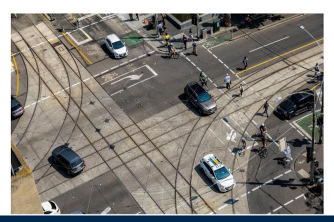
Infrastructure-based Detection and Localization of Road Users for Cooperative Autonomus Driving

Lance Bassett (Honors Capstone), Advisor: Henry Liu, Rusheng Zhang
UMTRI, University of Michigan, Ann Arbor, Michigan

ENGINEERING HONORS PROGRAM
UNIVERSITY OF MICHIGAN

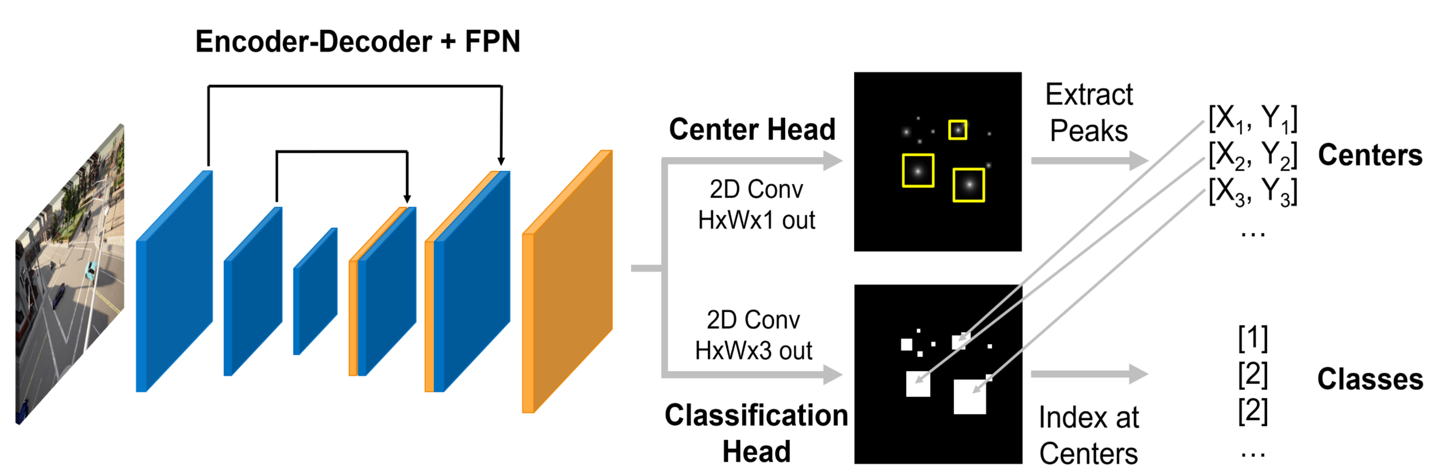
Introduction

Detection and localization of all road users is a difficult task to do well from a single on-road perspective (a vehicle's on-board sensors), but roadside units mounted in the infrastructure can provide a few distinct advantages. This project describes our detection pipeline and results, as well as some methods we use to improve results.

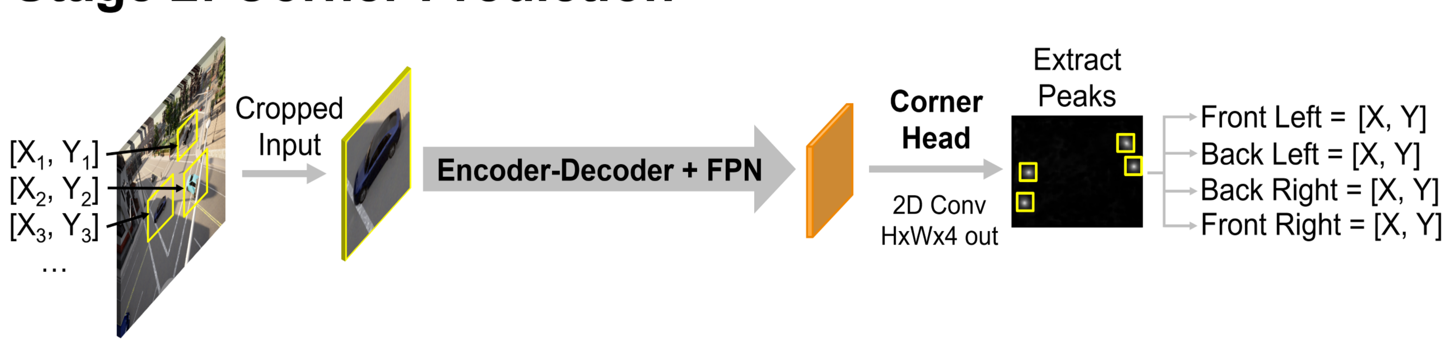

vs.


How it Works

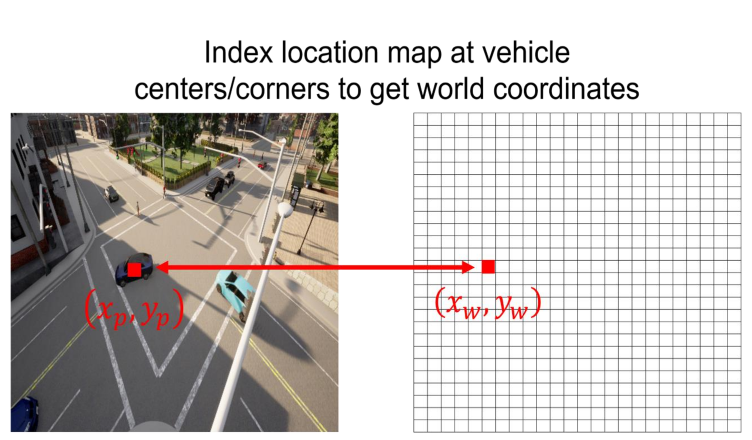
Stage 1: Center Prediction and Classification



Stage 2: Corner Prediction

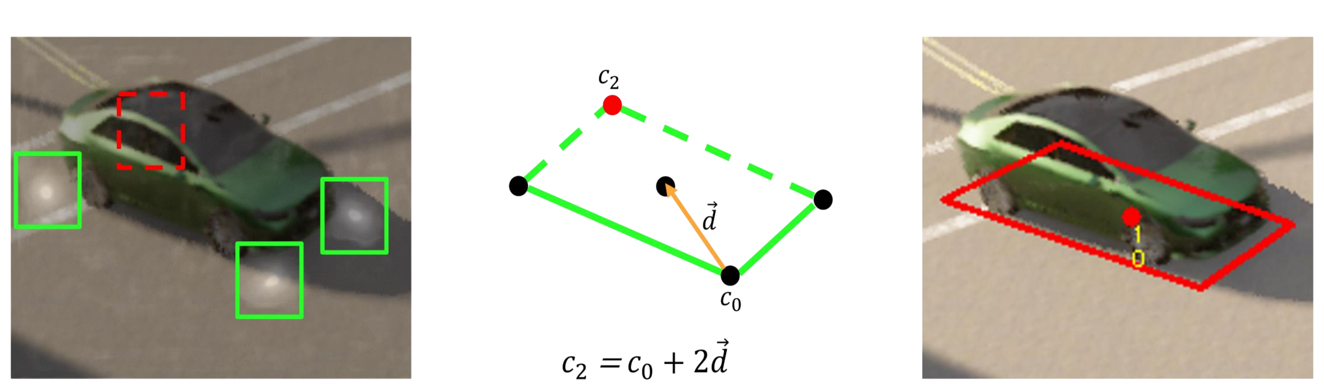


Stage 2.1: Pixel to World Coordinate Translation



$$map(x_{pixel}, y_{pixel}) = (x_{world}, y_{world})$$

Stage 2.2: Corner Completion

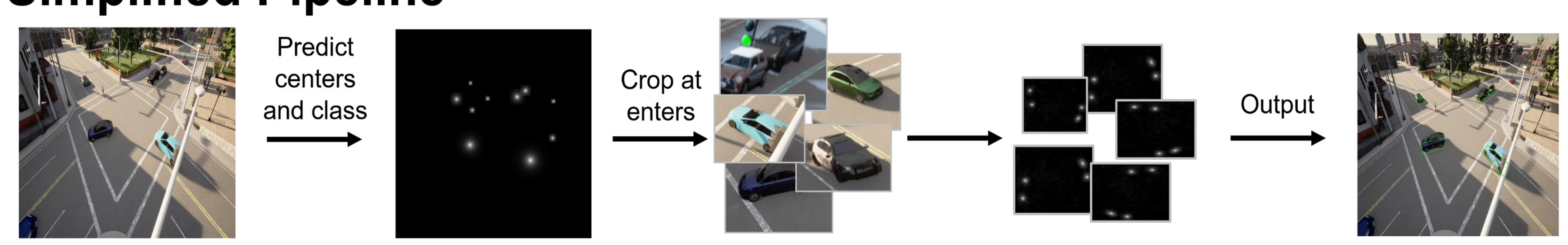


$$c_2 = c_0 + 2\vec{d}$$

Figure 1: Example of missing vehicle missing corner and subsequent completed result.

Results and Discussion

Simplified Pipeline



Sample Results

Figure 2: Sample results from two stage model. Center stage trained on 4.2k examples, crop stage trained on 35k examples.

Performance Metrics

Method	Center Pixel Error (px)	Corner Pixel Error (px)	Corner Global Error (m)
Two Stage	**(2.657, 1.239)**	**(0.984, 1.776)**	**(0.153, 0.170)**
Single Stage	(3.546, 1.197)	(1.471, 3.559)	(0.155, 0.207)

Predicting corners on crops allows for subpixel accuracy and lower error in both pixel location and global location

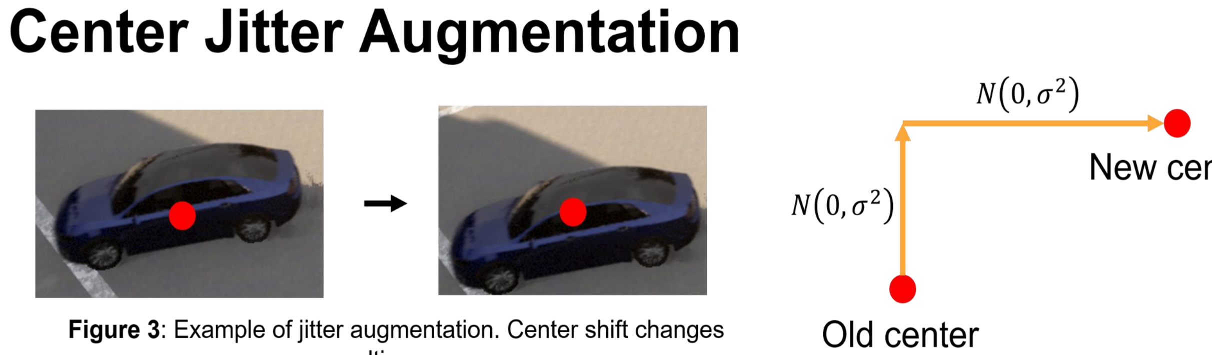
Table 1: Errors for our two different methods of detection. Single stage predicts corners and centers in the same step.

Method	Missed (n = 1936)	Detection Rate	Fixed
Two Stage	**26**	**.987**	72
Single Stage	48	.975	25

Table 2: Detection rate for our two different methods of detection.

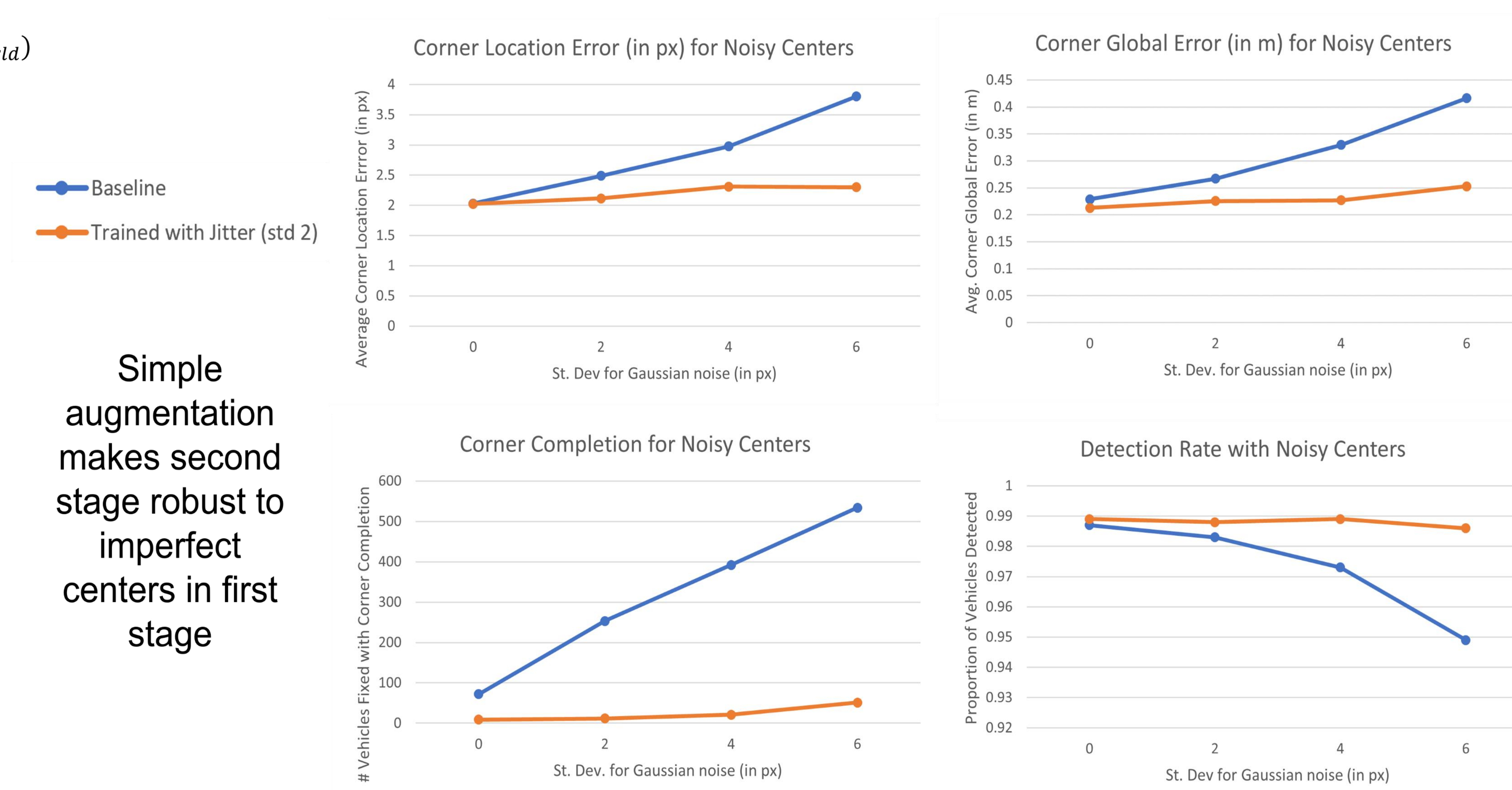
Missed come from corner overlap and mismatches, which limits benefits of corner completion

Center Jitter Augmentation



$$x_{jitter} = x_{center} + N(0, \sigma^2)$$
$$y_{jitter} = y_{center} + N(0, \sigma^2)$$

Figure 3: Example of jitter augmentation. Center shift changes resulting crop.



Simple augmentation makes second stage robust to imperfect centers in first stage

Advantages

- Second stage crop more universally deployable

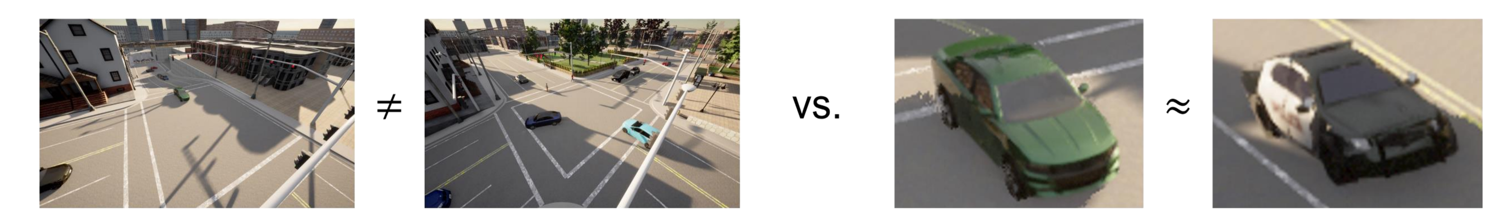
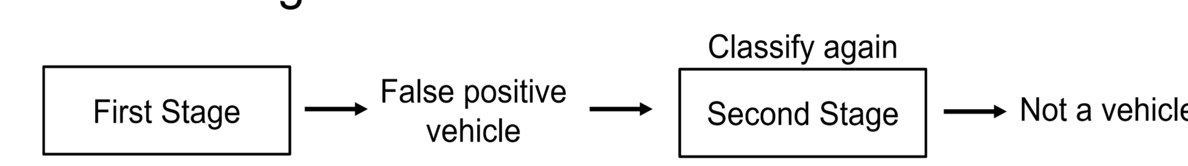


Figure 4: Comparison demonstrating low variability in crop of scene, despite high variability of actual scene.

- Second stage can refine first



- Capitalizes on fixed scene with preprocessed location maps
- Subpixel accuracy on cropped corners due to resize
- Avoids corner overlap and mismatches seen in single stage

Future Work

- Real intersection data
- Comprehensive Comparison
- Refining Predictions
- Multi-scene crop training
- Crop-size index map

References

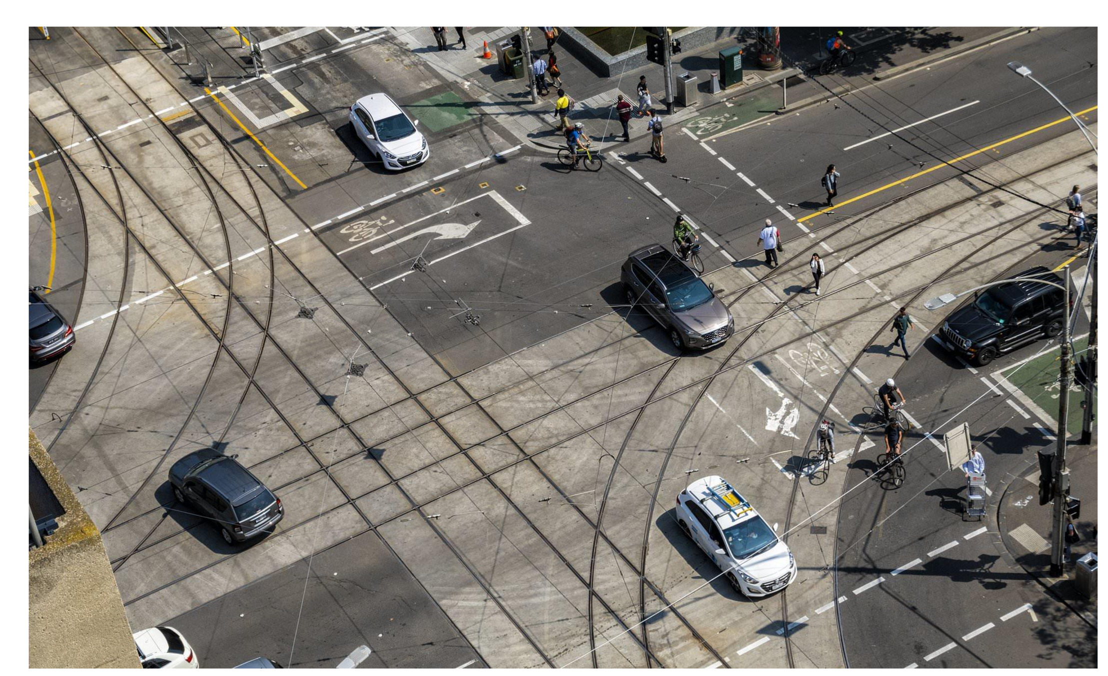
X. Zhou, D. Wang, and P. Krähenbühl, "Objects as points," *arXiv.org*, 25-Apr-2019. [Online]. Available: https://arxiv.org/abs/1904.07850v2. [Accessed: 19-Oct-2021].

ENGINEERING HONORS PROGRAM
UNIVERSITY OF MICHIGAN

Introduction

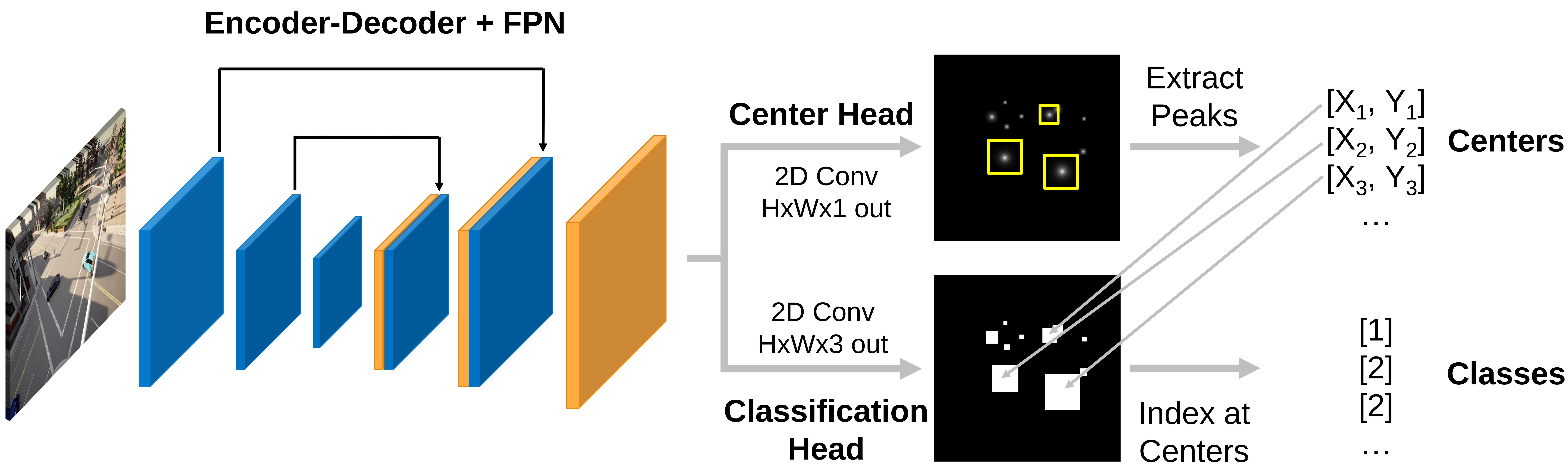
Detection and localization of all road users is a difficult task to do well from a single on-road perspective (a vehicle's on-board sensors), but roadside units mounted in the infrastructure can provide a few distinct advantages. This project describes our detection pipeline and results, as well as some methods we use to improve results.

vs.

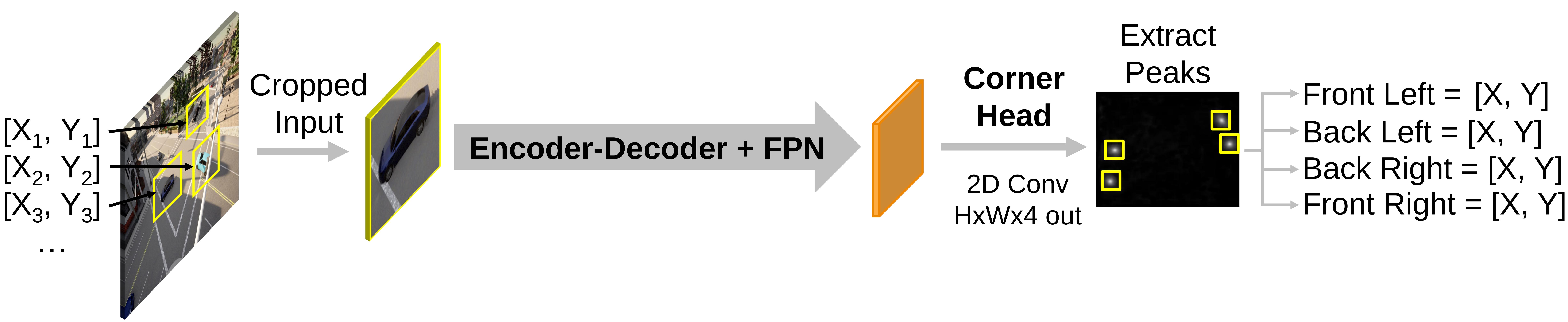


How it Works

Stage 1: Center Prediction and Classification

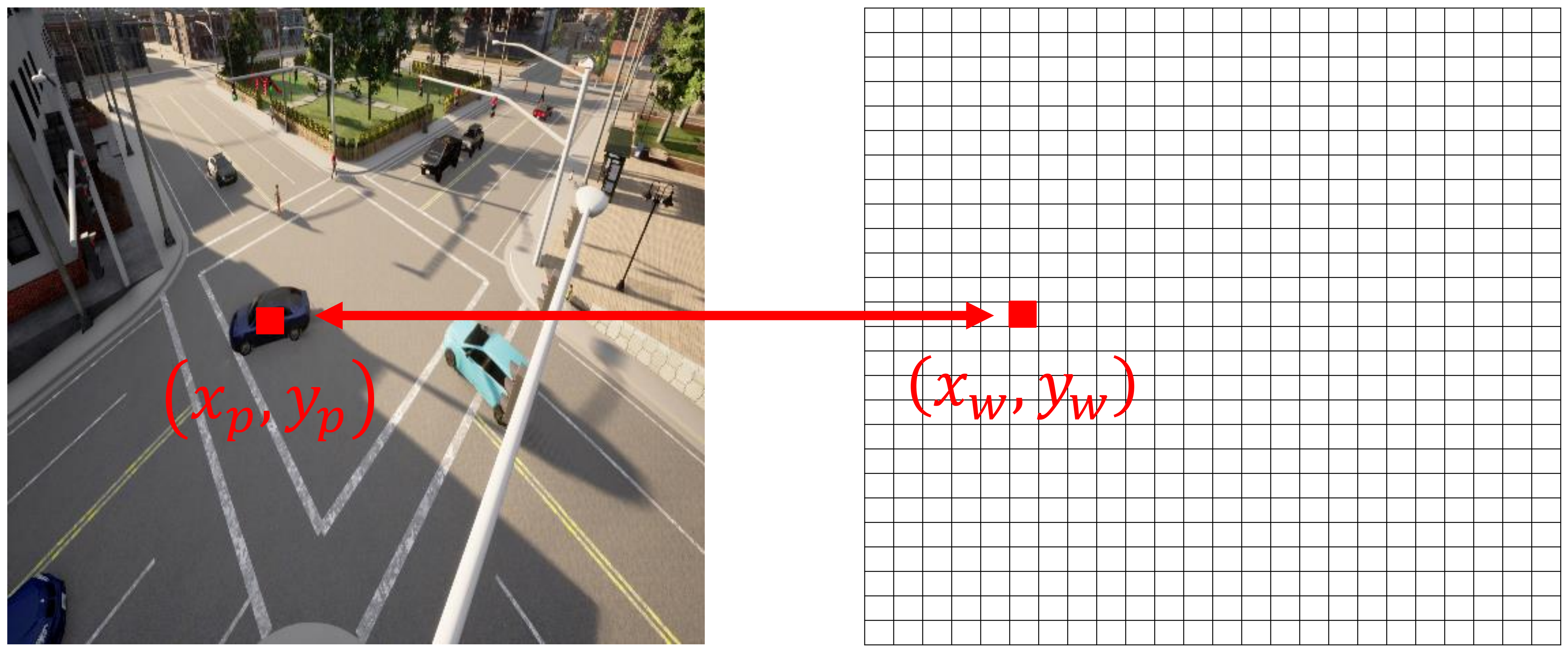


Stage 2: Corner Prediction



Stage 2.1: Pixel to World Coordinate Translation

Index location map at vehicle
centers/corners to get world coordinates



$$map(x_{pixel}, y_{pixel}) = (x_{world}, y_{world})$$

Stage 2.2: Corner Completion

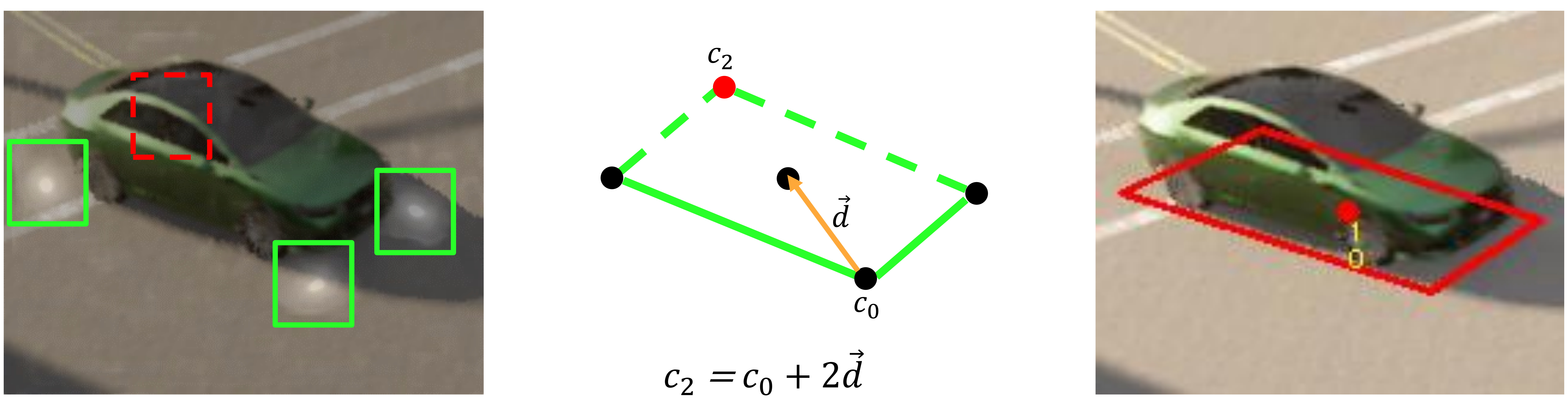
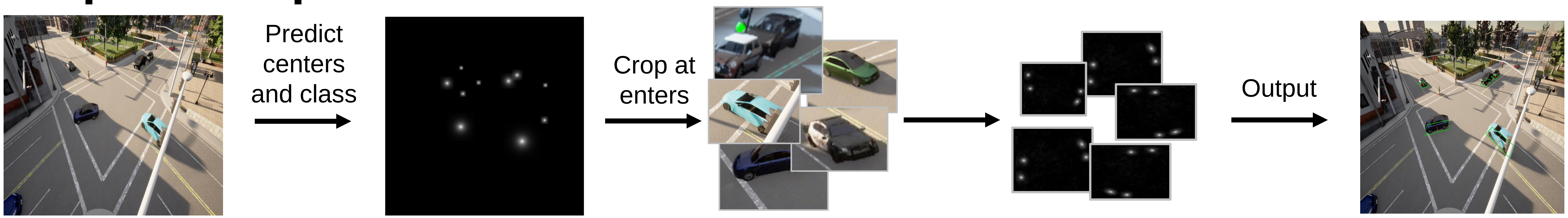


Figure 1: Example of missing vehicle missing corner and subsequent completed result.

Results and Discussion

Simplified Pipeline



Sample Results

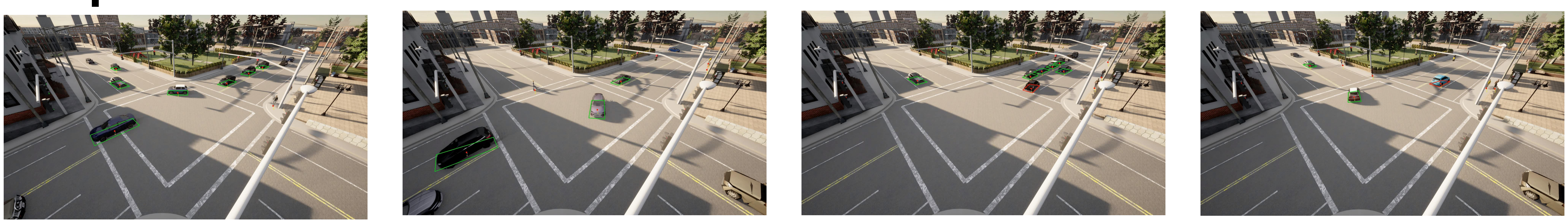


Figure 2: Sample results from two stage model. Center stage trained on 4.2k examples, crop stage trained on 35k examples.

Performance Metrics

Method	Center Pixel Error (px)	Corner Pixel Error (px)	Corner Global Error (m)
Two Stage	**(2.657, 1.239)**	**(0.984, 1.776)**	**(0.153, 0.170)**
Single Stage	(3.546, 1.197)	(1.471, 3.559)	(0.155, 0.207)

Predicting corners on crops allows for subpixel accuracy and lower error in both pixel location and global location

Table 1: Errors for our two different methods of detection. Single stage predicts corners and centers in the same step.

Method	Missed (n = 1936)	Detection Rate	Fixed
Two Stage	**26**	**.987**	72
Single Stage	48	.975	25

Missed come from corner overlap and mismatches, which limits benefits of corner completion

Table 2: Detection rate for our two different methods of detection.

Center Jitter Augmentation

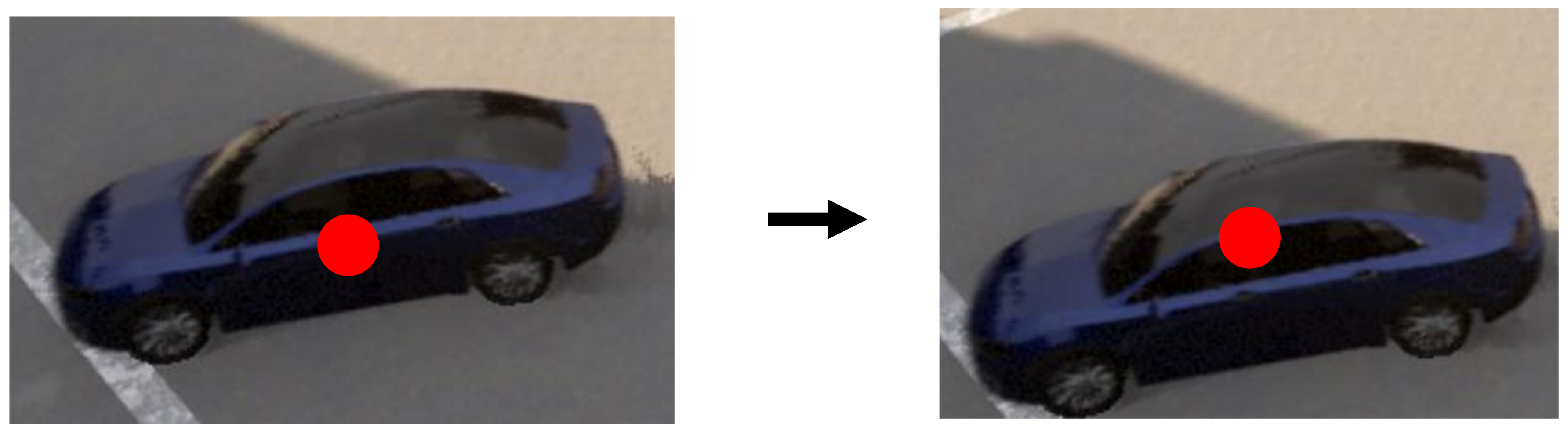
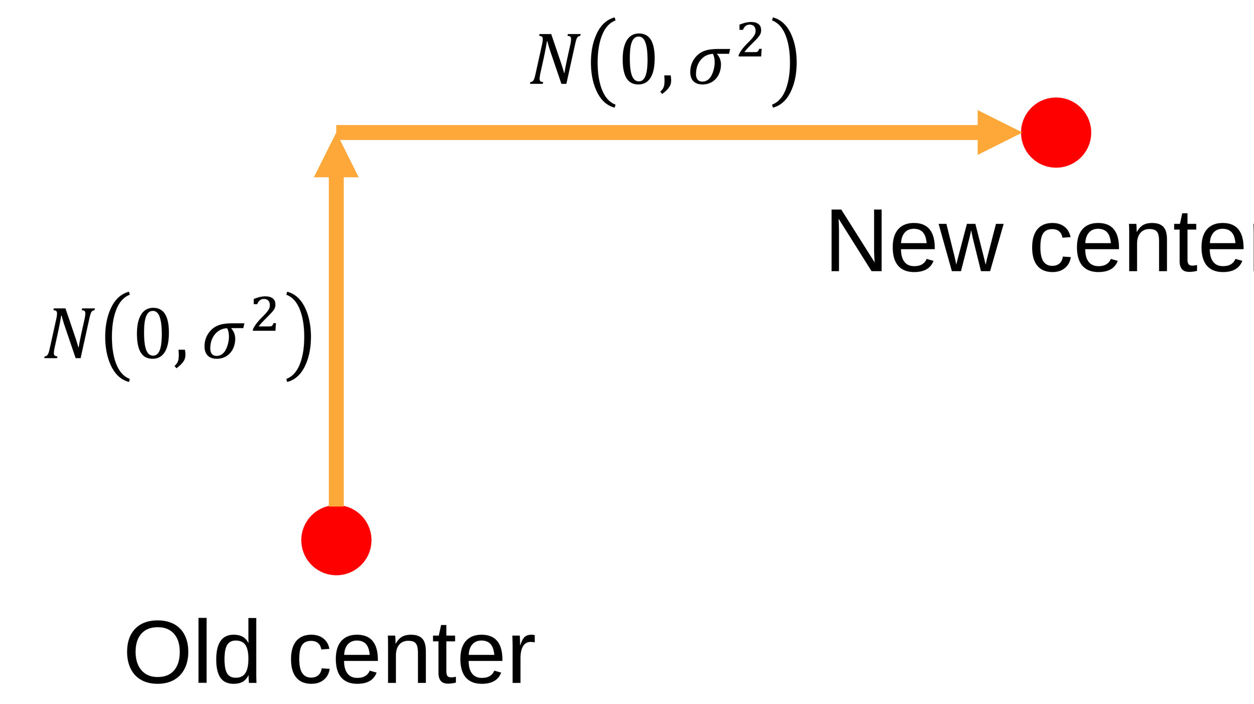
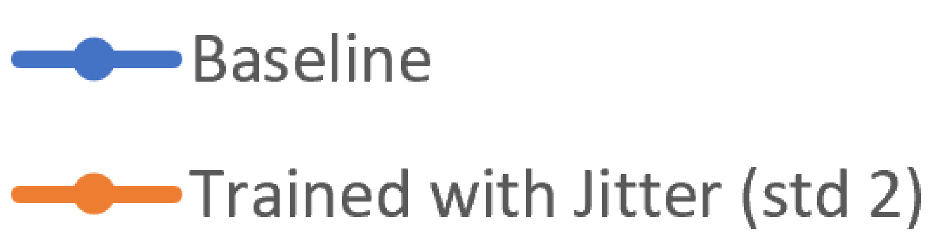
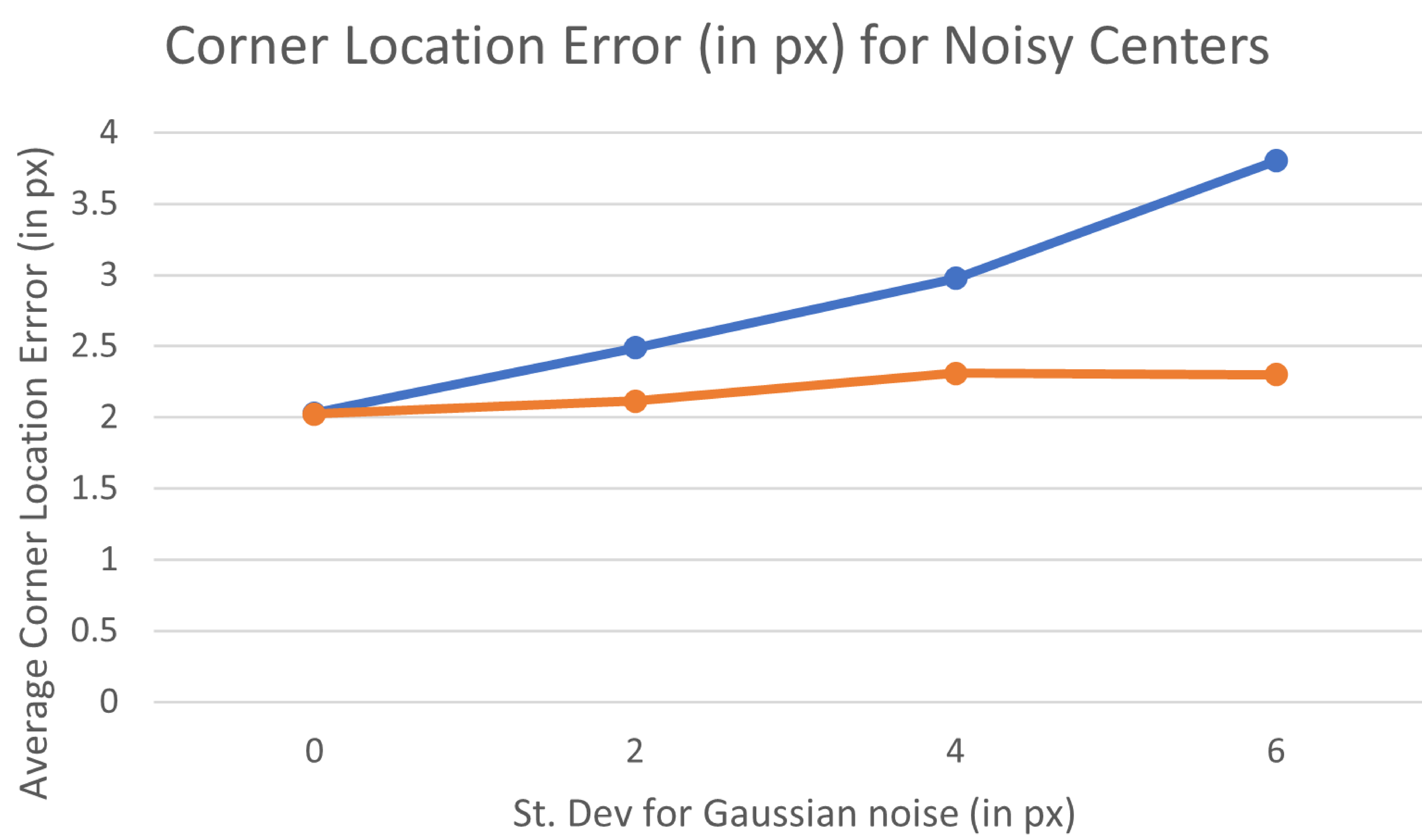


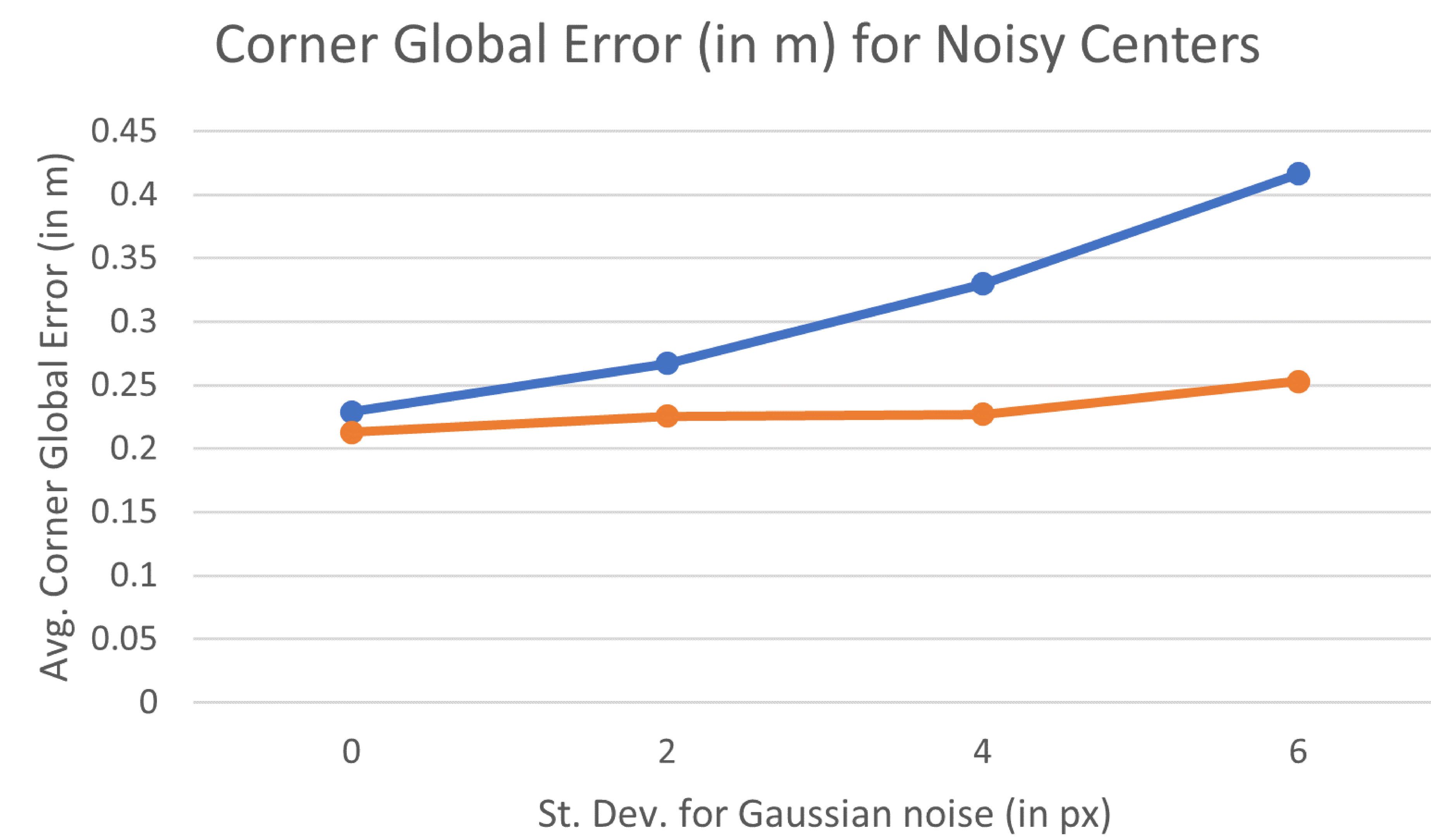
Figure 3: Example of jitter augmentation. Center shift changes resulting crop.



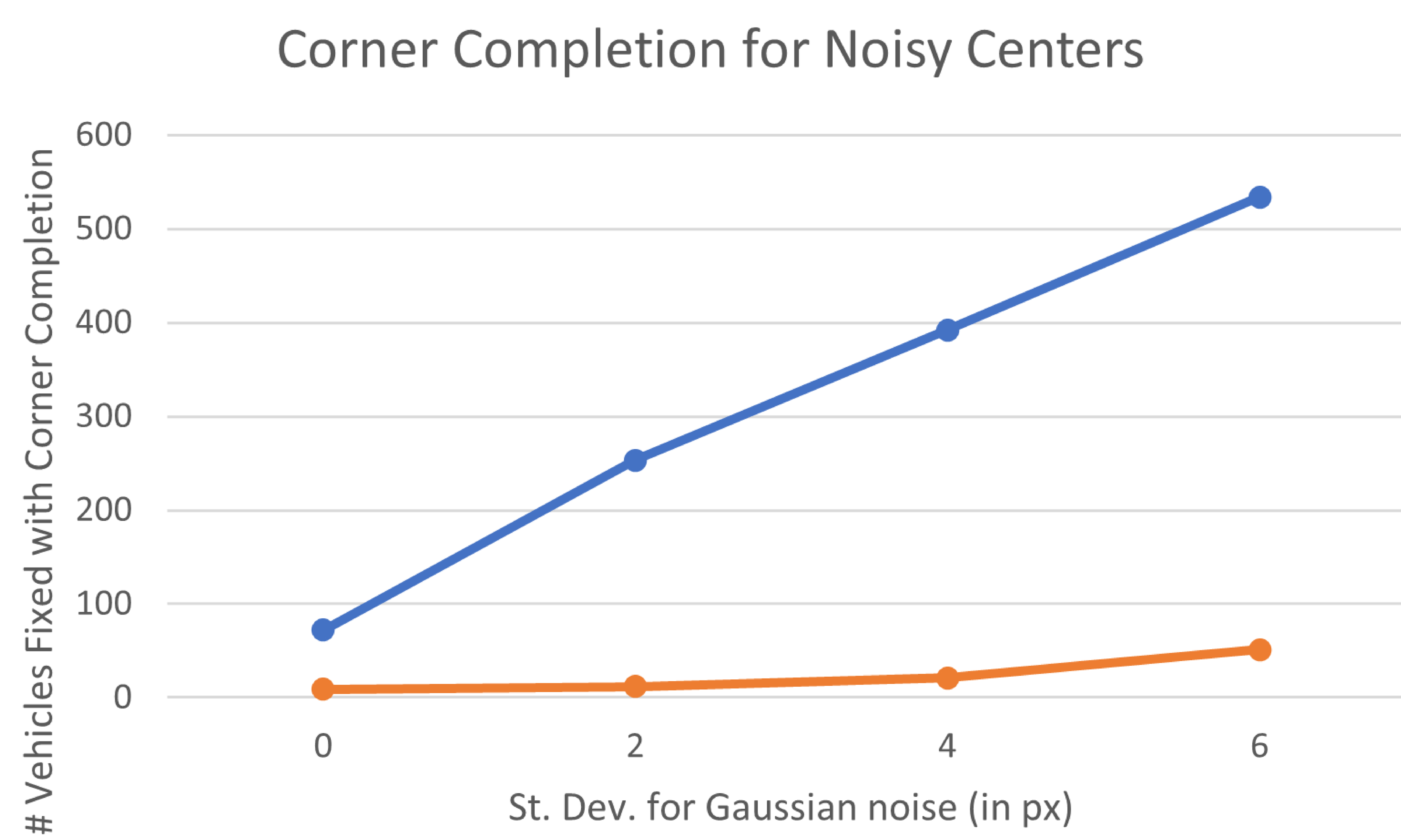
$$x_{jitter} = x_{center} + N\left(0, \sigma^2\right)$$
$$y_{jitter} = y_{center} + N\left(0, \sigma^2\right)$$

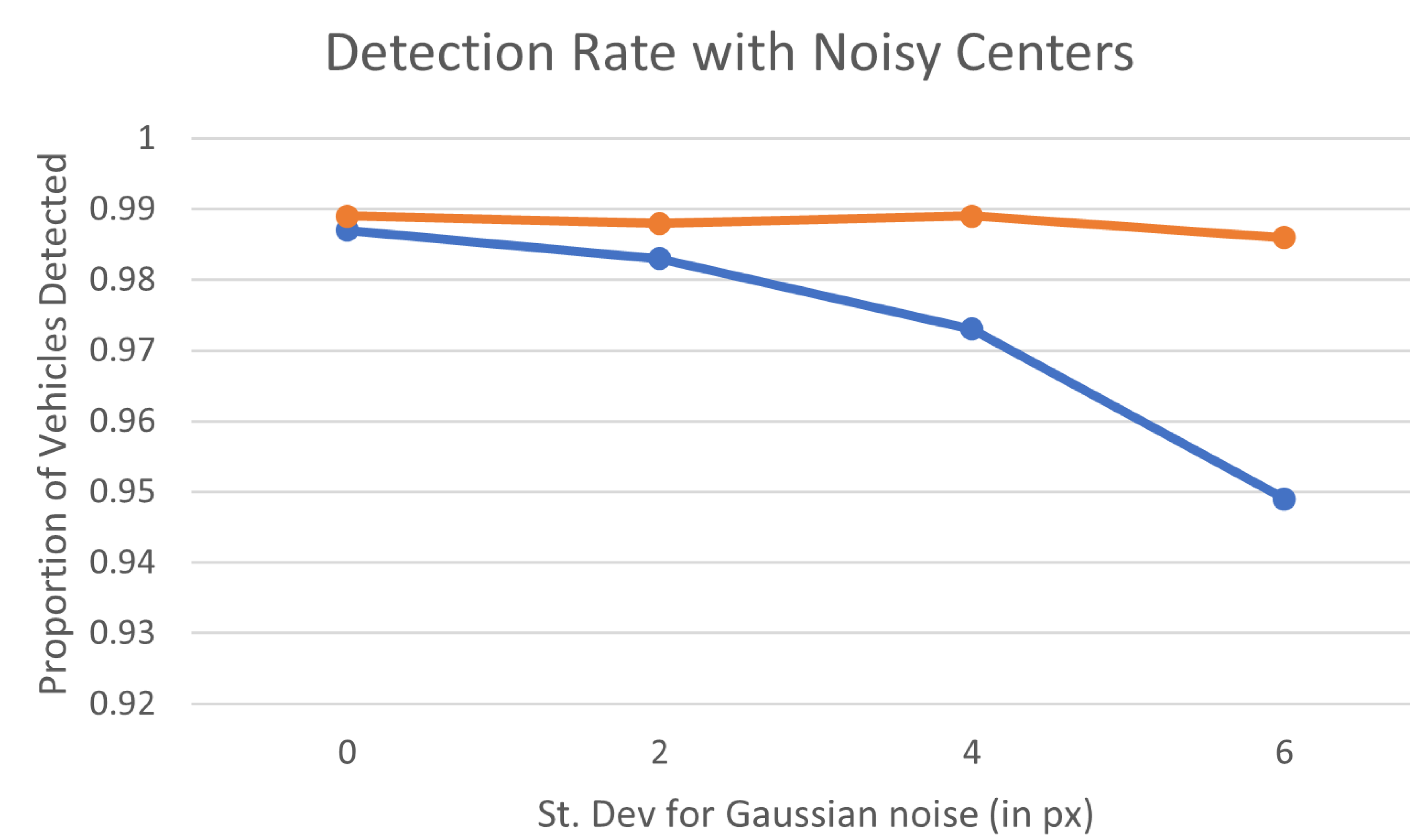






Simple augmentation makes second stage robust to imperfect centers in first stage



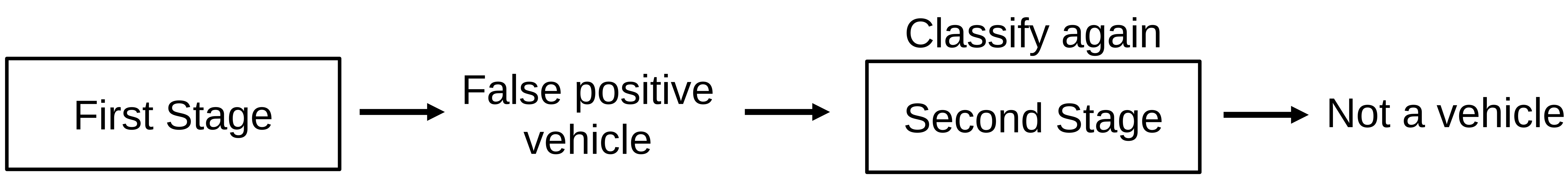


Advantages

- Second stage crop more universally deployable

Figure 4: Comparison demonstrating low variability in crop of scene, despite high variability of actual scene.

- Second stage can refine first



- Capitalizes on fixed scene with preprocessed location maps
- Subpixel accuracy on cropped corners due to resize
- Avoids corner overlap and mismatches seen in single stage

Future Work

- Real intersection data
- Comprehensive Comparison
- Refining Predictions
- Multi-scene crop training
- Crop-size index map

References

X. Zhou, D. Wang, and P. Krähenbühl, "Objects as points," *arXiv.org*, 25-Apr-2019. [Online]. Available: https://arxiv.org/abs/1904.07850v2. [Accessed: 19-Oct-2021].