The University of Texas Rio Grande Valley
Center for Multidisciplinary Research Excellence in Cyber-Physical Infrastructure Systems (MECIS)

Human-Centric Cities: A Digital Twin-Oriented Design of Interactive Autonomous Vehicles

National Science Foundation
NSF Award No. 2112650

Mario Camarena, Fatemeh Nazari, Ph.D., Mohamadhossein Noruzoliaee, Ph.D.

Introduction

Objective
- Develop a comprehensive perception system featuring a sematic segmentation models for autonomous vehicles (AVs) to contribute to a safer, more efficient, and sustainable transportation.

Motivation
- Improving road safety.
- Transforming transportation.
- Increase traffic efficiency.
- Focus on perception and semantic segmentation

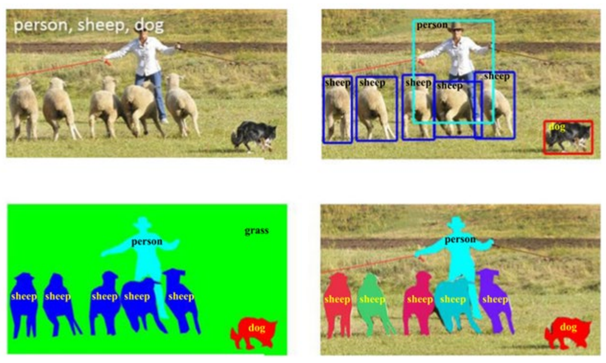
Background

Autonomous Vehicles
- Vehicles equipped with technology to navigate and operate independently, without human control

Importance of AVs
- Improve Safety
- Increase Productivity
- Reduce Traffic
- Create Independence

Autonomous Transportation Visual Representation.

Semantic Segmentation
- Classifies every pixel in an image.
- Utilizes deep learning for precision.
- Enables environment understanding.



Classification, Generic Segmentation, Sematic Segmentation, Object Segmentation.

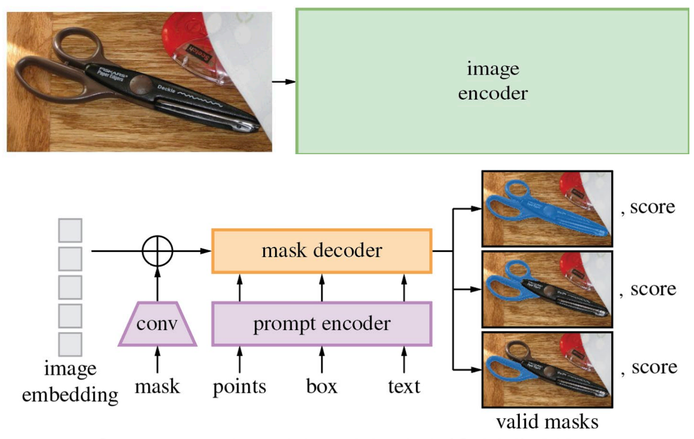
Methodology and Data

Data
- Cityscapes dataset [4]: Real-world urban street scenes with pixel-level annotations.

Preprocessing Data
- Resolution: Images resized to 1024x1024.
- 19 classes.

Hyperparameters

Learning Rate	Epochs	Batch Size	Activation	Loss Function
0.001	100	4	GELU	Focal

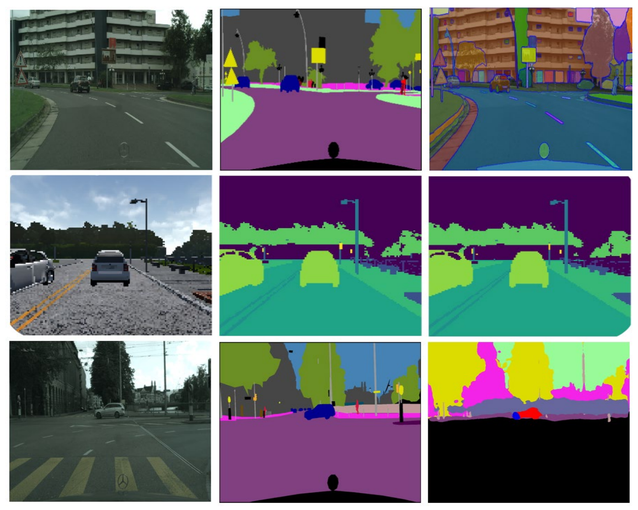


Segment Anything Model (SAM) [1]

Results

Models	SAM	U-Net	DeepLabV 3+
Global Accuracy	92.81%	85.46%	77.91%
Global Jaccard	87.36%	77.08%	64.16%
Global Precision	92.98%	85.16%	77.91%
Global F1 Score	92.76%	85.26%	75.15%
Global Recall	92.81%	85.46%	77.98%

Performance comparison of segmentation models (SAM [1], U-Net [2], and DeepLabV3+ [3]) trained on the Cityscapes dataset across key metrics



Segmentation comparisons across models: SAM [1] (top), U-Net [2] (middle), and DeepLabV3+ [3] (bottom).

Conclusions & Future Work

- Demonstrated high precision in segmenting urban elements like roads, sidewalks, vehicles, and pedestrians.
- Proven robust and generalizable with up to 93% accuracy on the training dataset.
- Showed great potential for real-world applications.

- Will explore end-to-end deep learning approaches, like reinforcement learning, for direct sensor-to-command mappings.
- Will extend to instance segmentation for detailed object class tracking.

Acknowledgments

This work was made possible through the support of the CREST Center for Multidisciplinary Research Excellence in Cyber-Physical Infrastructure Systems (MECIS) through NSF Award No. 2112650.

References

[1] Kirillov, A., et al. (2023). Segment Anything. arXiv:2304.02643.
[2] Ronneberger, O., et al. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. MICCAI, 234–241.
[3] Chen, L. C., et al. (2017). DeepLab: Semantic Image Segmentation with Deep Convolutional Nets. arXiv:1606.00915.
[4] Cordts, M., et al. (2016). Cityscapes dataset for semantic urban scene understanding. CVPR, 3213–3223.