2024

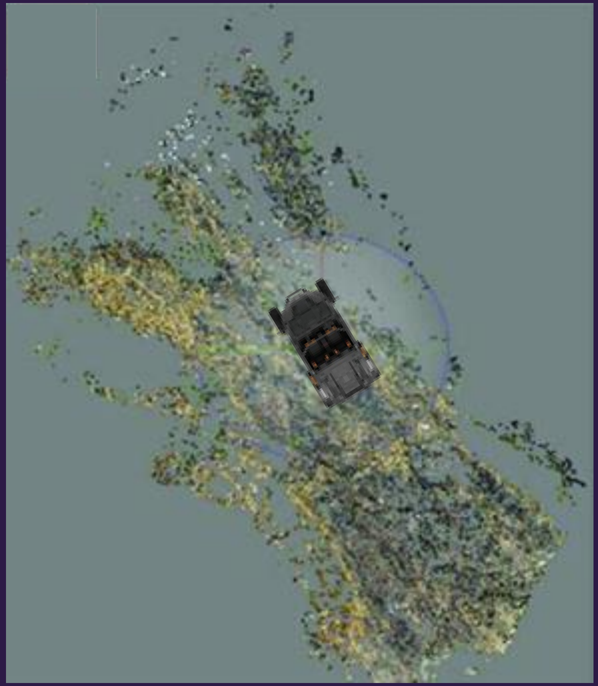
2.5-D Map Inpainting

Machine Learning Based Image Synthesis

Benjamin Johnson, Vasudev Purohit

CPSC 8810

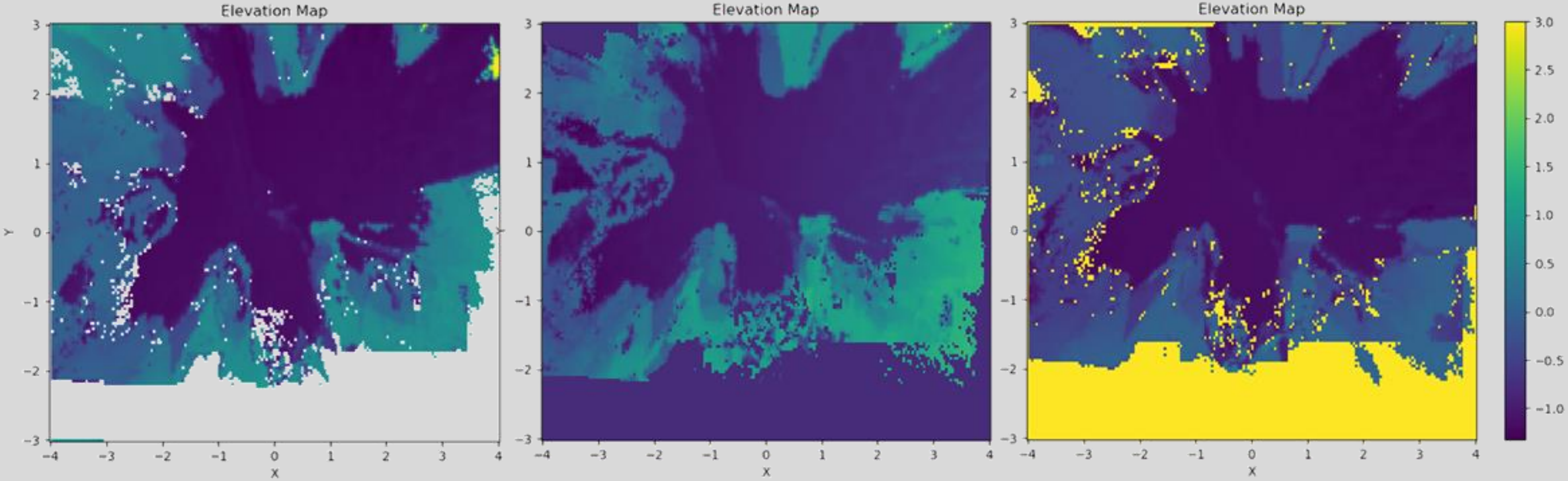
Department of Automotive Engineering



Introduction

- Cluttered environments lead to occlusion
- Traditional planners assume free or occupied, neither being ideal
- Data driven inpainting has been shown to help

Polaris RZR Pro-R 4



What can we do better? → **Costmap Inpainting**

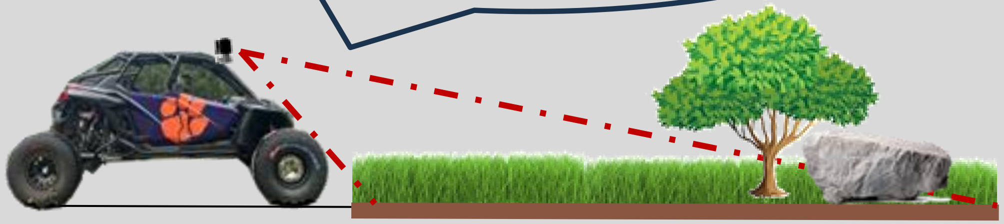
Introduction

Department of AUTOMOTIVE ENGINEERING

★ Why is uncertainty quantification in in-painting important?

Deterministic decision-making

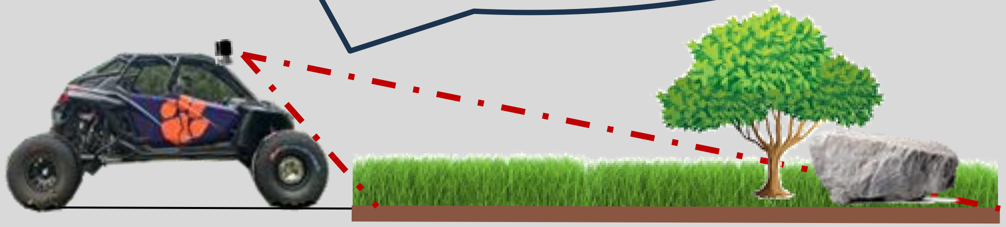
My map tells me that there is no rock behind the tree, I can move ahead with high speed!



Potentially Unsafe!

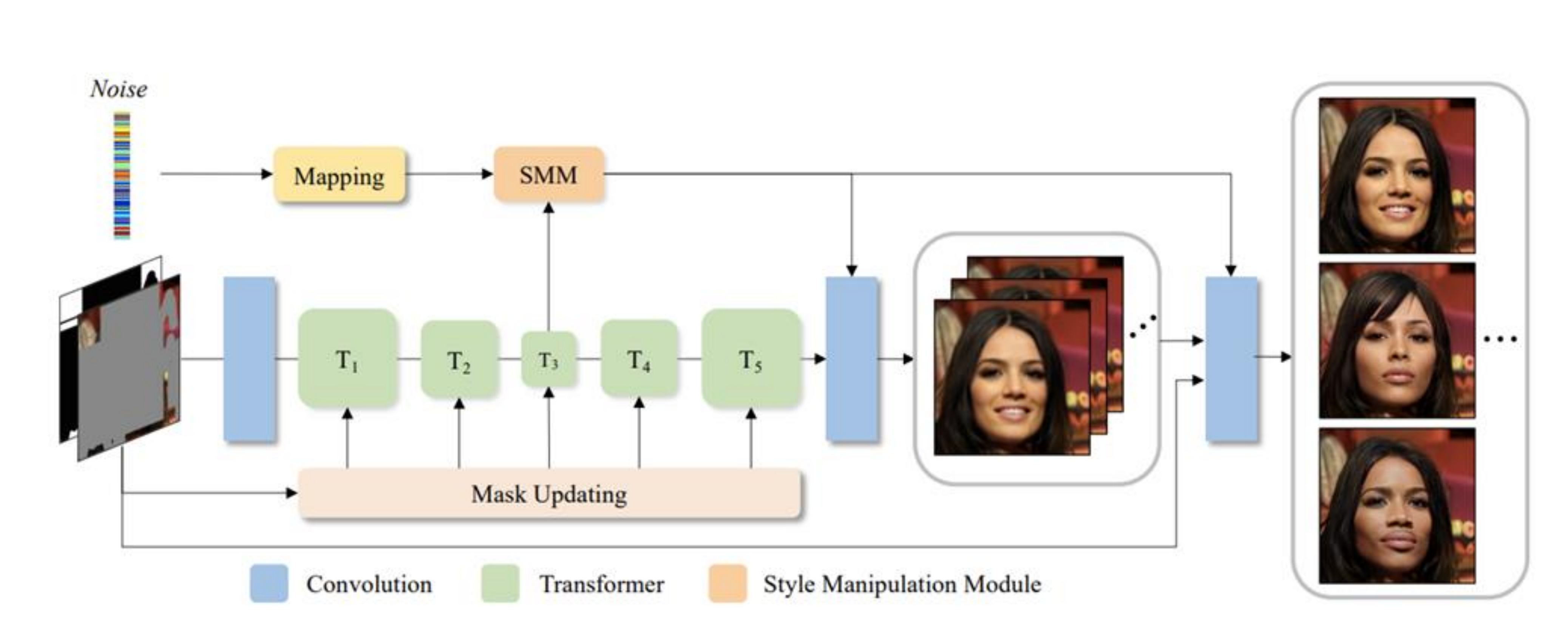
Stochastic decision-making

My map tells me that there is a 5% chance that there could be a rock behind the tree, I must move with caution!



Safe!

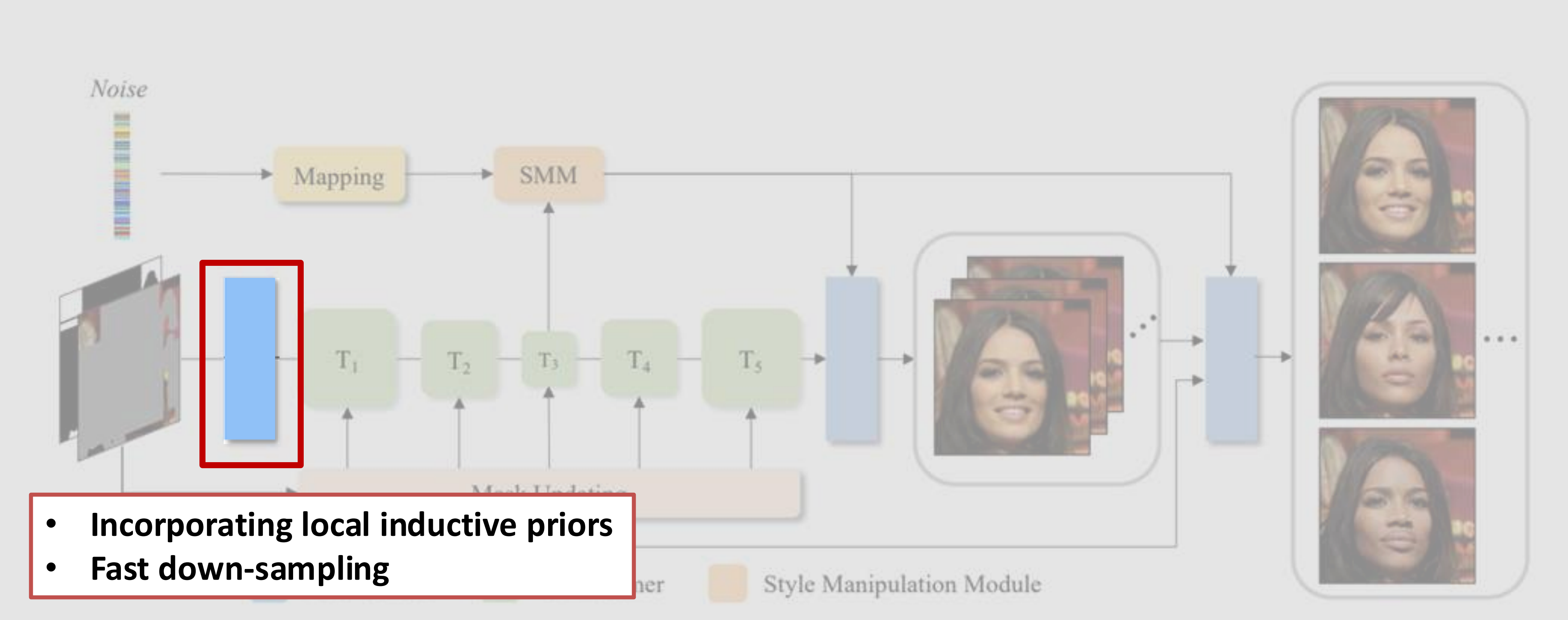
Model Architecture



MAT: Mask-Aware Transformer for Large Hole Image Inpainting [1]

[1] Li, W., Lin, Z., Zhou, K., Qi, L., Wang, Y., & Jia, J. (2022). Mat: Mask-aware transformer for large hole image inpainting. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 10758-10768).

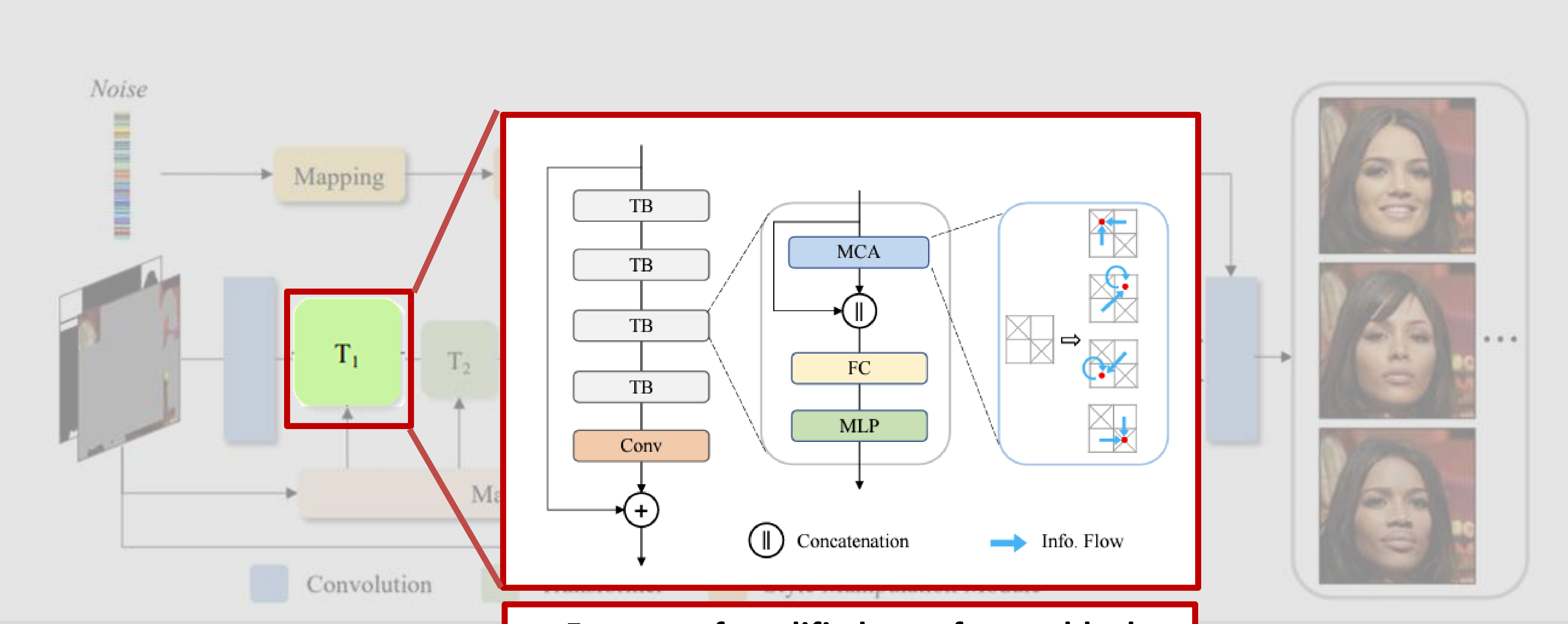
Model Architecture



- **Incorporating local inductive priors**
- **Fast down-sampling**

Model Architecture

Department of AUTOMOTIVE ENGINEERING

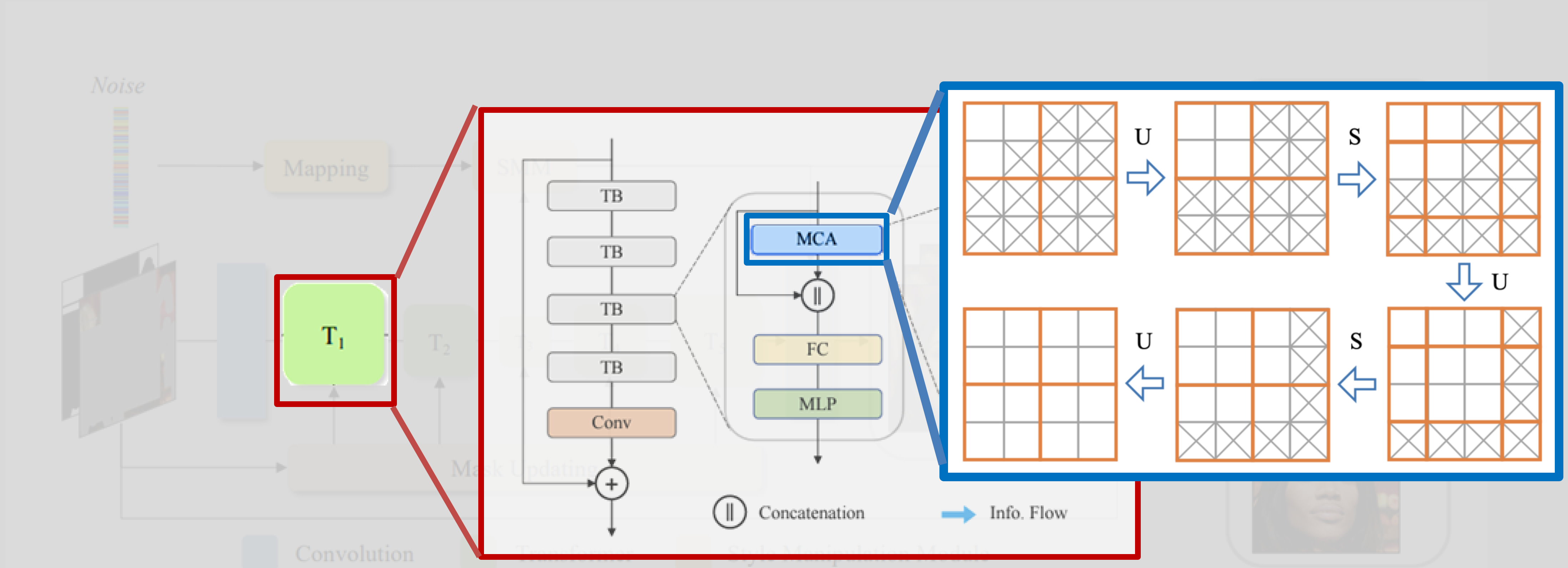


- 5 stages of modified transformer block
- Remove layer normalization
- Fusion learning i.l.o. residual learning, thus avoiding unstable optimization

$$\mathbf{X}'_{k,\ell} = \mathrm{FC}(\,[\mathrm{MCA}(\mathbf{X}_{k,\ell-1}),\ \mathbf{X}_{k,\ell-1}]\,)\,,$$

$$\mathbf{X}_{k,\ell} = \mathrm{MLP}(\mathbf{X}'_{k,\ell})\,.$$

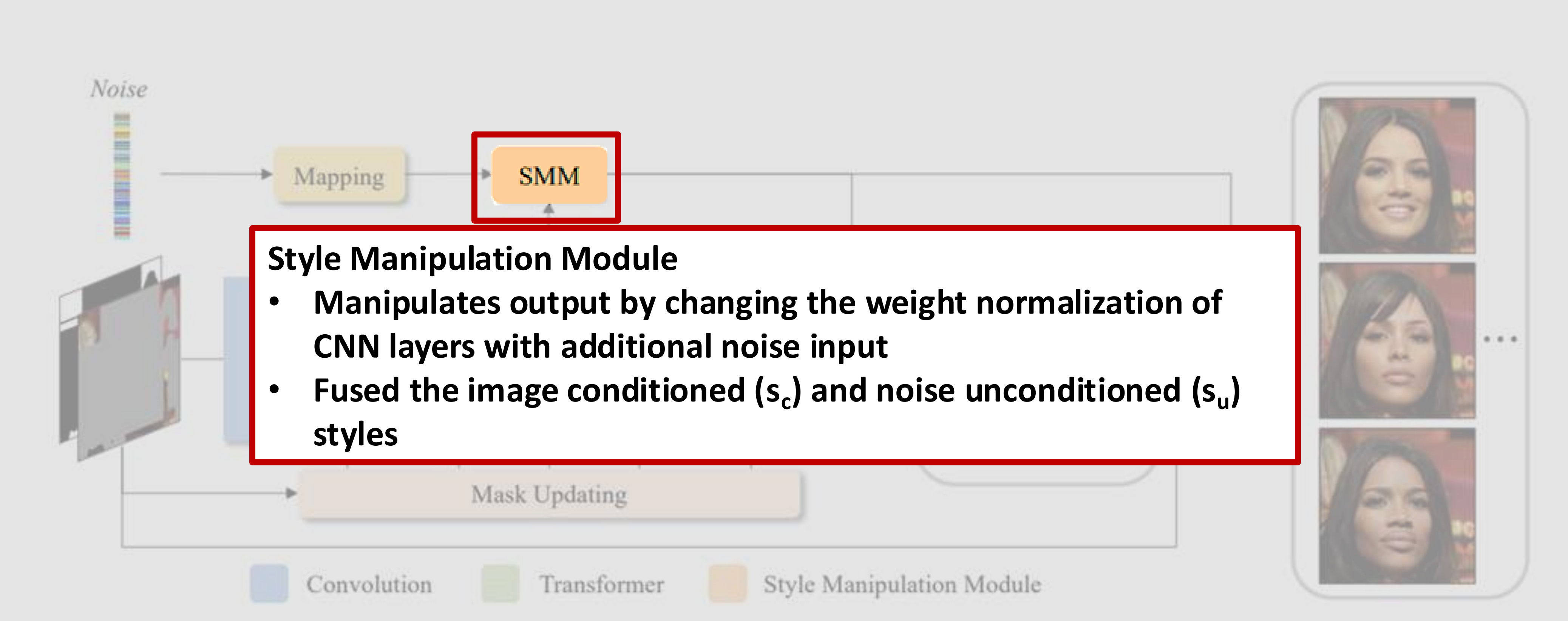
Model Architecture



MCA – Multi-Head Contextual Attention

- **Identifies valid tokens**
- **Does not process all valid tokens at once**
- **Overlapping masks to maintain global relationships, but computationally cheaper**

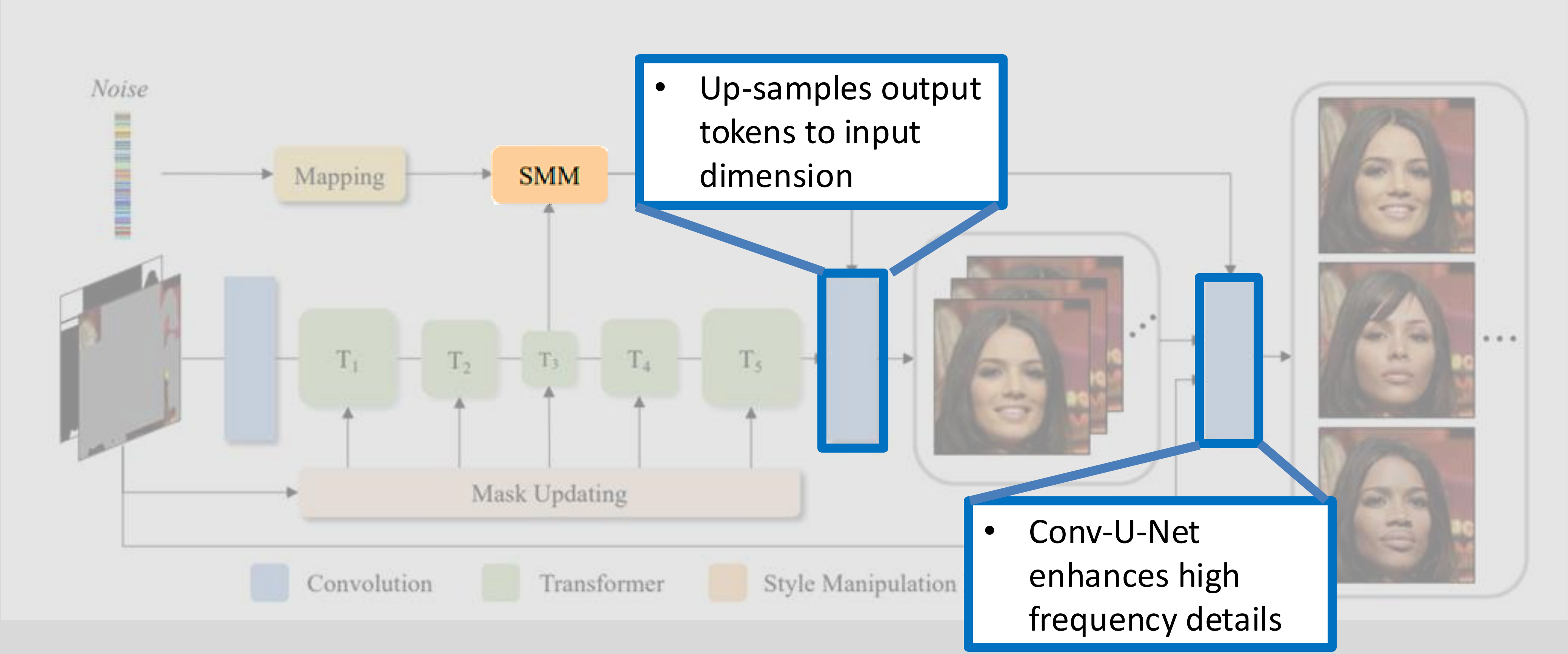
Model Architecture



Style Manipulation Module

- **Manipulates output by changing the weight normalization of CNN layers with additional noise input**
- **Fused the image conditioned (s_c) and noise unconditioned (s_u) styles**

Model Architecture

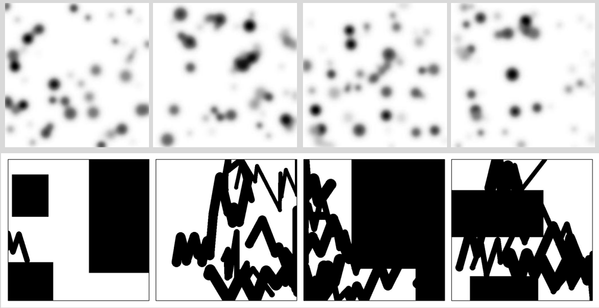


Preliminary Results – CelebHQ-256

Department of
AUTOMOTIVE ENGINEERING

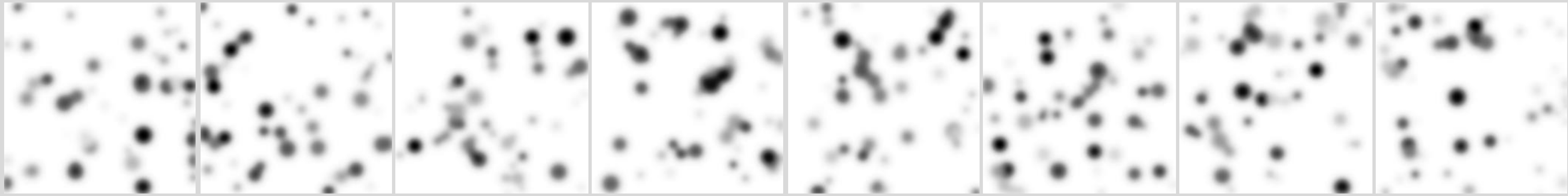
Training Data

Inference Data

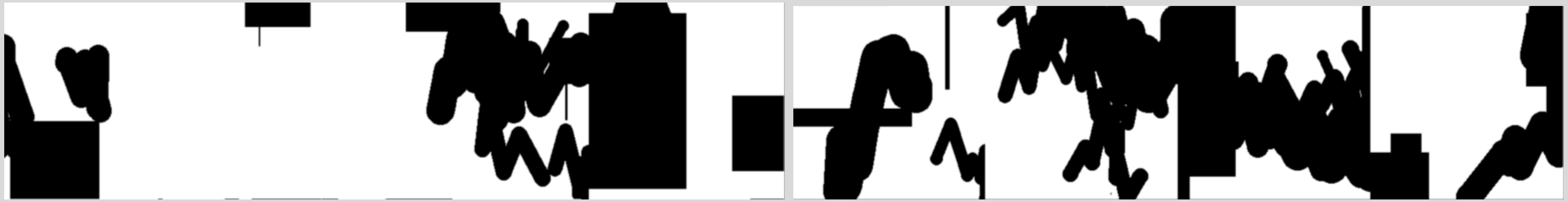


Custom elevation data: Random “trees” smoothed with Gaussian filter

Ground Truth



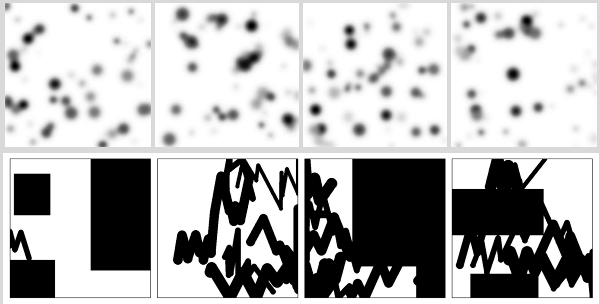
Masks (representative)



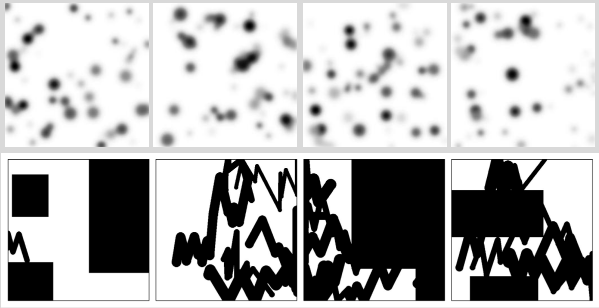
In-painted

Results – Model #1

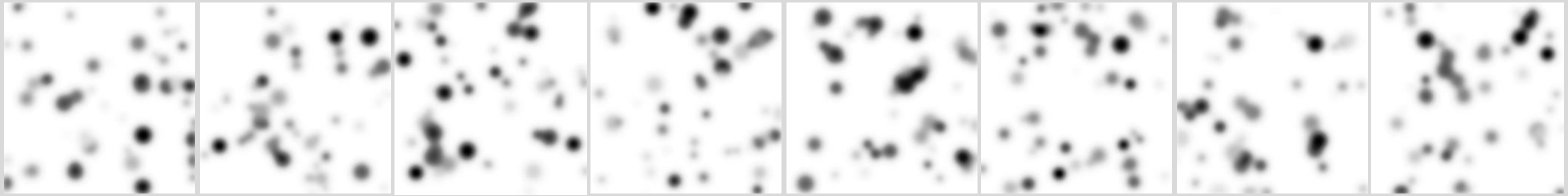
Training Data



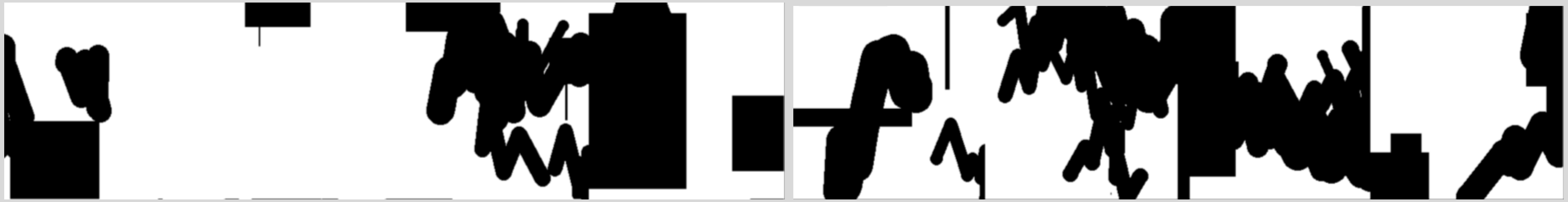
Inference Data



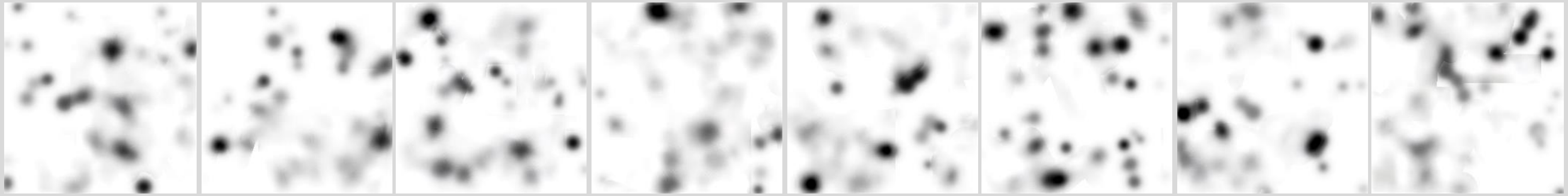
Ground Truth



Masks (representative)



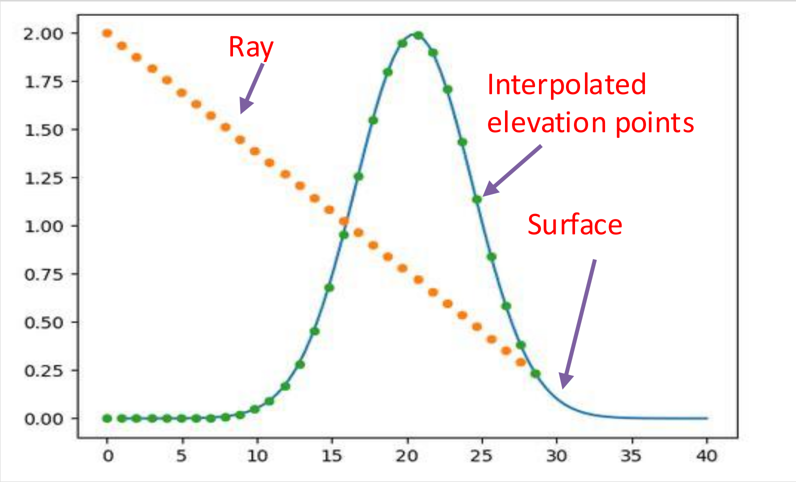
In-painted



Creating Custom Masks

- Original masks are randomly generated per image
- Random masks don't correlate well with "missing" data in elevation maps – typically due to occlusion
- The mask should be correlated with the image
- Use ray-casting to generate new masks

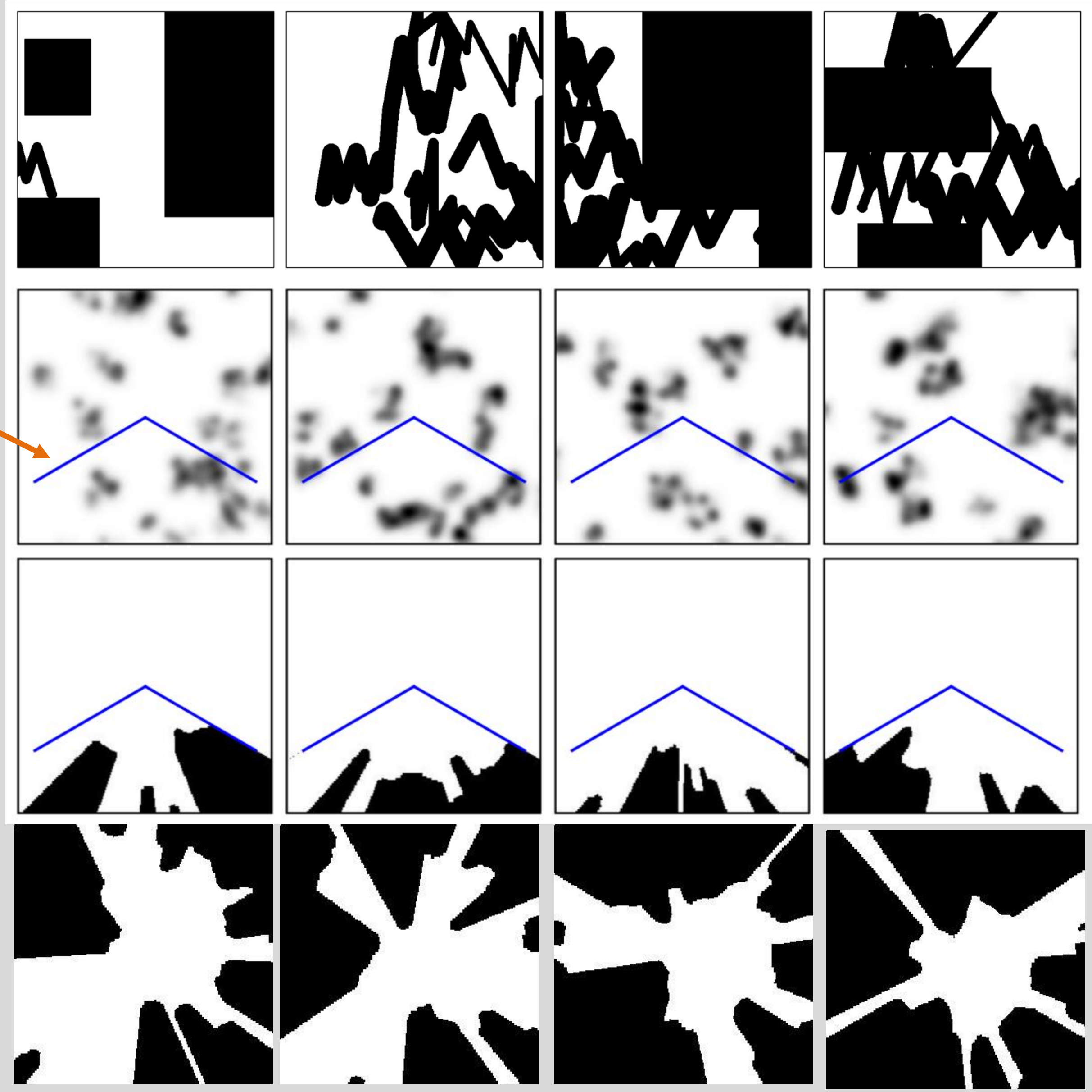
Ray Casting (2D example)



Vision cone

FWD Facing

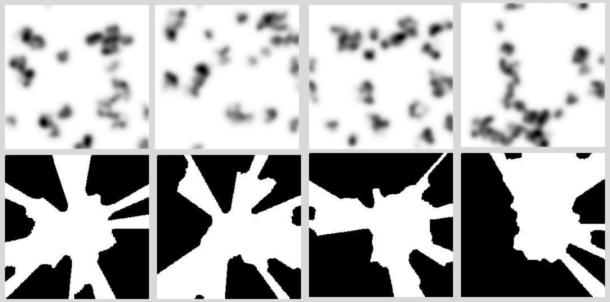
360°



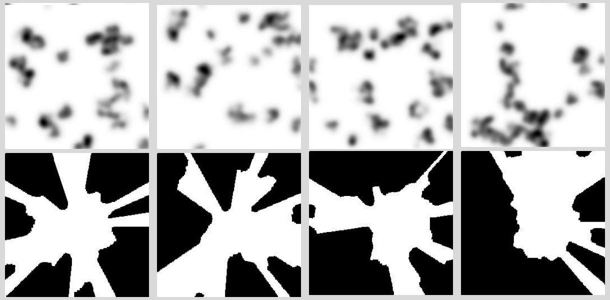
Results – Model #2.a

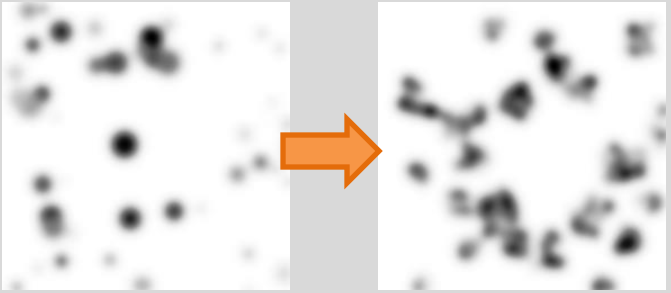
Department of AUTOMOTIVE ENGINEERING

Training Data



Inference Data

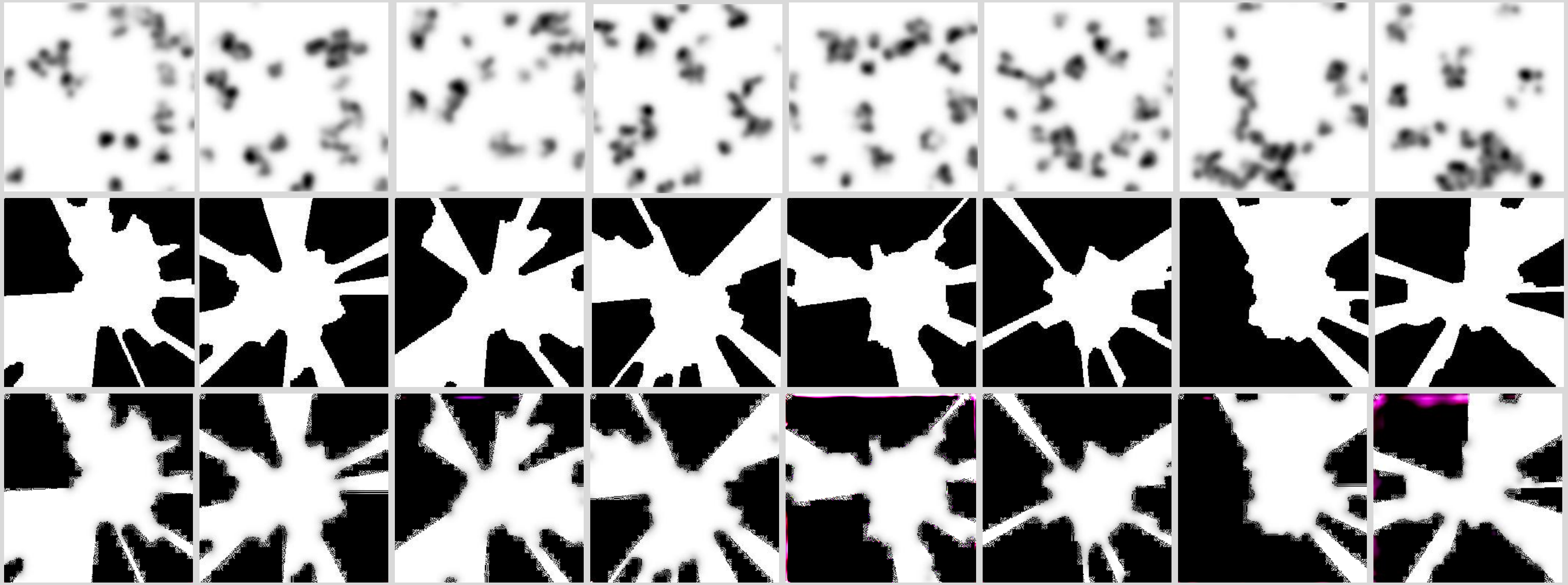




Ground Truth

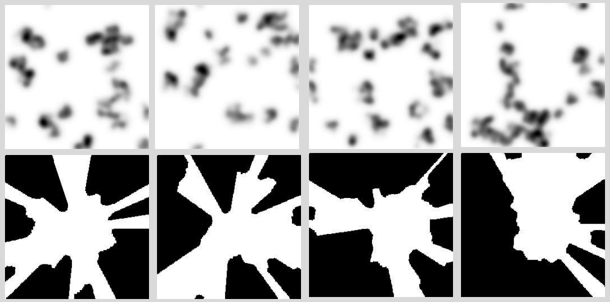
Masks (correlated)

In-painted

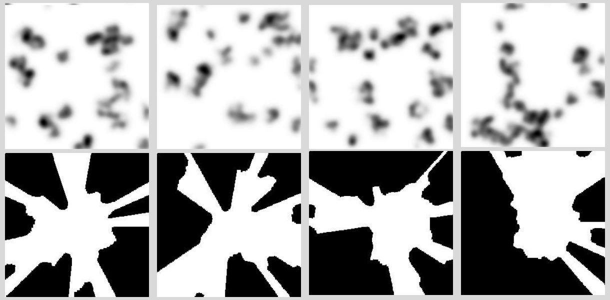


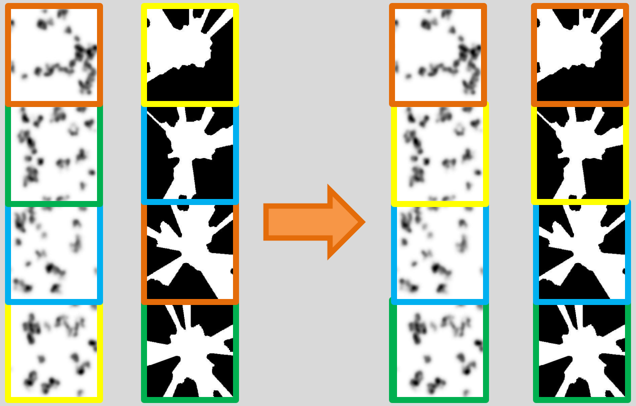
Results – Model #2.b

Training Data



Inference Data

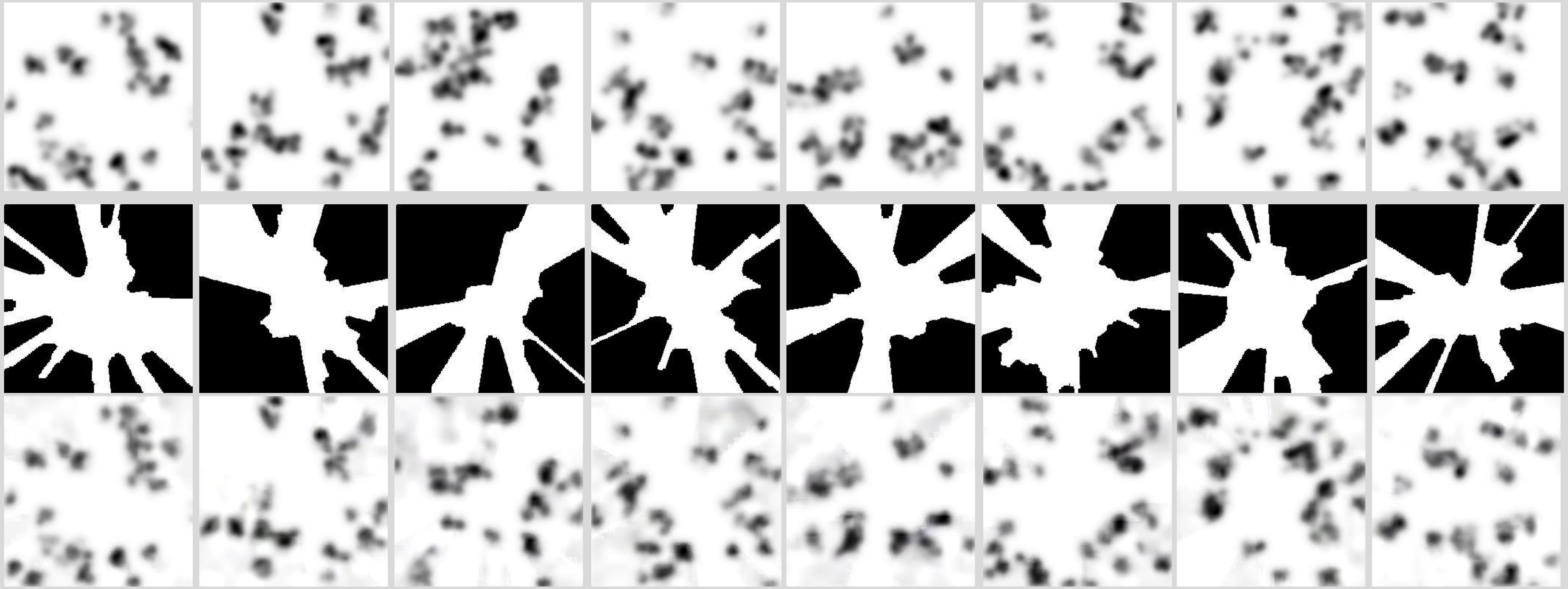




Ground Truth

Masks
(correlated)

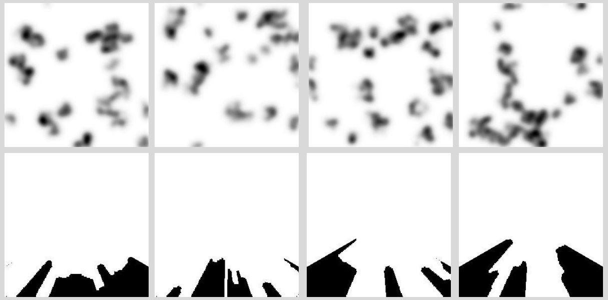
In-painted



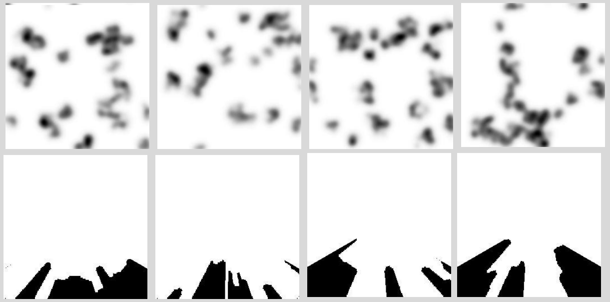
Department of
AUTOMOTIVE ENGINEERING

Results – Model 3

Training Data



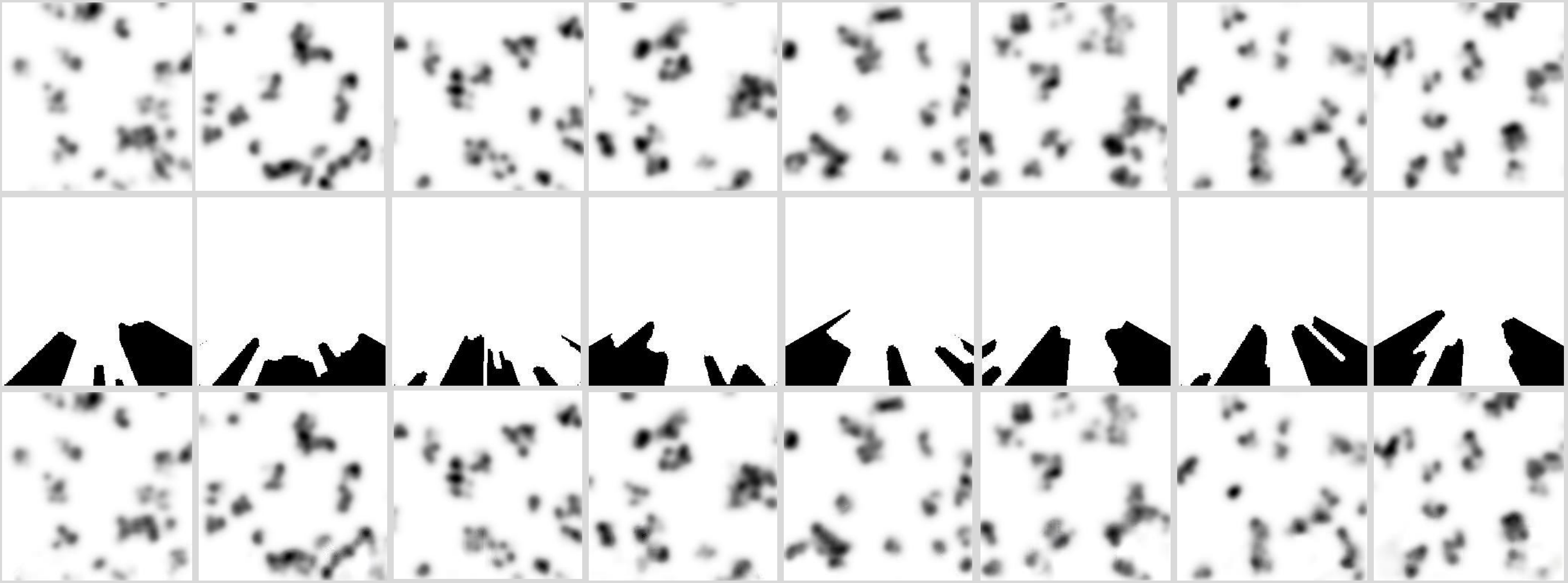
Inference Data



Ground Truth

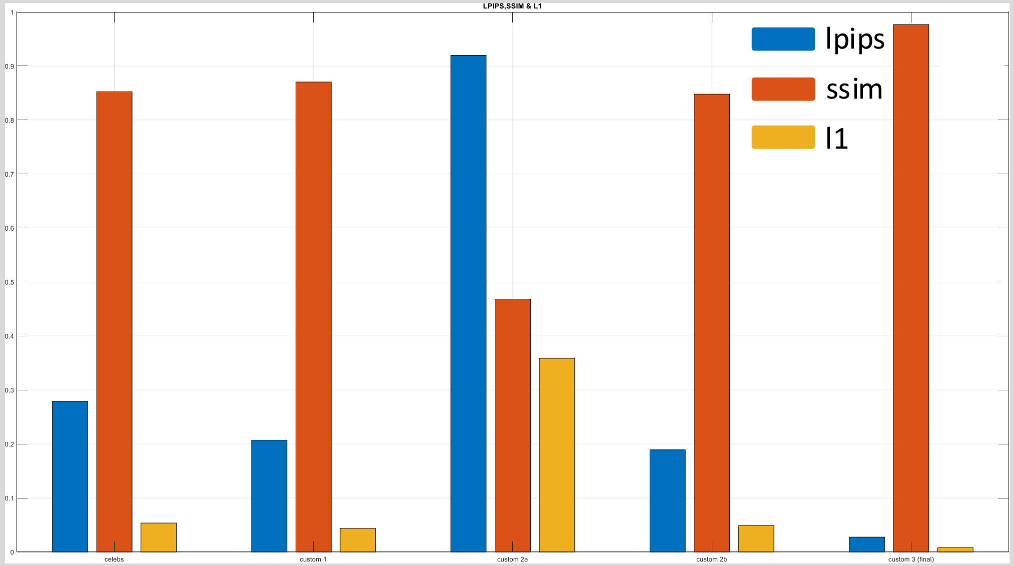
Masks
(correlated)

In-painted

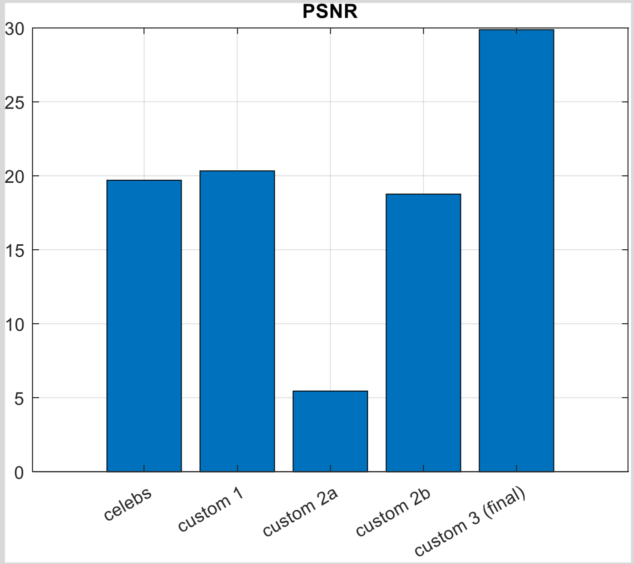


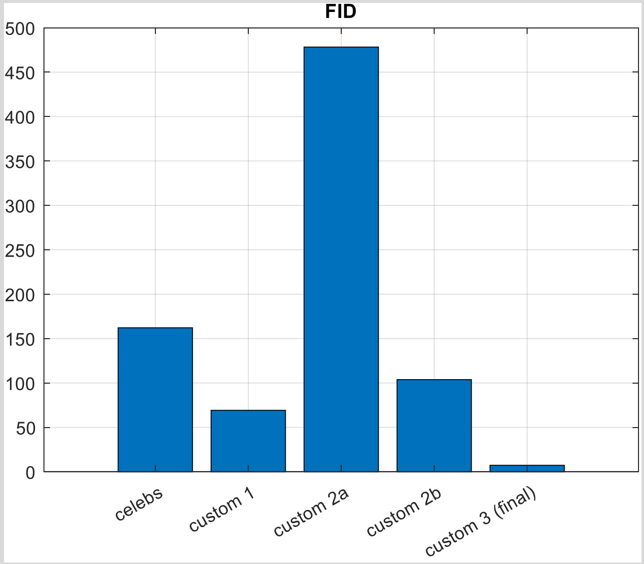
Statistics

Model	LPIPS	PSNR	SSIM	L1	FID
Celebs	0.2791	19.6982	0.8524	0.0536	162.12
Custom 1	0.2072	20.3282	0.8704	0.0436	69.20
Custom 2.a	0.9197	5.4545	0.4684	0.3587	478.22
Custom 2.b	0.1892	18.7609	0.8479	0.0486	103.78
Custom 3 (final)	**0.0276**	**29.868**	**0.9768**	**0.0078**	**7.31**



celebs custom 1 custom 2a custom 2b custom 3





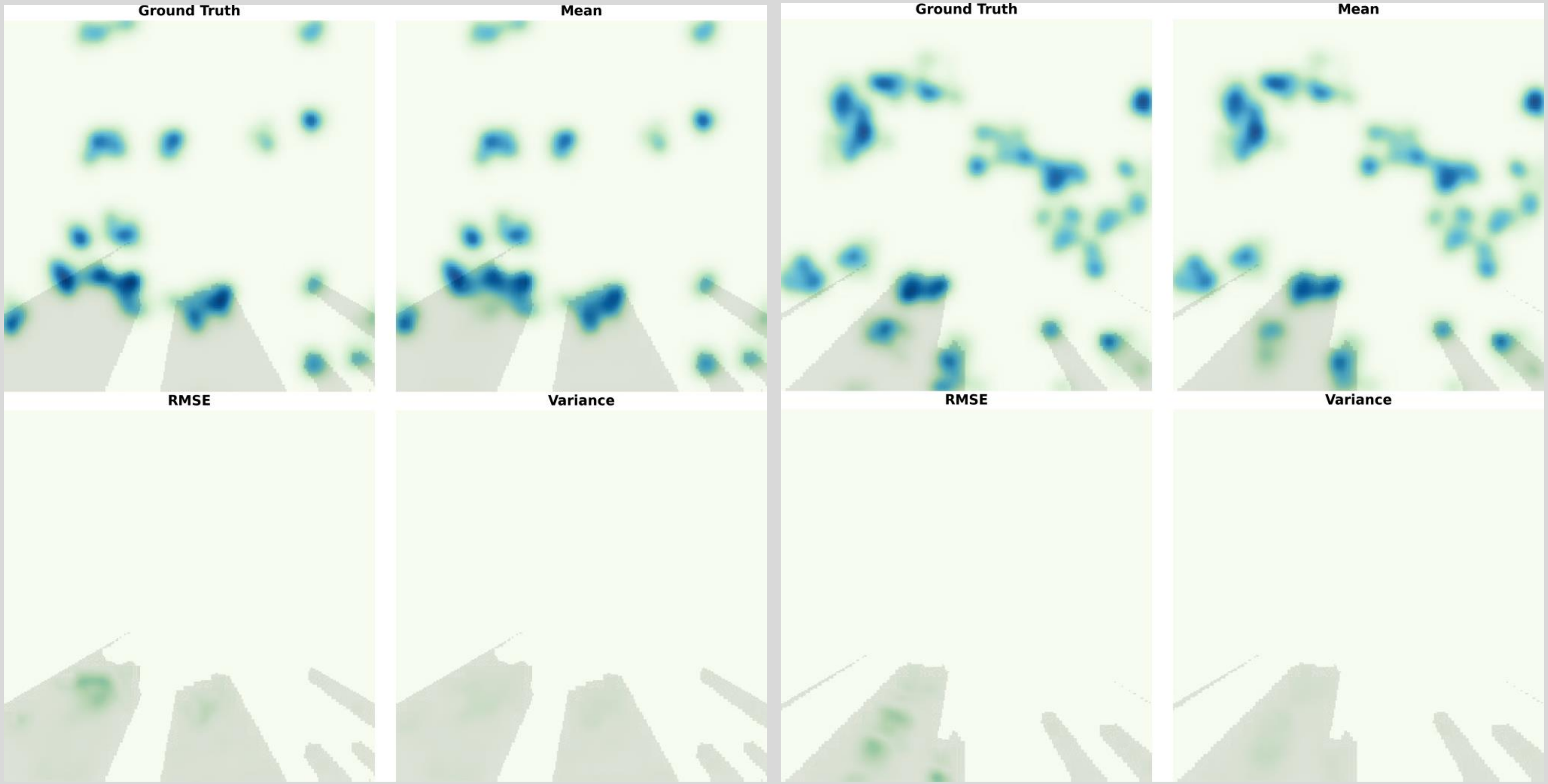
Inference Error and Uncertainty

Department of AUTOMOTIVE ENGINEERING

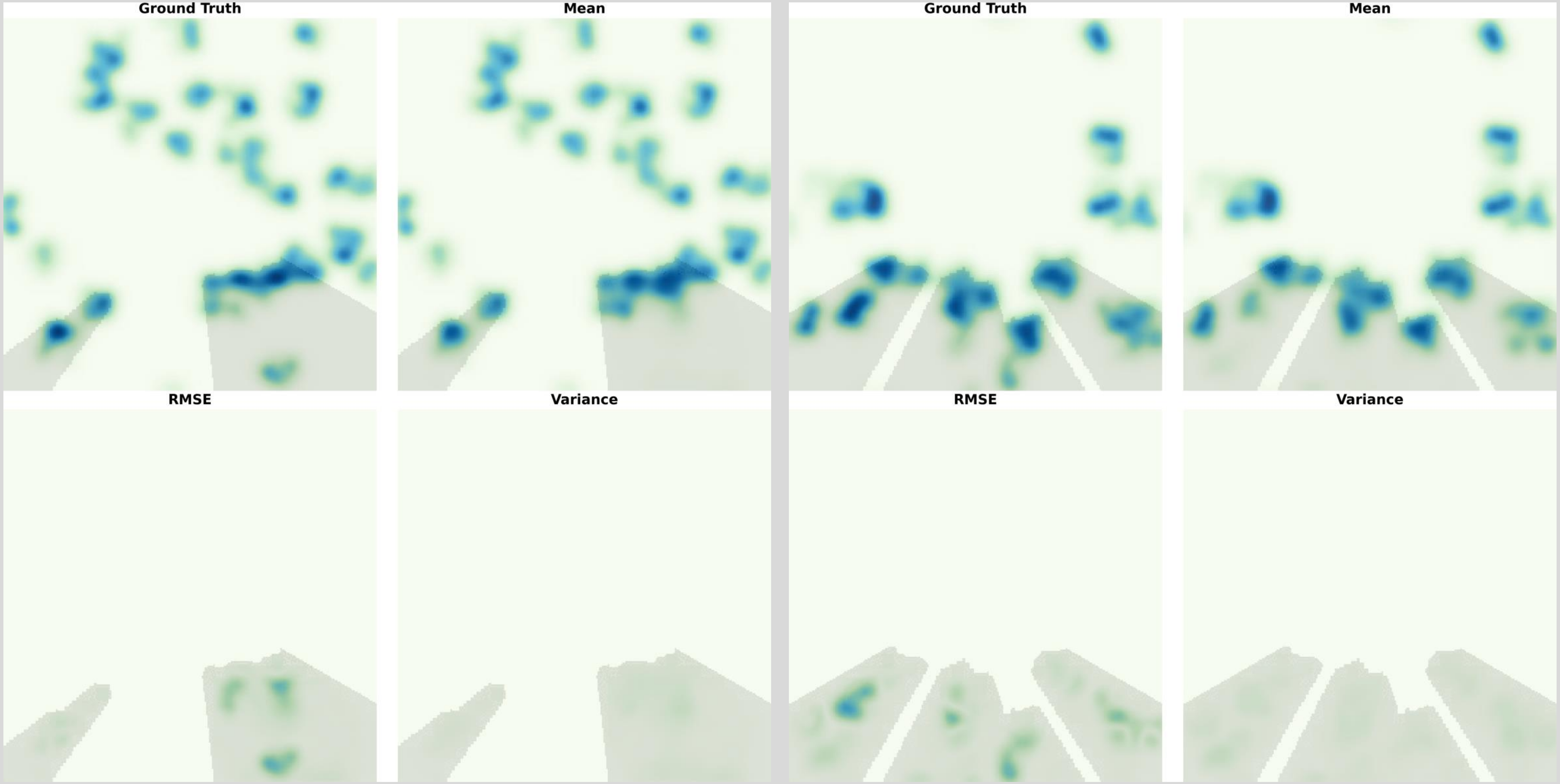
- Ensemble N=500 inferences
- Pixel wise error and uncertainty

$$Var = \sqrt{\sum_i \frac{\bar{x} - \widehat{x_i}}{N}}$$

$$RMSE = \sqrt{\sum_i \frac{x_i - \widehat{x_i}}{N}}$$



Inference Error and Uncertainty



Drawbacks and Next Steps

Department of AUTOMOTIVE ENGINEERING

- Variance or uncertainty is not necessarily a good measure of how good the reconstruction is. This could still lead to dangerous planning behavior.
- Train on real data
  - Original data was place holder data as access to ground truth data from test sight is not yet available
- Add a first-person perspective view image to condition the model in addition to the birds eye view and mask
  - Should help with contextual and semantic reasoning
- Validate on vehicle with planner in the loop