TopoAL: An Adversarial Learning Approach for Topology-Aware Road Segmentation

Supplementary material

Subeesh Vasu, Mateusz Kozinski, Leonardo Citraro, and Pascal Fua

CVLab, EPFL, Lausanne, Switzerland
firstname.lastname@epfl.ch

This supplementary material presents performance comparisons on the DeepGlobe Dataset [2] as well as ablation studies on the key ideas employed in our proposed method *TopoAL*.

Evaluation on DeepGlobe Dataset [2]

We perform additional experiments on the DeepGlobe Dataset [2] to demonstrate that the performance improvement achieved by our approach is generalisable to other datasets as well. In our experiments on DeepGlobe, we follow the dataset splits used by [1] to generate the train and test data. In DeepGlobe, the labels are pixel-based, wherein all the pixels belonging to the road are marked as the foreground. This results in ground truth road masks of varying width, which resulted in an imbalance between the pixel-wise loss and adversarial feedback under our settings. Therefore, we have first generated a new set of labels by dilating the skeletons to road masks that are 7 pixels wide and use them to establish the pixel-wise loss. With this new set of labels, the best performing pixel-wise loss was found to be a weighted combination of BCE loss and $\mathrm{SoftIoU}$ loss [1]. We, therefore, used the loss function of the following form to establish the pixel-wise loss for *UNet-TopoAL*as well as the baseline *UNet*.

$$\mathcal{L}_{pixel}(\hat{\mathbf{y}}, \mathbf{y}) = 0.2 * \mathcal{L}_{bce}(\hat{\mathbf{y}}, \mathbf{y}) - \mathrm{SoftIoU}(\hat{\mathbf{y}}, \mathbf{y}) \tag{1}$$

We use the following approaches as baselines.

- *UNet* [4]: Fully-convolutional network with skip connections.
- *MultiBranch* [1]: Recursive architecture jointly trained for road segmentation and orientation estimation.
- *D-LinkNet* [1]: An encoder-decoder architecture [5] jointly trained for road segmentation and orientation estimation [1].

To obtain the results of *MultiBranch*and *D-LinkNet*, we used the code provided by [1] with their default parameter settings and loss functions.

In Table 1, we compare *UNet*, *MultiBranch* and *D-LinkNet* against our approach *UNet-TopoAL* using the same metrics as the one used in the main paper. We can observe that the inclusion of *TopoAL* component onto the baseline network *UNet* enhances the performance on all the metrics, clearly demonstrating the potential generalisability of our approach to different datasets. As compared to the other baseline methods *MultiBranch*and *D-LinkNet*, *UNet-TopoAL* is a lightweight architecture and can, therefore, be not expected to outperform these methods significantly. Nevertheless, *UNet-TopoAL* is

		Pixel-based			Topology-aware			
		CCQ			TLTS	APLS	JUNCT	*H&M*
	Method	corr.	comp.	qual.	corr.		f1	f1
DeepGlobe	*MultiBranch* [1]	0.811	0.807	0.679	0.694	0.708	0.807	0.848
	D-LinkNet [1]	0.785	0.790	0.649	0.642	0.670	0759	0.824
	UNet [4]	0.822	0.745	0.642	0.638	0.643	0.773	0.827
	UNet-TopoAL (Ours)	0.825	0.764	0.658	0.671	0.666	0.792	0.838

Table 1: Quantitative comparison between baselines segmentation networks, and our approach *UNet-TopoAL*. Our *TopoAL* approach improves up on the baseline network *UNet* on all the metrics and yield competitive results as opposed to other methods that makes used of deeper networks.

able to deliver comparable performance to *D-LinkNet* [1], and one can hope to improve the performance further by making use of deeper network architectures.

Ablation Studies

To reveal the importance of each component used in our proposed approach *TopoAL*, we have conducted experiments with different discriminator training strategies. The methods that we compare here are alternative ways to train while moving from the idea of *VanillaGAN* to *TopoAL*. All the experiments are conducted with *UNet* as the generator. In Table 2, we report the values of *APLS* and *H&M*, for all the training schemes that we consider in our ablation study.

Method Name	STE	Multiplication by the dilated ground truth	Dynamic label assignment	Scale space labels	*APLS*	*H&M*
UNet-VanillaGAN				$k = 0$	0.607	0.748
UNet-PatchGAN				$k = 3$	0.616	0.726
Ablation 1				✓	0.626	0.720
Ablation 2	✓				0.588	0.722
Ablation 3	✓	✓	✓		0.63	0.750
Ablation 4			✓	✓	0.629	0.734
Ablation 5	✓		✓	✓	0.649	0.742
Ablation 6	✓	✓		✓	0.622	0.720
UNet-TopoAL	✓	✓	✓	✓	0.666	0.767

Table 2: Ablation studies on alternative ways to train the discriminator.

We use the following approaches in the ablation study reported in Table 2.

– *UNet-VanillaGAN*: We replace our sophisticated discriminator by a simple one that returns a binary flag for each input mask. We use the generator output (without any post-processing) as the prediction based input to the discriminator and is assigned with label 0. This method follows the same definition as the *UNet-VanillaGAN* reported in the main paper. Upon the visual inspection, we found this approach

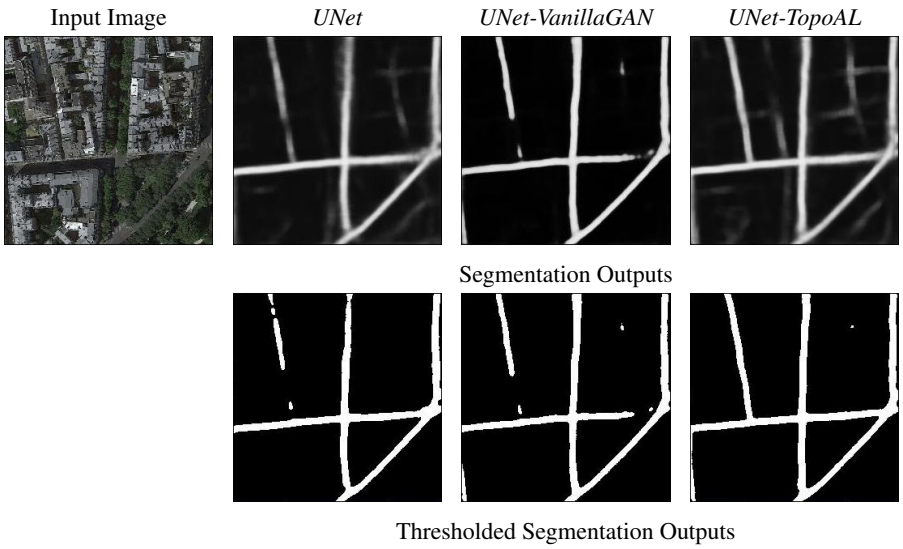

Fig. 1: **Visualisation of Segmentation Outputs.** In comparison to *UNet*, *UNet-VanillaGAN* generates segmentation outputs which has less variation in probability values, whereas *UNet-TopoAL* tends to focus on the connectivity of the resulting road network.

to deliver predicted road masks that are close-to-binary (as depicted in Figure. 1). This is not a surprise since the key factor that distinguishes the generator output from ground truth is the non-binary nature of predicted masks. Henceforth, the discriminator learns features related to the variations in the probability values of the generator output and motivates the generator to reduce the same. As we reported in the main paper, this approach does not result in any noticeable improvement over the baseline approach *UNet*.

- *UNet-PatchGAN*: We replace our sophisticated discriminator with that of the one used in *PatchGAN* [3]. This is done by removing the outputs corresponding to the levels $k \in \{0, 1, 2\}$ from the discriminator architecture of *TopoAL*. As in *UNet-VanillaGAN*, the *UNet-PatchGAN* makes use of predefined labels of 0 for output from the segmentation network. As is evident from Table 2 the performance of *UNet-PatchGAN* is only comparable to the baseline approach *UNet*.
- Ablation 1: Here, we use the same discriminator as that of *TopoAL*, which therefore uses a pyramid of labels to supervise discriminator. However, the labels used are predefined as in *UNet-VanillaGAN*, resulting in no significant performance improvement over the baseline approach *UNet*.
- Ablation 2: The generator output is passed through STE to form the prediction based input to the discriminator. All other settings are the same as *UNet-VanillaGAN*. Since *UNet-VanillaGAN* enforces binarization effect on the generator output, it will be interesting to see if we can remove this dilemma by using an STE, which will convert the non-binary generator output to the corresponding binary version. How-

ever, as reported in Table 2, the use of STE alone does not resolve the problem either.

- Ablation 3: Instead of the scale-space labels in *UNet-TopoAL*, here we use a single label value obtained via dynamic label assignment. As is evident from Table 2, the dynamic label assignment improves the scores, but the overall performance is lower than *UNet-TopoAL*.
- Ablation 4: The generator output is directly used as the prediction based input to the discriminator. Other settings are the same as *UNet-TopoAL*. The performance improvement under this setting indicates the effect of the proposed labeling scheme alone. This setting does not achieve significant improvement in scores. We believe that the drop in performance is because of the incompatibility (since generator output is non-binary in nature and has false positives) between prediction based input and the proposed labels.
- Ablation 5: From *UNet-TopoAL*, we remove the part of 'multiplication by the dilated ground truth' to generate prediction based input to the discriminator. The performance reduction (as compared to *UNet-TopoAL*) under this setting is an assessment of the amount by which the discriminator training gets influenced by the false positives present in the segmentation output.
- Ablation 6: From *UNet-TopoAL*, we remove the dynamic label assignment strategy alone to assess the impact of our key idea. Clearly, the performance of the resulting approach deteriorates significantly, underscoring the significance of dynamic label assignment on the success of *TopoAL*.
- *UNet-TopoAL*: Our proposed method that follows the same settings as we reported in the main paper.

Among various ways to train the discriminator, the training setup of *UNet-TopoAL* yields the best performance since *UNet-TopoAL* makes use of dynamic label assignment as well as scale space labels while keeping the inputs in a compatible form to the proposed label generation scheme.

Method Name	Levels in Scale space labels	*APLS*	*H&M*
UNet-TopoAL	$k \in \{0, 1, 2\}$	0.651	0.752
	$k \in \{0, 1, 2, 3\}$	0.666	0.767
	$k \in \{0, 1, 2, 3, 4\}$	0.665	0.759

Table 3: Ablation studies on the number of levels in the scale space labels used for discriminator training.

Another interesting aspect to investigate is the number of levels in scale space label based training that yields the best performance. The values corresponding to the number of levels varying from 3 to 5 are reported in Table 3. Clearly $k \in \{0, 1, 2, 3\}$ (*UNet-TopoAL*) results in the best performance. Furthermore, increasing the spatial-awareness to levels higher than four does not seem to improve the performance over *UNet-TopoAL*.

References

1. Batra, A., Singh, S., Pang, G., Basu, S., Jawahar, C., Paluri, M.: Improved Road Connectivity by Joint Learning of Orientation and Segmentation. In: Conference on Computer Vision and Pattern Recognition (June 2019)
2. Demir, I., Koperski, K., Lindenbaum, D., Pang, G., Huang, J., Basu, S., Hughes, F., Tuia, D., Raskar, R.: Deepglobe 2018: A Challenge to Parse the Earth through Satellite Images. In: Conference on Computer Vision and Pattern Recognition (June 2018)
3. Isola, P., Zhu, J., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: Conference on Computer Vision and Pattern Recognition (2017)
4. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Conference on Medical Image Computing and Computer Assisted Intervention. pp. 234–241 (2015)
5. Zhou, L., Zhang, C., Wu, M.: D-linknet: Linknet with pretrained encoder and dilated convolution for high resolution satellite imagery road extraction. In: CVPR Workshops (2018)