Artificial Intelligence Approach for Multi-modal Pavement Condition Evaluation

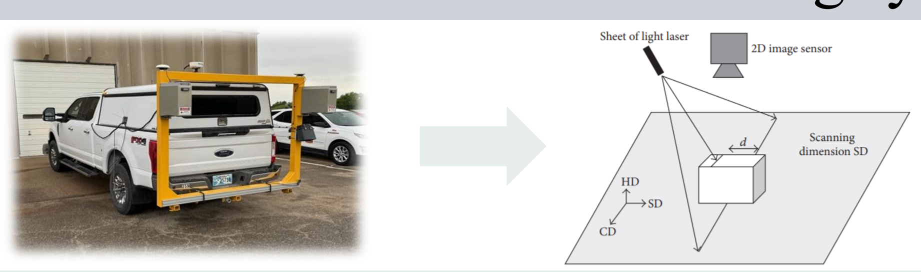
Junlong Liu[1], Joshua Q. Li[2], Fatemeh Ansari[3], Amir Golalipour[4], Kevin C.P. Wang[5]

1.Research Assistant, School of Civil and Environmental Engineering 2.Professor, School of Civil and Environmental Engineering, Oklahoma State University 3. Graduate Traffic Engineer, WSB Inc. 4.Office of Infrastructure & Development, Federal Highway Administration, 5.Professor, Western Transportation Institute, Montana State University

Introduction

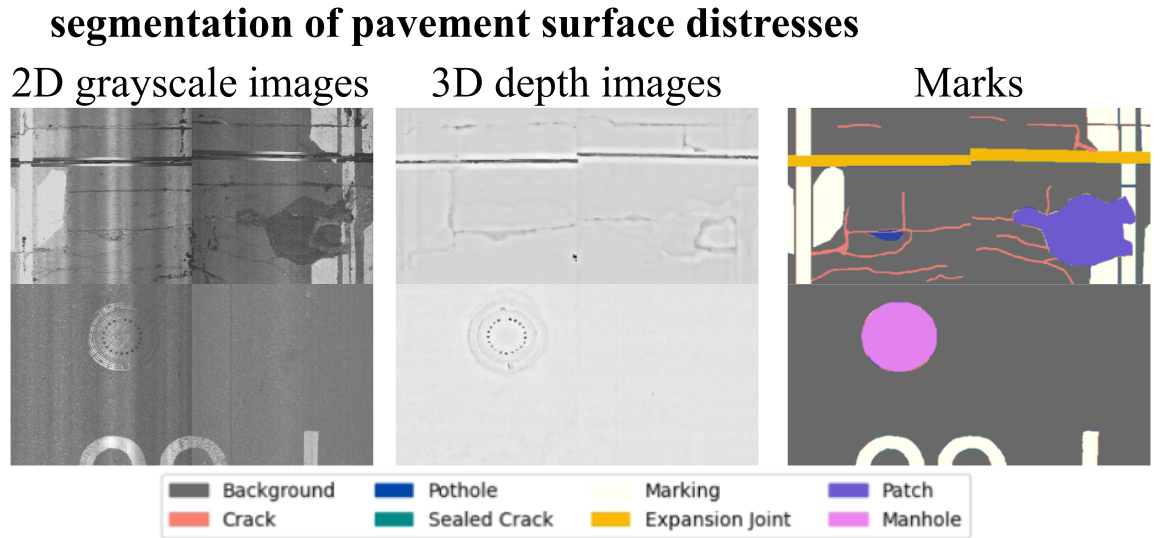
- Traditional methods: **time-consuming, labor-intensive**
- Digital image processing methods: improved automation but **remain sensitive to noise, lighting and complex backgrounds**
- Deep learning: offers superior feature extraction and segmentation, yet most deep learning methods
  - **Focus on single-modal and single-object object**
  - **Difficult to achieve complementarity between pixel and depth information**
  - **Unable to meet the requirements in real word**
- Multi-modal and multi-object methods: **combine 2D grayscale images and 3D depth images to achieve multi-object high-precision pavement condition evaluation**

Data collection: 3D laser scanning system



Objective

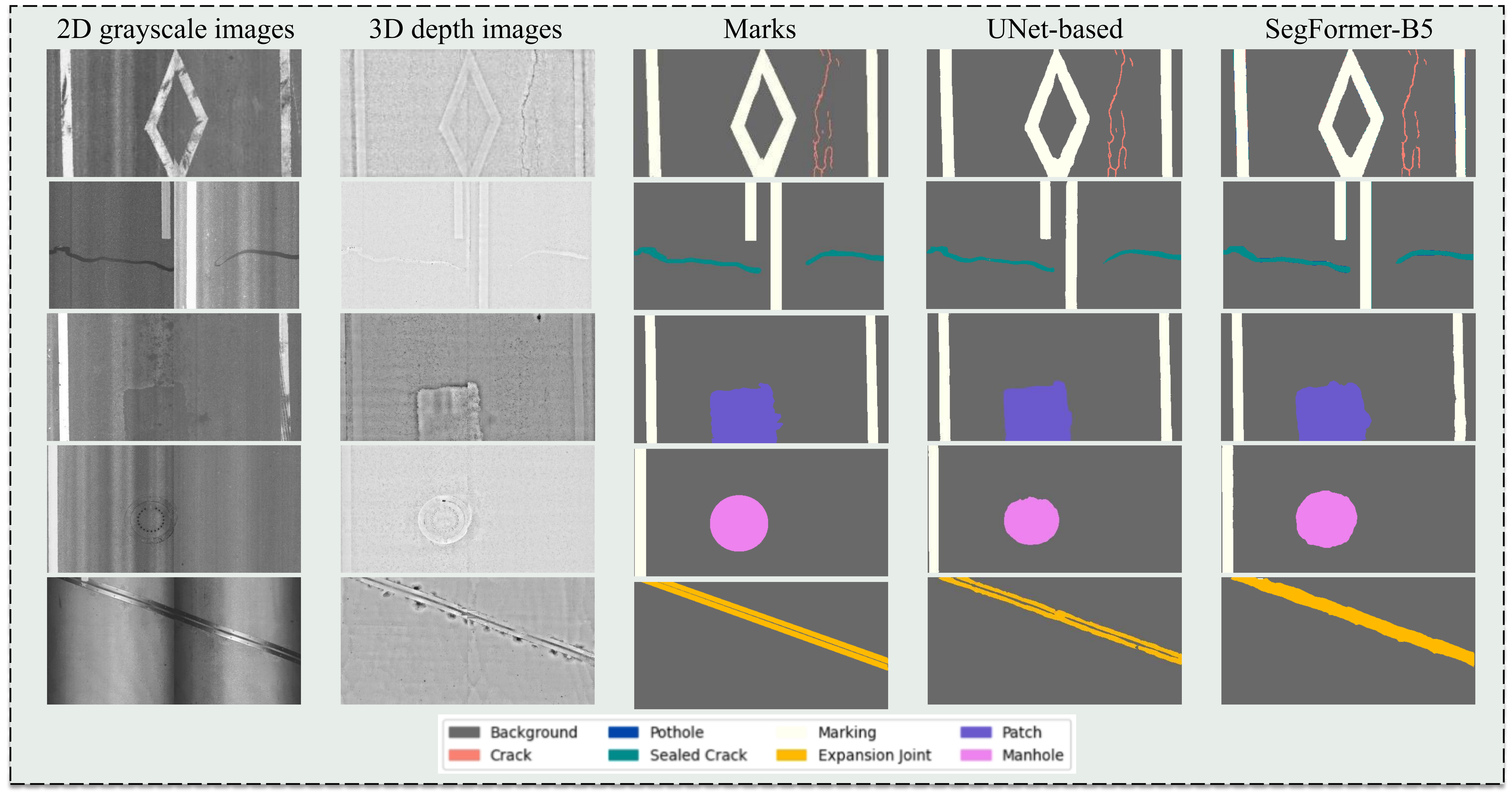
- **Develop UNet-based and SegFormer-B5 multi-modal models** for pavement condition evaluation
- **Compare proposed models for multi-object, pixel-level segmentation of pavement surface distresses**



Method

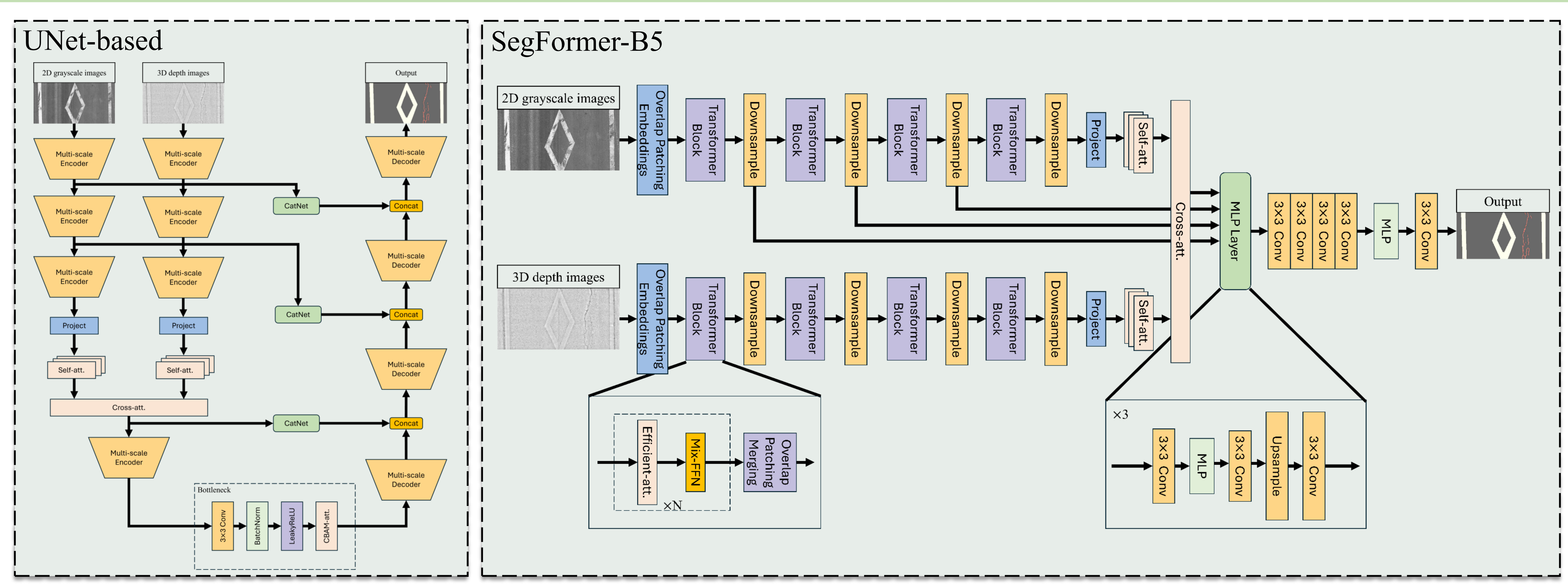
- Multi-modal Input Fusion
  - Dural independent encoder for 2D grayscale images and 3D depth images, self attention and cross attention to query and match each other
- UNet-based Architecture: strong local modeling, efficient
  - Encoder-decoder with skip connections for fine-grained pixel segmentation
  - Four multi-scale encoder & decoder + CBAM attention bottleneck
- SegFormer-B5 Architecture: strong global modeling, feature fusion
  - Encoder: efficient self-attention + overlapped patch merging for multi-scale features.
  - Decoder: MLP fusion, resolution-independent, parameter-efficient
- Train
  - Hardware: RTX 5090
  - Parameters: AdamW optimizer, epoch 300, batch size 8
  - Loss fuction = 2Cross-entropy loss + Dice loss + Focal loss

Results

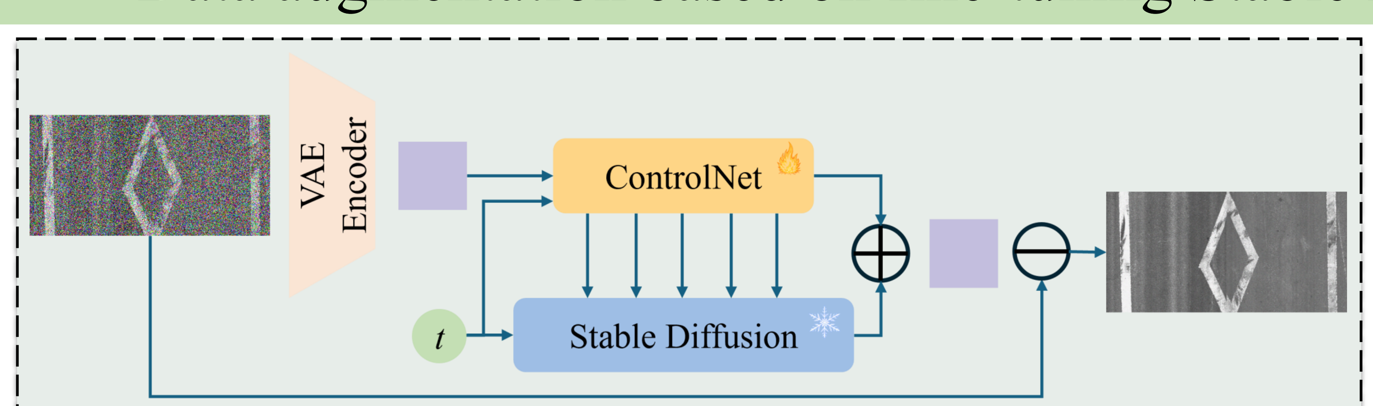


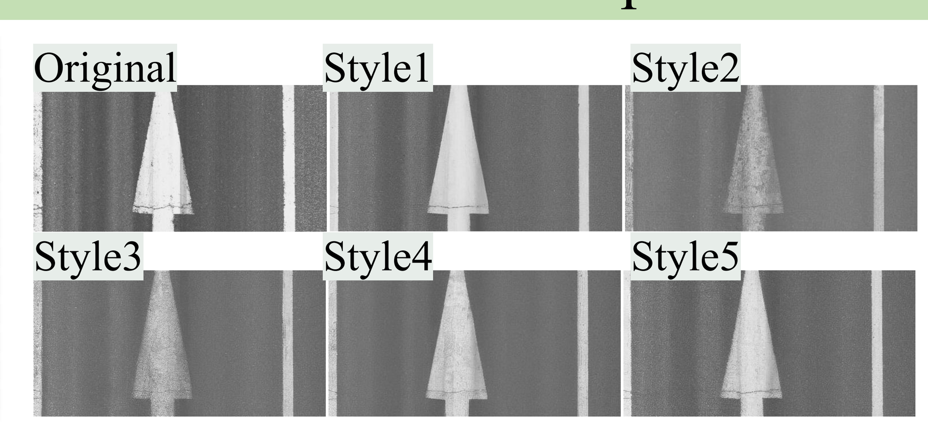
Class Name	UNet-based model		SegFormer-B5 model	
	F1-score	IoU	F1-score	IoU
Background	**99.09%**	**98.20%**	98.83%	97.69%
Crack	**73.41%**	**57.99%**	69.32%	53.05%
Pothole	80.44%	67.27%	**85.25%**	**74.29%**
Sealed crack	88.35%	79.14%	**88.70%**	**79.69%**
Marking	96.41%	93.07%	**96.83%**	**93.86%**
Expansion joint	93.19%	87.25%	**97.40%**	**94.93%**
Patch	**92.65%**	**86.31%**	90.23%	82.20%
Manhole	92.30%	85.70%	**96.35%**	**92.96%**
Overall	89.48%	81.87%	**90.36%**	**83.58%**

Multi-modal fusion model architecture



Data augmentation based on fine-tuning Stable Diffusion for crack and pothole





- By changing the random seed, each mask can generate 5 different styles of images
- The total number of pixels representing cracks and potholes was counted for each generated image. The top 30% of these images were selected as the training data, resulting in a total of 9514 synthetic images.

Discussions

- **Data imbalance**
  - Due to the inherent imbalance of class data, the accuracy of Crack and Pothole is relatively low
- **Improvement**
  - Class balancing strategy： data augmentation for specific object

Conclusions

- **Proposed two multi-modal and multi-object pavement condition evaluation models**
- **UNet**
  - Strengths: Ability of local feature modeling (crack, patch), and higher recall
  - Weaknesses: Struggles with global feature modeling
- **SegFormer-B5**
  - Strengths: Ability of global feature and irregular/complex modeling (expansion joint, marking), and higher precision
  - Weaknesses: Lower performance on rare classes (crack, pothole)
- **The two model architectures demonstrate complementary capabilities.**

Acknowledgements

This research was funded by the Federal Highway Administration under the Exploratory Advanced Research Program (Washington, DC: 2024)