

Foveated Imaging and Perception

Howard Xiao, Jan Ackermann, Boyang Deng, Gordon Wetzstein
Stanford University

Motivation

Ultra-high-resolution Input → Edge Devices **X**

sensor interfaces
sensor readout
memory access
downstream processing
...

Video Perception Performance **Boosts**

Emergent Dual-stream Sensors

Spatiotemporal Trade-off

(a) Object tracking (b) Text recognition (c) Robotics

Resolution reduction: 16x, 8x, 4x, 2x, 1x

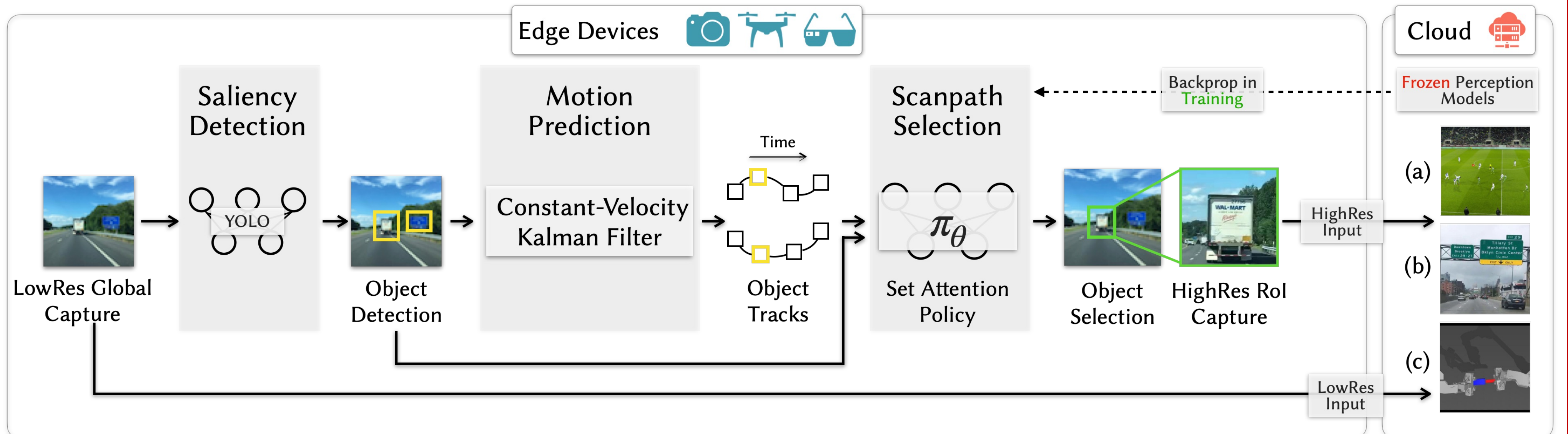
Frame rate reduction: 1x, 2x, 4x, 8x, 16x

IoU: 0.28, 0.04

Acc: 0.33, 0.01

Succ: 0.61, 0.02

Policy-based Foveated Imaging



Related Work

Foveated Computer Vision Active Vision Foveated Graphics

[1] Esteves et al. 2017

[6] Chuang et al. 2025

[3] Fan et al. 2024

[4] Wang et al. 2025

[5] Kerr et al. 2025

[2] Krajancich et al. 2023

Experimental Results

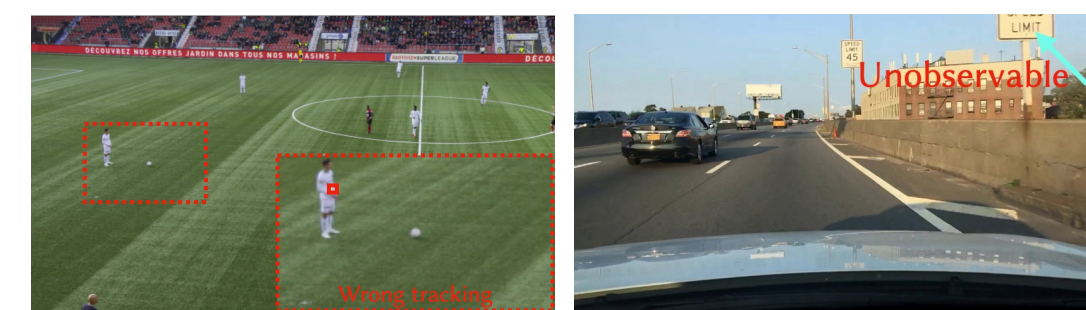
Quantitative Comparison Against Conventional Methods

Task	Metric	GT Oracle	Full-resolution	Spatial downsampling	Temporal downsampling	Foveated (Ours)
Object Tracking	IoU ↑	0.405	0.281	0.122	0.148	0.283
Text Recognition	Transcription Rate ↑	0.271	0.333	0.067	0.248	0.264
Robot Manipulation	Success Rate ↑ (Full Partial)	N/A	0.15 0.61	0.10 0.51	0.07 0.30	0.12 0.57

200 MP Imaging Prototype



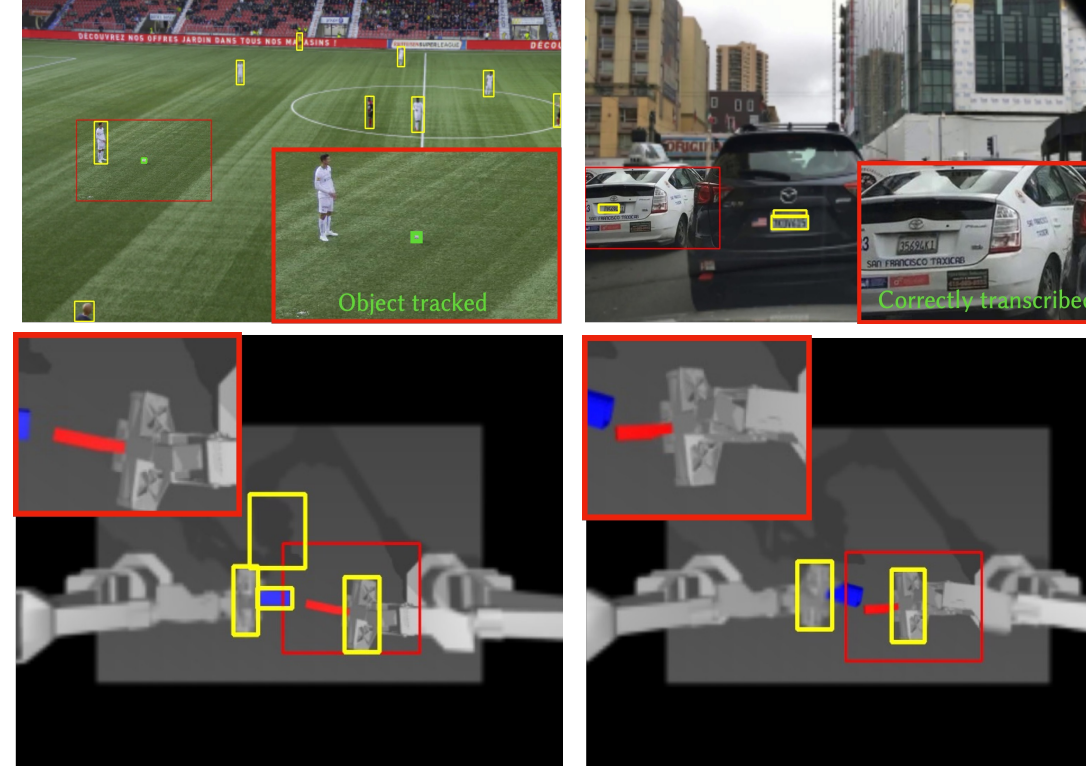
Conventional



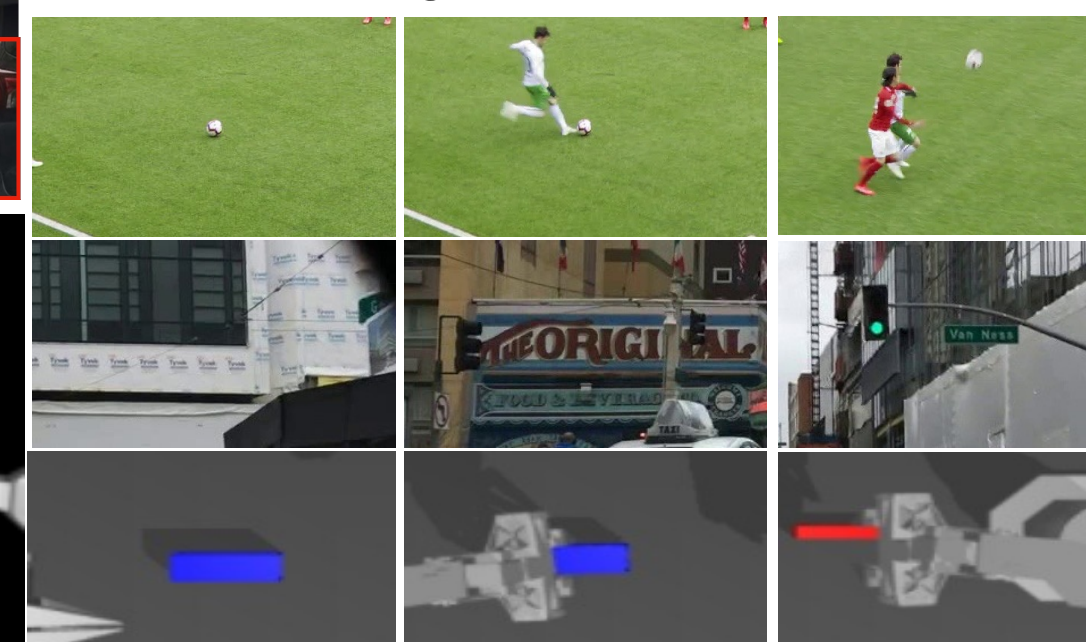
Other Foveation Methods

Variant	Soccer Tracking	Text Recognition	Robot Manipulation
Always-centered ROI selection	1.7%	3.6%	82.0%
Round-robin ROI scanning	3.5%	64.0%	72.1%
Ours	100.7%	79.3%	93.4%

Ours



Emergent Sensor Attention



References

- [1] Esteves, Allen-Blanchette, Zhou, Daniilidis. Polar Transformer Networks. ICLR, 2018.
- [2] Krajancich, Kellnhofer, Wetzstein. Towards Attention-aware Foveated Rendering. ACM Transactions on Graphics (TOG), 2023.
- [3] Fan, Shi, Wang, Ma, Wang. Scene-aware Foveated Rendering. IEEE Transactions on Visualization & Computer Graphics, 2024.
- [4] Wang, Yue, Yue, Wang, Jiang, Han, Ni, Pu, Shi, LU, et al. Emulating Human-like Adaptive Vision for Efficient and Flexible Machine Visual Perception. Nature Machine Intelligence, 2025.
- [5] Kerr, Hari, Weber, Kim, Yi, Bonnen, Goldberg, Kanazawa. 2025. Eye, Robot: Learning to Look to Act with a BC-RL Perception-Action Loop. CoRL, 2025.
- [6] Chuang, Zou, Lee, Gao, Soltani, Look, Focus, Act: Efficient and Robust Robot Learning via Human Gaze and Foveated Vision Transformers. Arxiv, 2025.