Rotation Matters: Generalized Monocular 3D Object Detection for Various Camera Systems

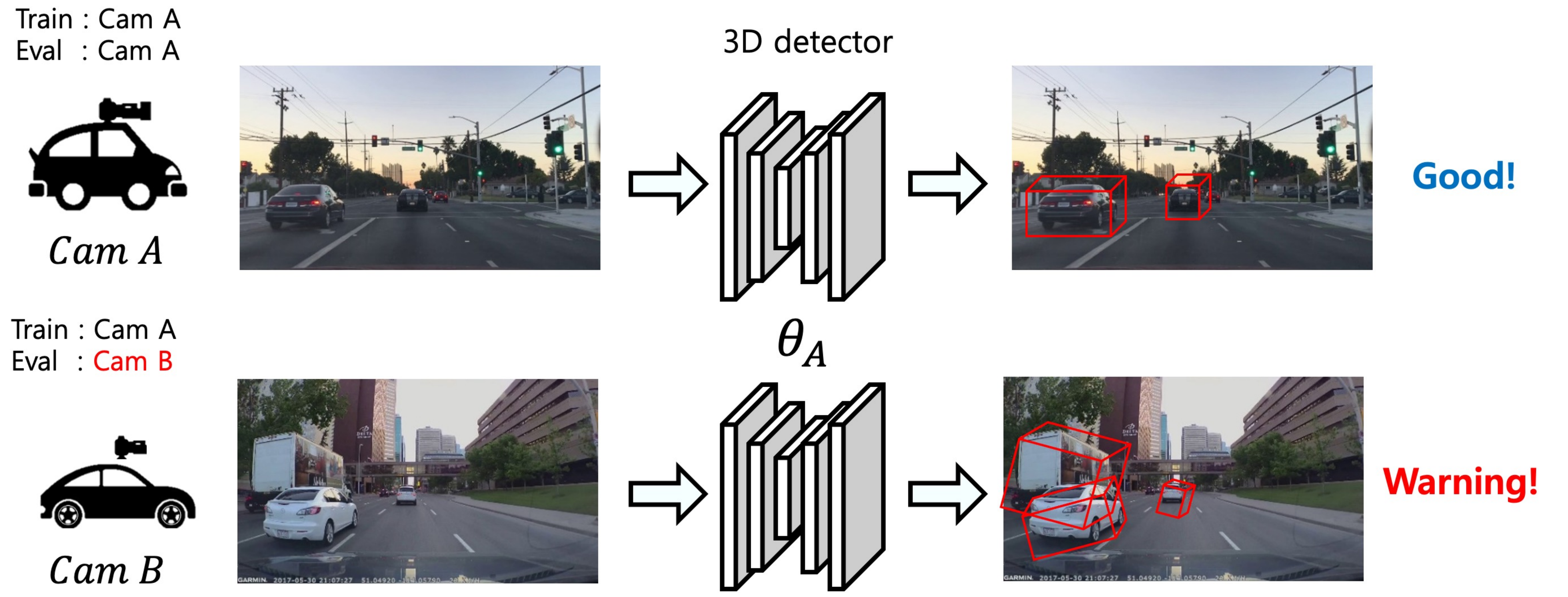
Sungho Moon, Jinwoo Bae, Sunghoon Im

Daegu Gyeongbuk Institute of Science & Technology

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

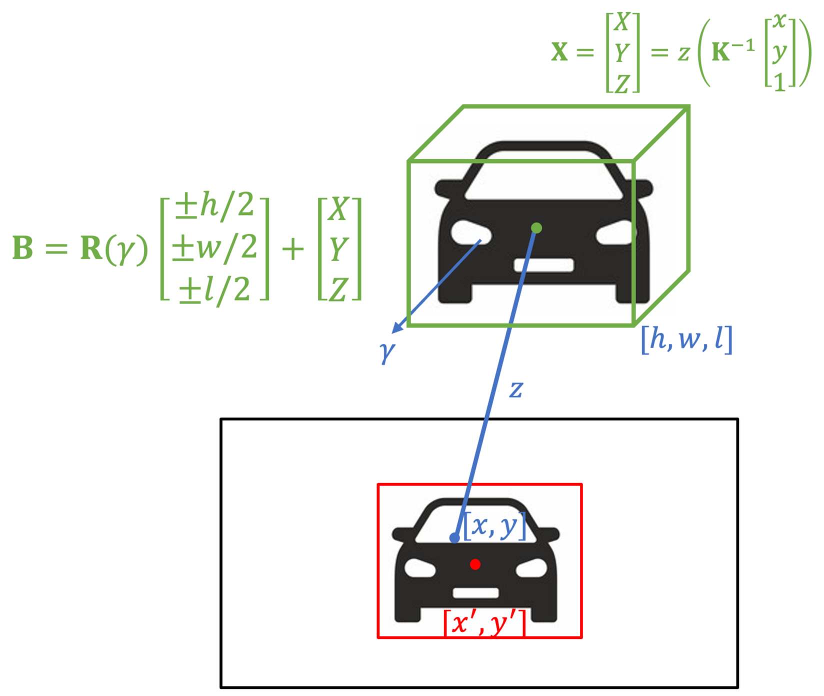
Problem

- **Existing method [1,2,3,4]** is trained with images captured by **only one camera**, which brings **poor results in other camera environments**.



Contribution

- A **generalized monocular 3D object detection method** trained on a specific camera system but **can be utilized in a variety of camera systems**.
- Figure out a factor leading performance degradation in a new camera system (Camera rotation w.r.t. a road plane)
- Our method achieves **the 6-to-10 times improvements** compared to state-of-the-art methods without training.



$$\mathbf{X} = \begin{bmatrix}X\\Y\\Z\end{bmatrix} = z\left(\mathbf{K}^{-1}\begin{bmatrix}x\\y\\1\end{bmatrix}\right)$$

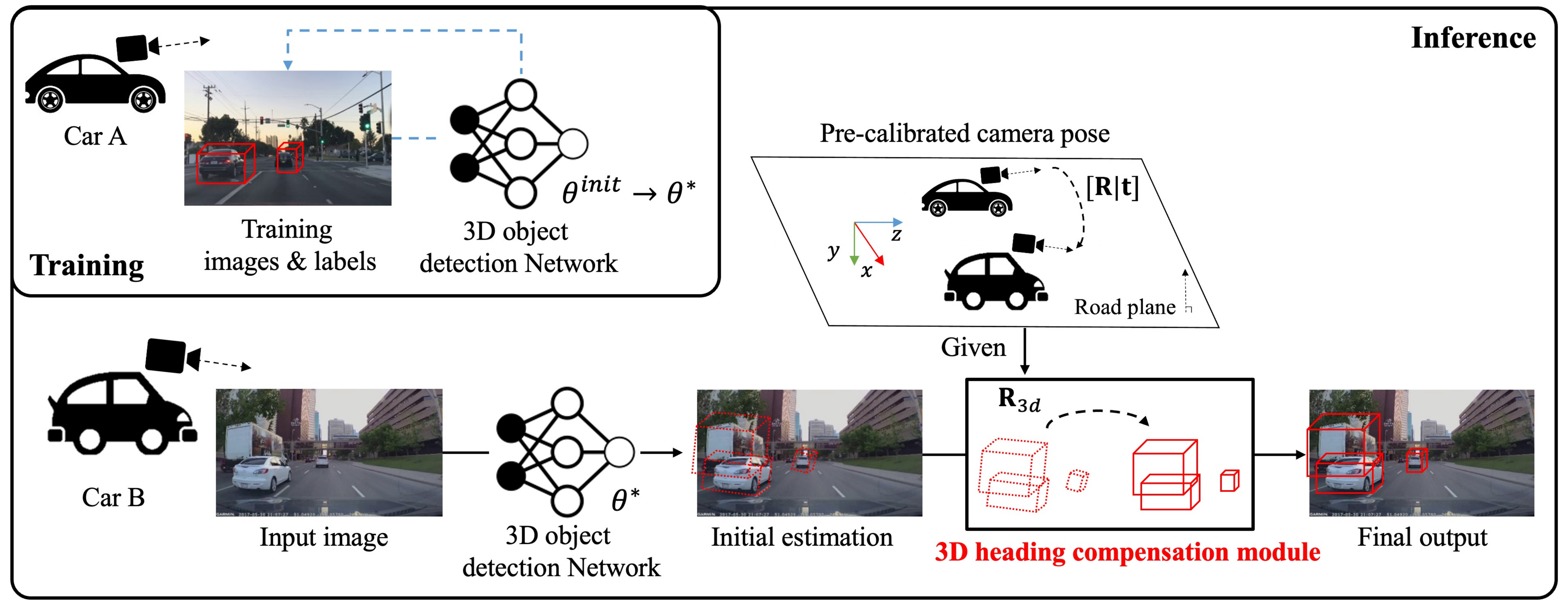
$$\mathbf{B} = \mathbf{R}(\gamma)\begin{bmatrix}\pm h/2\\ \pm w/2\\ \pm l/2\end{bmatrix} + \begin{bmatrix}X\\Y\\Z\end{bmatrix}$$

Acknowledgement

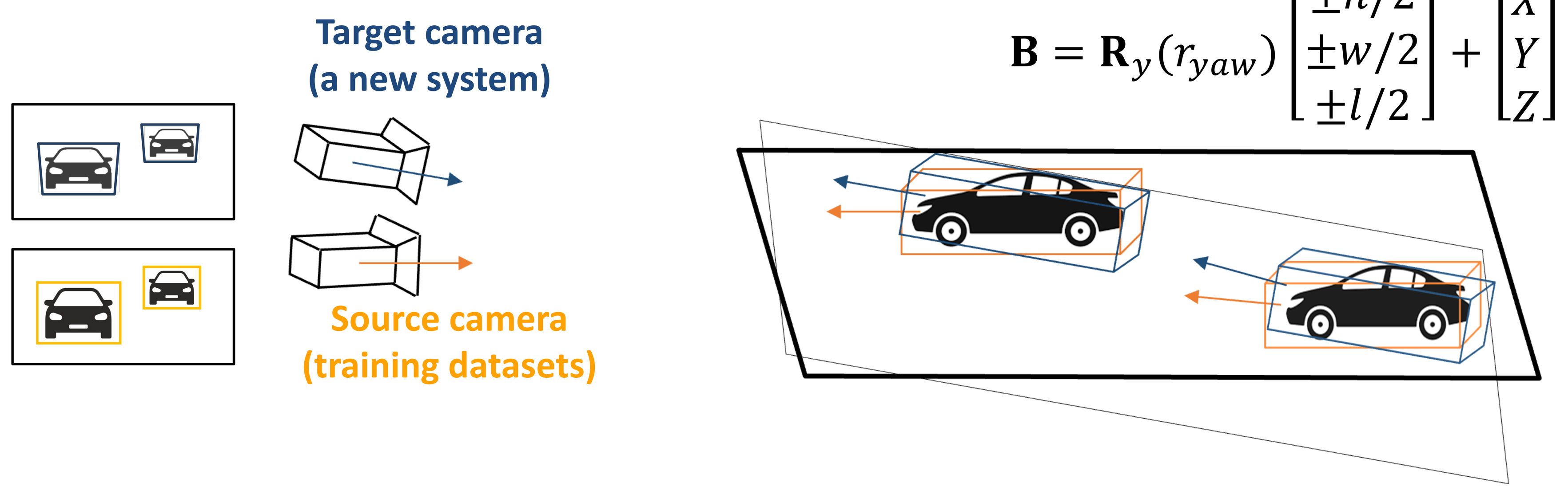
- This research was supported in part by Autonomous Driving Center, R&D Division, **Hyundai Motor Company**.

Method

- We use the existing 3D object detection method [1,2,3,4] to estimate 3D object detection and use **our 3D object detection method**.

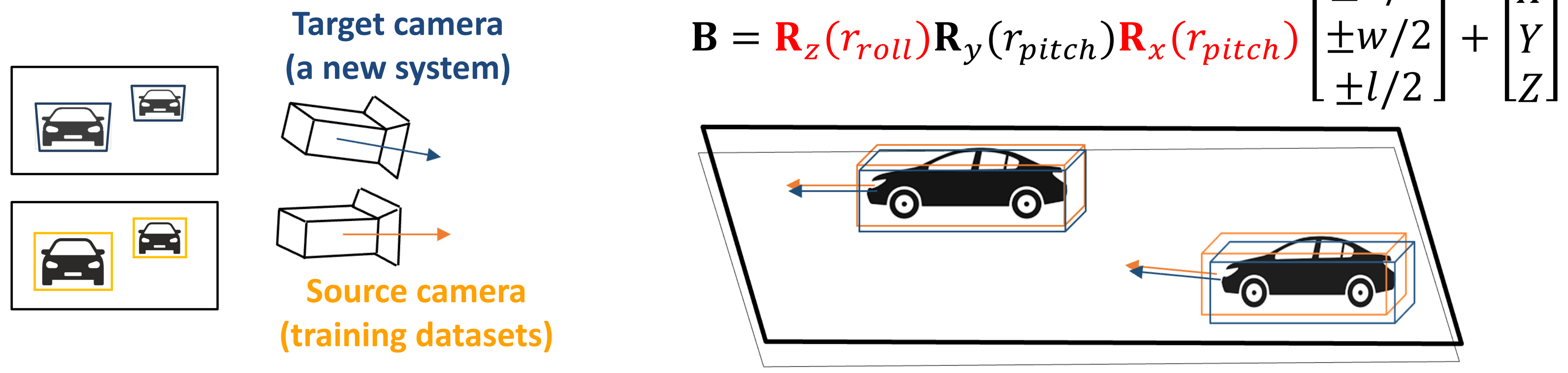


- Existing model [1,2,3,4] learned **with one camera** estimates **poorly** 3D object detection **in the target camera**.



$$\mathbf{B} = \mathbf{R}_y(r_{yaw})\begin{bmatrix}\pm h/2\\ \pm w/2\\ \pm l/2\end{bmatrix} + \begin{bmatrix}X\\Y\\Z\end{bmatrix}$$

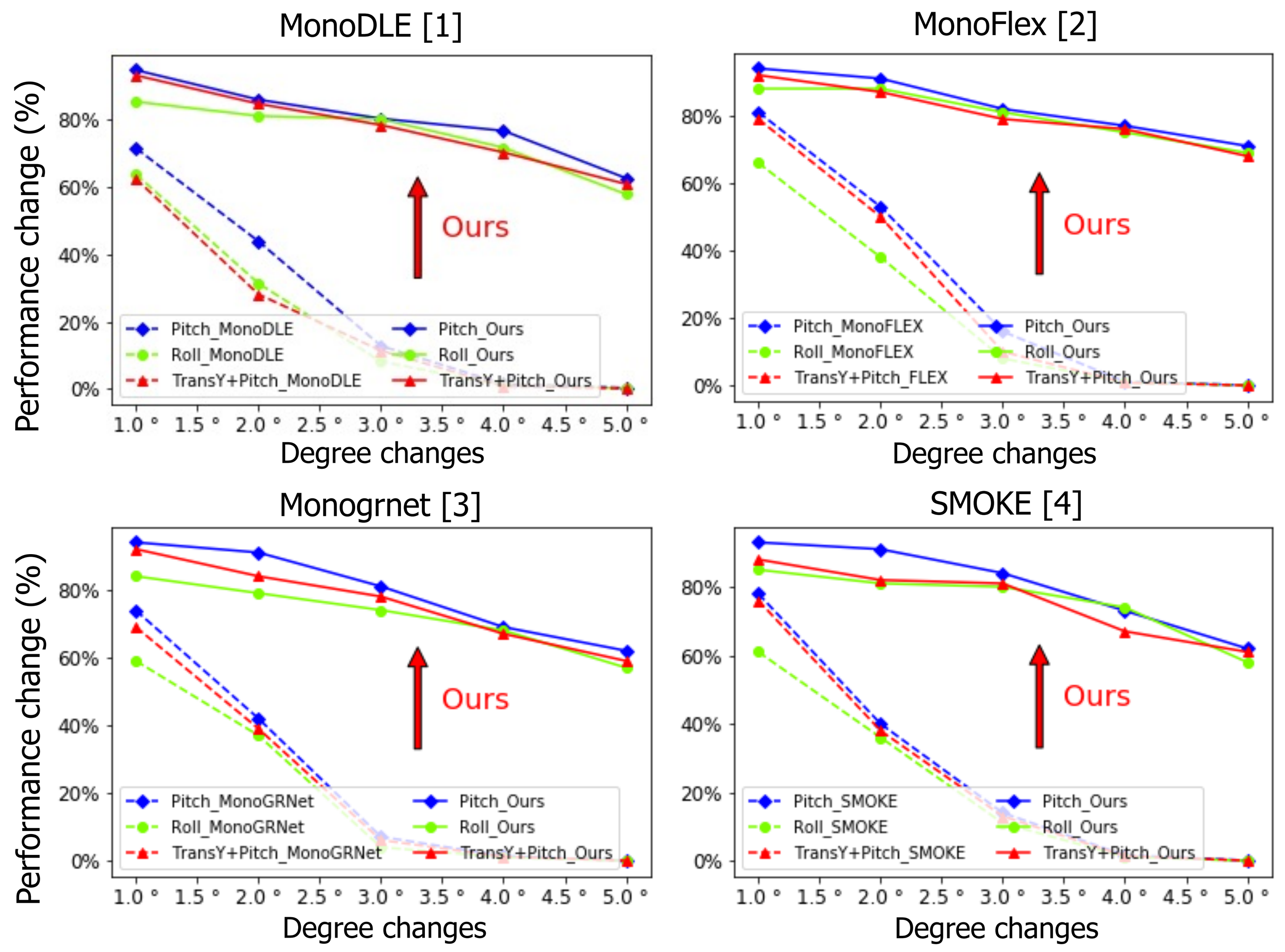
- **Our method** consisted of 3D object center and corners compensation estimates **properly** 3D object detection **in the target camera**.



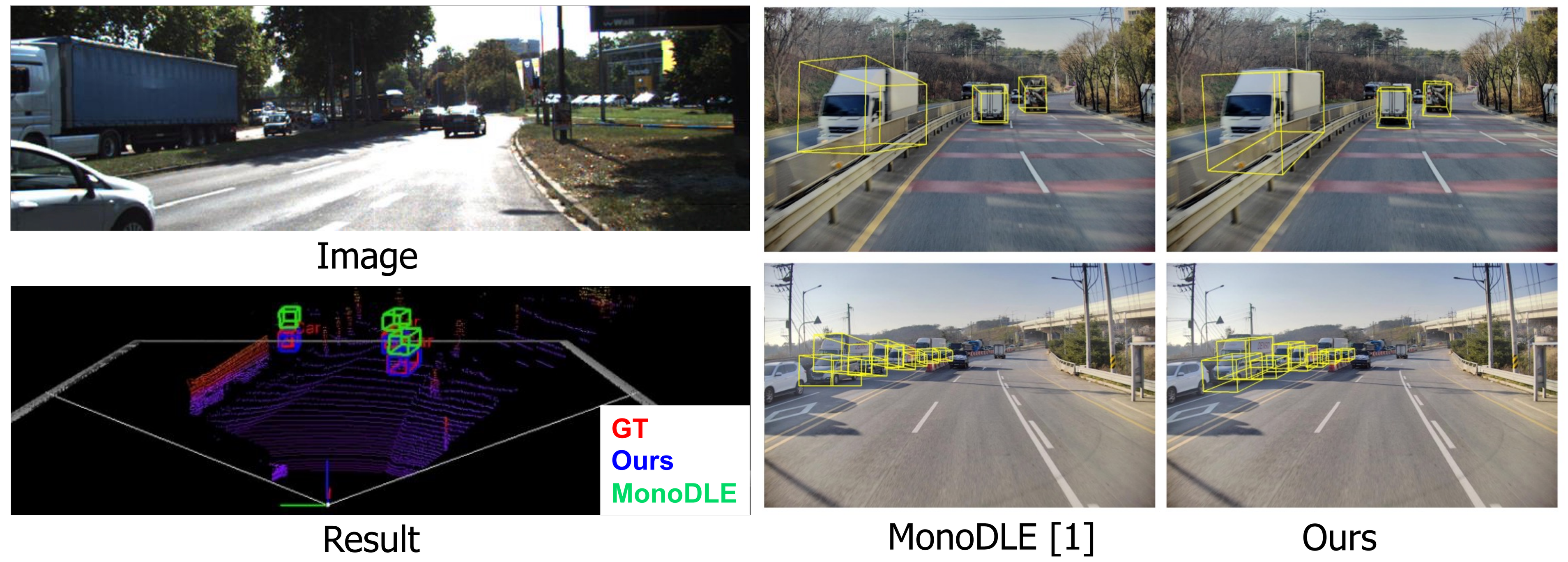
$$\mathbf{B} = \mathbf{R}_z(r_{roll})\mathbf{R}_y(r_{pitch})\mathbf{R}_x(r_{pitch})\begin{bmatrix}\pm h/2\\ \pm w/2\\ \pm l/2\end{bmatrix} + \begin{bmatrix}X\\Y\\Z\end{bmatrix}$$

Results

- Quantitative results: The performance of all existing models [1,2,3,4] with **our method** in various camera systems without additional training.



- Qualitative results



References

[1] X. Ma, et. al., "Delving into localization errors for monocular 3d object detection," CVPR21
[2] Y. Zhang, et. al., "Objects are different: Flexible monocular 3d object detection," CVPR21
[3] Z. Qin, et. al., "Monogrnet: A geometric reasoning network for monocular 3d object localization," AAAI19
[4] Z. Liu, et. al., "Smoke: Single-stage monocular 3d object detection via keypoint estimation," CVPRW20