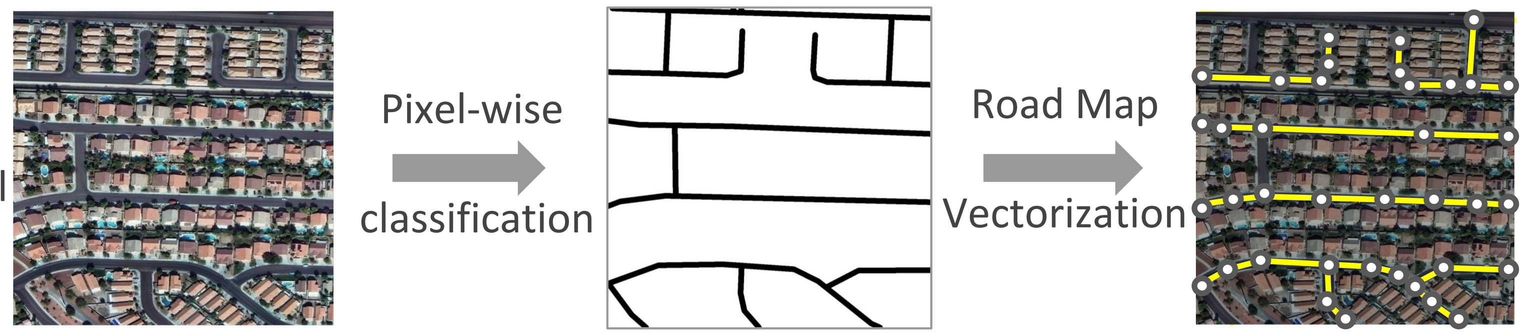
Improved Road Connectivity by Joint Learning of Orientation and Segmentation

Anil Batra† Suriya Singh† Guan Pang Saikat Basu C. V. Jawahar Manohar Paluri

CVIT · Mila · facebook · International Institute of Information Technology Hyderabad · Polytechnique Montréal

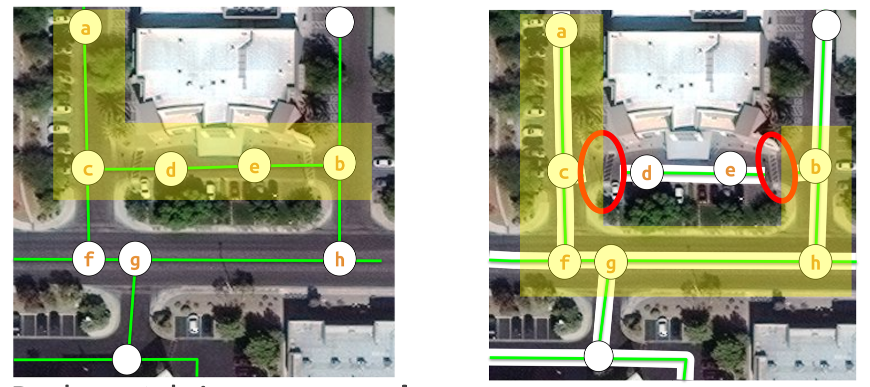
Overview

8 /16 bits RGB / hyperspectral images



Challenge

- Failures due to shadows, clutter, occlusion, and visual ambiguity.
- Difficult to enforce road topology supervision

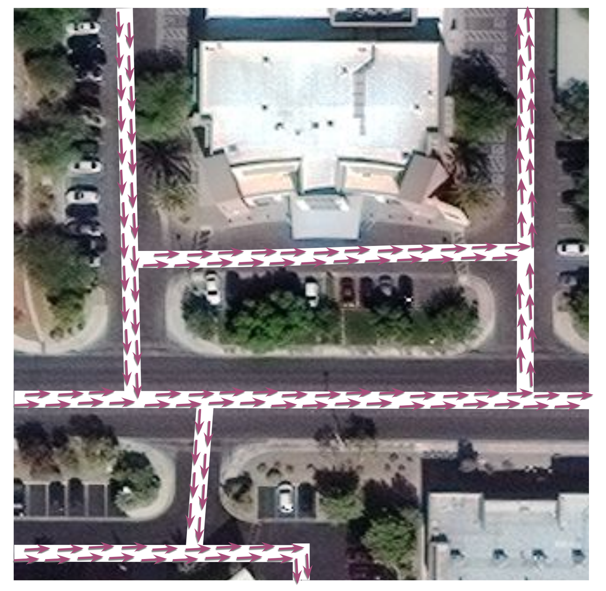


Path a ⇒ b in **connected** road maps

Path a ⇒ b in **broken** road maps

Task

- Predicts **topologically correct** and **connected** road mask.



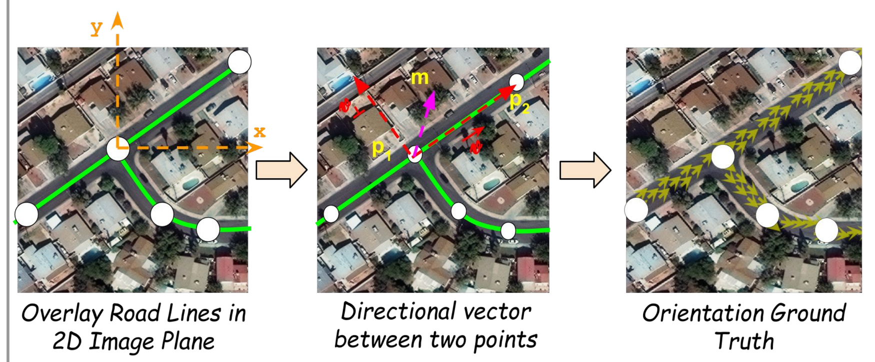
Joint learning of orientation and segmentation

Contributions

1. **Orientation Learning Task**
   - ✓ Auxiliary task to enforce connectivity
   - ✓ Free annotation from road linestrings
2. **Connectivity Refinement**
   - ✓ Modeled as inpainting task of semantic labels
   - ✓ Pre-trained with corrupted GT road masks
3. **Stacked Multi-branch Modules**
   - ✓ Fully convolutional module
   - ✓ Effectively utilize the mutual information between tasks

Road Orientation

- We want instance like feature which also ensure connectivity between segments
- Road segments lack the defined boundary and are interconnected
- Encode **Road Orientation** towards next pixel in same or adjacent road segment



$$\vec{v}_{ij}(x,y) = \frac{p_i(x,y)-p_j(x,y)}{||p_i(x,y)-p_j(x,y)||_2^2}$$

$$\vec{v}_{ij}(x,y) \equiv \langle 1 \quad \measuredangle o_r \rangle$$

$$o_{l_k}(m) = \begin{cases} o_r & \text{if } |\vec{v}_\perp \cdot \overrightarrow{(m-p_1)}| < \lambda_{orient} \\ o_b & \text{otherwise.} \end{cases}$$

Connectivity Supervision

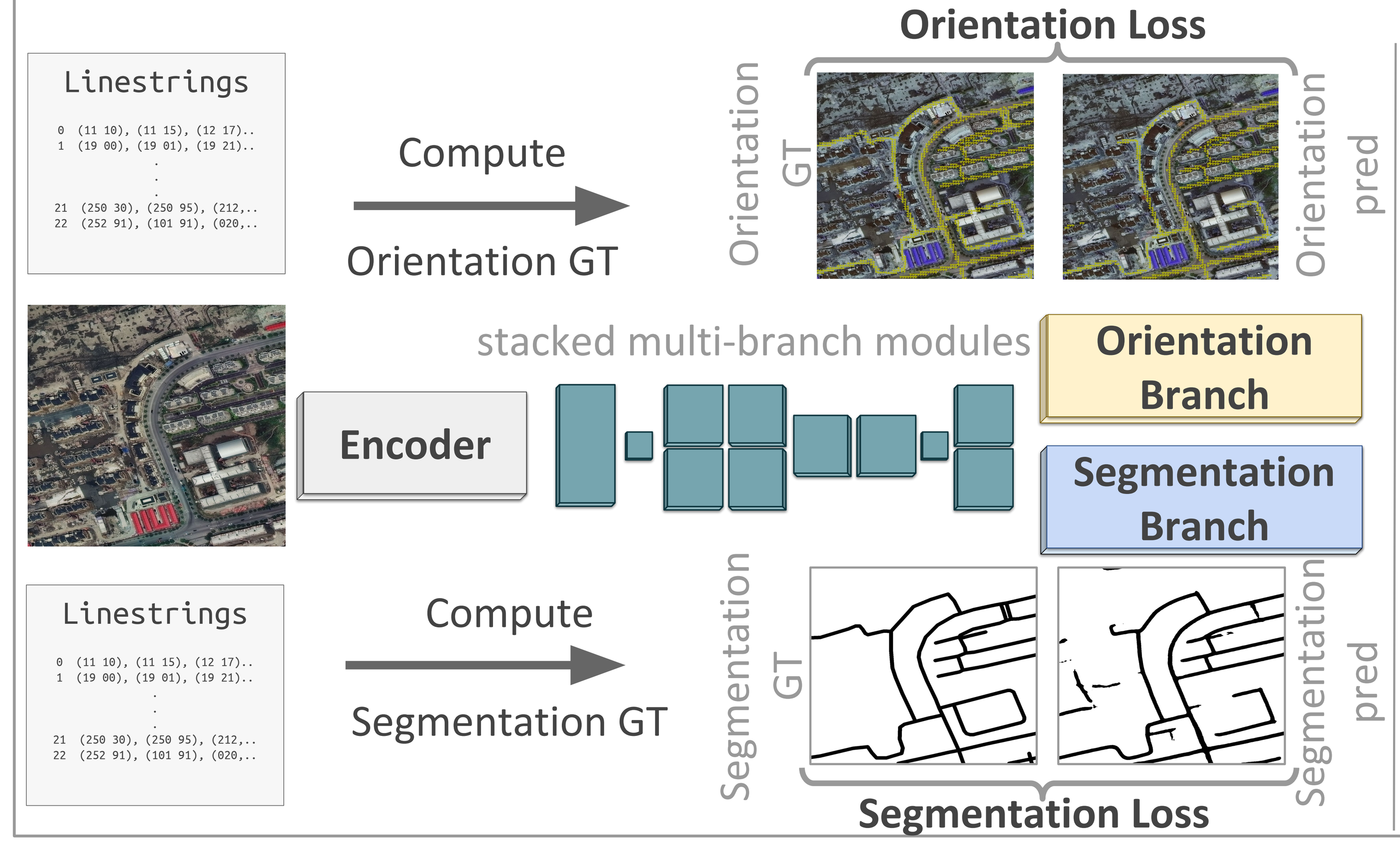
- Learning road orientations poses connectivity constraint and favours connected road segment.

† Equal contributions

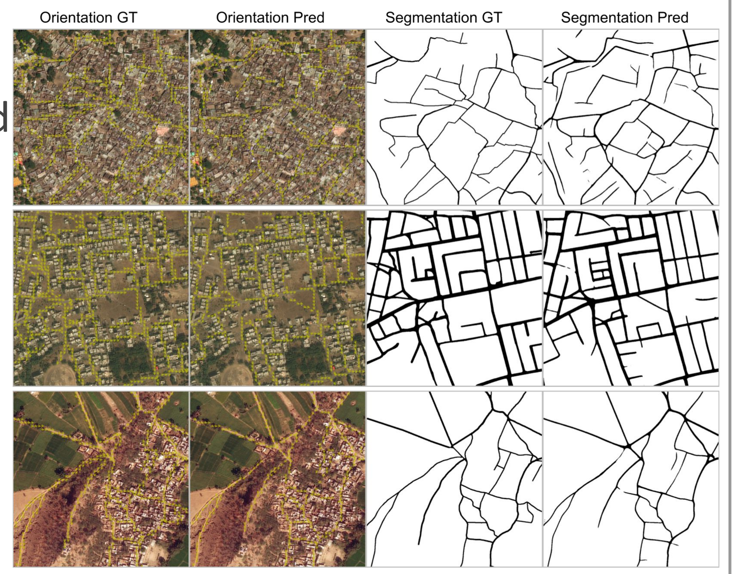
Codes : https://github.com/anilbatra2185/road_connectivity

We thank NSERC for partially supporting the work

Joint Learning of Orientation and Segmentation



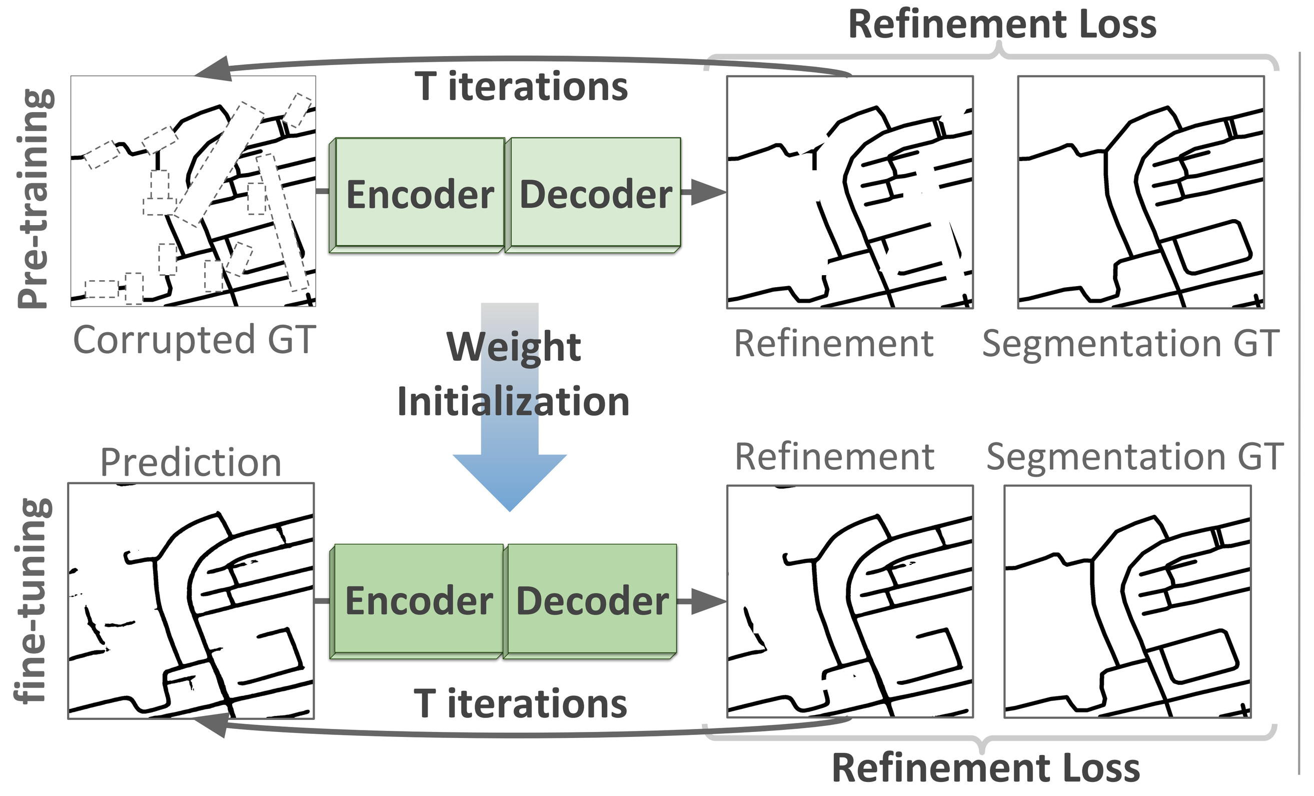
- Orientation and segmentation GT computed from road linestrings
- Predict orientation and segmentation through task specific decoders
- Consistent improvement across base architectures



Multitask Learning

$$Loss = \sum_s \Big(\underbrace{-\,\mathrm{SoftIoU}(f_{seg}(X), y)}_{\text{Segmentation Loss}} \underbrace{-\sum_{c=0}^{o_l} o_c \log\left(f_{orient}(X)\right)}_{\text{Orientation Loss}} \Big)$$

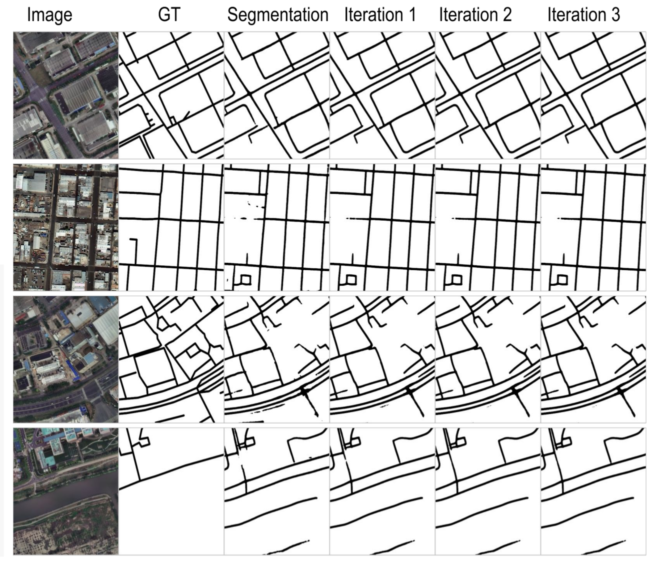
Connectivity Refinement



- Interpret segmentation failure as corruption and pre-train refinement network to restore the corrupted GT masks
- fine-tuned with output of base network to improve segmentation

$$\tilde{y}_t = g\Big([X, y', \tilde{y}_{t-1}]\Big) \quad t = 1, \ldots, T$$

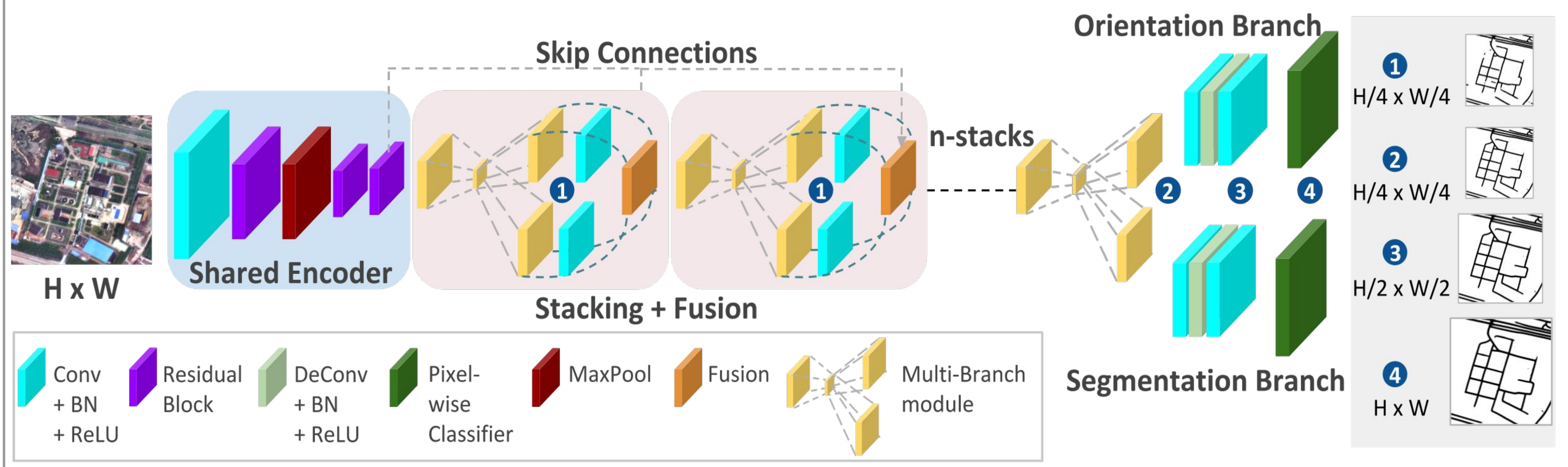
$$\hat{y}_t = g\Big([X, \hat{y}, \hat{y}_{t-1}]\Big) \quad t = 1, \ldots, T$$

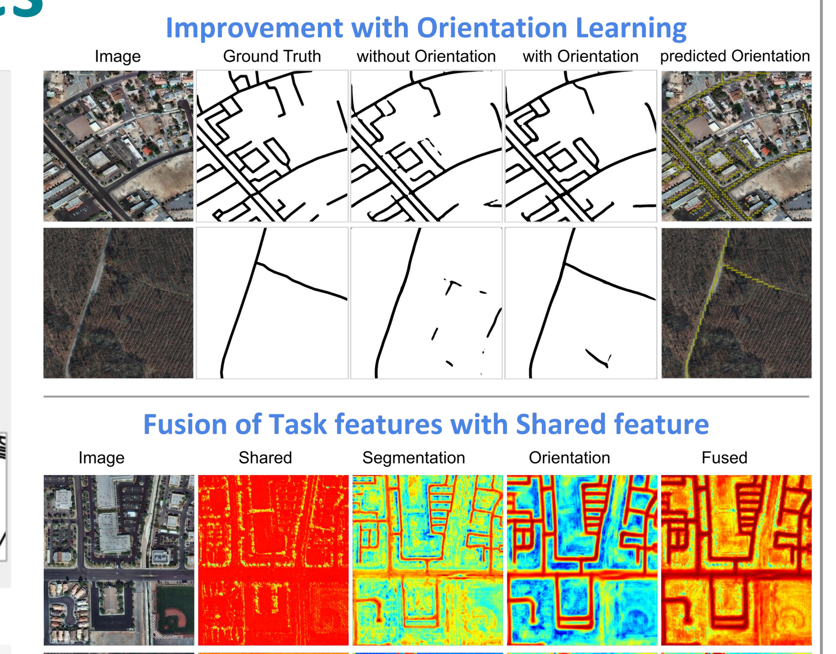


Mask corruption

- Erase real and add fake *road-like structures* in GT to simulate diverse segmentation failures.

Stacked Multi-branch Modules





Connectivity induced features

- Mutual information from both task is induced into shared feature representation via series of *stacked multi-branch modules*.

Results

APLS metric [5]

- Captures the deviations in shortest path distances between all pair of nodes in a graph.

$$S_{P\rightarrow T} = 1 - \frac{1}{|V|}\sum min\left(1, \frac{|L(a,b)-L(\hat{a},\hat{b})|}{L(a,b)}\right) \qquad APLS = \frac{1}{N}\sum_{(y,\hat{y})}\left(\frac{1}{\frac{1}{S_{P\rightarrow T}(G,\hat{G})}+\frac{1}{S_{T\rightarrow P}(\hat{G},G)}}\right)$$

$S_{P\rightarrow T}$ measures the sum of difference of shortest path for each node pair in ground truth graph.

Road Extraction Comparison

Dataset	Metric	DRM [1]	TL [2]	L34 [4]	L34 [4] + Orientation	MatAN [3]	Ours
Spacenet	*road IoU*	N/A	56.29	60.33	62.45	52.86	**63.75**
	APLS	50.59	49	55.69	60.76	46.44	**63.65**
DeepGlobe	*road IoU*	N/A	64.95	62.75	64.71	46.88	**67.21**
	APLS	61.66	56.91	65.33	68.71	47.15	**73.21**

Ablation Study

a) Consistent improvement with joint Orientation Learning

Method	Spacenet		DeepGlobe	
	road IoU	*APLS*	*road IoU*	*APLS*
Resnet -18	59.04	52.65	62.12	63.31
R18 + Orientation	61.90	59.06	**64.77**	**68.93**
R18 + Junctions*	58.41	52.76	63.54	66.20
Linknet - 34	60.33	55.69	62.75	65.33
L34 + Orientation	**62.45**	**60.77**	64.72	68.71
L34 + Junctions*	60.72	55.91	63.79	67.42

b) Architecture studies with Orientation Learning

Method	Spacenet		DeepGlobe	
	road IoU	*APLS*	*road IoU*	*APLS*
Resnet -18	61.90	59.06	64.77	68.93
Linknet 34	62.45	60.77	64.72	68.71
UNet	60.12	58.59	65.21	67.81
1 - Stack#	63.26	60.92	65.60	70.23
2 - Stack#	**63.75**	**63.65**	**67.21**	**73.21**
3 - Stack#	63.73	62.89	66.61	72.48

c) Improvement with Connectivity Refinement

Method	Spacenet		DeepGlobe	
	road IoU	*APLS*	*road IoU*	*APLS*
Base (Linknet 34)	60.33	55.69	62.75	65.33
+ Orientation	62.45	60.77	64.72	68.71
+ Refinement	**62.76**	**62.03**	**65.03**	**69.77**

d) Progressive improvement with each contribution

Method	Spacenet		DeepGlobe	
	road IoU	*APLS*	*road IoU*	*APLS*
Base (2-Stack)	61.51	58.70	64.23	67.98
+ Multi-Scale	61.80	58.49	64.44	67.92
+ Orientation	63.44	61.78	66.81	72.03
+ Feature Fusion	63.75	63.65	**67.21**	**73.21**
+ Refinement	**63.76**	**63.79**	67.02	73.20

* **Junctions**: Predict intersection points of road segments
\# **n-Stack:** Stacked Multi-branch model with n-stacks

References:

1. Máttyus *et. al.*, Deep-roadmapper: Extracting road topology from aerial images, ICCV 17
2. Mosinka *et. al.* Beyond the pixel-wise loss for topology aware delineation. CVPR 18
3. Máttyus *et. al.* Matching adversarial networks. CVPR 18
4. Chaurasia *et. al.* Linknet: Exploiting encoder representations for efficient semantic segmentation. VCIP 17
5. Spacenet: A remote sensing dataset and challenge series