WACV 2026
TUCSON, AZ 3/6 - 3/10

# JOCA: Task-Driven **J**oint **O**ptimisation of **C**amera Hardware and **A**daptive Camera Control Algorithms


Chengyang Yan


Mitch Bryson

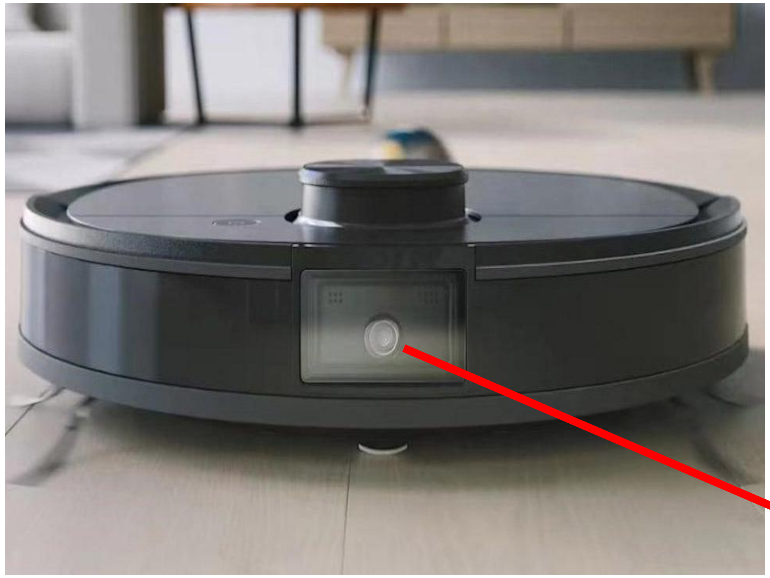

Donald G. Dansereau

THE UNIVERSITY OF SYDNEY
Australian Centre for Robotics

ROBOTIC IMAGING LAB

Project Page

# Motivation



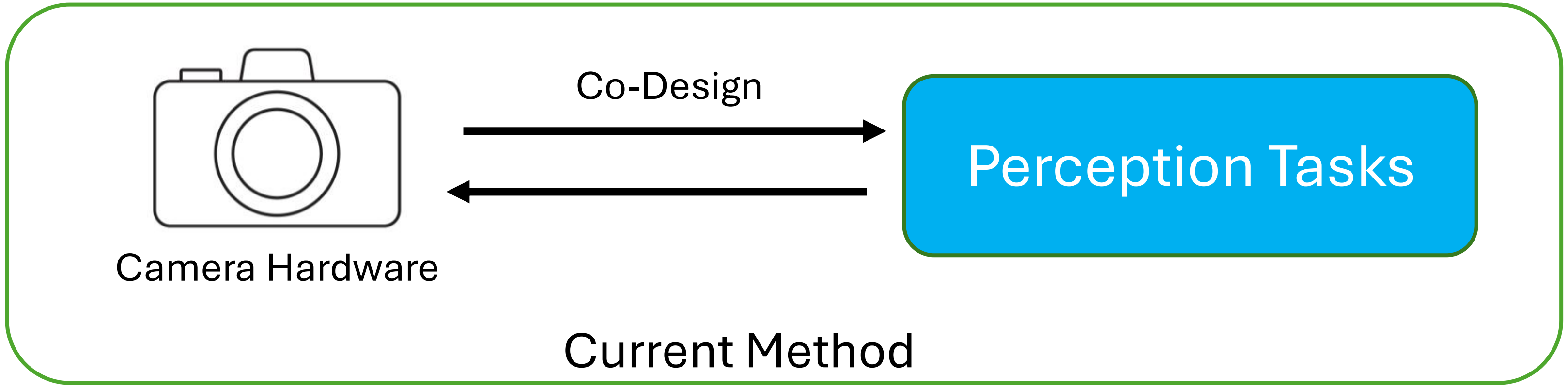
ecovacs.com

tesla.com

meta.com

CAMERAS

# Motivation

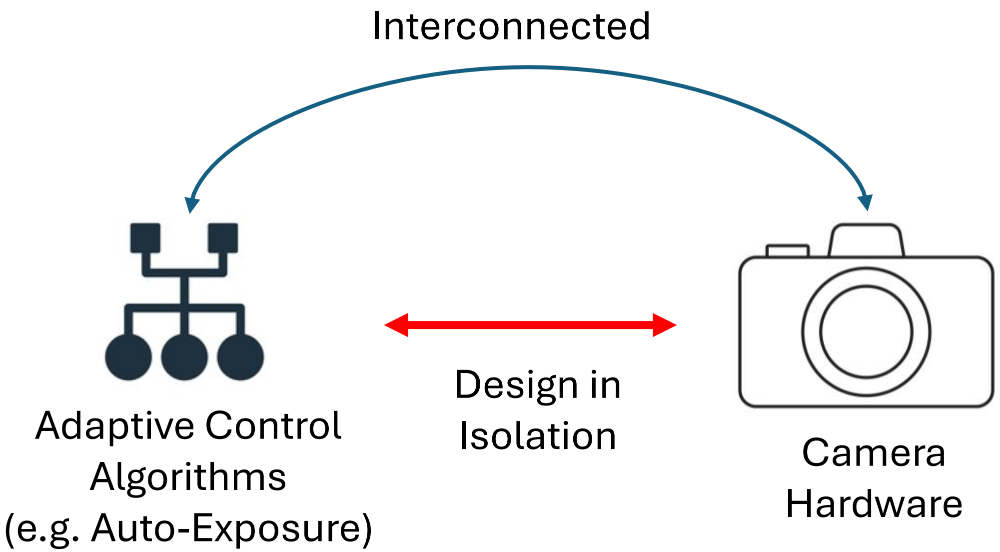


Co-Design

Perception Tasks

Camera Hardware

Current Method



Interconnected

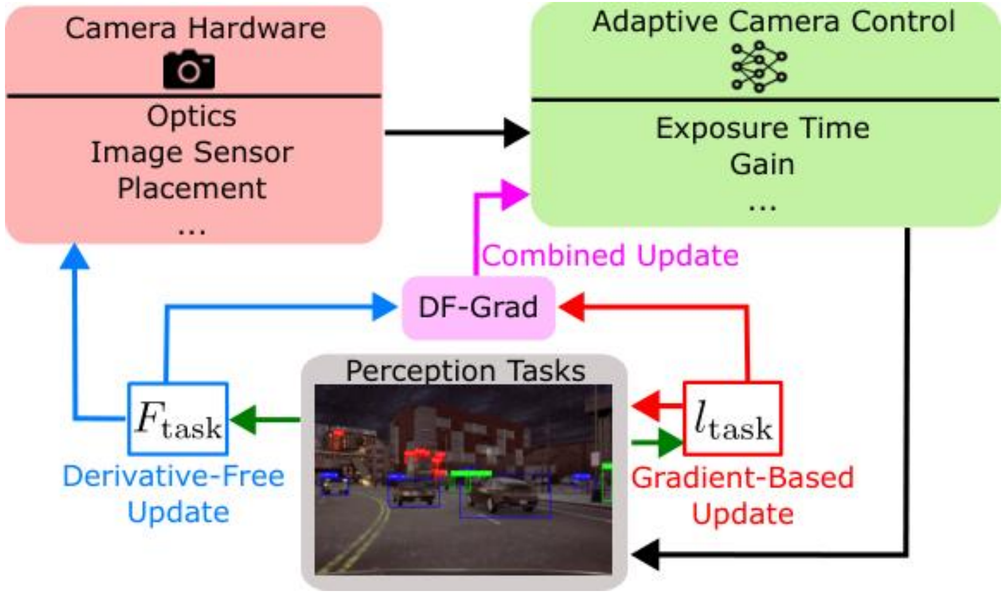
Adaptive Control Algorithms (e.g. Auto-Exposure)

Design in Isolation

Camera Hardware

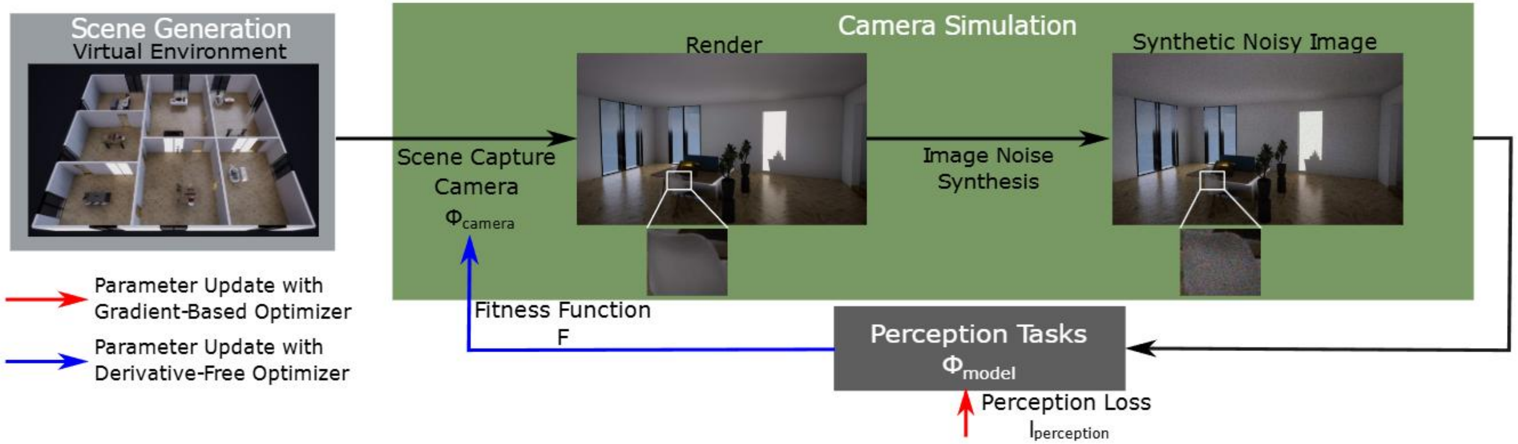
# Contributions

- The first end-to-end framework that jointly optimises camera hardware, adaptive camera control algorithms, and downstream perception tasks.
- A hybrid optimisation strategy, DF-Grad, which integrates derivative-free and gradient-based methods to train deep learning-based adaptive camera control algorithms under non-differentiable image effects rendering.
- Demonstration of camera design and camera control algorithms with improved performance under both standard and challenging conditions compared to prior methods.

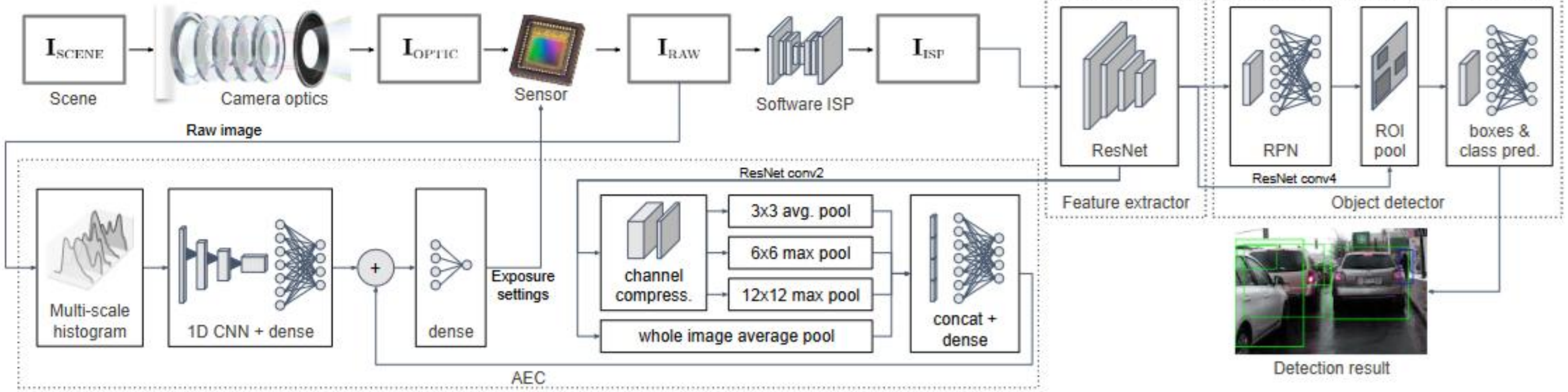


# Previous Works

- Task-Specific Camera Hardware Design (e.g. [Yan 2025]).
  - Focuses on hardware design and ignores dynamic camera parameters.

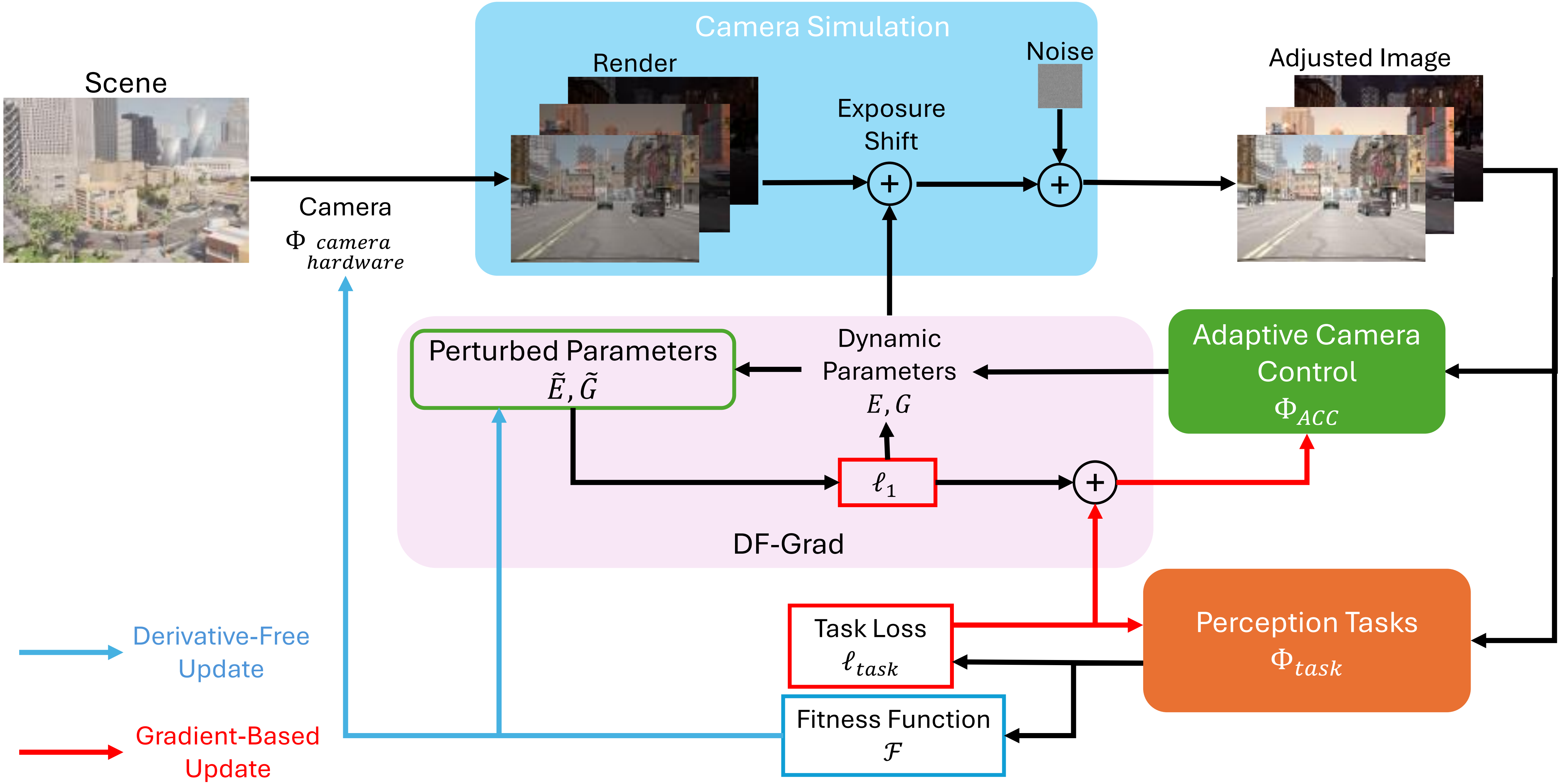


- Task-Specific Adaptive Camera Control Algorithm (e.g. [Onzon 2021]).
  - Adaptive control algorithms (auto-exposure) only.
  - Ignores image effects induced by varying dynamic parameters



[Yan 2025] TaCOS: Task-Specific Camera Optimization with Simulation, Chengyang Yan et al, WACV 2025.
[Onzon 2021] Neural Auto-Exposure for High-Dynamic Range Object Detection, Emmanuel Onzon et al, CVPR 2021.

# Overview



Scene

Camera

$\Phi_{camera\ hardware}$

Camera Simulation

Render

Exposure Shift

Noise

Adjusted Image

Dynamic Parameters

$E, G$

Perturbed Parameters

$\tilde{E}, \tilde{G}$

$\ell_1$

DF-Grad

Adaptive Camera Control

$\Phi_{ACC}$

Perception Tasks

$\Phi_{task}$

Task Loss

$\ell_{task}$

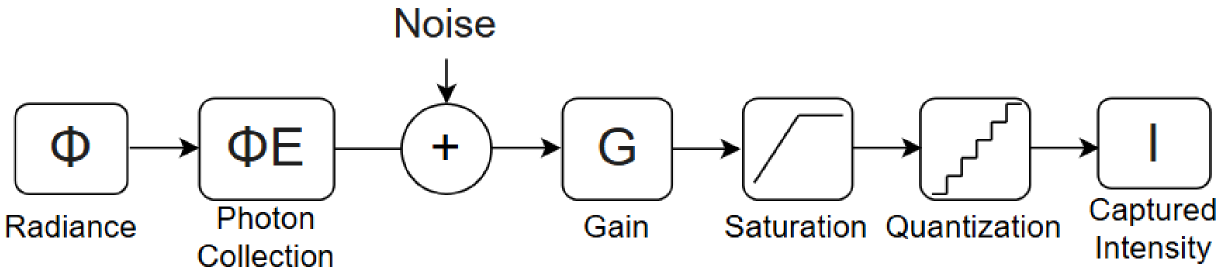
Fitness Function

$\mathcal{F}$

Derivative-Free Update

Gradient-Based Update

# Image Formation

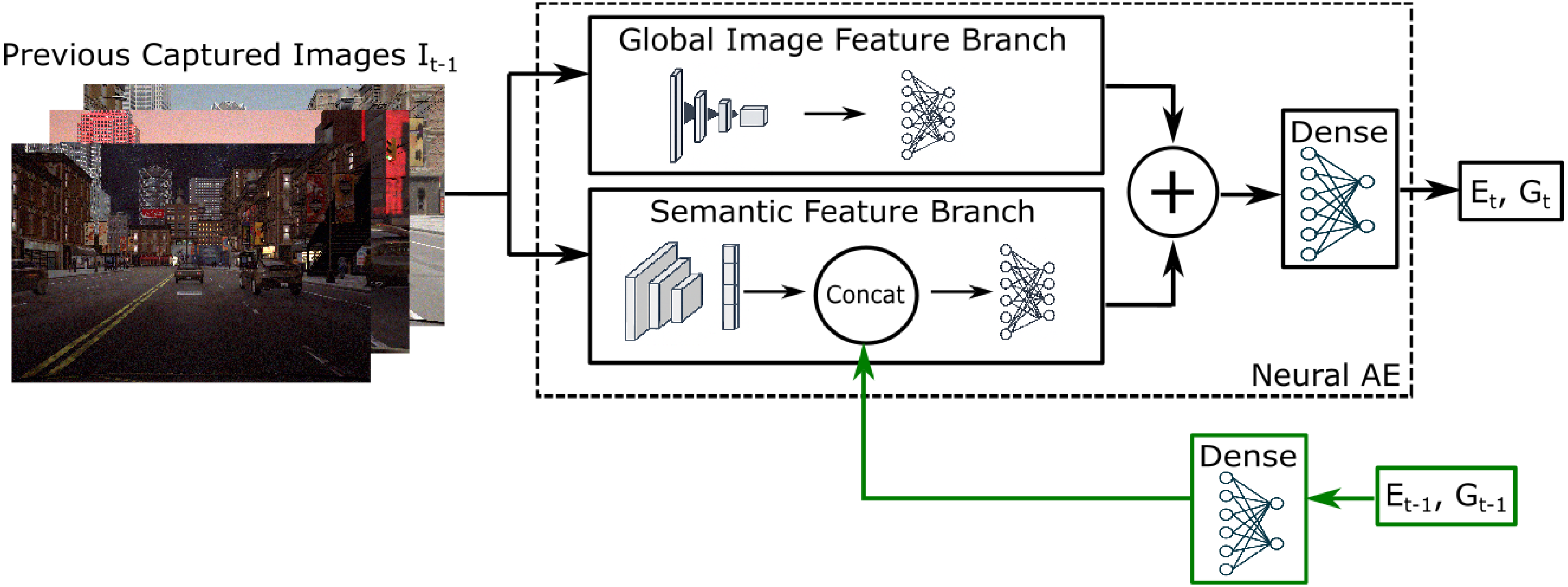


- Calibrate $E_0$ and $G_0$ in Data Images against a Physical Camera with know $E$ and $G$
- Exposure Shift: $I = I_0 \frac{E}{E_0} \frac{G}{G_0}$
- Affine Noise Model [Foi 2009]: $\boldsymbol{\sigma}^2 = \boldsymbol{\sigma}_p^2 \bar{I} + \boldsymbol{\sigma}_t^2$
- Generalise for Multiple Exposure and Image Sensor [Yan 2025]: $\sigma^2 = \frac{G}{G_0} \sigma_p^2 \bar{I} + \frac{G^2}{G_0^2} \sigma_r^2$

[Foi 2009] Clipped Noisy Images: Heteroskedastic Modeling and Practical Denoising, Alessandro Foi, Signal Processing 2009.
[Yan 2025] TaCOS: Task-Specific Camera Optimization with Simulation, Chengyang Yan et al, WACV 2025.

# Adaptive Camera Control Algorithm

- Adopt the Architecture of NeuralAE [Onzon 2021]



Previous Captured Images $I_{t-1}$

Global Image Feature Branch

Semantic Feature Branch

Concat

Dense

$E_t$, $G_t$

Neural AE

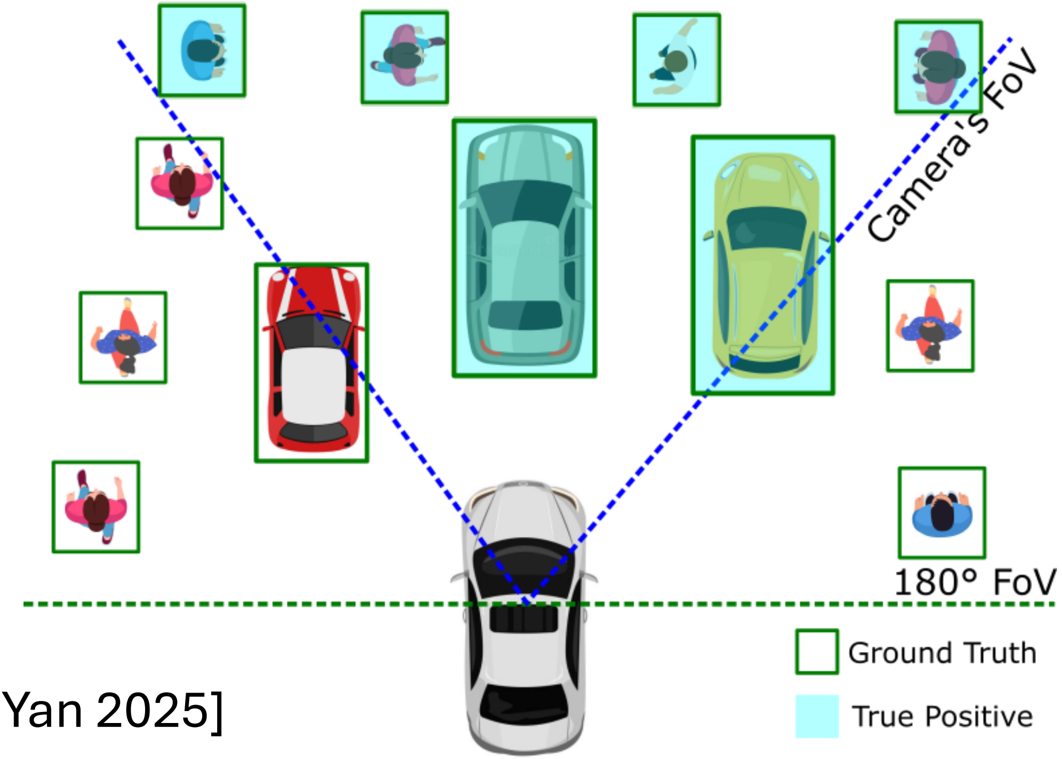
Dense

$E_{t-1}$, $G_{t-1}$

[Onzon 2021] Neural Auto-Exposure for High-Dynamic Range Object Detection, Emmanuel Onzon et al, CVPR 2021.

# Experiment Setup

- Object Detector
  - Faster R-CNN [Ren 2015]
  - AdamW [Loshchilov 2018]
  - Pretrained on ImageNet dataset
- Adaptive Camera Control
  - Adam [Kingma 2015]
  - Train from scratch
- Camera Hardware and Perturbations
  - Genetic Algorithm [Holland 1992] following TaCOS [Yan 2025]
  - AP@0.5IoU for perturbations
  - mAP@0.5-0.95IoU + $\frac{TP}{ALL@180°FoV}$ for camera parameters



[Ren 2015] Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks, Shaoqing Ren et al, NIPS 2015.
[Loshchilov 2018] Decoupled Weight Decay Regularization, Ilya Loshchilov et al, arXiv, 2018
[Kingma 2015] Adam: A Method for Stochastic Optimization, Diederik P Kingma et al, ICLR 2015
[Holland 1992] Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence, John Holland, MIT Press, 1992
[Yan 2025] TaCOS: Task-Specific Camera Optimization with Simulation, Chengyang Yan et al, WACV 2025.

# Comparison

## Baseline Methods

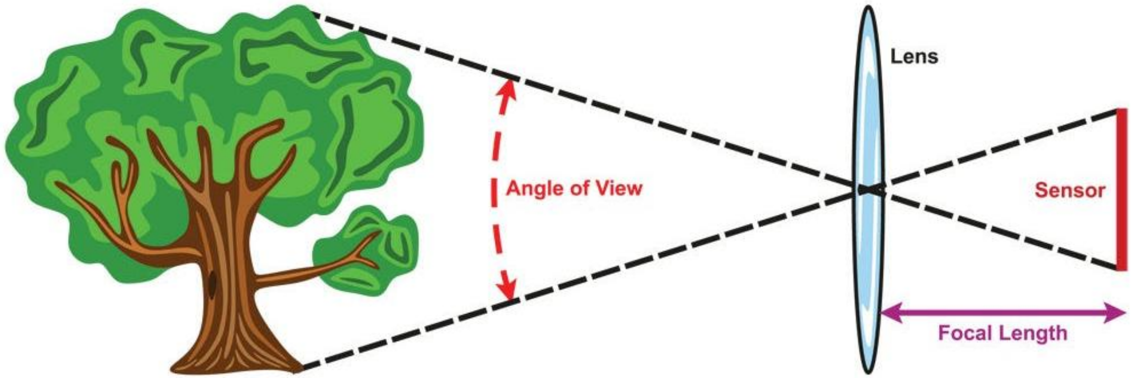
- Human-Designed Camera + Task-Specific Adaptive Camera Control
  - FLIR Flea 3 Camera with nuScene [Caesar 2020] Placement + NeuralAE [Onzon 2021]
  - Basler Dart DaA1280-54uc Camera with nuScene [Caesar 2020] Placement + NeuralAE [Onzon 2021]
- Task-Driven Camera Design + Non-Trainable Adaptive Camera Control
  - TaCOS [Yan 2025] + AverageAE [ARM 2020]

## Design Scenarios

- Standard Scenario with Calibrated Image Noise and Motion Blur
- High Noise Scenario for Darker Scenes and Stronger Image Noise
- High Motion Blur Scenario for Faster Motion

[Caesar 2020] nuScenes: A Multi-Modal Dataset for Autonomous Driving, Holger Caesar et al, CVPR 2020.
[Onzon 2021] Neural Auto-Exposure for High-Dynamic Range Object Detection, Emmanuel Onzon et al, CVPR 2021.
[Yan 2025] TaCOS: Task-Specific Camera Optimization with Simulation, Chengyang Yan et al, WACV 2025.
[ARM 2020] Mali-c71ae: Advanced ISP for Automotive and Industrial, ARM Ltd, 2020.

# Experiment: Synthetic Images



Focal Length, Image Sensor Dimension, and Pixel Size

Categorical

Camera Placement

Object Detection

CARLA Simulator [Dosovitskiy 2017]

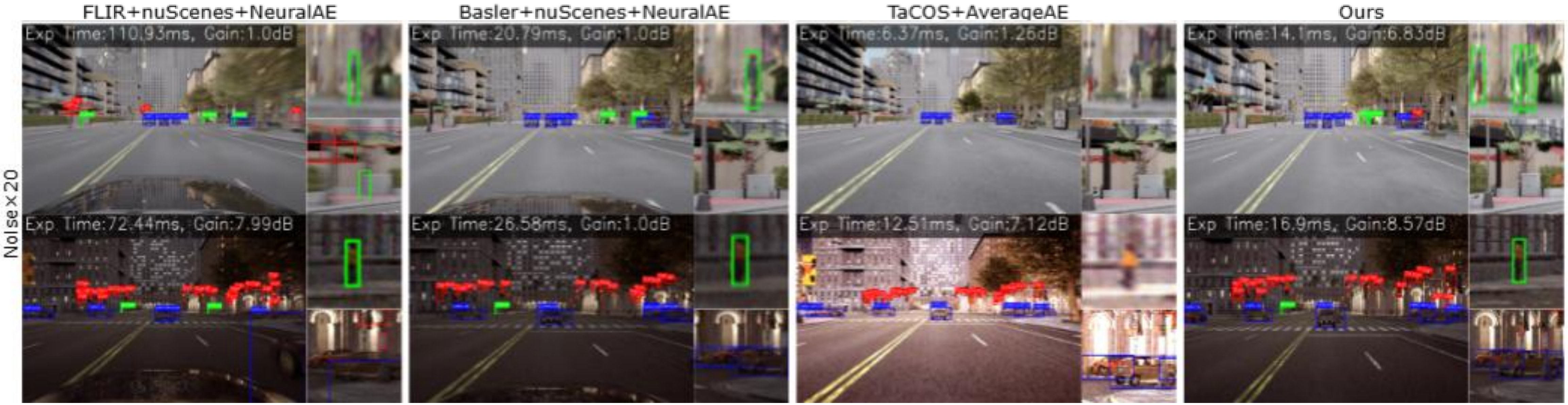
[Dosovitskiy 2017] CARLA: An Open Urban Driving Simulator, Alexey Dosovitskiy et al, CoRL 2017.

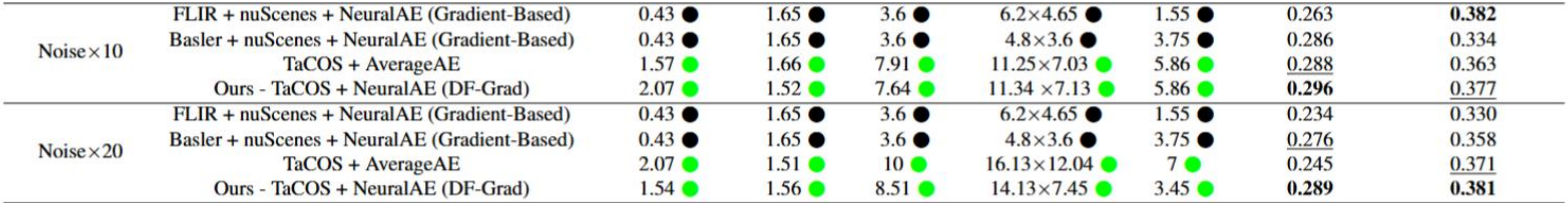
# Synthetic Images - Calibrated

| Scenario | Method | Camera Parameters | | | | | Performance | |
|---|---|---|---|---|---|---|---|---|
| | | Forward Position x $(m)$ | Height z $(m)$ | Focal Length $f$ (mm) | Sensor Size $w \times h$ (mm) | Pixel Size $p$ ($\mu$m) | Object Detect. mAP ↑ | True Positive Ratio TP/All@180° ↑ |
| Calibrated | FLIR + nuScenes + NeuralAE (Gradient-Based) | 0.43 ⚫ | 1.65 ⚫ | 3.6 ⚫ | 6.2×4.65 ⚫ | 1.55 ⚫ | 0.318 | 0.389 |
| | Basler + nuScenes + NeuralAE (Gradient-Based) | 0.43 ⚫ | 1.65 ⚫ | 3.6 ⚫ | 4.8×3.6 ⚫ | 3.75 ⚫ | 0.289 | 0.360 |
| | TaCOS + AverageAE | 2.43 🟢 | 1.7 🟢 | 10 🟢 | 14.44×9.9 🟢 | 4.5 🟢 | 0.303 | 0.358 |
| | Ours - TaCOS + NeuralAE (DF-Grad) | 1.35 🟢 | 1.3 🟢 | 5.19 🟢 | 7.37×4.92 🟢 | 2.4 🟢 | **0.325** | **0.390** |

🟢 - Optimised ⚫ - Fixed

# Synthetic Images – Higher Noise

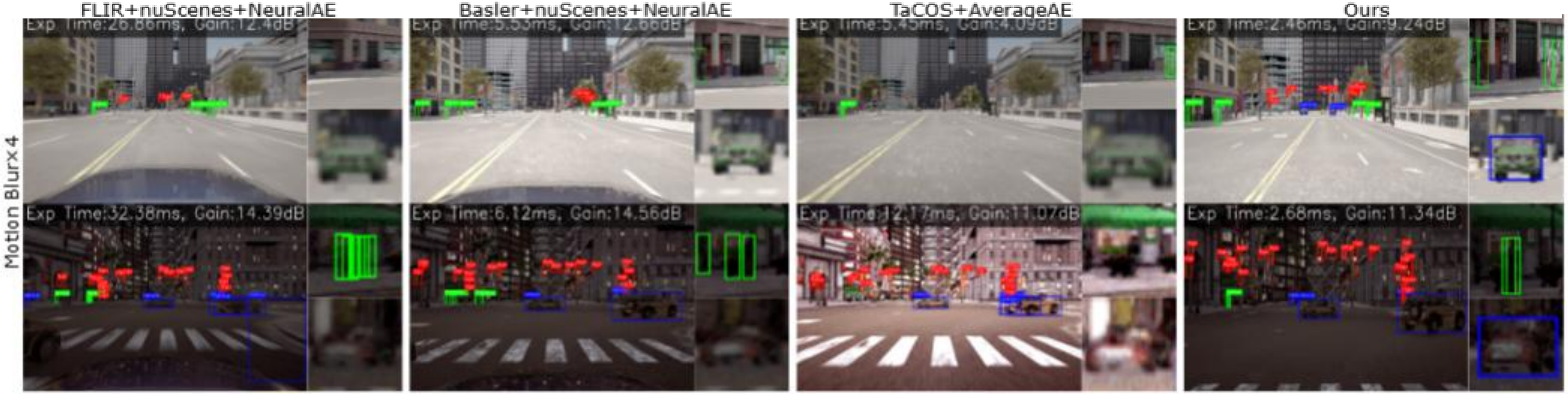


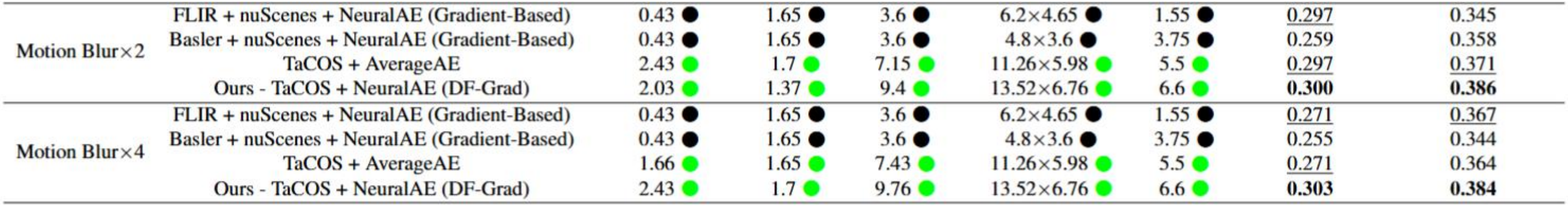


| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Noise × 10 | FLIR + nuScenes + NeuralAE (Gradient-Based) | 0.43 ● | 1.65 ● | 3.6 ● | 6.2×4.65 ● | 1.55 ● | 0.263 | **0.382** |
| | Basler + nuScenes + NeuralAE (Gradient-Based) | 0.43 ● | 1.65 ● | 3.6 ● | 4.8×3.6 ● | 3.75 ● | 0.286 | 0.334 |
| | TaCOS + AverageAE | 1.57 🟢 | 1.66 🟢 | 7.91 🟢 | 11.25×7.03 🟢 | 5.86 🟢 | <u>0.288</u> | 0.363 |
| | Ours - TaCOS + NeuralAE (DF-Grad) | 2.07 🟢 | 1.52 🟢 | 7.64 🟢 | 11.34 ×7.13 🟢 | 5.86 🟢 | **0.296** | <u>0.377</u> |
| Noise × 20 | FLIR + nuScenes + NeuralAE (Gradient-Based) | 0.43 ● | 1.65 ● | 3.6 ● | 6.2×4.65 ● | 1.55 ● | 0.234 | 0.330 |
| | Basler + nuScenes + NeuralAE (Gradient-Based) | 0.43 ● | 1.65 ● | 3.6 ● | 4.8×3.6 ● | 3.75 ● | <u>0.276</u> | 0.358 |
| | TaCOS + AverageAE | 2.07 🟢 | 1.51 🟢 | 10 🟢 | 16.13×12.04 🟢 | 7 🟢 | 0.245 | <u>0.371</u> |
| | Ours - TaCOS + NeuralAE (DF-Grad) | 1.54 🟢 | 1.56 🟢 | 8.51 🟢 | 14.13×7.45 🟢 | 3.45 🟢 | **0.289** | **0.381** |

🟢 - Optimised ● - Fixed

# Synthetic Images – Higher Motion Blur





| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Motion Blur×2 | FLIR + nuScenes + NeuralAE (Gradient-Based) | 0.43 ⚫ | 1.65 ⚫ | 3.6 ⚫ | 6.2×4.65 ⚫ | 1.55 ⚫ | 0.297 | 0.345 |
| | Basler + nuScenes + NeuralAE (Gradient-Based) | 0.43 ⚫ | 1.65 ⚫ | 3.6 ⚫ | 4.8×3.6 ⚫ | 3.75 ⚫ | 0.259 | 0.358 |
| | TaCOS + AverageAE | 2.43 🟢 | 1.7 🟢 | 7.15 🟢 | 11.26×5.98 🟢 | 5.5 🟢 | 0.297 | 0.371 |
| | Ours - TaCOS + NeuralAE (DF-Grad) | 2.03 🟢 | 1.37 🟢 | 9.4 🟢 | 13.52×6.76 🟢 | 6.6 🟢 | **0.300** | **0.386** |
| Motion Blur×4 | FLIR + nuScenes + NeuralAE (Gradient-Based) | 0.43 ⚫ | 1.65 ⚫ | 3.6 ⚫ | 6.2×4.65 ⚫ | 1.55 ⚫ | 0.271 | 0.367 |
| | Basler + nuScenes + NeuralAE (Gradient-Based) | 0.43 ⚫ | 1.65 ⚫ | 3.6 ⚫ | 4.8×3.6 ⚫ | 3.75 ⚫ | 0.255 | 0.344 |
| | TaCOS + AverageAE | 1.66 🟢 | 1.65 🟢 | 7.43 🟢 | 11.26×5.98 🟢 | 5.5 🟢 | 0.271 | 0.364 |
| | Ours - TaCOS + NeuralAE (DF-Grad) | 2.43 🟢 | 1.7 🟢 | 9.76 🟢 | 13.52×6.76 🟢 | 6.6 🟢 | **0.303** | **0.384** |

🟢 - Optimised ⚫ - Fixed

# Experiment: Real-World Images

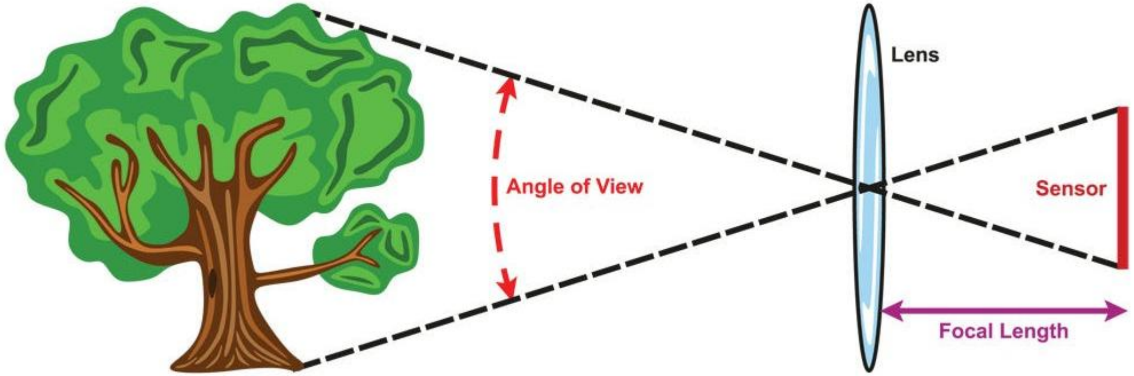


Image Sensor Dimension, and Pixel Size

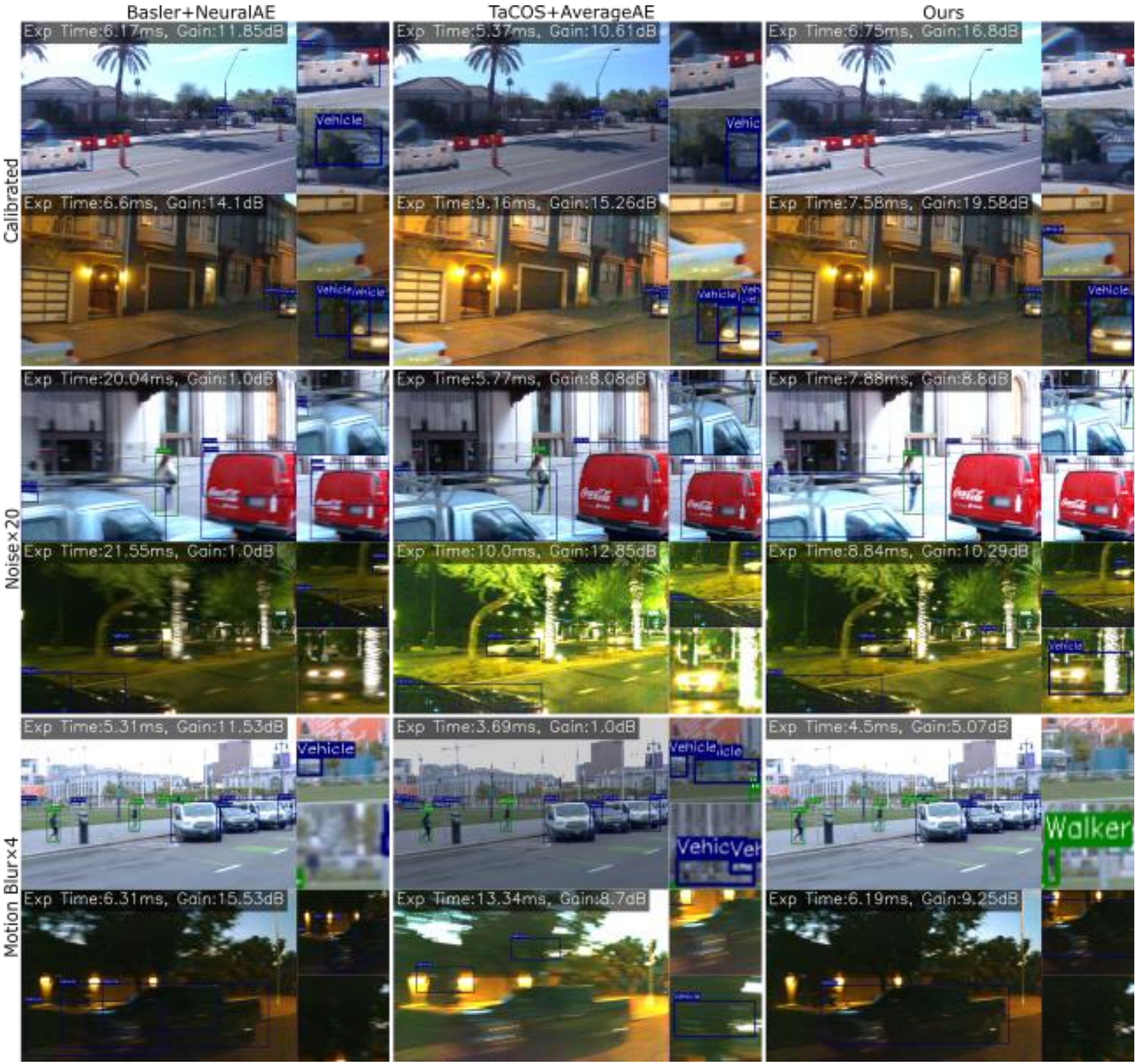
Categorical

Object Detection

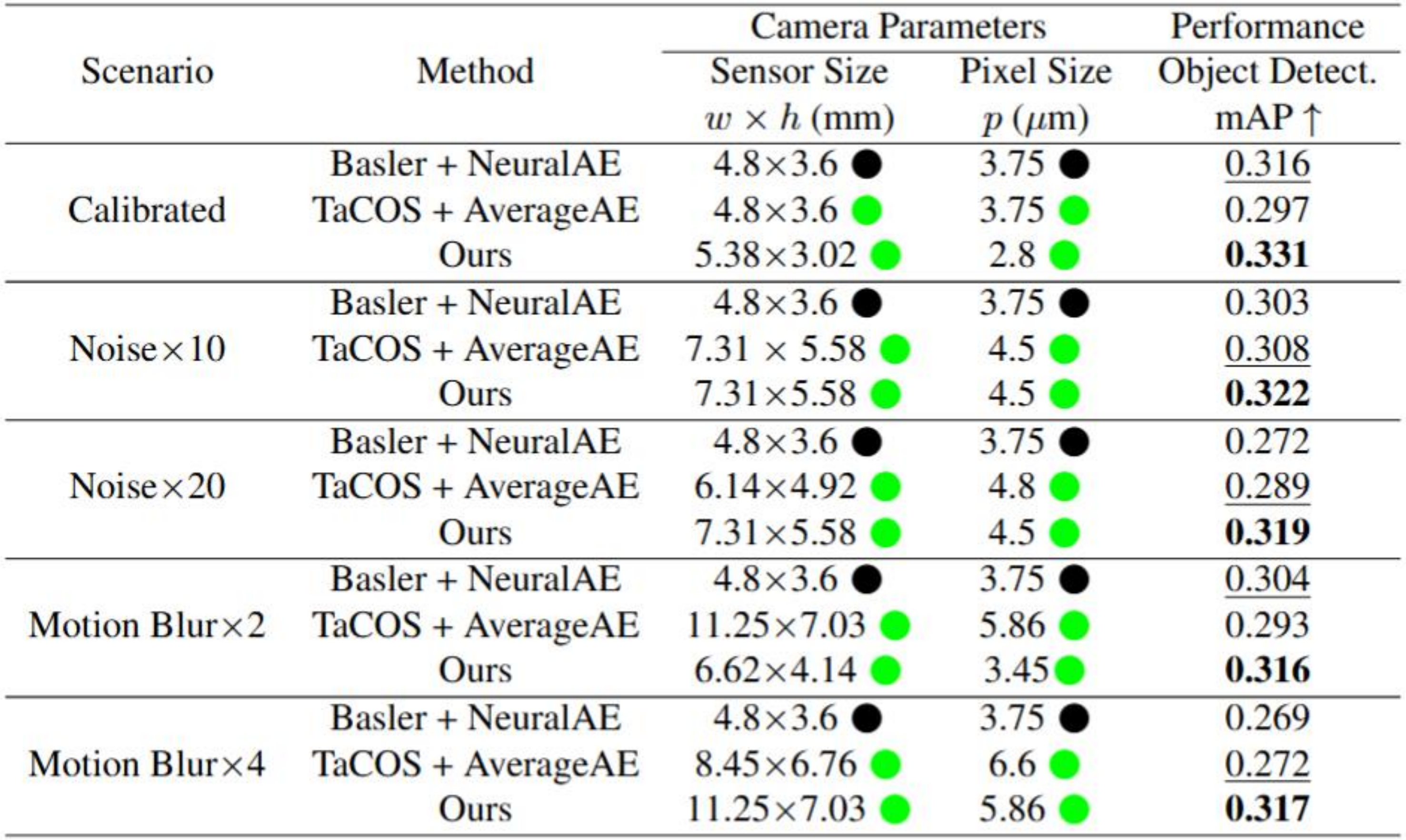
Waymo Open [Sun 2020]

[Sun 2020] Scalability in Perception for Autonomous Driving: Waymo Open Dataset, Pei Sun et al, CVPR 2020.

# Real-World Images



# Real-World Images

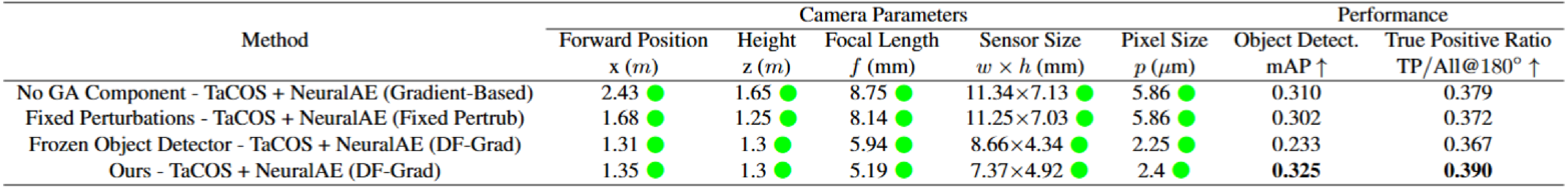


| Scenario | Method | Camera Parameters | | Performance |
|---|---|---|---|---|
| | | Sensor Size $w \times h$ (mm) | Pixel Size $p$ ($\mu$m) | Object Detect. mAP ↑ |
| Calibrated | Basler + NeuralAE | 4.8×3.6 ● | 3.75 ● | 0.316 |
| | TaCOS + AverageAE | 4.8×3.6 🟢 | 3.75 🟢 | 0.297 |
| | Ours | 5.38×3.02 🟢 | 2.8 🟢 | **0.331** |
| Noise×10 | Basler + NeuralAE | 4.8×3.6 ● | 3.75 ● | 0.303 |
| | TaCOS + AverageAE | 7.31 × 5.58 🟢 | 4.5 🟢 | 0.308 |
| | Ours | 7.31×5.58 🟢 | 4.5 🟢 | **0.322** |
| Noise×20 | Basler + NeuralAE | 4.8×3.6 ● | 3.75 ● | 0.272 |
| | TaCOS + AverageAE | 6.14×4.92 🟢 | 4.8 🟢 | 0.289 |
| | Ours | 7.31×5.58 🟢 | 4.5 🟢 | **0.319** |
| Motion Blur×2 | Basler + NeuralAE | 4.8×3.6 ● | 3.75 ● | 0.304 |
| | TaCOS + AverageAE | 11.25×7.03 🟢 | 5.86 🟢 | 0.293 |
| | Ours | 6.62×4.14 🟢 | 3.45 🟢 | **0.316** |
| Motion Blur×4 | Basler + NeuralAE | 4.8×3.6 ● | 3.75 ● | 0.269 |
| | TaCOS + AverageAE | 8.45×6.76 🟢 | 6.6 🟢 | 0.272 |
| | Ours | 11.25×7.03 🟢 | 5.86 🟢 | **0.317** |

🟢 - Optimised ● - Fixed

# Ablation Study

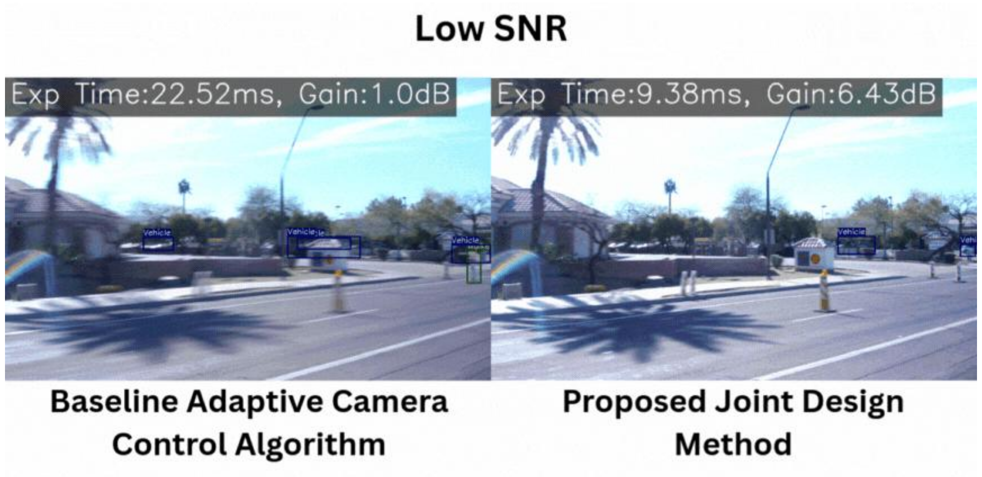
- Without Derivative-Free Component for Adaptive Control Algorithm
- Fixed Perturbations – Perturbations not otpimised by GA
- Frozen Object Detector



| Method | Camera Parameters | | | | | Performance | |
|---|---|---|---|---|---|---|---|
| | Forward Position x ($m$) | Height z ($m$) | Focal Length $f$ (mm) | Sensor Size $w \times h$ (mm) | Pixel Size $p$ ($\mu$m) | Object Detect. mAP ↑ | True Positive Ratio TP/All@180° ↑ |
| No GA Component - TaCOS + NeuralAE (Gradient-Based) | 2.43 ● | 1.65 ● | 8.75 ● | 11.34×7.13 ● | 5.86 ● | 0.310 | 0.379 |
| Fixed Perturbations - TaCOS + NeuralAE (Fixed Pertrub) | 1.68 ● | 1.25 ● | 8.14 ● | 11.25×7.03 ● | 5.86 ● | 0.302 | 0.372 |
| Frozen Object Detector - TaCOS + NeuralAE (DF-Grad) | 1.31 ● | 1.3 ● | 5.94 ● | 8.66×4.34 ● | 2.25 ● | 0.233 | 0.367 |
| Ours - TaCOS + NeuralAE (DF-Grad) | 1.35 ● | 1.3 ● | 5.19 ● | 7.37×4.92 ● | 2.4 ● | **0.325** | **0.390** |

# Conclusion & Future Work

- A novel end-to-end method that jointly optimises camera hardware, adaptive camera control algorithms, and perception tasks.
- A hybrid optimization method, DF-Grad, combining derivative-free and gradient-based optimisers to train neural network-based adaptive camera control algorithms under non-differentiable image effects rendering.
- JOCA outperforms existing isolated camera design and adaptive camera control methods under both standard and challenging conditions.



**Baseline Adaptive Camera Control Algorithm**

**Proposed Joint Design Method**

## Future Work

- Extend to more complex camera design problems involving unconventional cameras and multi-camera designs, more challenging design scenarios involving adverse weather conditions.

# Acknowledgments

This research was supported in part through the NVIDIA Academic Grant Program.

# Thanks!

Project Page

THE UNIVERSITY OF SYDNEY

Australian Centre for Robotics

ROBOTIC IMAGING LAB