Self-Supervised Feature Learning for Semantic Segmentation of Overhead Imagery

facebook.

Suriya Singh [1*], Anil Batra [1*], Guan Pang [2], Lorenzo Torresani [2,3], Saikat Basu [2], Manohar Paluri [2], and C. V. Jawahar [1]

[1] IIIT Hyderabad, India   [2] Facebook   [3] Dartmouth College
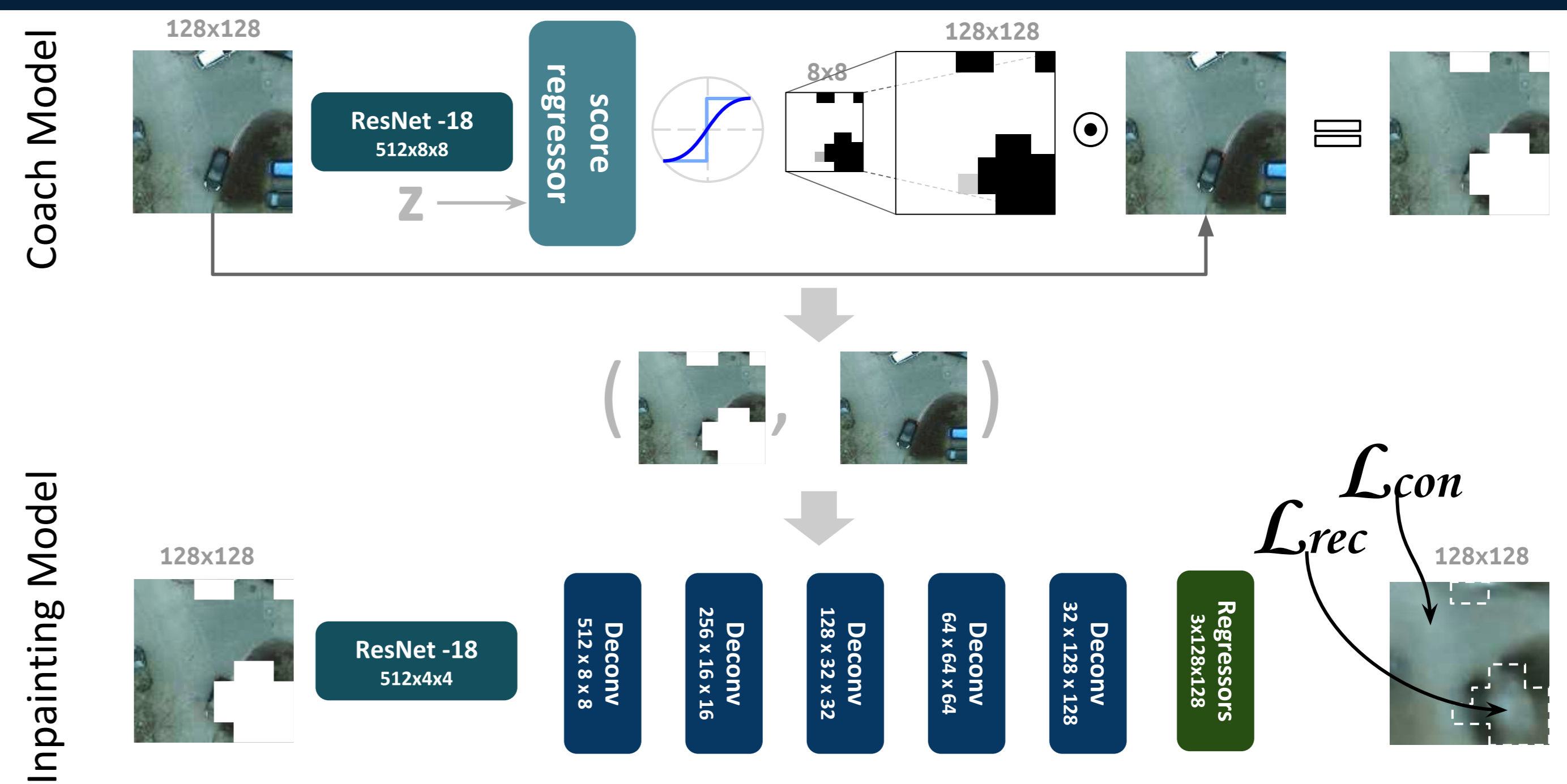
1. Overview

Overhead Imagery Tasks: Scene parsing, Road networks extraction, Land use estimation

- ❏ Abundance of unlabeled overhead imagery
- ❏ Large domain gap from ground imagery
- ❏ Overhead imageries lack specific subject

2. Self-Supervised Learning Approach

Predicted Mask

Adversarial Training

Semantic Inpainting
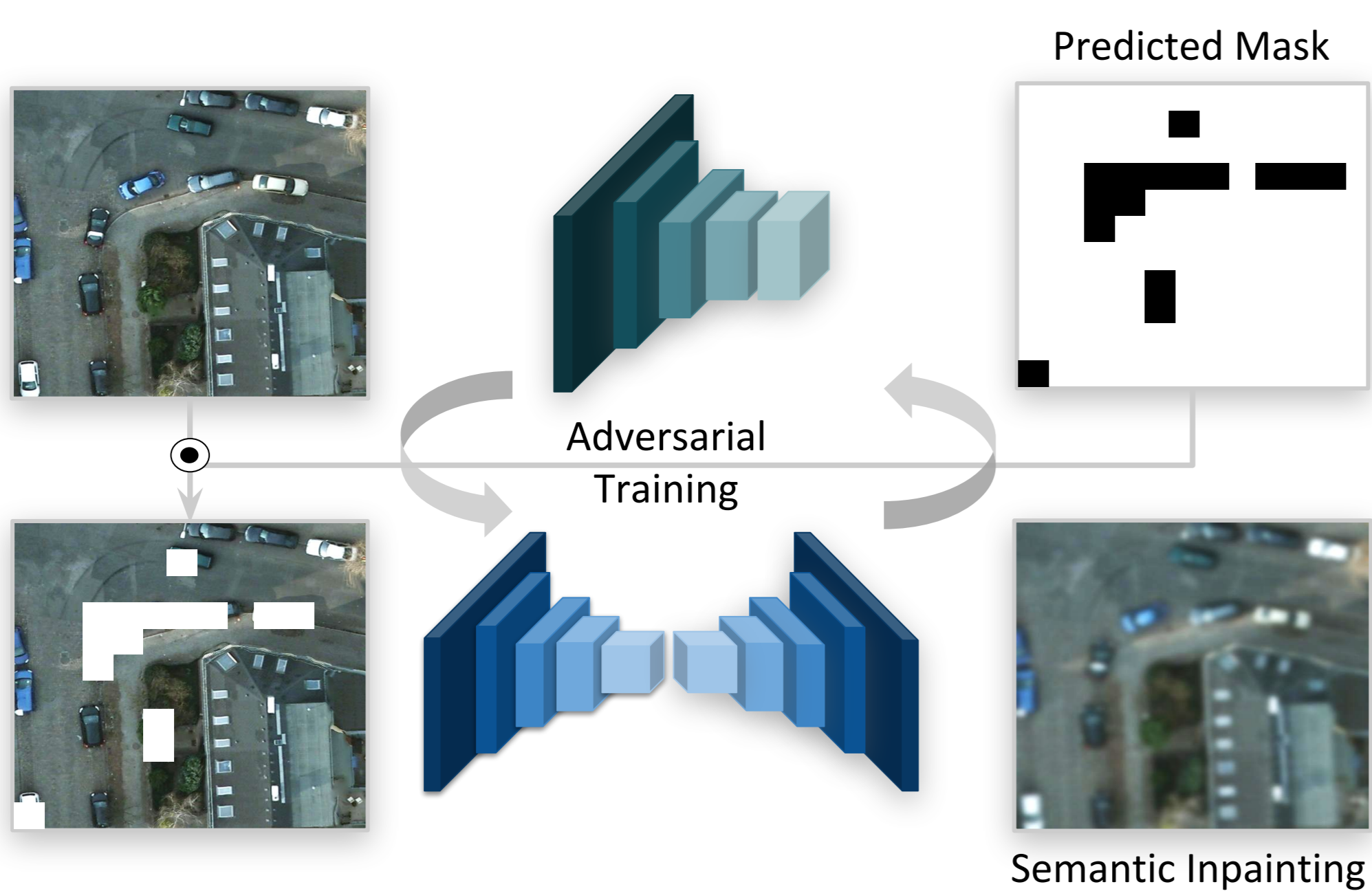
- ★ **Semantic image inpainting [2]**
  - ○ Reconstruct regions erased from the image
  - ○ Task close to semantic segmentation
  - ○ Encoder-decoder model with bottleneck
  - ○ Re-use pre-trained encoder for various tasks

Contributions

- ★ **Unified approach for overhead imagery tasks**
- ★ **Architectural Improvements for [2]**
  - ✔ Encoder-decoder network w/o bottleneck
  - ✔ High capacity ResNet-18 as backbone
  - ✔ Re-use pre-trained encoder and decoder
- ★ **Adversarial mask prediction**

Random Masks (Context Encoders [2])   Adversarial Masks (Coach Model)

3. Adversarial Mask Prediction
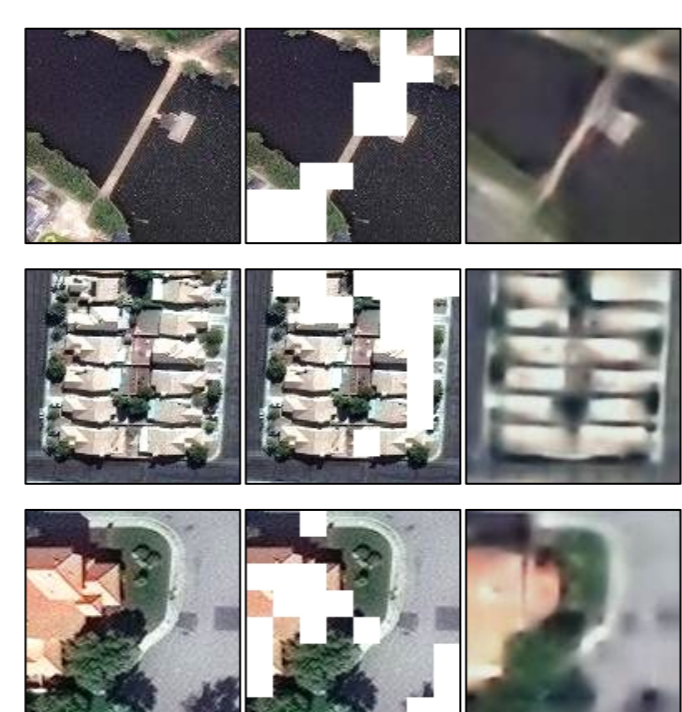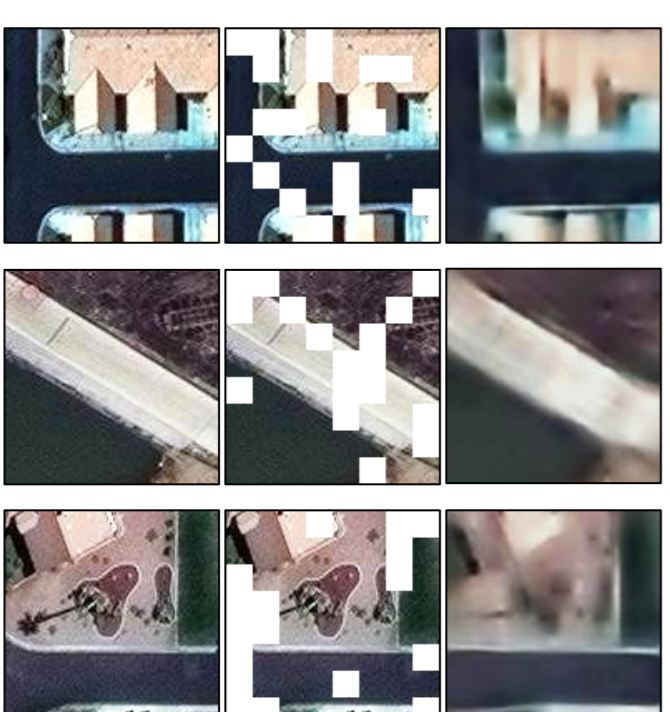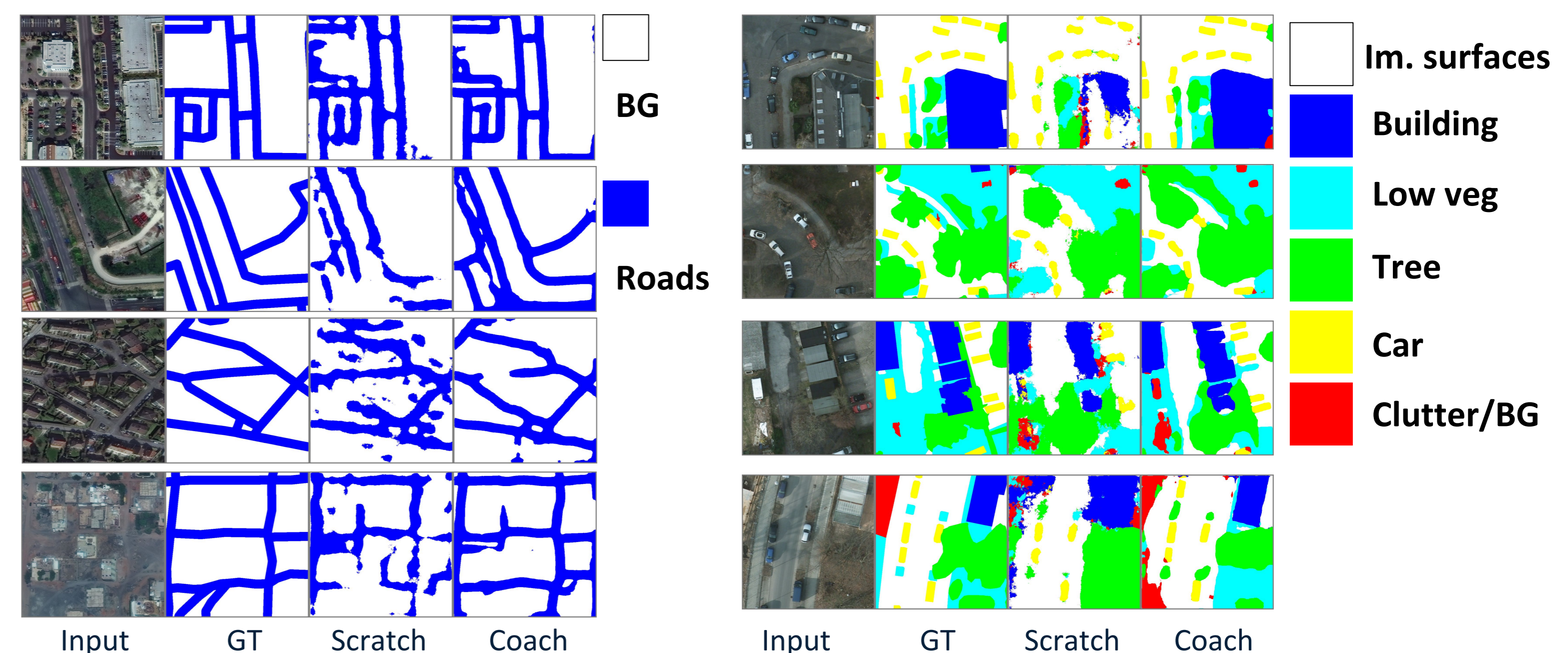

Coach Model

Inpainting Model

$$L_{inpainting} = \mathbf{w}_{rec}\mathcal{L}_{rec} + \mathbf{w}_{con}\mathcal{L}_{con}$$
$$L_{coach}(x) = 1 - \mathcal{L}_{rec}(x \odot M) = 1 - \mathcal{L}_{rec}(x \odot C(x))$$

- ❏ Predicts increasingly difficult mask for inpainting.
- ❏ This mask is used to erase regions of the image.
- ❏ Trained in adversary to the inpainting network.

4. Results

mIOU with 10% data for supervised training

Method	Arch Improv	Potsdam	SpaceNet	DG Roads	DG Lands
Context Pred[1]	✗	0.273	0.593	0.478	0.257
Context Enc[2]	✗	0.298	0.610	0.478	0.339
Splitbrain[3]	✗	0.265	0.641	0.482	0.411
ImageNet (ResNet)	✗	0.493	0.701	0.669	**0.575**
Scratch	✔	0.418	0.661	0.607	0.507
Autoencoder	✔	0.499	0.742	0.742	0.499
Context Enc (Ours)	✔	0.562	0.762	0.759	0.503
Coach Mask (Ours)	✔	**0.568**	**0.770**	**0.768**	0.529

Dataset	Method	(a) Labeled				(b) Unlabeled					
		10%	25%	50%	100%	1K	2K	5K	10K	500K	100K
Potsdam	CE	0.562	0.628	0.668	0.698	0.432	0.453	0.537	0.561	0.548	0.562
	Coach	**0.568**	**0.637**	**0.674**	**0.705**	0.446	0.469	0.541	0.563	0.566	0.565
SpaceNet	CE	0.762	0.781	0.795	0.804	0.696	0.731	0.754	0.759	0.763	0.765
	Coach	**0.770**	**0.786**	**0.797**	**0.806**	0.709	0.731	0.757	0.770	0.774	0.774

BG   Roads

Input   GT   Scratch   Coach

Im. surfaces, Building, Low veg, Tree, Car, Clutter/BG

Input   GT   Scratch   Coach

References

1. Doersch et al. Unsupervised visual representation learning by context prediction. ICCV15
2. Pathak et al. Context encoders: Feature learning by inpainting. CVPR16
3. Zhang et al. Split-brain AE: Unsupervised learning by cross-channel prediction. CVPR17

* Equal contributions

Codes are available at
https://github.com/suriyasingh/Self-supervision-for-segmenting-overhead-imagery

BMVC 2018
Newcastle upon Tyne