

LangDA: Language-guided Domain Adaptive Semantic Segmentation

1/21/2025

Chang Liu¹ Saad Hossain¹ C Thomas² Kwei-Herng Lai²
Raviteja Vemulapalli² Sirisha Rambhatla¹ Alexander Wong^{1,2}

¹University of Waterloo ²Apple

{chang.liu, s42hossa, sirisha.rambhatla, alexander.wong}@uwaterloo.ca
{c.thomas, khlai, r_vemulapalli}@apple.com



UNIVERSITY OF
WATERLOO

FACULTY OF
ENGINEERING

Complex Scene: How Should We Understand It?

Classification

Detection

Semantic Segmentation



This Is... A Road?

Classification

Detection

Semantic Segmentation



Locate Cars & Bikes

Classification

Detection

Semantic Segmentation



Knows exactly where each class is

Classification

Detection

Semantic Segmentation



Road

Sidewalk

Car

Pole

Building

Sign

Fence

Tram

Vegetation

Static

Sky

Wall

Dynamic

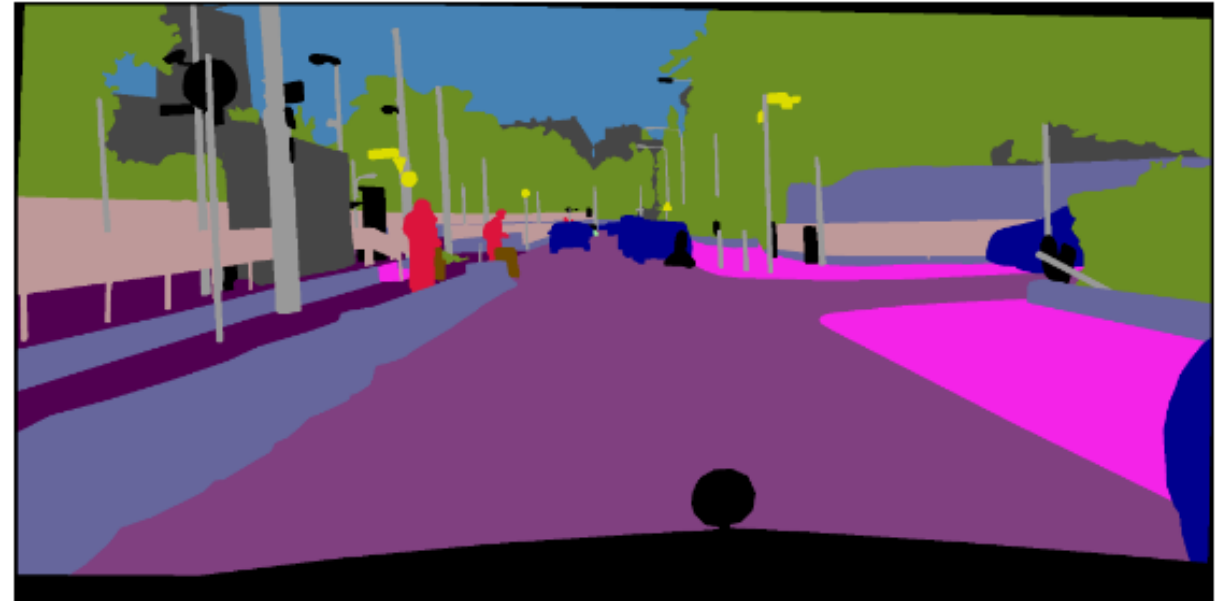
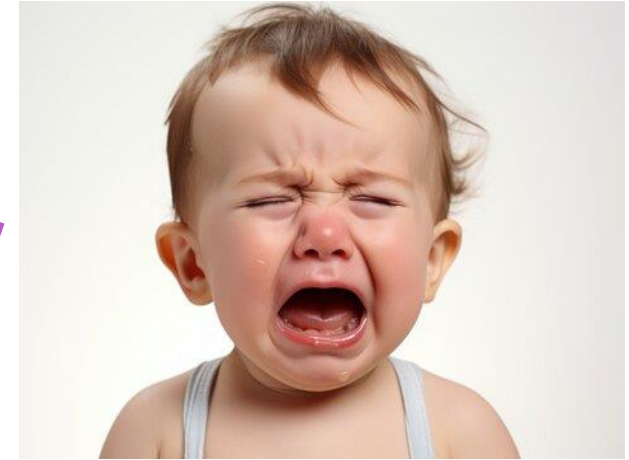
Person



However... Manual Annotation



However... Manual Annotation



Deep Learning Methods

Semi-supervised learning

Weakly-supervised learning

Transfer learning

Unsupervised domain adaptation

Learning from Synthetic Data

Zero-shot learning and few-shot learning

Active learning

Deep Learning Methods

Semi-supervised learning

Weakly-supervised learning

Transfer learning

Unsupervised domain adaptation

Learning from Synthetic Data

Zero-shot learning and few-shot learning

Active learning

Transfer Learning: Transfer and Adaptation

- Learn on one task, transfer to another
- Learn on one labelled distribution, test on another unlabeled distribution

Transfer Learning: Transfer and Adaptation

- Learn on one task, transfer to another
- Learn on one labelled distribution, test on another unlabeled distribution

Unsupervised Domain Adaptation

Domain Gap

Related data distributions still have differences

Source domain -> Unlabeled Target domain

Domain Gap

Related data distributions still have differences

Source domain -> Unlabeled Target domain



- Different weather

Domain Gap

Related data distributions still have differences

Source domain -> Unlabeled Target domain



- Different weather

Domain Gap

Related data distributions still have differences

Source domain -> Unlabeled Target domain



- Different weather, lighting

Domain Gap

Related data distributions still have differences

Source domain -> Unlabeled Target domain



- Different weather, lighting
- Synthetic vs. real

Domain Gap

Related data distributions still have differences

Source domain -> Unlabeled Target domain



- Different weather, lighting
- Synthetic vs. real

Domain Gap

Related data distributions still have differences

Source domain -> Unlabeled Target domain

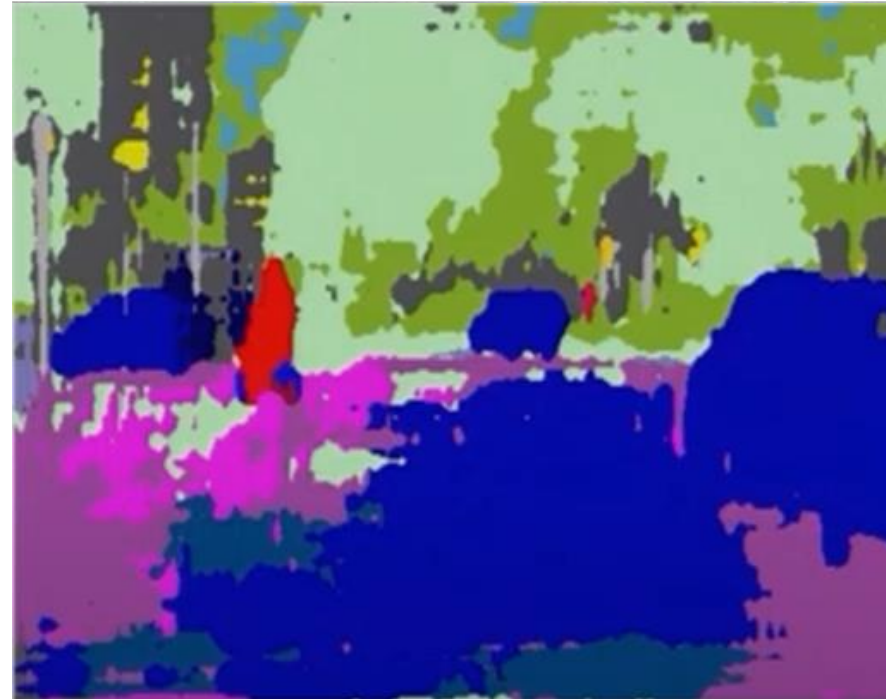


- Different weather, lighting
- Synthetic vs. real

Domain Gap

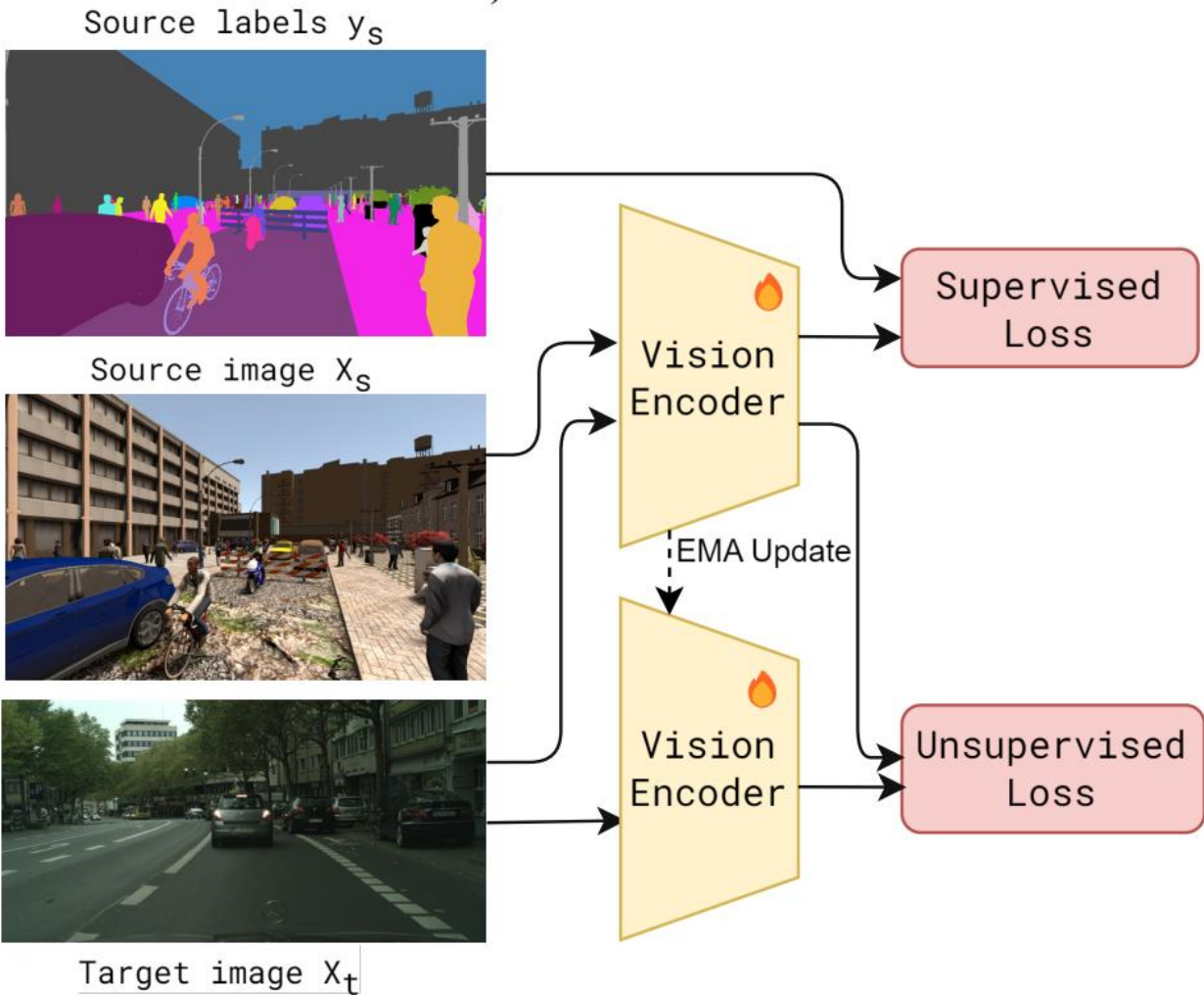
Related data distributions still have differences

Source domain -> Unlabeled Target domain



- Different weather, lighting
- Synthetic vs. real

Prior Work: Traditional UDA



- Knowledge distillation

Images



Give the model more information?

Images



Give the model more information?

Images



Language information?

Give the model more information?

Images



Vision Language Models!

Give the model more information?

Images



Vision Language Models!

“The image shows a busy city scene with road filled with cars, motorcycles, and bicycles. Sidewalk has pedestrians and riders. Buildings, fence, and pole are in the background. Vegetation is near the edge, and sky above. People are scattered across the scene, some near the road, others by the buildings.”

Give the model more information?

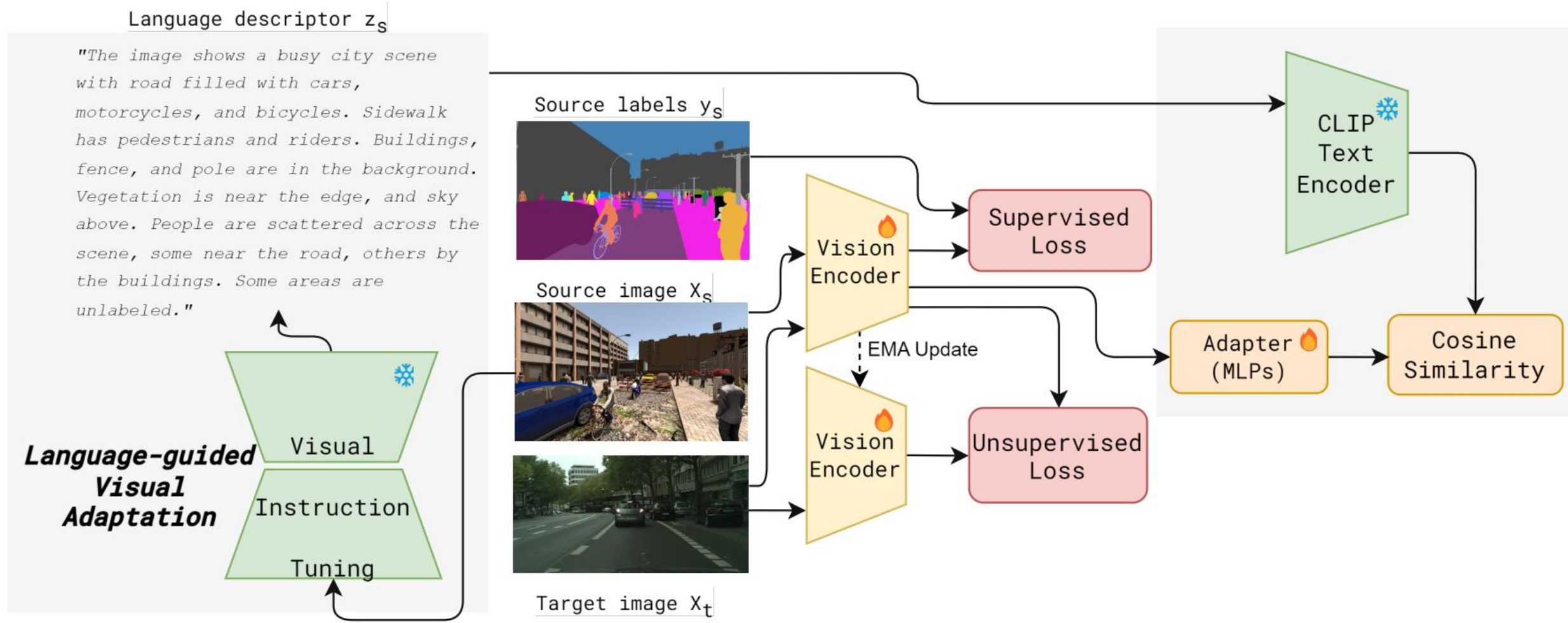
Images



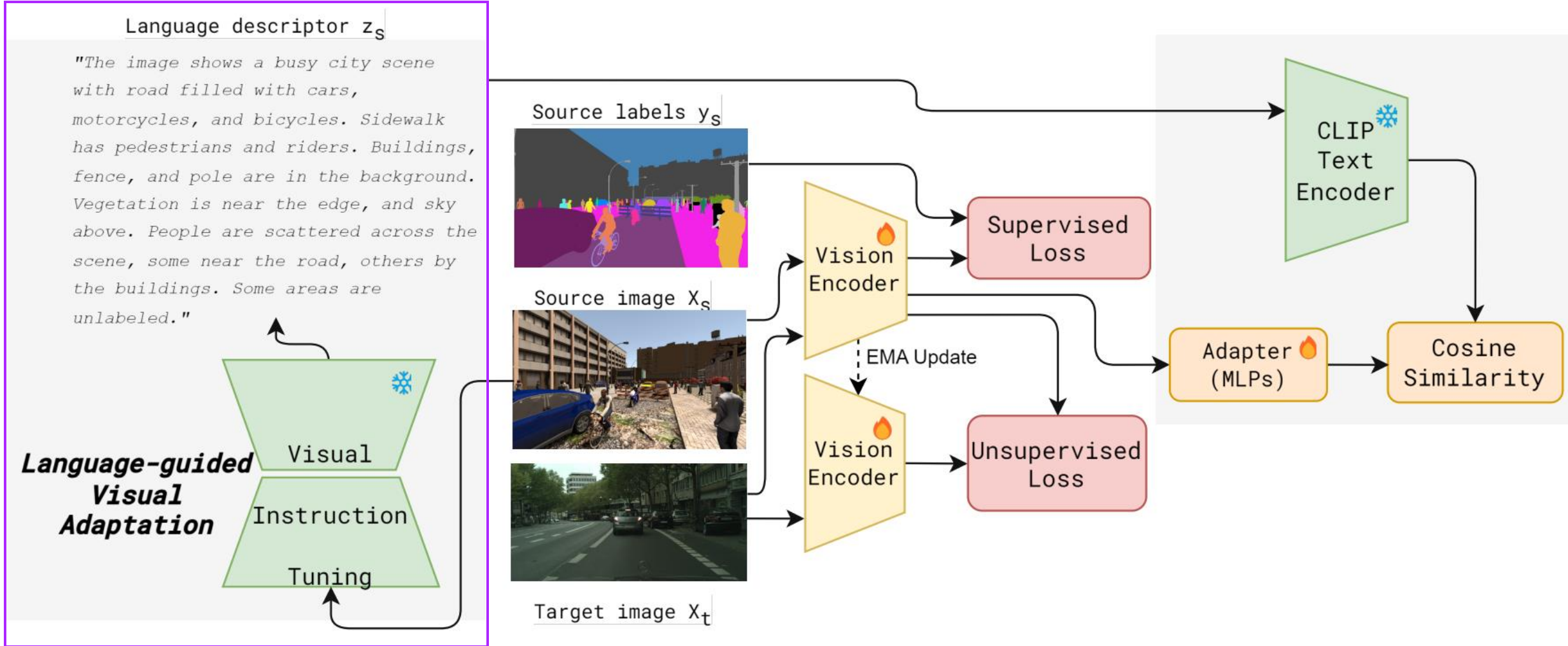
Vision Language Models!

“The image shows a busy city scene with road filled with cars, motorcycles, and bicycles. Sidewalk has pedestrians and riders. Buildings, fence, and pole are in the background. Vegetation is near the edge, and sky above. People are scattered across the scene, some near the road, others by the buildings.”

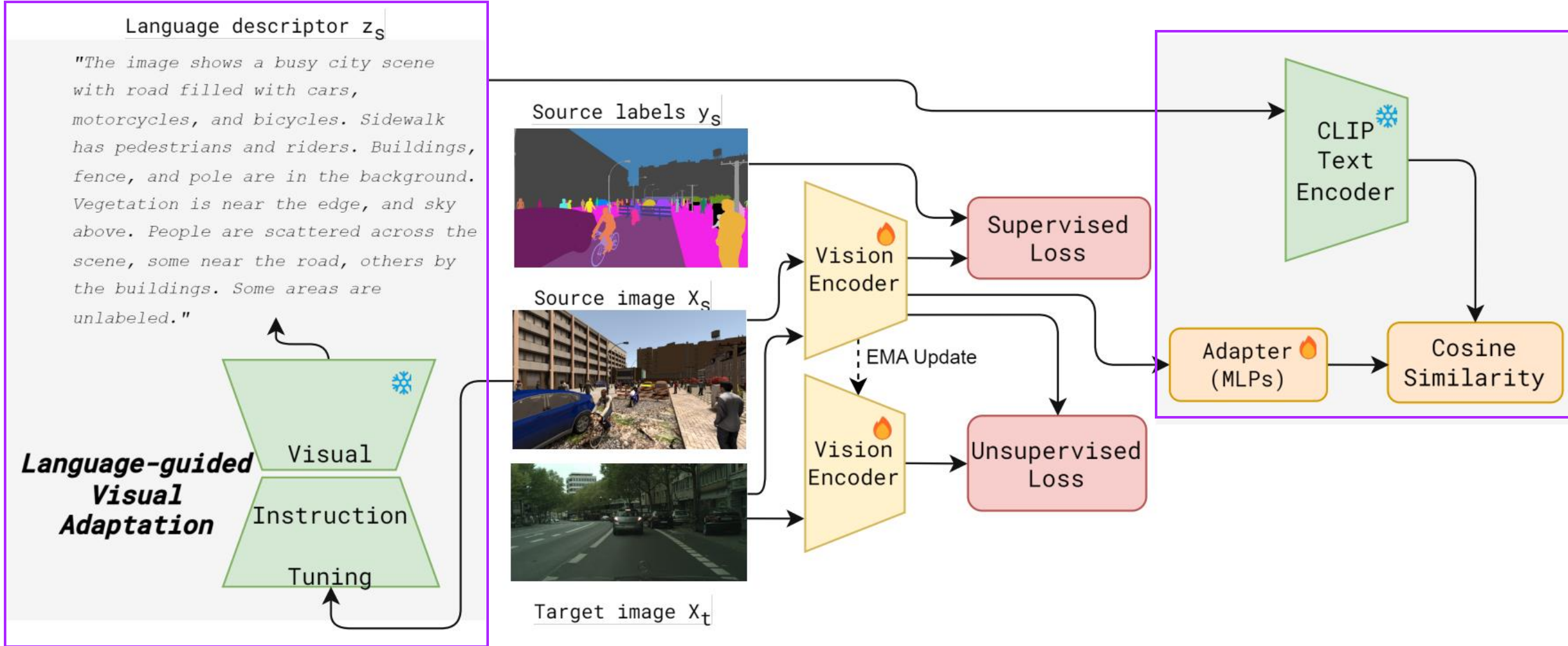
Proposed Method: LangDA



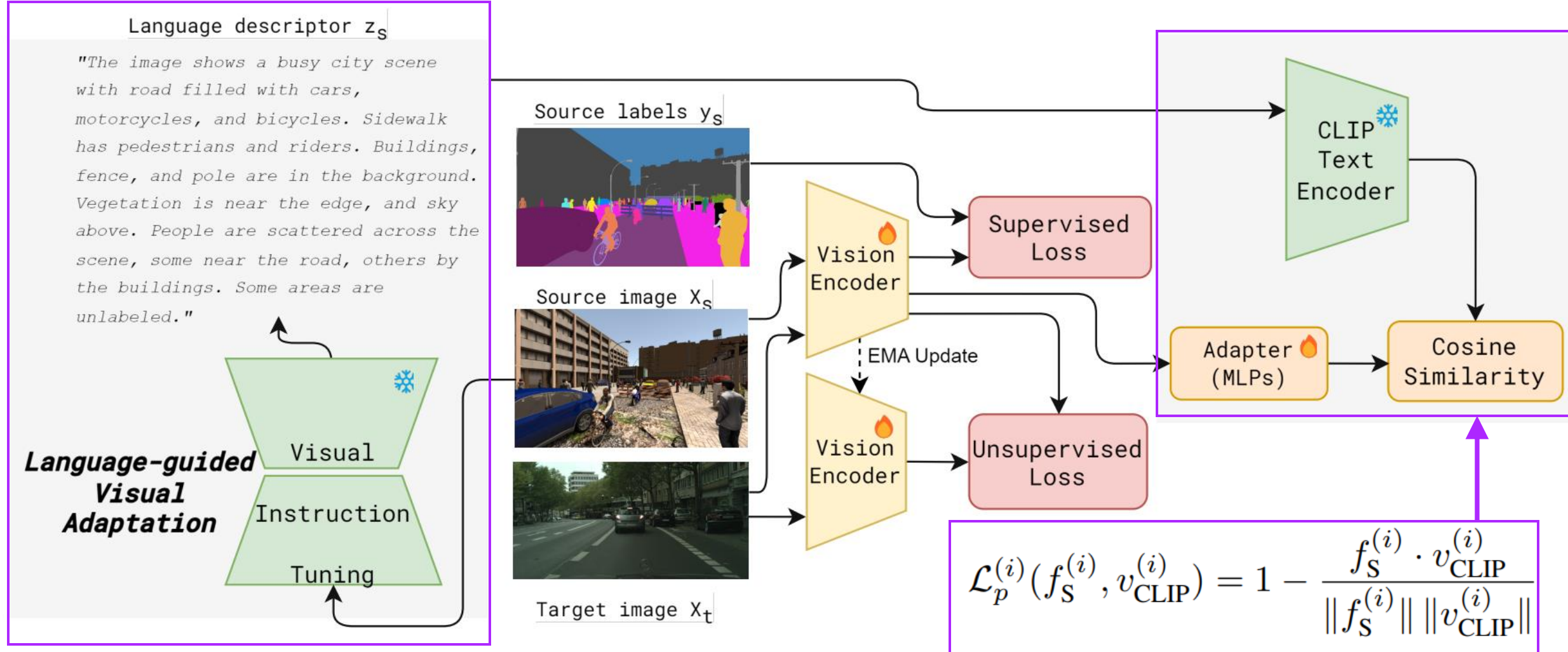
Proposed Method: LangDA



Proposed Method: LangDA

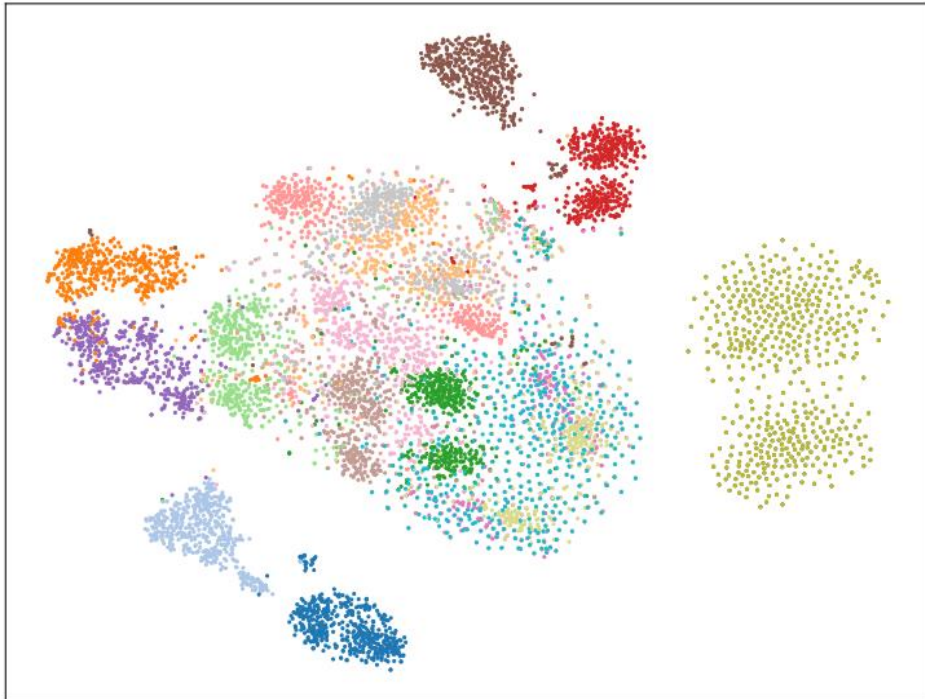


Proposed Method: LangDA

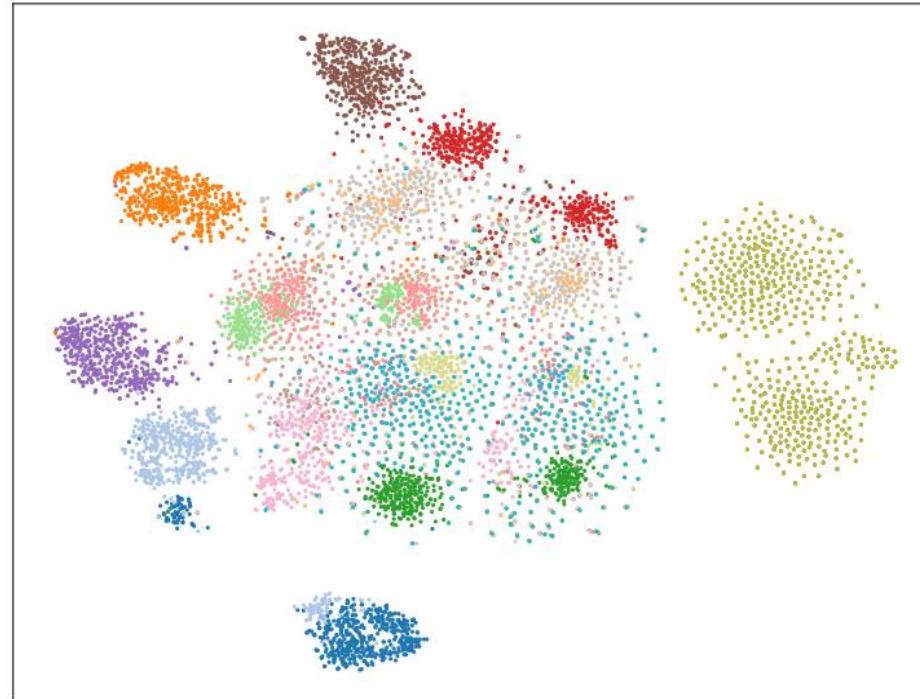


t-SNE

Left: Prior Work (DAFormer)



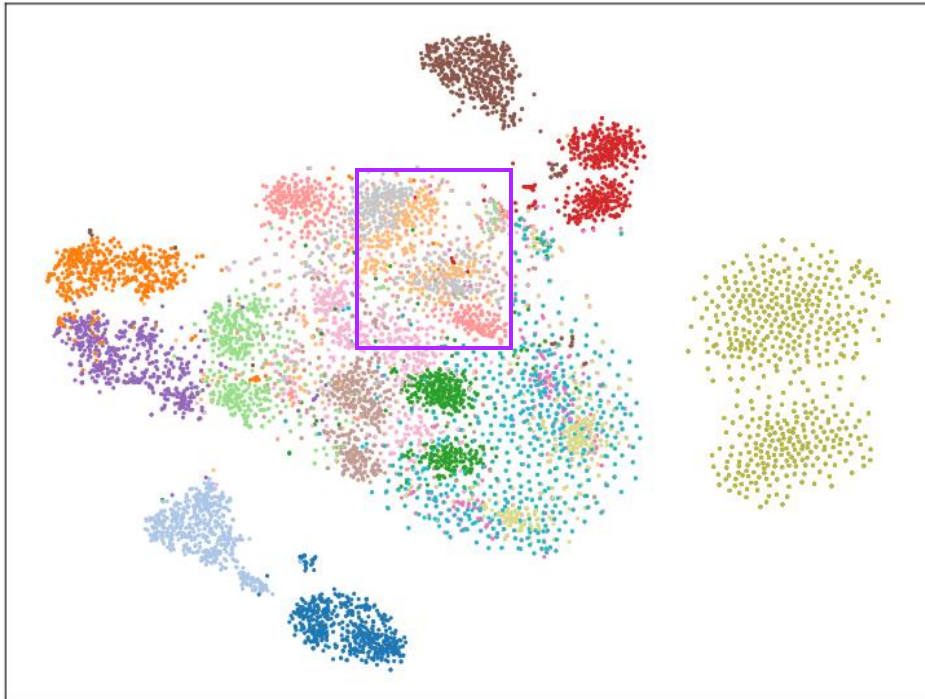
Right: LangDA (Ours)



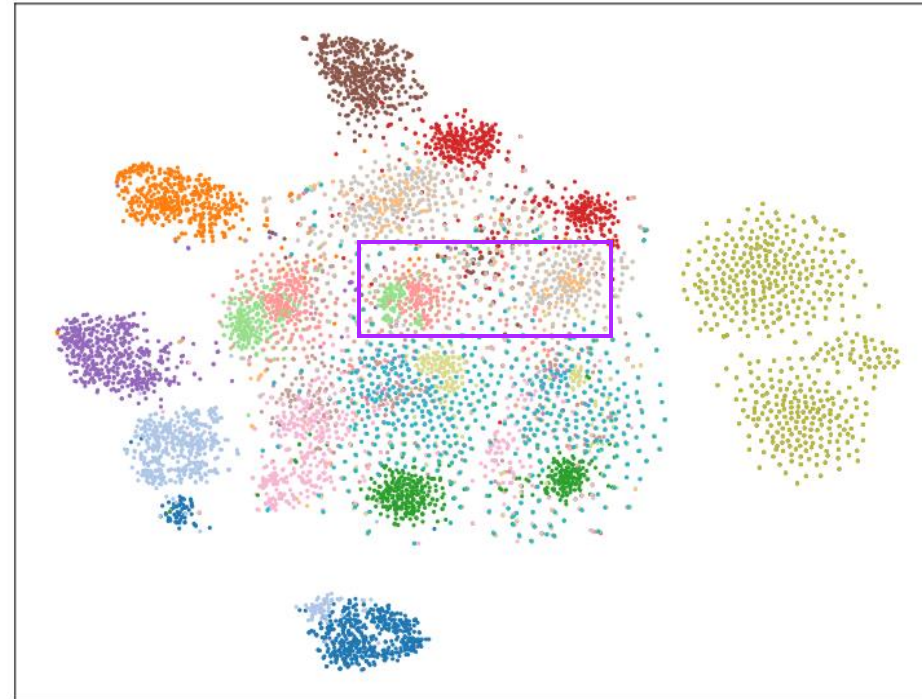
- road
- sidewalk
- building
- wall
- fence
- pole
- traffic light
- traffic sign
- vegetation
- terrain
- sky
- person
- rider
- car
- truck
- bus
- train
- motorcycle
- bicycle

t-SNE

Left: Prior Work (DAFormer)



Right: LangDA (Ours)



- road
- sidewalk
- building
- wall
- fence
- pole
- traffic light
- traffic sign
- vegetation
- terrain
- sky
- person
- rider
- car
- truck
- bus
- train
- motorcycle
- bicycle

Quantitative Result

Method	Backbone	Unlabeled Target Data	Prompt Description	% mIoU \uparrow
Source only	ResNet-50			29.3
PODA [13]	ResNet-50		✓	29.5
ULDA [11]	ResNet-50		✓	30.8
Source only	ResNet-101			29.4
ADVENT [3]	ResNet-101	✓		41.2
CBST [25]	ResNet-101	✓		42.6
DACS [4]	ResNet-101	✓		48.3
CorDA [26]	ResNet-101	✓		55.0
ProDA [27]	ResNet-101	✓		55.5
DAFormer [2]	SegFormer	✓		61.1
LangDA (Ours)	SegFormer	✓	✓	62.0 (+0.9%)

What's Next?

More datasets:
Day -> Night,
Normal ->
Adverse Weather

More baselines &
existing works

Ablation studies

More Quantitative Result

Method	Backbone	Unlabeled Target Data	Text Prompts	% mIoU \uparrow
Source only	ResNet-50			29.3
PODA [†] [8]	ResNet-50		✓	29.5
ULDA [†] [40]	ResNet-50		✓	30.8
Source only	ResNet-101			29.4
ADVENT [37]	ResNet-101	✓		41.2
CBST [43]	ResNet-101	✓		42.6
DACS [36]	ResNet-101	✓		48.3
CorDA [38]	ResNet-101	✓		55.0
ProDA [42]	ResNet-101	✓		55.5
DAFormer [†] [11]	SegFormer	✓		61.1
LangDA(Ours) + DAFormer	SegFormer	✓	✓	62.0
HRDA [12]	SegFormer	✓		65.8
LangDA (Ours) + HRDA	SegFormer	✓	✓	66.3
MIC [13]	SegFormer	✓		67.3
CoPT [24]	SegFormer	✓	✓	67.4
LangDA (Ours) + MIC	SegFormer	✓	✓	70.0

Qualitative Results: Synthetic -> Real

Image



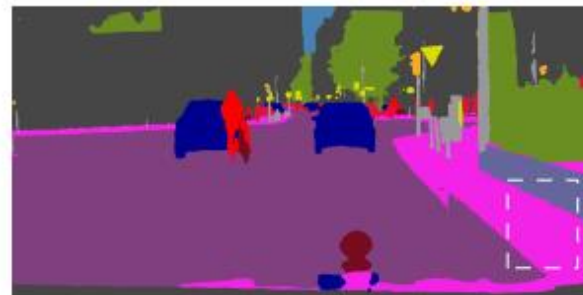
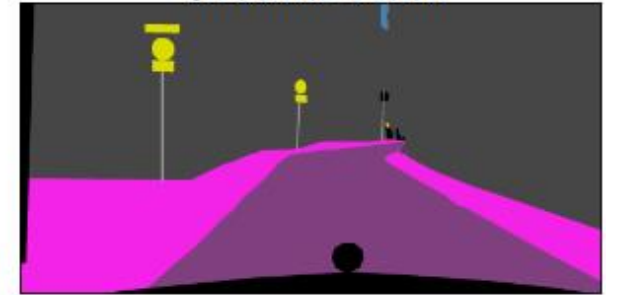
MIC [13]



LangDA (Ours)



Ground Truth



Qualitative Results: Normal -> Adverse Weather

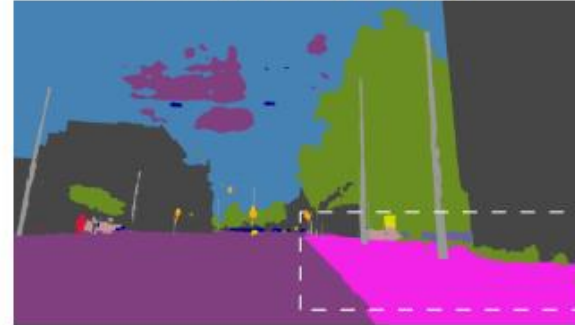
Image



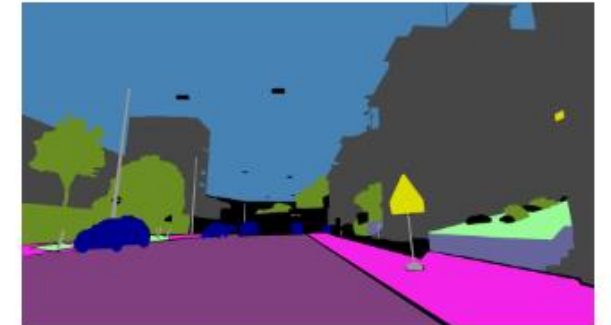
MIC [3]



LangDA (Ours)



Ground Truth

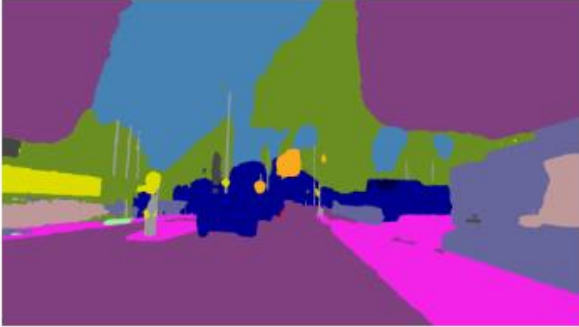


Qualitative Results: Day -> night

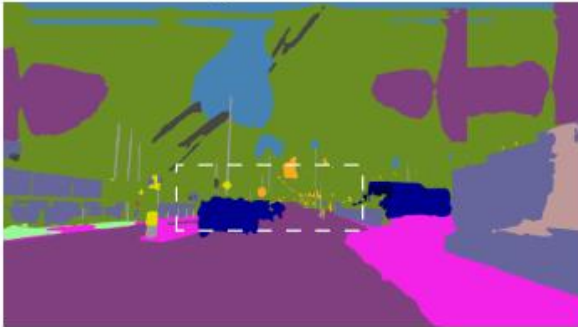
Image



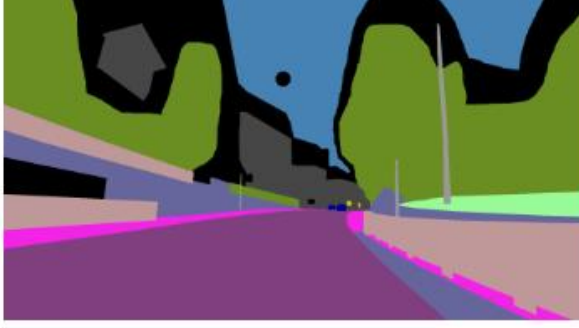
MIC [3]



LangDA (Ours)



Ground Truth



UNIVERSITY OF
WATERLOO



FACULTY OF ENGINEERING

Questions?