between design attribute and design features as a L1 regularized regression model. Given the design feature representation

$\mathbf{H}_n$ for Design $\mathbf{D}_n$ as well as its design attribute value a_n, we assume that there is a linear relationship between a_n and $\mathbf{H}_n$:

$$a_n = \mathbf{H}_n \beta + \epsilon_n \qquad (4.1)$$

where ϵ_n is a Gaussian distributed random variable. To determine the coefficient vector β, we minimizes a loss consisting of the distance between the design attribute value and its estimation as well as a L1 regularization on β, as given in Equation (4.1), in which the parameter is determined by cross validation as is common in L1 regularization methods (Equation (4.1))

$$\beta = \arg\min_{\beta_0} \sum_{n=1}^{N} \|a_n - \mathbf{H}_n \beta_0\|_2 + \alpha |\beta_0|_1 \qquad (4.2)$$

The role of the L1 regularization is to reduce the dimensionality of β according to the shrinkage parameter α.

4.3.3 Salient feature selection using attribute prediction model

The L1 regularized linear regression in attribute prediction estimates the coefficient vector β, where only some of its elements are non-zero. The features corresponding to those nonzero coefficients are modeled to influence the attribute. Based on this idea, we define the salient coefficient set

$$S = \left\{ p | \beta^p \neq \mathbf{0}, \beta = \left[\beta^1, \beta^2, \ldots, \beta^m\right] \right\} \qquad (4.3)$$

where m is the number of features, and the salient feature representation for design image $\mathbf{D}_n$ is:

$$\hat{\mathbf{H}}_n^q = (\mathbf{0}, \ldots, \mathbf{0}, \mathbf{h}_n^q, \mathbf{0}, \ldots, \mathbf{0}) \in \mathcal{H}, q \in S \qquad (4.4)$$

88

This representation contains only one influential factor of the attribute. Using this feature representation, we are then able to apply the following feature visualization method to separately visualize the influential factors.

4.3.4 Feature visualization using deconvolutional neural network

A deconvolutional neural network may be considered an approximated inverse mapping of a convolutional neural network (*Zeiler and Fergus*, 2014). This inverse mapping is achieved by inverting the operations in the original convolutional neural network in the reverse sequence. In our model, the salient feature representation $\hat{\mathbf{H}}_n^q$ is passed as input to the deconvolutional neural network attached to AlexNet. Successive layers are inverted until we reach the input pixel space. This operation allows us to obtain a feature image $\mathbf{V}_n^q$ in the design image space $\mathcal{D}$, which contains only the pixel information that influence the salient feature, which itself most influences the desired design attribute. The attraction region $\mathbf{V}_n^q \in \mathcal{D}$ consists of those pixels in $\mathbf{V}_n^q$ that have a larger value than a pre-specified threshold. This threshold can be set by the designer to leverage the concentration of the attraction region, in which a higher threshold indicates a more concentrated attraction region.

There are three basic operations in the typical convolutional neural network: (i) max pooling, which means that only the maxima of a small region is passed to the next layer; (ii) ReLU rectification, which is a nonlinear function $f(\mathbf{x}) = max(\mathbf{x}, \mathbf{0})$; and (iii) convolution, whose key parameters are its weight matrix $\mathbf{W}$ and bias vector $\mathbf{b}$. In a deconvolutional neural network, the corresponding inverse operations are: (i) Unpooling: Though the max pooling operation is non-invertible, we can approximately invert it by recording the locations of the maxima in a set of switch variables when the input image is processed in the convolutional neural network. The value from the layer above is placed into the locations of the maxima according to the corresponding switch variable, such that the structure of maxima is preserved. These
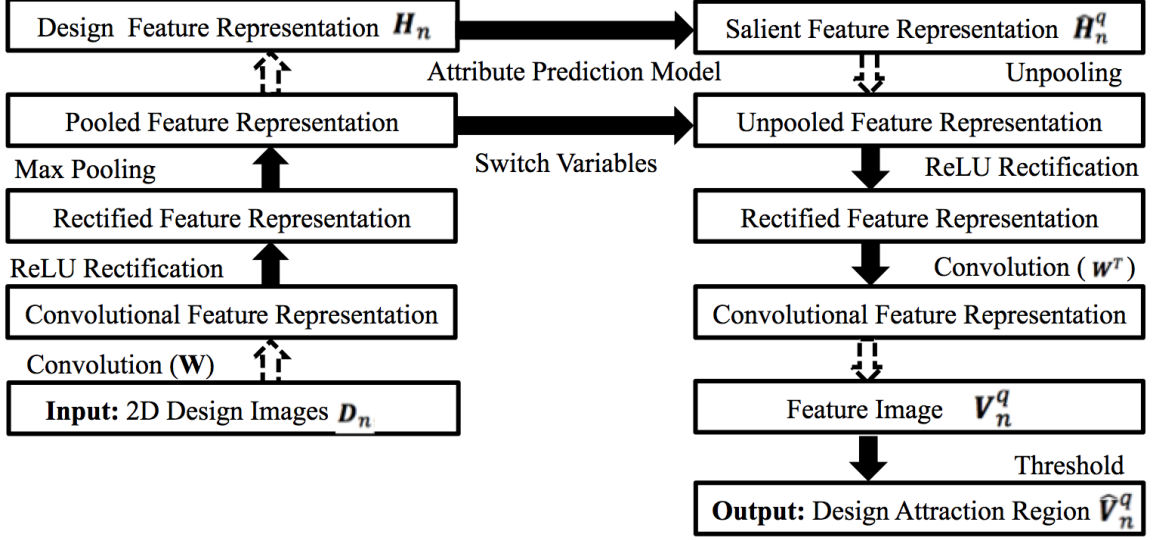
Figure 4.4: Deconvolutional neural network method flow.

Table 4.2: Ten design attributes used for partial ranking evaluation for 2D vehicle images.

Low Attr.	Awkward	Weak	Conservative	Basic	Conventional
High Attr.	Well Proportioned	Powerful	Sporty	Luxurious	Distinctive
Low Attr.	Passive	Traditional	Understated	Friendly	Mature
High Attr.	Active	Innovative	Expressive	Aggressive	Youthful

maxima are analogous to the salient information in the forward design communication framework, as salient information in the design is likely to be conveyed to humans perceptual processing units. (ii) Rectification: The approximate inverse operation of ReLU rectification is itself. (iii) Convolution: To approximately invert the convolution operation, the convolution operator with transposed weight matrix $\mathbf{W}^T$ is used.

4.4 Experiment

We conducted an experiment composed of four parts: (i) Estimate the feature representation of 110 vehicle images through a convolutional neural network, which

shares the same structure as AlexNet and is trained by both the ImageNet and Flickr datasets. We use pretrained parameters obtained from a verified deep learning platform Caffe (*Jia et al.*, 2014). (ii) Develop prediction models for the ten aesthetic attributes listed in Table 4.2. These attributes are used by design teams in the automotive industry (*Burnap et al.*, 2016a). Each prediction model is from the same feature representation to an aesthetic attribute. The value of an aesthetic attribute is obtained through the Markov Chain modeled in the crowdsourced human feedback as detailed in Section 4.4.1. (iii) Conduct feature selection based on the attribute prediction model. The L1 regresssion model allows us to select the influential predictors. Its regularization parameter α is adaptive through cross validation. (iv) Visualize the salient features selected from the previous step and empirically choose a threshold for visual attraction region that can reflect the desired concentration. In our study, we choose the threshold $\gamma = \mu + \sigma$, where μ is the mean pixel value in feature image and is the standard deviation of pixel values.

4.4.1 Crowdsourcing for design attribute values

A databased-backed web application was developed to crowdsource partial rankings of the 110 vehicle images for the set of 10 design attributes from Table 4.2. These partial rankings were then aggregated using the Markov chain described in Section 4.3.2 to obtain the values of all 10 design attribute for each of the 110 vehicle models. We gathered 361 participants through the crowdsourcing platform Amazon Mechanical Turk. Participants were directed to an introduction page, where they were given instructions on ranking vehicles according to a semantic differential for a randomly assigned design attribute from Table 4.2. This semantic differential consisted of only one of the ten attributes from low to high value or vice versa to act as a counterbalance for ordering biases. Over the entire interactive survey, a participant was always given the design attribute semantic differential in the same direction (either "low value"
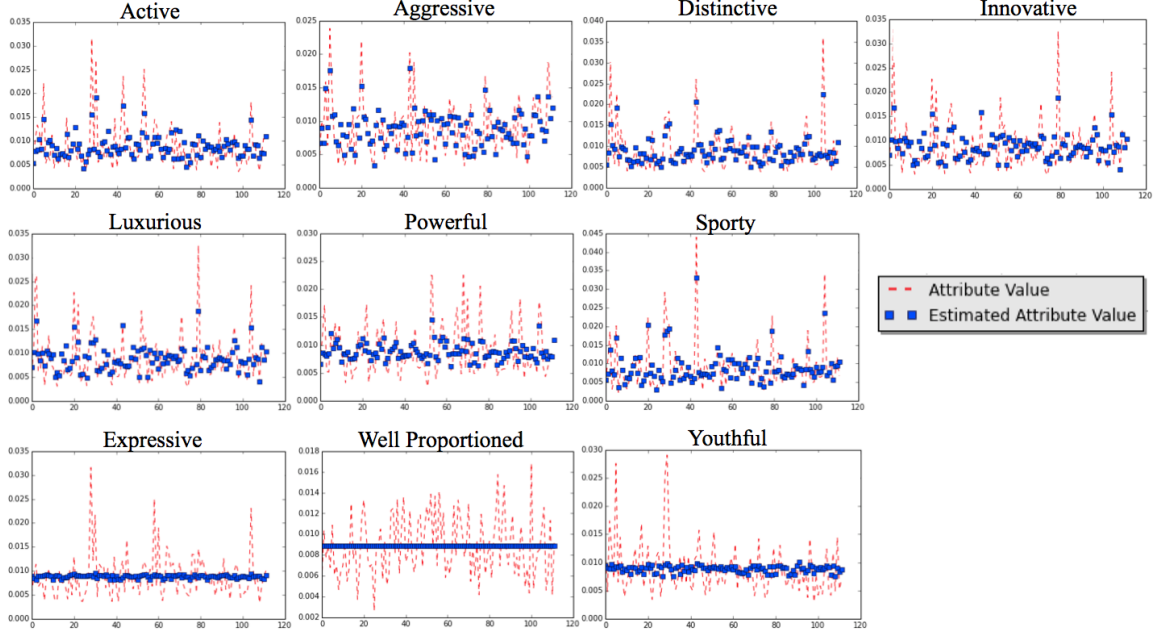
Figure 4.5: L1 regression prediction performance for all 10 design attributes with the x axis representing the vehicle ID and the y axis representing the attribute values and estimated values.

to "high value" or vice versa) to reduce participant burden, though direction was randomized across participants. Next, participants were directed to the 2D design partial ranking page, with four vehicles chosen from the set of 110 vehicles in a top row and four outlined placeholders in a bottom row. Instructions on the page were given to drag-and-drop the four designs from the top row to the bottom row using the mouse, including the possibility of reordering the partial ranking.

4.5   Results, Discussion and Limitations

The attribute prediction performance is given in Figure 4.5. Seven out of ten prediction models provide attribute estimations that are similar to the attribute values obtained through crowdsourcing. This indicates that the features from the 5th layer in the AlexNet contain the important information for those seven attributes, and thus visualizing these features is a meaningful way to predict the attraction regions. Model

fitness further validates the model for forward communication from the design to these seven attributes. However, the prediction model fails to predict three attributes including 'expressive', 'well-proportioned,' and 'youthful'. A possible reason is that the influence of the 5th layer features may not be as meaningful, and thus these three attributes are not included in our analysis and visualization.

Figure 4.6 shows the visual attraction region for the design attribute 'active'. We cover subsections of the design images with two groups of attraction regions. Each group corresponds to one salient feature. The images in the same row show the predicted attraction regions of the same feature for different cars. The predicted attraction regions focus on the same region of the car (front light in the first row) despite other variations in the image space such as vehicle shape, color, and viewpoint. The images in the same column show the predicted attraction regions of different features in the same car. In this case, different attraction regions are shown for different features of the same car. It is important to point out that these attraction regions are estimated from our model without using eye-tracker data.

A limitation of the present work is lack of validation. While qualitatively we can see that the predicted areas of visual attraction indeed only occupy subsections of the 2D vehicle images on the vehicle itself, we do not have an objective or independent measure. That is, while we capture a function in the forward direction from images to attribute values and visualize projections of its inverse from attribute features back to images, we do not have a way to assess whether the projected inverse mapping is correct. There are two difficulties here: (i) Defining an error metric for validity, and (ii) obtaining "ground truth" values for validity. Defining an error metric may be best addressed with assumptions from visual attraction models from psychobiological human attraction models. While it may be possible to ask customers to give their "stated response" by clicking on regions of interest, it may be more fruitful to instead compare our predicted visual attraction with empirically-derived revealed response"

93
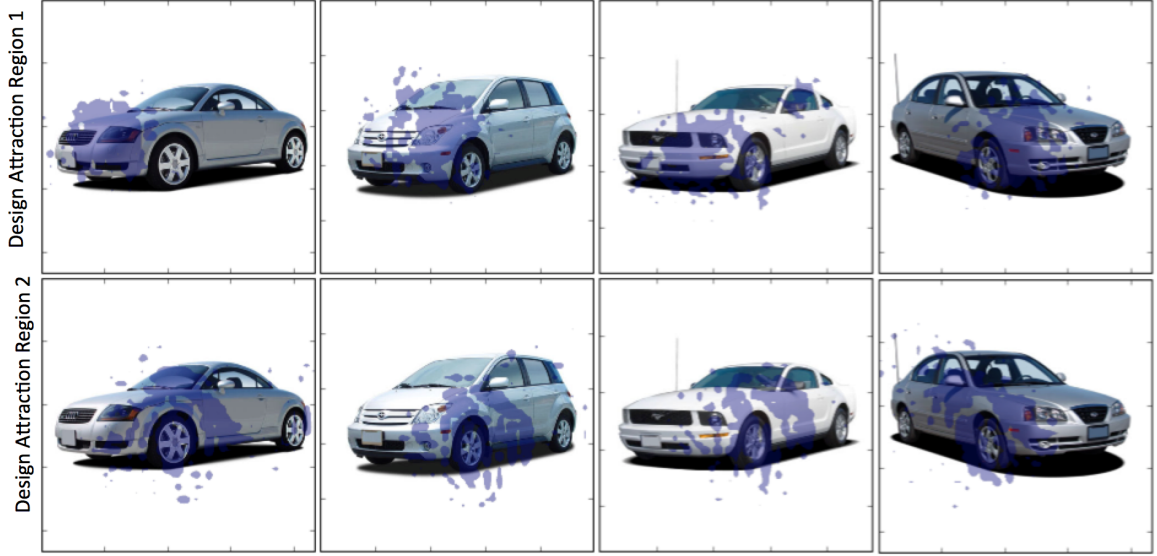
Figure 4.6: Examples of predicted attraction regions for design attribute 'Active'. The top row corresponds to an unknown design feature describing and 'Active' car, seemingly focused on vehicle headlights, while the bottom row corresponds to a sepearte unknown design feature, seemingly focused on the front quarter-panel and door.

cues such as customer eye-tracking data (*Du and MacDonald*, 2014; *Marshall et al.*, 2014; *Tovares et al.*, 2014) and implicit dimensionality reduction (*Yumer et al.*, 2015).

One important future direction is testing different feature representation, especially different convolutional neural networks architectures such as VGGNet (*Simonyan and Zisserman*, 2014). The second future direction is better feature visualization to reveal more design details in the attention region. These design details may be valuable clues for designers to improve the aesthetic appeal of designed artifacts, for example, (*Selvaraju et al.*, 2016; *Li et al.*, 2018; *Zhou et al.*, 2016). Another future direction is the study of the approximately inverting process used in the model describing the backward communication, such as when it will fail and the bounds of errors. This is more theoretical work and beyond the scope of this paper.

Future research to improve this method include validation using data-specific metrics, correlations with other methods such as eye-tracking that captures visual attrac-

tion in the design space, better feature visualization, and theoretical analysis of the proposed algorithm.

4.6   Summary

While visual attraction, in both descriptive and predictive senses, has a long history of study in the design community, the present work contributes a mathematical predictive model to identify specific regions that may attract the user.

We introduced a data-driven method building on the framework of design as a communication process. We extended this method to include four stages in a forward - backward pipeline: (i) design feature learning, (ii) design attribute prediction, (iii) design feature selection, and (iv) design feature visualization.

This method is novel in that it is data-driven and does not require humans to provide the attention data. The modeling tools we used for the data-driven method include a convolutional neural network, L1 regression, a crowdsourced partial ranking Markov chain, and a deconvolutional neural network. This work is a first step toward data-driven predictions on how portions of the design space (regions of attraction) affect various design attributes via features learned using large-scale image data and selectively weighted crowdsourced perceptual responses.

CHAPTER V

Deep Design: Product Aesthetics for Heterogeneous Markets

5.1 Introduction

In this chapter, we present an approach to visualize the possible factors affecting customers' perception of product aesthetic appeal in heterogeneous market. Aesthetic appeal is of critical importance for product design, as it not only attracts customer attention, but assists in conveying design attributes (e.g., 'luxurious,' 'sporty,' 'well-proportioned') that are meaningful to the customer (*Berlyne*, 1971; *Coates*, 2003). Conveying these aesthetic attributes is particularly important for the automotive industry, as underpinned by the most respected industry assessments (e.g., J.D. Power Initial Quality Study (*Tews*, 2016), J.D. Power APEAL Study (*Dobrian*, 2016)) and internal confidential studies at General Motors.

Specifically, exterior styling is always in the top two or three reasons for purchase, year after year. It is also a prominent reason for not considering a vehicle for purchase or for rejecting it as a finalist. This pattern has been found not just in developed markets, such as the US, but in emerging markets such as India (*Motors*, 2014). Understanding these aesthetic preferences remains an important and ongoing challenge for product designers.

The major challenge behind this understanding is the "heterogeneity" of diverse customers across various markets and the inherent subjectivity of their aesthetic perceptions. This challenge is especially vital for customer-centered product designs such as automobiles, as these designs require product differentiation across market segments. This challenge is further exacerbated given the globalized nature of modern automobile design, with customers often geographically and culturally distant from the designers. Accordingly, product designers use a variety of qualitative and quantitative methods to assess aesthetic preferences across market segments, with examples including design theme clinics, focus groups, customer surveys, design reviews, and Kansei engineering (*Nagamachi*, 1995). The primary goal of these methods is to understand the reasons "why" the customer perceives a design concept as being aesthetically appealing or unappealing. Ideally, a designer could identify specific regions of the physical product design that contribute to the customer's perception of design attributes. Identification of these regions, called "salient design regions" (*Du and MacDonald*, 2014), can provide valuable insight during the design process.

While these methods may capture in-depth customer rationale for aesthetic perceptions, they have two main drawbacks. First, customers often cannot articulate accurately why they like or dislike a design (*Nisbett and Wilson*, 1977; *Silvera et al.*, 2002). Second, they are not scalable due to being labor and resource intensive, particularly as multinational enterprises often deal with hundreds or thousands of heterogeneous markets.

In this chapter, we aim to understand perceptions of aesthetic design attributes across customers from heterogeneous markets, and to do this at the scale consistent with a global company. Specifically, we aim to answer three fundamental questions in the context of product design:

1. Does the product design achieve the desired aesthetic design attributes for a given market segment?

2. What are the product's salient design regions for a given design attribute?

3. How do salient design regions differ across different market segments?

We propose a deep learning approach for prediction of aesthetic design attributes for a given customer, that allows interpretation of the reasons "why" a design is perceived as appealing or unappealing. This approach uses a deep learning architecture that captures the heterogeneity of customer perceptions across aesthetic design attributes. Moreover, this approach has the capacity to analyze large-scale data, such that customer heterogeneity may be accurately modeled across market segments. Importantly, this approach enables visual interpretation of results by identifying regions of the product design that are relevant for a given design attribute.

We conduct an study to test this deep learning approach using 179,000 2D images of vehicles in the last decade, 3,302 customer profiles as well as 33,020 data points of customer perceptions of aesthetic design attributes crowdsourced using an online web application. Our results show that we are indeed able to predict diverse customer perceptions over design attributes, as well as visually interpret the reasons underlying customer perceptions.

The main contribution of this research is providing an approach to interpreting aesthetic design appeal for design concepts across heterogeneous markets. This approach is scalable to hundreds or thousands of markets, an important consideration for multinational enterprises engaged in product design. Methodological contributions include a novel deep Siamese neural network architecture using conditional generative adversarial networks, trained using multimodal data including 2D images, numerical labels, and large-scale crowdsourced aesthetic response data.

The rest of this chapter is structured as follows: Section 2 discusses previous efforts that quantitatively analyze product aesthetics. Section 3 introduces the research approach as well as the deep learning model and its interpretation algorithm. Section 4 details the experimental setup, describes the data sets and presents results showing

aesthetic perceptions across market segments. Section 5 discusses how this work contributes to the product design, as well as its limitations. Section 6 provides a summary.

5.2 Related Work

We review related work from the engineering and product design communities, as well as recent advances in deep learning for aesthetic styling.

5.2.1 Product Design Aesthetics

The engineering and product design communities have studied factors affecting the styling attributes of designs using both experimental approaches and modeling approaches.

Experimental approaches to understand implicit customer perceptions of aesthetic design attributes often employ eye-tracking. The earliest pioneering work with such eye-tracking dates back to 1935, when an experiment recorded eye-gaze fixation across regions in artistic pictures (*Buswell*, 1935). Eye-tracking methods have been successfully applied in various domains such as optimizing the layout of product placement in advertisements (*Duchowski*, 2002), web page layouts (*Wang et al.*, 2014; *Buscher et al.*, 2009), and consumer choice under pressure (*Reutskaja et al.*, 2011). For product design specifically, eye-tracking has been applied to design representations (*Reid et al.*, 2013; *Marshall et al.*, 2014), relations with vehicle face components (*Du and MacDonald*, 2014; *Windhager et al.*, 2010), and design diagram assessment (*Ruckpaul et al.*, 2015).

Aesthetic modeling approaches have conventionally relied on hand-crafted design representations such as a set of parametric control points to manipulate vehicle silhouettes (*Petiot and Dagher*, 2011; *Reid et al.*, 2010; *Poirson et al.*, 2013; *Orsborn et al.*, 2006), representations created implicitly using finite shape grammars (*Pugliese*
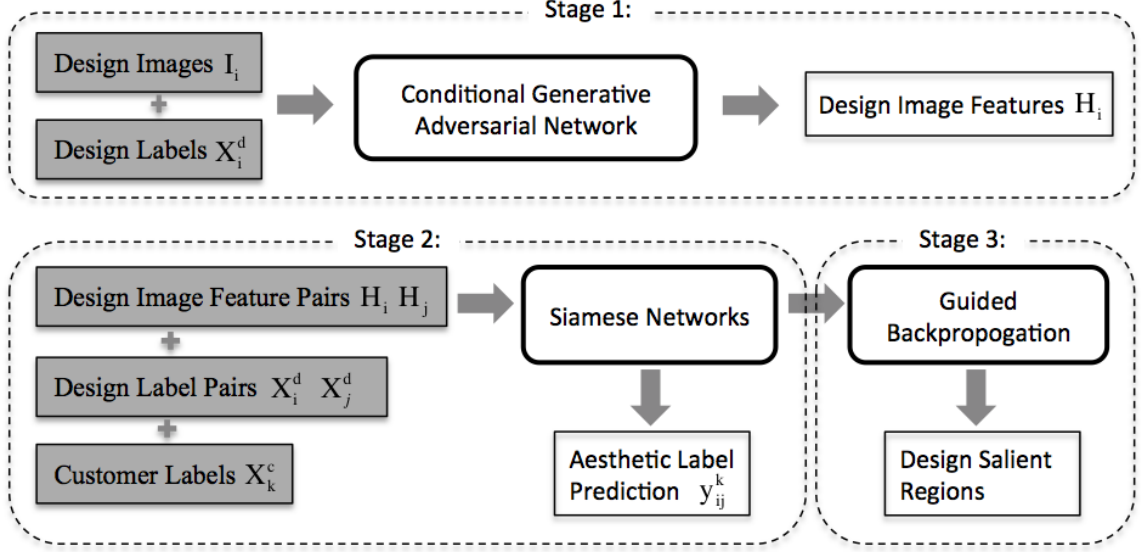
Figure 5.1: Overview of the proposed deep learning approach for aesthetic design appeal prediction for heterogeneous customers. Grey boxes represent the inputs, white boxes represent outputs, and rounded corner boxes represent the model or algorithm.

and Cagan, 2002; *McCormack et al.*, 2004), representations corresponding to a change in intensity (*Quercia et al.*, 2014), or representation manipulated by a set of geometrical handles (*Yumer et al.*, 2015; *Burnap et al.*, 2016a). In spite of their high interpretability and successful applications, the fidelity of these models are bounded by the realism and flexibility of the design representation (*Burnap et al.*, 2016b). Recently, design research has hybridized modeling approaches with experimental based approaches such as assessments of vehicle face attributes with Kansei engineering and eye-tracking (*Chang et al.*, 2013). These aesthetic models has been used to optimize automobile design; examples include 2D vehicle side view silhouettes (*Reid et al.*, 2010; *Orsborn et al.*, 2009) and 2D vehicle faces (*Petiot and Dagher*, 2011; *Ranscombe et al.*, 2012). In addition, recent work has extended this research into 3D (*Ren et al.*, 2013) and virtual reality representations (*Tovares et al.*, 2014).
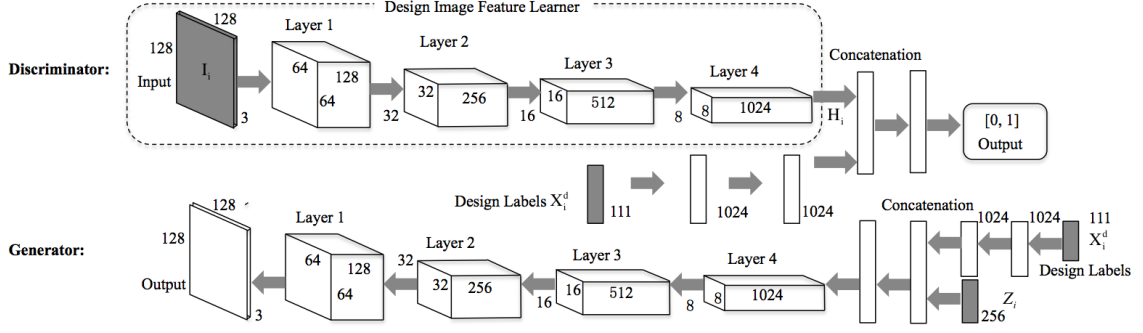
Figure 5.2: Discriminator and generator in conditional generative adversarial network. Grey boxes represent inputs and white boxes represents convolutional layers in discriminator and upsampling layers.

5.2.2 Deep Learning for Aesthetics

Deep learning has emerged as a state-of-art approach to model large datasets that include hierarchical data such as 2D images, including recent work in aesthetic styling. These studies frame styling problems as a supervised learning task by using the object's style as labels. Examples include image style recognition (*Karayev et al.*, 2013), comparing illustrative style using stylistic labels (*Garces et al.*, 2014), and learning high-level judgments of urban perception (*Ordonez and Berg*, 2014; *Naik et al.*, 2014). Unsupervised learning approaches have also covered cases where stylistic labels are not available, such as style comparison (*Furuya et al.*, 2015), and style transfer to generate paintings in styles of artists from Van Gogh to Picasso (*Gatys et al.*, 2015). These works validate the possibility of capturing aesthetic-related information using features in deep learning models.

At the same time, however, deep learning has interpretation challenges as these methods create highly nonlinear interactions among the input variables. From a technical standpoint, commonly used optimization objectives such as predictive accuracy or likelihood may no longer provide sufficient criteria for interpretation of deep learning prediction models.

101

These interpretation challenges have been the focus of recent work in deep learning. A widely-used approach to interpreting a deep learning model is showing the patches of the 2D image with highest response on feature detectors. Researchers have successfully built a deep learning model to predict safety of urban scenes and interpreted the patterns of safety by visualizing those patches (*Porzi et al.*, 2015). Unlike urban scenes, small differences in curves or shapes can dramatically change the aesthetic of a product design, thus image patches are too coarse to reveal the salient design regions. A deep convolutional neural network can also be interpreted by computing an approximate inverse mapping of the neural network such as deconvnet, which uses "switch" variables to record the position of pooling operations so that the irreversible pooling operator can be approximately reversed (*Zeiler and Fergus*, 2014). This method has been used to visualize aesthetic attributes for product design (*Pan et al.*, 2016), however, this approach is limited because it relies on the selection of neurons to be visualized. Interpretation of deep learning classifiers can also be accomplished by learning an interpretable model locally around the prediction (*Ribeiro et al.*, 2016). However, this approach may fail in interpreting product aesthetics because the underlying model is likely highly non-linear even in the locality of the prediction and this may lead to a biased interpretation. The interpretation technique used in our approach, Guided Backpropagation (*Springenberg et al.*, 2014), facilitates interpretation by visualizing the salience map. This saliency map captures the salient region in any shape and does not rely on the selection of neurons. Moreover, backpropagation can handle a high degree of nonlinearity.

5.3   Research Approach

We introduce a research approach that develops a deep learning model to predict and interpret how a customer or market segment perceives a product design concept according to a given aesthetic design attribute (e.g., 'Appealing', 'Sporty'). As shown

in Figure 5.1, the overall research approach consists of three stages:

1. Stage 1 converts design images to a lower-dimensional feature representation by training a conditional generative adversarial network (cGAN) on the distribution of design images given the design labels (e.g., brand, bodytype).

2. Stage 2 trains a Siamese network with a pair of cGANs to predict how a customer will perceive a design for a given design attribute, for example, whether a ('Rich,' 'Male') perceives a '2014 Range Rover' as 'Sporty.'

3. Stage 3 uses guided backpropagation to obtain the saliency map of the Siamese network, then filters the saliency map to discover salient design regions for the given design attribute. This stage allows visual interpretation of predictions of the deep learning model; for example, providing an account for why a ('Rich', 'Male') perceives a '2014 Range Rover' as 'Sporty.'

We next formalize customers, design concepts, and aesthetic design attributes, followed by discussing these three stages of the research approach in detail. Denote the i-th design $\mathbf{D}_i$ as represented by its image $\mathbf{I}_i$ and its design labels $\mathbf{X}_i^d$, i.e. $\mathbf{D}_i = \{\mathbf{I}_i, \mathbf{X}_i^d\}$. Denote the k-th customer $\mathbf{X}_k^c$ as a one-hot encoded vector of customer variables. For each tuple $(\mathbf{D}_i, \mathbf{D}_j, \mathbf{X}_k^c)$, there is a corresponding label y_{ij}^k, with $y_{ij}^k = 1$ having the interpretation that customer k prefers design i over design j for the given design attribute. For example, if the design attribute is 'Sportiness,' then $y_{ij}^k = 1$ corresponds to customer k perceiving design i as more sporty than design j, and $y_{ij}^k = 0$ corresponds to customer k perceiving design i as less sporty than design j.

5.3.1 Conditional Generative Adversarial Network

The generative adversarial network (GAN) is a generative model consisting of two components, a *discriminator* $\mathcal{D} : \mathbf{I} \rightarrow [0, 1]$ and a *generator* $\mathcal{G} : \mathbf{Z} \rightarrow \mathbf{I}$, where $\mathbf{Z} \in \mathbb{R}^{n_z}$ is a noise vector used to seed the generator. The value of $\mathbf{Z}$ is sampled from a

noise distribution $p_{\mathbf{z}}(\mathbf{z})$, which is a standard Gaussian distribution in this work. The *discriminator* and *generator* are posed in an adversarial game. The *discriminator* aims to distinguish between real samples from the training data and fake samples generated by the *generator*, while the *generator* aims to generate samples that can not be distinguished by the *discriminator*. This adversarial game is obtained by using a min-max value function as the objective:

$$
\min_{\mathcal{G}} \max_{\mathcal{D}} \quad \Big(\mathbb{E}_{\mathbf{I} \sim p_I(I)} \left[\log \mathcal{D}\left(\mathbf{I}\right) \right] + \\
\mathbb{E}_{\mathbf{Z} \sim p_{\mathbf{z}}(\mathbf{z})} \left[\log \left(1 - \mathcal{D}\left(\mathcal{G}\left(\mathbf{Z}\right)\right) \right) \right] \Big)
\tag{5.1}
$$

Our work extends conventional generative adversarial networks with a conditional architecture, termed a conditional generative adversarial network (cGAN). In this architecture, there is a set of variables $\mathbf{X}^d$ that are believed to be relevant to the image $\mathbf{I}$, and the cGAN aims to capture the relationship between the image and this external information. The *generator* $\mathcal{G}$ and *discriminator* $\mathcal{D}$ in the cGAN model can be redefined as following:

$$
\begin{aligned}
\mathcal{G}: & \quad (\mathbf{Z} \times \mathbf{X}^d) \rightarrow \mathbf{I} \\
\mathcal{D}: & \quad (\mathbf{I} \times \mathbf{X}^d) \rightarrow [0, 1]
\end{aligned}
\tag{5.2}
$$

The *generator* $\mathcal{G}$ defines a conditional distribution $p_g(\mathbf{I}|\mathbf{X}^d)$, enabling conditioning of the generative model with contextual information $\mathbf{X}^d$. In this work, $\mathbf{X}^d$ are the design labels of an automobile, such as brand, body type, color, and viewpoints. By varying the value of $\mathbf{X}^d$ in the generator, the design labels of the generated sample can be explicitly controlled. More importantly, conditioning on design labels will prompt the model to focus on learning the features describing the appearance of the design instead of the known semantic features of the design labels. These design labels are detailed in Table 5.1. In this way, the features extracted from cGAN are more relevant to our later predictive task of capturing aesthetic appeal as will be discussed later.

Similar to the GAN, the generator $\mathcal{G}$ and discriminator $\mathcal{D}$ in cGAN are posed in an adversarial game by a minmax value function:

$$\min_{\mathcal{G}} \max_{\mathcal{D}} \quad \Big(\mathbb{E}_{(\mathbf{I}, \mathbf{X}^d) \sim p_I(I, \mathbf{X}^d)} \left[\log \mathcal{D} \left(\mathbf{I}, \mathbf{X}^d \right) \right] + $$
$$\mathbb{E}_{\mathbf{X}^d \sim p_{\mathbf{X}^d}, \mathbf{z} \sim p_{\mathbf{z}}(\mathbf{z})} \left[\log \left(1 - \mathcal{D} \left(\mathcal{G} \left(\mathbf{Z}, \mathbf{X}^d \right), \mathbf{X}^d \right) \right) \right] \Big) \tag{5.3}$$

Conventionally, the discriminator aims to assign a positive label to the training samples $(\mathbf{I}_i, \mathbf{X}_i^d)$, and a negtive label to generated samples $\mathcal{G}(\mathbf{Z}_i, \mathbf{X}_i^d)$, $i = 1, 2, \ldots, n$. The objective function of a conventional discriminator is:

$$J_{\mathcal{D}} = -\frac{1}{n} \Big(\sum_{i=1}^{n} \log \mathcal{D}(\mathbf{I}_i, \mathbf{X}_i^d) + $$
$$\sum_{i=1}^{n} \log(1 - \mathcal{D}(\mathcal{G}(\mathbf{Z}_i, \mathbf{X}_i^d), \mathbf{X}_i^d)) \Big) \tag{5.4}$$

We extend this formulation by forcing the discriminator to capture the link between the design images and their design labels by modifying the loss function (*Reed et al.*, 2016). Specifically, we penalize the discriminator when it assigns a positive label to an incorrect training sample $(\mathbf{I}_i, \mathbf{X}_r^d)$, where r is randomly drawn from $[1, 2, \ldots, n]$ and $r \neq i$.

$$J_{\mathcal{D}} = -\frac{1}{n} \big(\sum_{i=1}^{n} \log \mathcal{D}(\mathbf{I}_i, \mathbf{X}_i^d) + \frac{1}{2} \big(\sum_{i=1}^{n} \log(1 - \mathcal{D}(\mathbf{I}_i, \mathbf{X}_r^d)) $$
$$+ \sum_{i=1}^{n} \log(1 - \mathcal{D}(\mathcal{G}(\mathbf{Z}_i, \mathbf{X}_i^d), \mathbf{X}_i^d)))) \tag{5.5}$$

Moreover, we maximize the probability assigned by the discriminator to the sample generated by the generator, resulting in the following generator loss function:

$$J_{\mathbf{G}} = -\frac{1}{n} \sum_{i=1}^{n} \log \mathcal{D} \left(\mathcal{G}(\mathbf{Z}_i, \mathbf{X}_i^d), \mathbf{X}_i^d \right) \tag{5.6}$$

We use deep neural networks for the discriminator and generator. Their architectures are similar to each other as shown in Figures 5.2. In the discriminator, the input design images and design labels are processed separately by several layers before they are concatenated together. The grey boxes represent the inputs, white boxes represent the output of fully connected layers, and rectangular prisms represent either the output of convolutional layers with filter size 5×5 and ReLu layers in the *discriminator* or deconvolutional layers with filter size 5×5 in the *generator*. The output of the fourth layer in the discriminator (denoted by the dotted box in Figure 5.2) is then used as the feature representation of the design images.

Though there are simpler models to extract image features, cGAN is used here for several reasons. First, cGAN is a generative model that provides a visual sanity check of whether the cGAN is capturing the distribution of vehicles in the 2D image space, while non-generative models may fall to provide such a visualization. Second, a distinguishing difference between cGAN and other generative models is that cGAN learns the conditional distribution of the vehicle images given the design labels. In this way, the cGAN can focus more on the image features other than design labels whose relationship with product aesthetic preference can be predicted and interpreted using simpler models (*Nagamachi*, 1995). Third, using such a generative model allows the generation of new product designs with desired aesthetic attributes.

5.3.2 Siamese Network

Siamese neural networks are a class of neural network architectures that contain two or more identical subnetworks (*Chopra et al.*, 2005). These identical subnetworks share the same architecture as well as the same parameters and weights. Siamese neural networks are common for modeling similarity or a relation between two com-

Figure 5.3: Siamese network of identical conditional generative adversarial networks, with conditioning on design and customer labels. This structure is used to model a customer's aesthetic perception y_{ij}^k for a given design attribute.
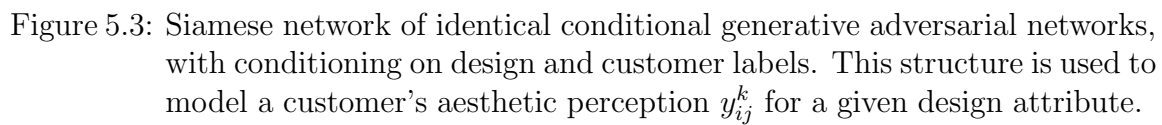
parable inputs, for example, verifying handwritten signatures. The Siamese structure offers several technical advantages, including requiring fewer parameters to estimate so is less likely to overfit the data.

The structure of the Siamese network used in this work is given in Figure 5.3. The "design image feature learner," or the feature representation given by the bottom four layers of the discriminator from the cGAN (see in Figure 5.2), are used as the Siamese network's subnetworks. This 2D image feature representation is then connected with a feature representation of the design labels. After subtracting between concatenated 2D image and design label features, the result is then concatenated with features of customer labels. This concatenated feature vector is then passed through two fully connected layers before the binary prediction task. The objective we used to train the entire model is the cross entropy $\mathbf{J}_s$ of this task.

$$
\begin{aligned}
\mathbf{J}_s = -\frac{1}{n}\Big(& \sum_{i,j,k} y_{ij}^k \log(\sigma(f(\mathbf{I}_i, \mathbf{I}_j, \mathbf{X}_i^d, \mathbf{X}_j^d, \mathbf{X}_k^c)))+ \\
& (1 - y_{ij}^k)\log(1 - \sigma(f(\mathbf{I}_i, \mathbf{I}_j, \mathbf{X}_i^d, \mathbf{X}_j^d, \mathbf{X}_k^c)))\Big)
\end{aligned}
\tag{5.7}
$$

where $\sigma(\cdot)$ is the sigmoid function, and $f(\cdot)$ represents the Siamese network.

5.3.3 Guided Backpropagation

Guided backpropagation computes a saliency map for a trained neural network (*Springenberg et al.*, 2014). This saliency map is used to visualize which pixels/regions of an input image are most important for a a neural network's prediction. The key idea behind guided backpropagation is to compute the gradient of the neural network's prediction with respect to the input image with fixed weights. This determines which pixels in the design image are sensitive to the prediction label, or in other words, which pixels can significantly affect the prediction even with small perturbations.

Compared with other visualization methods, guided backpropagation has the ability to produce sharp visualizations of salient image regions. This sharpness is particularly important for our task as shapes and edges of product designs are a major contributor to a customer's aesthetic perception (*Orbay et al.*, 2015). Accordingly, we use guided backpropagation to visualize the trained Siamese network from Section 5.3.2. This allows interpretation of which regions of a design most contribute to a customer's perceptual response over aesthetic design attributes.

Guided backpropagation is an extension of conventional backpropagation. The primary difference is how the gradient is backpropagated through "neurons," in which we always assume as linear rectifier units, $y(x) = max(x, 0) = x \cdot [x > 0]$, where $[\cdot]$ is the indicator function. In conventional backpropagation, the gradient of the rectifier's output with respect to its input is defined as follows: $\frac{dy}{dx} y(x) = [x > 0]$. Backpropagation of the error signal δ_i through the rectifier is $\delta_{i-1} = \delta_i \cdot [x > 0]$. Instead, in guided backpropagation, the error signal is $\delta_{i-1} = \delta_i \cdot [x > 0] \cdot [\delta_i > 0]$ when passing through the rectifier. This results in guided backpropagation only passing positive error to positive inputs, such that the error signal is guided not only by the input from the layer below the rectifier, but also by the error signal from layers above the rectifier.

Based on the obtained saliency map, we define salient regions by thresholding on saliency map values. This threshold is a hyperparameter chosen by designers, who have domain expertise in this area. A higher threshold results in salient regions with higher levels of sensitivity, while lower thresholds allow more holistic visualization of salient regions.

5.4 Study

We conducted a study to test whether the deep learning research approach introduced in Section 5.3 can be used to understand aesthetic perceptions of customers in

Table 5.1: Design labels

Label Names	Dim.	Label Value
Year	15	2000-2014
Make	48	Land Rover, Nissan, etc.
Model	23	Range Rover Sport, Rogue Select, etc.
Body type	20	SUV, Sedan, etc.
View Point	2	$[\sin\theta, \cos\theta]$, where θ is the angle.
Color	3	RGB

Table 5.2: Customer labels

Label Names	Dim.	Label Value
Age	1	0 - 99
Gender	3	Female, Male, Prefer not to say.
Income Level	20	$0 - $200,000+
House Region	5	Metropolitan, Suburban, etc.
Family Size	10	0 - 20
Current Car Brand	48	Audi, Cadiliac, BMW, etc.

heterogeneous markets. Specifically, we aim to capture customer perceptions of four pairs of aesthetic design attributes: 'Sporty' vs. 'Conservative', 'Luxurious' vs. 'Basic', 'Innovative' vs. 'Traditional', and 'Appealing' vs. 'Unappealing', for the sport utility vehicles (SUV) designed on the U.S. market from 2010 to 2014, followed by visual interpretation of salient regions of these SUVs according to customer perceptions of sportiness.

5.4.1 Data

Four data sources are used from different modalities: (i) 2D design images, (ii) design labels (e.g., bodytype), (iii) customer labels crowdsourced using an online interactive web application, and (iv) aesthetic perception data for a given customer and set of designs.
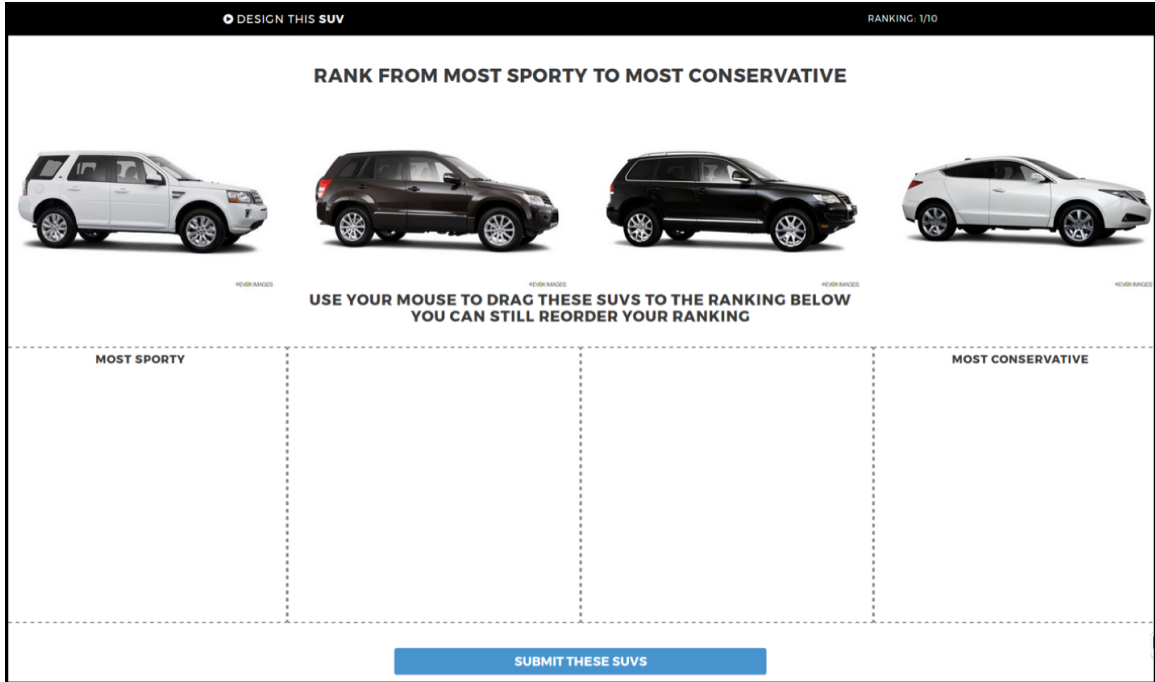
Figure 5.4: A snapshot of the ranking page in the crowdsourcing web application.

5.4.1.1   Design Data

The full design data set consists of 2D images and design labels corresponding to semantic information about these images. This data set contains 179,702 2D images of vehicle designs on the U.S. market from 2000 to 2014. Each design image has corresponding design labels as listed in Table 5.1. The full design data set was used to train the conditional generative adversarial network described in subsection 5.3.1.

The SUV data set consists of 13,464 2D images and labels of SUV design on the U.S. market from 2010 to 2014, which covers 373 SUV models from 29 brands. This data set was used to collect customer aesthetic perceptions.

5.4.1.2   Customer Data

A crowdsourcing web application was developed to collect customer aesthetic perceptions for the four pairs of design attributes. Customers first landed on a home page that described the aesthetic perception task. They were then directed to a data

111

collection page as shown in Figure 5.4, in which they were asked to rank four randomly selected SUVs from the same viewpoint along one randomly selected semantic differential such as 'Sporty' vs. 'Conservative.' The order of this semantic differential was randomly flipped for each customer to counterbalance for ordering biases; however, a single customer always saw the same semantic differential and the same ordering.

Customers were asked to complete 10 rankings, with different viewpoints and SUVs for each ranking. Upon completion of 10 rankings, they were redirected to a survey page, in which they were asked to answer questions about themselves, populating customer labels as listed in Table 5.2.

A total of 3,302 respondents were collected through General Motors' respondent panels. These respondents had bought an SUV in the past 5 years. Respondents were drawn from several sources and compensated in a variety of ways ranging from no compensation to a Sweepstakes entry to win a $500 gift card. Figure 5.5 shows the statistics of these customer data. Demographics include 61.6% of the respondents being female. We note that this data has a broad distribution of customers labels, suggesting it contains numerous heterogeneous market segments.

5.4.2 Procedure

The procedure involved three steps: (1) data preprocessing, (2) training of the cGAN to estimate the feature representation of design images, and (3) training of the Siamese network, containing two cGANs, as well as incorporating customer label and aesthetic perceptions data.

5.4.2.1 Data Preprocessing

For each attribute (e.g., "Sporty"), respondents who evaluated that attribute were split into training and validation datasets at an 80%/20% ratio. Splitting on the users
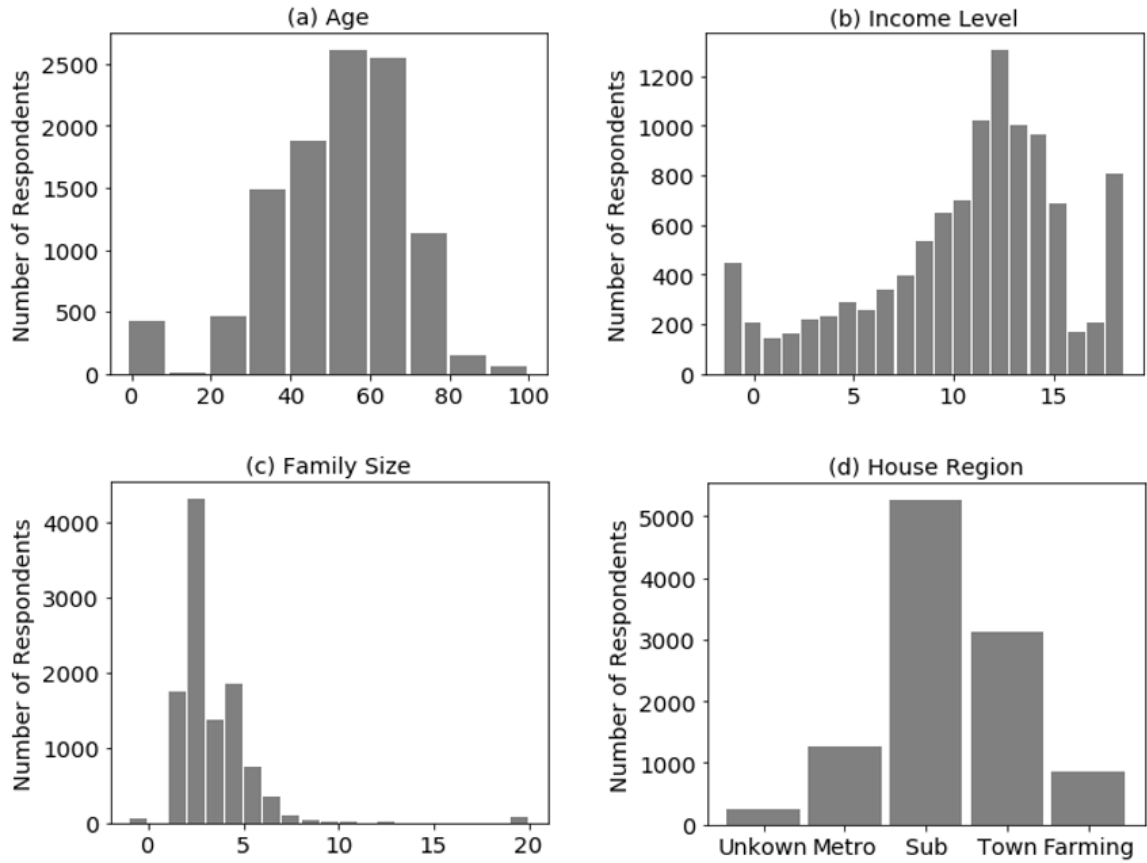
Figure 5.5: The customer data distribution of (a) Age, (b) Income Level, (c) Family Size, and (d) Housing/Living Region, where "Metro" means "Metropolitan", "Sub" means "Suburban", "Town" means "Small Town", and "Farming" means "Farming Area".

themselves ensured that the task being assessed was more general than splitting on the evaluations of the respondents. The aesthetic perception data obtained during crowdsourcing was converted from a ranking of four designs to a binary comparison format. This conversion generated a single binary choice pair for each ranking, by taking the first and last design from the ranking. These binary comparisons of design were assigned the label '1' to the pair $[(\mathbf{I}_i, \mathbf{X}_i^d), (\mathbf{I}_j, \mathbf{X}_j^d)]$ if vehicle i was ranked higher than vehicle j, otherwise '0.' Only the "first" rank and "last" rank are used instead of all pairwise binary choices with a ranking of 4 SUVs.

5.4.2.2 Conditional Adversarial Network Training

Though the study focuses on predicting design attributes such as 'Sportiness' of SUVs for customers in heterogeneous markets, we trained the cGAN using the design data containing 2D images from vehicles of 20 body types (e.g., sedans, trucks) and conditioned on design labels listed in Table 5.1. This captures the notion that there are commonalities in aesthetic appearance among all types of vehicles. For example, all vehicles have headlights and wheels. As a result, features learned from all vehicles may better capture the appearance of SUVs as opposed to only training the cGAN with images of SUVs.

The cGAN was trained using the ADAM optimizer (*Kingma and Ba*, 2014) to minimize the loss functions of the *discriminator* and *generator*, i.e $\mathbf{J}_\mathcal{D}$ and $\mathbf{J}_\mathcal{G}$ in Equations (5.4) and (5.5). Specifically, we propagate the gradients of the loss function of the *discriminator* $\mathbf{J}_\mathcal{D}$ once, then propagate the gradients of the loss function of the *generator* $\mathbf{J}_\mathcal{G}$ twice. This sequential training procedure aims at avoiding the discriminator improving too quickly relative to the generator. Moreover, we maintain disentanglement of 2D images and their conditioned labels by training on combinations of real/wrong images with real/wrong labels as described in (*Reed et al.*, 2016).

Figure 5.6 shows randomly generated vehicle images using the cGAN generator.

Figure 5.6: Randomly generated vehicle designs from the cGAN generator. These images provide evidence the cGAN is capturing the data distribution of vehicles, particularly with more realism than similar approaches by the authors such as variational autoencoders.

Figure 5.7: Visualization of salient design regions for the 2014 Range Rover Sport. The first row shows salient regions for 'Suburban' 'Women,' while the second row shows salient regions for 'Rich' 'Men' 'Over 40."

Though not the focus on this work, these images provide a sanity check that the cGAN is capturing the distribution of vehicles in the 2D image space. Note that these images, while plausibly real, do not exist in the training data set. The model required cGAN hyperparameter tuning to achieve aesthetic realism, noting that conventional metrics such as sample loss or pixel-wise distance have been shown to produce images of poor aesthetic quality (*Theis et al.*, 2015).

5.4.2.3 Siamese Network Training

We trained the Siamese network by minimizing the negative log likelihood given in Equation (5.7), using the ADAM optimizer over minibatches of training data (*Kingma and Ba*, 2014). Training was improved by updating only portions of the Siamese network to maintain relative information flow between portions of the cGAN and the randomly initialized portions of the Siamese network. Moreover, we applied batch normalization for every convolutional layer in Figure 5.3.

	Siamese Net with Image Features, Design labels, and Customer labels	Siamese Net with Design labels and Customer labels
Attribute	Accuracy (Std.Dev)	Accuracy (Std. Dev)
Sporty	75.07 (0.33)	69.17 (0.15)
Appealing	67.29 (0.18)	64.82(0.24)
Innovative	75.44 (0.39)	74.89(0.09)
Luxurious	75.09 (0.13)	74.53 (0.18)

Table 5.3: Averaged prediction accuracy and its standard deviation on hold-out test data using the Siamese Net with image features, design labels, customer labels or only with the design and customer labels. Average and standard deviation were calculated from 5 random training and testing splits common to each method.

5.4.3 Model Accuracy

The Siamese network achieves different testing accuracies depending on the design attribute as given in Table 5.3. As a sanity check, a Siamese network with the same architecture as shown in Figure 5.3, but *without* pretrained 2D image features from the cGAN, achieves lower prediction accuracy on all four design attributes. This suggests the Siamese network architecture is learning how a given customers perceives SUV design attributes.

5.4.4 Visualization of Aesthetic Saliency

We turn our attention to visualizing the model in order to interpret "why" a customer perceives a SUV across aeshtetic design attributes such as 'Sporty.' Moreover, we demonstrate that we are able to perform this visual interpretation for customers in differing market segments.

In particular, we analyze salient regions of a 2014 Land Rover Range Rover Sport for the design attribute: 'Sporty'. From internal research in General Motors, one market segment of the Range Rover Sport is suburban women who opt for a classy SUV. Another market segment is rich men over 40 who want to project proclivities

for off-road adventures. By filtering our customer data according to these criteria, we obtain two separate datasets, one for each predefined market segment. Among customers who ranked the 'Sportiness' of the 2014 Range Rover Sport, there were 15 women living in suburban regions with a family size larger than 2. Similarly, there were 12 men with an age greater than 40 and annual income more than $50,000.

Figure 5.7 shows salient regions for each market segment, corresponding to the 'Sportiness' of the 2014 Range Rover Sport. To obtain these regions, we computed the saliency map of the market segments using guided backpropagation as detailed in Section 5.3.3, then filtered pixels in the saliency map using a threshold of $[-3\sigma, 3\sigma]$, where σ is the standard deviation of pixel values in the saliency map. In other words, only pixels with an absolute value larger than 3σ are considered salient pixels.

5.5 Contributions and Limitations

5.5.1 Contributions to Product Design

The high-level goal of this research is to address the three design questions introduced in Section 5.1, in the context of the proposed deep learning model. These design questions are addressed below using quantitative metrics, as well as qualitative interpretation using input from designers and marketers at General Motors:

(1) Does the product design achieve desired aesthetic design attributes for a given market segment?

As detailed in Section 5.4.3, the Siamese network was able to predict the design attribute 'Sporty' to 75.07% accuracy, the design attribute 'Appealing' to 67.29%, the design attribute 'Innovative' to 75.44 %, and the design attribute 'Luxurious' to 75.09%, using a hold-out testing dataset. This is evidence that the proposed approach has utility in helping product designers and executives understand whether given design achieves a desired aesthetic design attribute.

Many design decisions rely on the designer's ability to predict how those choices will affect the perceived design attributes. Along with the many decisions each designer makes in developing the design, these decisions also include executive design reviews and selection of designs. The prediction obtained using our approach not only has relatively high prediction accuracy but also captures the heterogeneity of the market, which can help company decision makers understand how each design will be perceived in the multiple markets which it is aimed. Moreover, our approach allows high capacity and flexibility of testing a large number of new designs within a brief period of time, while traditional market researches (e.g., focus groups, surveys) require much more time and resources and introduce confidentiality issues, all of which our approach avoids.

(2) Where do salient design regions exist on the product for a given design attribute?
As shown in Figure 5.7, we are able to visualize salient regions of a 2014 Range Rover for the aesthetic design attribute 'Sportiness.' These regions are shown to differ depending on the perceiver's demographics and presumably, viewpoint. Identification of these salient design regions can help designers interpret and better understand which elements of the design are most responsible for the customer's perception. Such information is incredibly important to designers as they relate physical design details to psychological customer reactions.

(3) How do salient design regions differ across different market segments?
As shown in Figure 5.7, there are some commonalities between the salient regions for suburban women and rich men over 40. For example, the design of the lower front face (shown in the front view) and the side mirrors (shown in the front and rear view) are common salient regions for both market segments. There are also common regions which are not salient regions for both market segments such as the lower part of the side doors (shown in the side view).

There are also interesting differences between the two market segments. In gen-

eral, the salient regions of suburban women cover a larger proportion of the design than those of rich men over 40. This indicates either that suburban women are more sensitive to design appearance details than rich men over 40, or that they are processing the stimuli more as Gestalts than as individual elements.

These design details include the shape of the back of the car, as shown in the images in the second column (from the 30 degree isometric viewpoint). Also, in line with lifestyle differences, these suburban women with a family seem to be more attentive to rear seat headroom (see the third and fourth columns of Figure 5.7); these customers may be more likely to have rear seat passengers and may be assessing this functionality as part of their overall assessment. Salient design regions help the designer learn the general relationships between his or her design actions and the perceptual results, which adds to the long term skill development of the design community. When coupled with other marketing analysis and cognitive study (e.g., eye tracking), our approach will enable deeper insights into the market segments and what differentiates their aesthetic reactions.

5.5.2 Limitations

The aesthetic design processes at global enterprises use a number of approaches to understand perceptions of design attributes for heterogeneous markets, with approaches related to the current work including design theme studies and focus groups. In these approaches, customers from various market segments around the world assess baseline and concept designs on design attributes using in-person design stimuli, with follow-up discussion in focus groups.

While these approaches are able to gather rich customer response data, they are not scalable to hundreds or thousands of distinct market segments across the world. This offers promising opportunity for the proposed research as a complement to existing aesthetic design approaches at multinational product design companies. Advanc-

ing this deep learning approach into practice, however, requires overcoming a number of limitations.

First, in contrast to many machine learning tasks focused on increasing prediction accuracy, such as optimal ad placement for advertising companies such as Google, understanding the underlying factors affecting heterogeneous use perceptions are most important for this work. For example, how the ordering of perceptual stimuli affect the construction of customer preferences, which may suggest the layout of such information presenting design options to customers. This provides an opportunity for machine learning algorithms such as the one used here to inform design process.

This work may thus provide a test bed for design-specific and psychological questions, such as differences between binary choice and partial ranking tasks under various mediums. This may be particularly relevant given the ongoing shift to internet-based information seeking of customers. Important in this direction are consistency metrics that measure the difference between the salient regions of known similar customers. Such metrics have proven challenging, due to the mismatch between common quantitative metrics for model evaluation and the realism of designs encoded using generative models (*Theis et al.*, 2015).

Second, the prediction accuracy of the current work must be increased before designers may have full confidence in predicted answers to aesthetic design questions. Increasing this prediction accuracy may take a number of directions. Collecting more customer data may significantly improve accuracy, both by simply having more data, but perhaps more importantly, by having more customer variables. One can imagine customer variables, such as 'hobbies' and 'environmental consciousness,' may provide a much richer representation of customers with regards to their aesthetic preferences.

Architectural changes to this deep learning approach, beyond the Siamese network, may provide additional opportunities for improved accuracy. While details are not reported, a number of similar approaches were attempted before the current ar-

chitecture was selected. Pretrained neural networks did not result in useful image features, and in fact, reduced aesthetic prediction accuracy below baseline levels due to the increase in parameters. Similarly, generative approaches such as the use of variational autoencoders did not improve prediction accuracies. The authors suggest this is likely due to the low realism of the current state-of-the-art of generative modeling. In this direction, recent results in stacking multiscale generative adversarial networks has shown impressive capture of the underlying data distribution (*Zhang et al.*, 2016). Moreover, changing the prediction task itself to better capture the human perception process will likely improve accuracy; for example, changing to a ranking output task.

Third, validation of the proposed deep learning approach requires additional study. High prediction accuracy does not necessarily lead to valid answers to design questions (*Theis et al.*, 2015). For example, learned feature representations may lead to highly distributed encodings that are efficient for separation of data in the feature space rather than localized encodings that more representative of human perceptions over design. A possible direction to validate our approach is to cross-reference findings from design theme clinics and focus groups, or use experiment-based methods such as eye-tracking (*Reid et al.*, 2013; *Marshall et al.*, 2014). The generalizability of the proposed approach can be validated by the studies on other products besides vehicles.

There are many interesting future directions. For example, the generative model used in our approach provides a possibility of using the deep learning model to generate new designs with the desired aesthetic attributes.

5.6   Summary

Aesthetic appeal is of critical importance to customer-centric product designs such as automobiles. This creates an ongoing challenge for designers that aim to understand the factors influencing a customer's aesthetic perception over design attributes. Exacerbating this challenge is the scale at which such an understanding is undertaken,

with global enterprises designing for hundreds or thousands of heterogeneous market segments.

We introduced a research approach to predict and interpret customer perceptions of design attributes for heterogeneous markets. Specifically, we build on recent advances in deep learning and develop a Siamese neural network containing a pair of conditional generative adversarial networks. This model takes as input 2D design images and associated labels, customer data corresponding to heterogeneous market segments, and the perceptions of these customers across aesthetic design attributes.

A study was conducted to assess the utility of this research approach. A dataset consisting of automotive vehicles from 2000-2014, as well as customer data collected using an online crowdsourcing web application, was used to train the Siamese network. Our results show that this research approach is indeed able to predict design attributes across customers belonging to heterogeneous market segments. Further, we show visual interpretation of customer perceptions of design attributes for various market segments.

While this approach shows that the proposed research approach is viable in the context of scalable understanding of customer perceptions to aesthetic product design, a number of limitations must be overcome before this approach may be advanced to practice. At the same time, many of these limitations may be mitigated by recent advances in other areas of deep learning, as well as complementary approaches already used at multinational design enterprises.

CHAPTER VI

Conclusion

6.1 Dissertation Review

The general goal of this dissertation is to develop data-driven approaches that help designers understand and predict more reliably the product preference for a future product in a heterogeneous market, so that this understanding can inform the designers' decision-making. Specifically, we addressed four research questions in the context of visual aesthetics preference in previous chapters that are summarized below.

1. Does the product achieve the desired aesthetics design attributes for a given market segment?

   We proposed three methods that designers can use to answer this question. First, as detailed in Chap III, we developed an algorithm to aggregate crowd-sourced rankings into a numerical assessment of aesthetics design attributes for the given design image. This method provides an objective quantification of aesthetics design attributes so that designers can use this method to evaluate whether the product achieved the desired aesthetics design attributes in design review stage. Second, Chap IV introduced a model that can predict the aesthetics design attribute for a new design image using an L1 regularized linear

regression with features from a deep neural network. This method allows designers to quantify the aesthetic design attributes of a new design concept in the early stage of the design process. Third,in Chap V, we presented a Siamese network that was able to predict aesthetics design attributes for heterogeneous market with a high accuracy in held-out datasets. In summary, we developed three methods to help product designers and executives understand whether a design achieves a desired aesthetic design attribute for existing and new designs in homogeneous and heterogeneous markets.

2. How important are aesthetics design attributes when compared with functional attributes and price?

   As detailed in Chap III, we proposed an approach to determine the relative importance of aesthetic design attributes when both aesthetic design attributes and functional attributes are considered by customers. We test the proposed approach on an SUV aesthetics study. The results shows that aesthetics is the second important design attribute and only inferior to price, but superior to functional attributes.

3. What are the possible factors affecting customers perception of product aesthetics?

   In Chap IV and Chap V, we presented two approaches to visualize salient design regions for given aesthetic attributes and design images. These regions are shown to differ depending on the perceivers demographics and viewpoint. Identification of these salient design regions can help designers interpret and better understand which elements of the design are most responsible for the customers perception. Such information is important to designers as they relate physical design details to customers' aesthetic reactions.

4. How do these factors differ across different market segments?

In Chap V, the method can model the heterogeneous market by including customers' demographics information. Using this model, we were able to visualize the salient design regions for different market segments. Our SUV aesthetics study showed that the salient design regions differ depending on the respondents' demographics. Both commonalities and differences in salient design regions can be observed between groups. Qualitative interpretation using input from designers and marketers at General Motors suggested that these findings were interpretable and enabled deeper insights into the market segments and what differentiates their aesthetic reactions.

6.2   Dissertation Contributions

The main contribution of this dissertation is to demonstrate that we can quantify product preference using a purely data-driven approach in a way that has value for the practicing designer. We aim to provide designers a method to objectively measure the product preference. In addition, we aim to develop a number of quantitative aesthetics models that interpret how portions of the design space affect product aesthetics perception and preference in both homogeneous and heterogeneous markets. Moreover, these models are scalable to hundreds or thousands of markets, an important consideration for enterprises engaged in product design across globally dispersed markets. This contribution may be expanded as follows:

- We have quantitatively modeled the design process including aesthetic preferences. Using the proposed data-driven approaches, we have extracted objective measurement about product aesthetics and have visualized the salient design regions that contribute to the customers perception of design attributes. Those measurement and visualization contributions are important to designers as they relate physical design to psychological customer reactions.

126

- The proposed approach has utility in helping product designers and executives make design decisions more reliably. Moreover, those findings obtained via our approach can help the designer learn the general relationships between his or her design actions and the perceptual results, which adds to the long term skill development of the design community.

- This approach allows high capacity and flexibility of testing a large number of new designs within a brief period of time, while traditional market research (e.g., focus groups, surveys) requires more time and resources and introduces confidentiality issues, all of which our approach avoids.

Methodological contributions include introducing feature representation as an intermediate step in aesthetic preference modeling. This contribution may be expanded as follows:

- We have investigated various feature learning methods to model customers' preference. Specifically, we have used the features that reflect marketing insights and demographic data (Chap II), the features that are hand-engineered (Chap III), the features that are originally designed for an object classification (Chap IV), and the features that are from a generative model (Chap V). We have shown that using feature learning methods can increase the preference prediction accuracy and offer deeper insights to inform design decisions. Moreover, we have shown that not all feature learning methods can be used to model preference (e.g. PCA and RBM in Chap II), as these features may not be able to capture the preference rationale or cannot be interpreted by designers.

- We have demonstrated four approaches to interpret the features and preference function: identifying the optimal design with preference as the objective (Chap II), used controlled experiments (Chap III), approximated the inverse of the preference function (Chap IV), and visualized salient design regions (Chap

V). Our experiments using real design and marketing data have shown that those approaches can obtain operational useful insights into customer aesthetic perception and preference.

- We have addresses how to use feature representation to include multimodal data as input. In particular, we have demonstrated training the aesthetic preference model using large-scale multimodal data including 2D images, numerical labels, and crowdsourced aesthetic response data.

6.3 Future work

We have discussed a number of limitations existing in our research as well as a number of future directions in previous chapters. We summarized three major future directions of this dissertation below:

The first major direction of future work is the rigorous validation of the quantitative aesthetics model. There are three aspects for validation: theoretical analysis, empirical study, and cross-reference. In spite of its high predictive accuracy, it is important to validate the proposed quantitative model via theoretical analysis, which can provide designers statistical evidence on how confident designers should be toward the design decision that affected by the quantitative model. In addition, the operational usefulness of the proposed models should be tested with empirical studies, for example, an experiment on whether designers are able to improve the product aesthetics appealing using the insights from the proposed models. Moreover, these insights can be compared with descriptive research in aesthetics as cross-reference.

The second major direction of future work is to generate new design concepts using quantitative aesthetics preference model. In this way, we may obtain novel design concepts that have desired aesthetics. These generative design concepts may be a promising approach to exploring design space and inspiring design ideation. The

deep learning approach presented in chapter 5 has started this direction by including a conditional generative model that can generate new design images accordingly. Much more work needs to be done to increase the realism and the semantic meaning of the generated designs.

The third major direction of future work is to generalize the data-driven aesthetics preference model to senses other than the visual sense, such as hearing, touch, smell, taste, and motion. In particular, research in the machine learning community has developed feature learning methods to encode audio input into feature representations. These feature representations have been successfully applied to tasks involving audio information.

BIBLIOGRAPHY

BIBLIOGRAPHY

Abernethy, J., T. Evgeniou, O. Toubia, and J.-P. Vert (2008), Eliciting consumer preferences using robust adaptive choice questionnaires, *IEEE Transactions on Knowledge and Data Engineering*, *20*(2), 145–155.

Althaus, S. L. (2003), *Collective preferences in democratic politics: Opinion surveys and the will of the people*, Cambridge University Press.

Ben-Akiva, M., et al. (1999), Extended framework for modeling choice behavior, *Marketing letters*, *10*(3), 187–203.

Berg, L. P., and J. M. Vance (2017), Industry use of virtual reality in product design and manufacturing: a survey, *Virtual reality*, *21*(1), 1–17.

Berkovec, J., and J. Rust (1985), A nested logit model of automobile holdings for one vehicle households, *Transportation Research Part B: Methodological*, *19*(4), 275–285.

Berlyne, D. E. (1971), *Aesthetics and psychobiology*, vol. 336, JSTOR.

Birkhoff, G. D. (1933), *Aesthetic measure*, vol. 38, Harvard University Press Cambridge.

Birol, E., K. Karousakis, and P. Koundouri (2006), Using a choice experiment to account for preference heterogeneity in wetland attributes: the case of cheimaditida wetland in greece, *Ecological economics*, *60*(1), 145–156.

Bloch, P. H. (1995), Seeking the ideal form: Product design and consumer response, *The Journal of Marketing*, pp. 16–29.

Bodenhofer, U., and F. Klawonn (2004), A formal study of linearity axioms for fuzzy orderings, *Fuzzy Sets and Systems*, *145*(3), 323–354.

Bottou, L. (2010), Large-scale machine learning with stochastic gradient descent, in *Proceedings of COMPSTAT'2010*, pp. 177–186, Springer.

Brin, S., and L. Page (2012), Reprint of: The anatomy of a large-scale hypertextual web search engine, *Computer networks*, *56*(18), 3825–3833.

Burnap, A., Y. Ren, H. Lee, R. Gonzalez, and P. Y. Papalambros (2014), Improving preference prediction accuracy with feature learning, in *ASME 2014 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, pp. V02AT03A012–V02AT03A012, American Society of Mechanical Engineers.

Burnap, A., Y. Ren, R. Gerth, G. Papazoglou, R. Gonzalez, and P. Y. Papalambros (2015), When crowdsourcing fails: A study of expertise on crowdsourced design evaluation, *Journal of Mechanical Design*, *137*(3), 031,101.

Burnap, A., J. Hartley, Y. Pan, R. Gonzalez, and P. Y. Papalambros (2016a), Balancing design freedom and brand recognition in the evolution of automotive brand styling, *Design Science*, *2*.

Burnap, A., Y. Liu, Y. Pan, H. Lee, R. Gonzalez, and P. Y. Papalambros (2016b), Estimating and exploring the product form design space using deep generative models, in *ASME 2016 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, pp. V02AT03A013–V02AT03A013, American Society of Mechanical Engineers.

Burnap, A., Y. Pan, Y. Liu, Y. Ren, H. Lee, R. Gonzalez, and P. Y. Papalambros (2016c), Improving design preference prediction accuracy using feature learning, *Journal of Mechanical Design*, *138*(7), 071,404.

Buscher, G., E. Cutrell, and M. R. Morris (2009), What do you see when you're surfing?: using eye tracking to predict salient regions of web pages, in *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 21–30, ACM.

Buswell, G. T. (1935), How people look at pictures: a study of the psychology and perception in art.

Chang, Y.-M., C. Chu, and M. Ma (2013), Exploring the visual cognitive features on the design of car based on the theory of eye-tracking technology, *Przeglad Elektrotechniczny*, *89*(1), 143–146.

Chapelle, O., and Z. Harchaoui (2004), A machine learning approach to conjoint analysis, *Advances in Neural Information Processing Systems*, pp. 257–264.

Chen, W., C. Hoyle, and H. J. Wassenaar (2013), *Decision-Based Design*, Springer London, London.

Chopra, S., R. Hadsell, and Y. LeCun (2005), Learning a similarity metric discriminatively, with application to face verification, in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 1, pp. 539–546, IEEE.

Chrome Systems Inc. (2008), Chrome New Vehicle Database, Information inline at: http://www.chrome.com.

Coates, D. (2003), *Watches tell more than time: Product design, information, and the quest for elegance*, McGraw-Hill London.

Collobert, R., and J. Weston (2008), A unified architecture for natural language processing: Deep neural networks with multitask learning, in *Proceedings of the 25th international conference on Machine learning*, pp. 160–167, ACM.

Crilly, N., J. Moultrie, and P. J. Clarkson (2004), Seeing things: consumer response to the visual domain in product design, *Design studies*, *25*(6), 547–577.

Dalal, N., and B. Triggs (2005), Histograms of oriented gradients for human detection, in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 1, pp. 886–893, IEEE.

Deng, J., W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei (2009), Imagenet: A large-scale hierarchical image database, in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pp. 248–255, IEEE.

Dobrian, J. (2016), 2016 u.s. apeal study: Safety features score big, boosting new-vehicle appeal, http://www.jdpower.com/cars/articles/jd-power-studies/2016-us-apeal-study-results.

Du, P., and E. F. MacDonald (2014), Eye-tracking data predict importance of product features and saliency of size change, *Journal of Mechanical Design*, *136*(8), 081,005.

Duchowski, A. T. (2002), A breadth-first survey of eye-tracking applications, *Behavior Research Methods, Instruments, & Computers*, *34*(4), 455–470.

Dukerich, J. M., B. R. Golden, and S. M. Shortell (2002), Beauty is in the eye of the beholder: The impact of organizational identification, identity, and image on the cooperative behaviors of physicians, *Administrative Science Quarterly*, *47*(3), 507–533.

Durgee, J. F. (1988), Product drama., *Journal of Advertising Research*.

Dym, C. L., A. M. Agogino, O. Eris, D. D. Frey, and L. J. Leifer (2005), Engineering design thinking, teaching, and learning, *Journal of Engineering Education*, *94*(1), 103–120.

Evgeniou, T., C. Boussios, and G. Zacharia (2005), Generalized robust conjoint estimation, *Marketing Science*, *24*(3), 415–429.

Evgeniou, T., M. Pontil, and O. Toubia (2007), A convex optimization approach to modeling consumer heterogeneity in conjoint estimation, *Marketing Science*, *26*(6), 805–818.

Fazel, M. (2002), Matrix rank minimization with applications, Ph.D. thesis.

Feick, L., and R. A. Higie (1992), The effects of preference heterogeneity and source characteristics on ad processing and judgements about endorsers, *Journal of Advertising*, *21*(2), 9–24.

Friedman, J., T. Hastie, and R. Tibshirani (2001), *The elements of statistical learning*, vol. 1, Springer series in statistics Springer, Berlin.

Fuge, M. (2015), A scalpel not a sword: On the role of statistical tests in design cognition, in *ASME 2015 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, pp. 1–11, American Society of Mechanical Engineers.

Furuya, T., S. Kuriyama, and R. Ohbuchi (2015), An unsupervised approach for comparing styles of illustrations, in *Content-Based Multimedia Indexing (CBMI), 2015 13th International Workshop on*, pp. 1–6, IEEE.

Gagniuc, P. A. (2017), *Markov Chains: From Theory to Implementation and Experimentation*, John Wiley & Sons.

Garces, E., A. Agarwala, D. Gutierrez, and A. Hertzmann (2014), A similarity measure for illustration style, *ACM Transactions on Graphics (TOG)*, *33*(4), 93.

Gatys, L. A., A. S. Ecker, and M. Bethge (2015), A neural algorithm of artistic style, *arXiv preprint arXiv:1508.06576*.

Girshick, R., J. Donahue, T. Darrell, and J. Malik (2014), Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation, pp. 580–587, IEEE, doi:10.1109/CVPR.2014.81.

Goldstein, E. B., and J. Brockmole (2016), *Sensation and perception*, Cengage Learning.

Gonzalez, R., and G. Wu (1999), On the shape of the probability weighting function, *Cognitive psychology*, *38*(1), 129–166.

Green, P. E., and V. Srinivasan (1990), Conjoint analysis in marketing: new developments with implications for research and practice, *The journal of marketing*, pp. 3–19.

Guyon, I., and A. Elisseeff (2003), An introduction to variable and feature selection, *The Journal of Machine Learning Research*, *3*, 1157–1182.

Hauser, J. R., and V. R. Rao (2004), Conjoint analysis, related modeling, and applications, *Advances in Marketing Research: Progress and Prospects*, pp. 141–68.

He, L., M. Wang, W. Chen, and G. Conzelmann (2014), Incorporating social impact on new product adoption in choice modeling: A case study in green vehicles, *Transportation Research Part D: Transport and Environment*, *32*, 421–434.

Hekkert, P. (2006), Design aesthetics: principles of pleasure in design, *Psychology science*, *48*(2), 157.

Hekkert, P., and H. Leder (2008), Product aesthetics, in *Product experience*, pp. 259–285, Elsevier.

Hekkert, P., D. Snelders, and P. C. Wieringen (2003), most advanced, yet acceptable: Typicality and novelty as joint predictors of aesthetic preference in industrial design, *British journal of Psychology*, *94*(1), 111–124.

Hekkert, P. P. M. (1995), Artful judgements: A psychological inquiry into aesthetic preference for visual patterns, Ph.D. thesis, TU Delft, Delft University of Technology.

Hinton, G., et al. (2012), Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups, *Signal Processing Magazine, IEEE*, *29*(6), 82–97.

Hinton, G. E. (2002), Training products of experts by minimizing contrastive divergence, *Neural computation*, *14*(8), 1771–1800.

Holt, C. A. (1986), Preference reversals and the independence axiom, *The American Economic Review*, *76*(3), 508–515.

Hubel, D. H., and T. N. Wiesel (1962), Receptive fields, binocular interaction and functional architecture in the cat's visual cortex, *The Journal of physiology*, *160*(1), 106–154.

Jansen, S., H. Boumeester, H. Coolen, R. Goetgeluk, and E. Molin (2009), The impact of including images in a conjoint measurement task: evidence from two small-scale studies, *Journal of housing and the built environment*, *24*(3), 271–297.

Jia, Y., E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell (2014), Caffe: Convolutional architecture for fast feature embedding, in *Proceedings of the 22nd ACM international conference on Multimedia*, pp. 675–678, ACM.

Karayev, S., M. Trentacoste, H. Han, A. Agarwala, T. Darrell, A. Hertzmann, and H. Winnemoeller (2013), Recognizing image style, *arXiv preprint arXiv:1311.3715*.

Kelly, J. C., P. Maheut, J.-F. Petiot, and P. Y. Papalambros (2011), Incorporating user shape preference in engineering design optimisation, *Journal of Engineering Design*, *22*(9), 627–650.

Khalid, H. M., and M. G. Helander (2006), Customer emotional needs in product design, *Concurrent Engineering*, *14*(3), 197–206.

Kingma, D., and J. Ba (2014), Adam: A method for stochastic optimization, *arXiv preprint arXiv:1412.6980*.

Kingma, D. P., and M. Welling (2013), Auto-encoding variational bayes, *arXiv preprint arXiv:1312.6114*.

Krizhevsky, A., I. Sutskever, and G. E. Hinton (2012), Imagenet classification with deep convolutional neural networks, in *Advances in neural information processing systems*, pp. 1097–1105.

Kumar, D., C. Hoyle, W. Chen, N. Wang, G. Gomez-Levi, and F. Koppelman (2007), Incorporating customer preferences and market trends in vehicle packaging design, *International Journal of Production Design*.

Leder, H., B. Belke, A. Oeberst, and D. Augustin (2004), A model of aesthetic appreciation and aesthetic judgments, *British journal of psychology*, *95*(4), 489–508.

Lee, H., A. Battle, R. Raina, and A. Y. Ng (2006), Efficient sparse coding algorithms, in *Advances in neural information processing systems*, pp. 801–808.

Lee, H., C. Ekanadham, and A. Y. Ng (2008), Sparse deep belief net model for visual area v2, in *Advances in neural information processing systems*, pp. 873–880.

Lee, H., R. Grosse, R. Ranganath, and A. Y. Ng (2011), Unsupervised learning of hierarchical representations with convolutional deep belief networks, *Communications of the Association for Computing Machinery*, *54*(10), 95–103.

Lenk, P. J., W. S. DeSarbo, P. E. Green, and M. R. Young (1996), Hierarchical bayes conjoint analysis: recovery of partworth heterogeneity from reduced experimental designs, *Marketing Science*, *15*(2), 173–191.

Lewis, K. E., W. Chen, L. C. Schmidt, and A. Press (2006), *Decision making in engineering design*, ASME Press New York.

Li, K., Z. Wu, K.-C. Peng, J. Ernst, and Y. Fu (2018), Tell me where to look: Guided attention inference network, *arXiv preprint arXiv:1802.10171*.

Li, L.-J., H. Su, L. Fei-Fei, and E. P. Xing (2010), Object bank: A high-level image representation for scene classification & semantic feature sparsification, in *Advances in neural information processing systems*, pp. 1378–1386.

Liu, G., and S. Yan (2014), Scalable low-rank representation, in *Low-Rank and Sparse Modeling for Visual Analysis*, edited by Y. Fu, pp. 39–60, Springer International Publishing.

Livingstone, M. S., and D. H. Hubel (1987), Psychophysical evidence for separate channels for the perception of form, color, movement, and depth, *The Journal of Neuroscience*, *7*(11), 3416–3468.

Mairal, J., J. Ponce, G. Sapiro, A. Zisserman, and F. R. Bach (2009), Supervised dictionary learning, in *Advances in neural information processing systems*, pp. 1033–1040.

Marĉelja, S. (1980), Mathematical description of the responses of simple cortical cells, *JOSA*, *70*(11), 1297–1300.

Maritz Research Inc. (2007), Maritz Research 2006 new vehicle customer satisfactions survey, Information online at: http://www.maritz.com.

Marshall, B. H., S. Sareen, J. A. Springer, and T. Reid (2014), Eye tracking data understanding for product representation studies, in *Proceedings of the 3rd annual conference on Research in information technology*, pp. 3–8, ACM.

McCormack, J. P., J. Cagan, and C. M. Vogel (2004), Speaking the buick language: capturing, understanding, and exploring brand identity with shape grammars, *Design studies*, *25*(1), 1–29.

McFadden, D., and K. Train (2000), Mixed MNL models for discrete response, *Journal of Applied Econometrics*, *15*(5), 447–470.

McManus, I., A. L. Jones, and J. Cottrell (1981), The aesthetics of colour, *Perception*, *10*(6), 651–666.

Meikle, J. (2010), *Twentieth century limited: Industrial design in America 1925-1939*, Temple University Press.

Michalek, J. J., F. M. Feinberg, and P. Y. Papalambros (2005), Linking marketing and engineering product design decisions via analytical target cascading*, *Journal of Product Innovation Management*, *22*(1), 42–62.

Mittelman, R., H. Lee, B. Kuipers, and S. Savarese (2013), Weakly supervised learning of mid-level features with Beta-Bernoulli process restricted Boltzmann machines, in *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pp. 476–483, IEEE.

Monö, R. G., M. Knight, and R. Monö (1997), *Design for product understanding: The aesthetics of design from a semiotic approach*, Liber.

Morrow, W. R., M. Long, and E. F. MacDonald (2014), Market-system design optimization with consider-then-choose models, *Journal of Mechanical Design*, *136*(3), 031,003.

Motors, G. (2014), *General Motors Internal Research*.

Moulson, T., and G. Sproles (2000), Styling strategy, *Business Horizons*, *43*(5), 45–52.

Nagamachi, M. (1995), Kansei engineering: a new ergonomic consumer-oriented technology for product development, *International Journal of Industrial Ergonomics*, *15*(1), 3–11.

Naik, N., J. Philipoom, R. Raskar, and C. Hidalgo (2014), Streetscore-predicting the perceived safety of one million streetscapes, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 779–785.

Netzer, O., et al. (2008), Beyond conjoint analysis: Advances in preference measurement, *Marketing Letters*, *19*(3-4), 337–354.

Nisbett, R. E., and T. D. Wilson (1977), Telling more than we can know: Verbal reports on mental processes., *Psychological review*, *84*(3), 231.

Norman, D. A. (2005), *Emotional design: Why we love (or hate) everyday things*, Basic books.

Orbay, G., L. Fu, and L. B. Kara (2015), Deciphering the influence of product shape on consumer judgments through geometric abstraction, *Journal of Mechanical Design*, *137*(8), 081,103.

Ordonez, V., and T. L. Berg (2014), Learning high-level judgments of urban perception, in *European Conference on Computer Vision*, pp. 494–510, Springer.

Orsborn, S., J. Cagan, R. Pawlicki, and R. C. Smith (2006), Creating cross-over vehicles: Defining and combining vehicle classes using shape grammars, *AIE EDAM: Artificial Intelligence for Engineering Design, Analysis, and Manufacturing*, *20*(03), 217–246.

Orsborn, S., J. Cagan, and P. Boatwright (2009), Quantifying aesthetic form preference in a utility function, *Journal of Mechanical Design*, *131*(6), 061,001.

Pan, Y., A. Burnap, Y. Liu, H. Lee, R. Gonzalez, and P. Papalambros (2016), A quantitative model for identifying regions of design visual attraction and application to automobile styling, in *Proceedings of the 2016 Internation Design Conference*.

Pan, Y., A. Burnap, J. Hartley, R. Gonzalez, and P. Y. Papalambros (2017), Deep design: Product aesthetics for heterogeneous markets, in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1961–1970, ACM.

Panchal, J. (2015), Using crowds in engineering designtowards a holistic framework, in *Proceedings of the 2015 International Conference on Engineering Design*, pp. 1–10, Design Society.

Papalambros, P. Y., and D. J. Wilde (2000), *Principles of optimal design: modeling and computation*, Cambridge university press.

Parikh, N., and S. Boyd (2013), Proximal algorithms, *Foundations and Trends in Optimization*, *1*(3).

Pazzani, M. J., and D. Billsus (2007), Content-based recommendation systems, in *The adaptive web*, pp. 325–341, Springer.

Petiot, J.-F., and A. Dagher (2011), Preference-oriented form design: application to cars headlights, *International Journal on Interactive Design and Manufacturing (IJIDeM)*, *5*(1), 17–27.

Poirson, E., J.-F. Petiot, L. Boivin, and D. Blumenthal (2013), Eliciting user perceptions using assessment tests based on an interactive genetic algorithm, *Journal of Mechanical Design*, *135*(3), 031,004.

Porzi, L., S. Rota Bulò, B. Lepri, and E. Ricci (2015), Predicting and understanding urban perception with convolutional neural networks, in *Proceedings of the 23rd ACM international conference on Multimedia*, pp. 139–148, ACM.

Pugliese, M. J., and J. Cagan (2002), Capturing a rebel: modeling the harley-davidson brand through a motorcycle shape grammar, *Research in Engineering Design*, *13*(3), 139–156.

Quercia, D., N. K. O'Hare, and H. Cramer (2014), Aesthetic capital: what makes london look beautiful, quiet, and happy?, in *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*, pp. 945–955, ACM.

Ranscombe, C., B. Hicks, G. Mullineux, and B. Singh (2012), Visually decomposing vehicle images: Exploring the influence of different aesthetic features on consumer perception of brand, *Design Studies*, *33*(4), 319–341.

Reber, R., N. Schwarz, and P. Winkielman (2004), Processing fluency and aesthetic pleasure: Is beauty in the perceiver's processing experience?, *Personality and social psychology review*, *8*(4), 364–382.

Reed, S., Z. Akata, X. Yan, L. Logeswaran, B. Schiele, and H. Lee (2016), Generative adversarial text to image synthesis, in *Proceedings of The 33rd International Conference on Machine Learning*, vol. 3.

Reid, T. N., R. D. Gonzalez, and P. Y. Papalambros (2010), Quantification of perceived environmental friendliness for vehicle silhouette design, *Journal of mechanical design*, *132*(10), 101,010.

Reid, T. N., B. D. Frischknecht, and P. Y. Papalambros (2012), Perceptual attributes in product design: Fuel economy and silhouette-based perceived environmental friendliness tradeoffs in automotive vehicle design, *Journal of Mechanical Design*, *134*, 041,006.

Reid, T. N., E. F. MacDonald, and P. Du (2013), Impact of product design representation on customer judgment, *Journal of Mechanical Design*, *135*(9), 091,008.

Ren, Y., A. Burnap, and P. Papalambros (2013), Quantification of perceptual design attributes using a crowd, in *DS 75-6: Proceedings of the 19th International Conference on Engineering Design (ICED13), Design for Harmonies, Vol. 6: Design Information and Knowledge, Seoul, Korea, 19-22.08. 2013*.

Reutskaja, E., R. Nagel, C. F. Camerer, and A. Rangel (2011), Search dynamics in consumer choice under time pressure: An eye-tracking study, *The American Economic Review*, *101*(2), 900–926.

Ribeiro, M. T., S. Singh, and C. Guestrin (2016), Why should i trust you?: Explaining the predictions of any classifier, in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1135–1144, ACM.

Ruckpaul, A., T. Fürstenhöfer, and S. Matthiesen (2015), Combination of eye tracking and think-aloud methods in engineering design research, in *Design Computing and Cognition'14*, pp. 81–97, Springer.

Salakhutdinov, R., and G. E. Hinton (2009), Deep boltzmann machines, in *International conference on artificial intelligence and statistics*, pp. 448–455.

Salakhutdinov, R., A. Mnih, and G. Hinton (2007), Restricted boltzmann machines for collaborative filtering, in *Proceedings of the 24th international conference on Machine learning*, pp. 791–798, ACM.

Selvaraju, R. R., M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra (2016), Grad-cam: Visual explanations from deep networks via gradient-based localization, *See https://arxiv. org/abs/1610.02391 v3*, *7*(8).

Shocker, A. D., M. Ben-Akiva, B. Boccara, and P. Nedungadi (1991), Consideration set influences on consumer decision-making and choice: Issues, models, and suggestions, *Marketing Letters*, *2*(3), 181–197.

Silvera, D. H., R. A. Josephs, and R. B. Giesler (2002), Bigger is better: The influence of physical size on aesthetic preference judgments, *Journal of Behavioral Decision Making*, *15*(3), 189–202.

Simonyan, K., and A. Zisserman (2014), Very deep convolutional networks for large-scale image recognition, *arXiv preprint arXiv:1409.1556*.

Smolensky, P. (1986), Parallel distributed processing: Explorations in the microstructure of cognition, vol. 1, chap. Information Processing in Dynamical Systems: Foundations of Harmony Theory, pp. 194–281, MIT Press, Cambridge, MA, USA.

Springenberg, J. T., A. Dosovitskiy, T. Brox, and M. Riedmiller (2014), Striving for simplicity: The all convolutional net, *arXiv preprint arXiv:1412.6806*.

Srivastava, N., and P. R. Schrater (2012), Rational inference of relative preferences, in *Advances in neural information processing systems*, pp. 2303–2311.

Sylcott, B., J. J. Michalek, and J. Cagan (2013), Towards understanding the role of interaction effects in visual conjoint analysis, in *ASME 2013 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, pp. V03AT03A012–V03AT03A012, American Society of Mechanical Engineers.

Tews, J. (2016), J.d. power initial quality study 2016, http://www.jdpower.com/press-releases/2016-us-initial-quality-study-iqs.

Theis, L., A. v. d. Oord, and M. Bethge (2015), A note on the evaluation of generative models, *arXiv preprint arXiv:1511.01844*.

Tomioka, R., T. Suzuki, M. Sugiyama, and H. Kashima (2010), A fast augmented lagrangian algorithm for learning low-rank matrices, in *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, pp. 1087–1094.

Torrance, G. W., W. Furlong, D. Feeny, and M. Boyle (1995), Multi-attribute preference functions, *Pharmacoeconomics*, *7*(6), 503–520.

Toubia, O., D. I. Simester, J. R. Hauser, and E. Dahan (2003), Fast polyhedral adaptive conjoint estimation, *Marketing Science*, *22*(3), 273–303.

Tovares, N., P. Boatwright, and J. Cagan (2014), Experiential conjoint analysis: An experience-based method for eliciting, capturing, and modeling consumer preference, *Journal of Mechanical Design*, *136*(10), 101,404.

Tuarob, S., and C. S. Tucker (2015), Automated discovery of lead users and latent product features by mining large scale social media networks, *Journal of Mechanical Design*, *137*(7), 071,402.

United States Census Bureau (2006), 2006 U.S. Census estimates, Information online at: http://www.census.gov.

van der Maaten, L. (2008), Visualizing data using t-sne, *Journal of Machine Learning Research*, *9*, 2579–2605.

Van Horn, D., and K. Lewis (2015), The use of analytics in the design of sociotechnical products, *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, *29*(01), 65–81.

Von Neumann, J., and O. Morgenstern (2007), *Theory of Games and Economic Behavior (60th Anniversary Commemorative Edition)*, Princeton University Press.

Wang, M., and W. Chen (2015), A data-driven network analysis approach to predicting customer choice sets for choice modeling in engineering design, *Journal of Mechanical Design*, *137*(7), 071,410.

Wang, Q., S. Yang, M. Liu, Z. Cao, and Q. Ma (2014), An eye-tracking study of website complexity from cognitive load perspective, *Decision support systems*, *62*, 1–10.

Wassenaar, H., W. Chen, J. Cheng, and A. Sudjianto (2005), Enhancing discrete choice demand modeling for decision-based design, *Journal of Mechanical Design*, *127*(4), 514–523.

Wassenaar, H. J., and W. Chen (2003), An approach to decision-based design with discrete choice analysis for demand modeling, *Journal of Mechanical Design*, *125*(3), 490–497.

Windhager, S., F. Hutzler, C.-C. Carbon, E. Oberzaucher, K. Schaefer, T. Thorstensen, H. Leder, and K. Grammer (2010), Laying eyes on headlights: Eye movements suggest facial features in cars, *Collegium antropologicum*, *34*(3), 1075–1080.

Yumer, M. E., S. Chaudhuri, J. K. Hodgins, and L. B. Kara (2015), Semantic shape editing using deformation handles, *ACM Transactions on Graphics (TOG)*, *34*(4), 86.

Zajonc, R. B. (1968), Attitudinal effects of mere exposure., *Journal of personality and social psychology*, *9*(2p2), 1.

Zeiler, M. D., and R. Fergus (2014), Visualizing and understanding convolutional networks, in *European conference on computer vision*, pp. 818–833, Springer.

Zhang, H., T. Xu, H. Li, S. Zhang, X. Huang, X. Wang, and D. Metaxas (2016), Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks, *arXiv preprint arXiv:1612.03242*.

Zhou, B., A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba (2016), Learning deep features for discriminative localization, in *Computer Vision and Pattern Recognition (CVPR), 2016 IEEE Conference on*, pp. 2921–2929, IEEE.