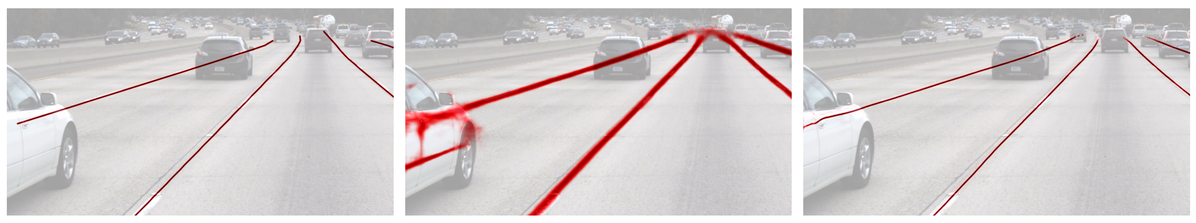
EL-GAN: Embedding Loss Driven Generative Adversarial Networks for Lane Detection

Mohsen Ghafoorian, Cedric Nugteren, Nóra Baka, Olaf Booij and Michael Hofmann
TomTom, Amsterdam, The Netherlands

Overview: Structured Prediction

ConvNets have been successfully applied to semantic segmentation problems. However, there are many problems that are inherently not pixel-wise classification problems but are frequently formulated as semantic segmentation. In dense prediction (e.g for lane marking detection) certain structures/qualities often need to be preserved. E.g. convergence in the vanishing point, smoothness and continuity, at reasonable distance to each other, consistent with the representations of other objects in the image, etc.



labels regular CNN EL-GAN

Pixel level losses are not inherently able to model and enforce these qualities.
other methods in the literature often address this with:

- **Hand-crafted post-processing**: ad-hoc and domain specific, often computationally expensive.
- **Conditional random fields**: partial coverage of consistencies, extra computational burden at inference time.
- **Additional engineered loss terms**: Ad-hoc and domain specific, often tricky to formulate differentiable loss terms.

Baseline: Adversarial Training for Dense Prediction

Add an extra adversarial loss term to represent how plausible (structure preserving) the predictions are [1, 2]:

$$\mathcal{L}_{\text{gen}}(x, y; \theta_{\text{gen}}, \theta_{\text{disc}}) = \mathcal{L}_{\text{fit}}(G(x; \theta_{\text{gen}}), y) + \lambda \mathcal{L}_{\text{adv}}(G(x; \theta_{\text{gen}}); x, \theta_{\text{disc}}), \quad (1)$$

where $\mathcal{L}_{\text{fit}}$ represents pixel-wise binary cross entropy loss and $\mathcal{L}_{\text{adv}}$ is formulated with $\mathcal{L}_{\text{bce}}(D(G(x; \theta_{\text{gen}}); \theta_{\text{disc}}), 1)$. At the same time, the discriminator minimizes the following loss:

$$\mathcal{L}_{\text{disc}}(x, y; \theta_{\text{gen}}, \theta_{\text{disc}}) = \mathcal{L}_{\text{bce}}(D(G(x; \theta_{\text{gen}}); \theta_{\text{disc}}), 0) + \mathcal{L}_{\text{bce}}(D(y; \theta_{\text{disc}}), 1). \quad (2)$$

Two issues with the above adversarial formulation:

1. **Direct dependence on discriminator's interpretations**: The feedback given to the generator only stems from the discriminator's representations of notions of fakeness and reality, which might be misleading if the learned discrimination is not realistic.
2. **Ignoring the image/label pairing information**: The adversarial training is not leveraging the the valuable image/label pairing information available in the supervised learning scenarios.

Our Contribution: Embedding Loss for Adversarial Training

The idea is to leverage the labels to steer the adversarial training and base the adversarial loss on high-level structures/characteristics of labels:

$$\mathcal{L}_{\text{gen}}(x, y; \theta_{\text{gen}}, \theta_{\text{disc}}) = \mathcal{L}_{\text{fit}}(G(x; \theta_{\text{gen}}), y) + \lambda \mathcal{L}_{\text{adv}}(G(x; \theta_{\text{gen}}), y; x, \theta_{\text{disc}}), \quad (3)$$

$$\mathcal{L}_{\text{adv}}(G(x; \theta_{\text{gen}}), y; x, \theta_{\text{disc}}) = \|D_e(y; x, \theta_{\text{disc}}) - D_e(\hat{y}; x, \theta_{\text{disc}})\|_2, \quad (4)$$

where D_e represents embeddings extracted from a certain layer in the discriminator network.

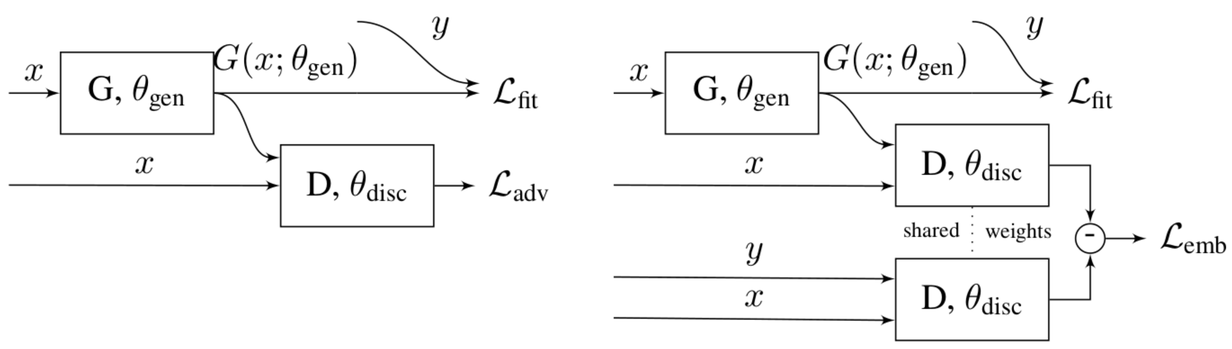


Figure: Illustration of the novel training set-up for the generator loss: left for a conventional GAN (Equations 1, 2), right when using the embedding loss (Equations 3, 4)

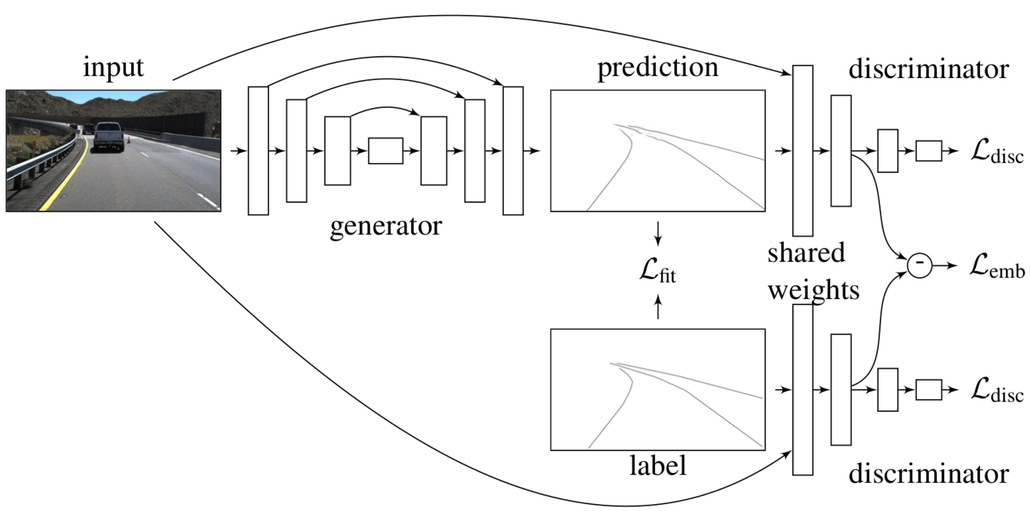
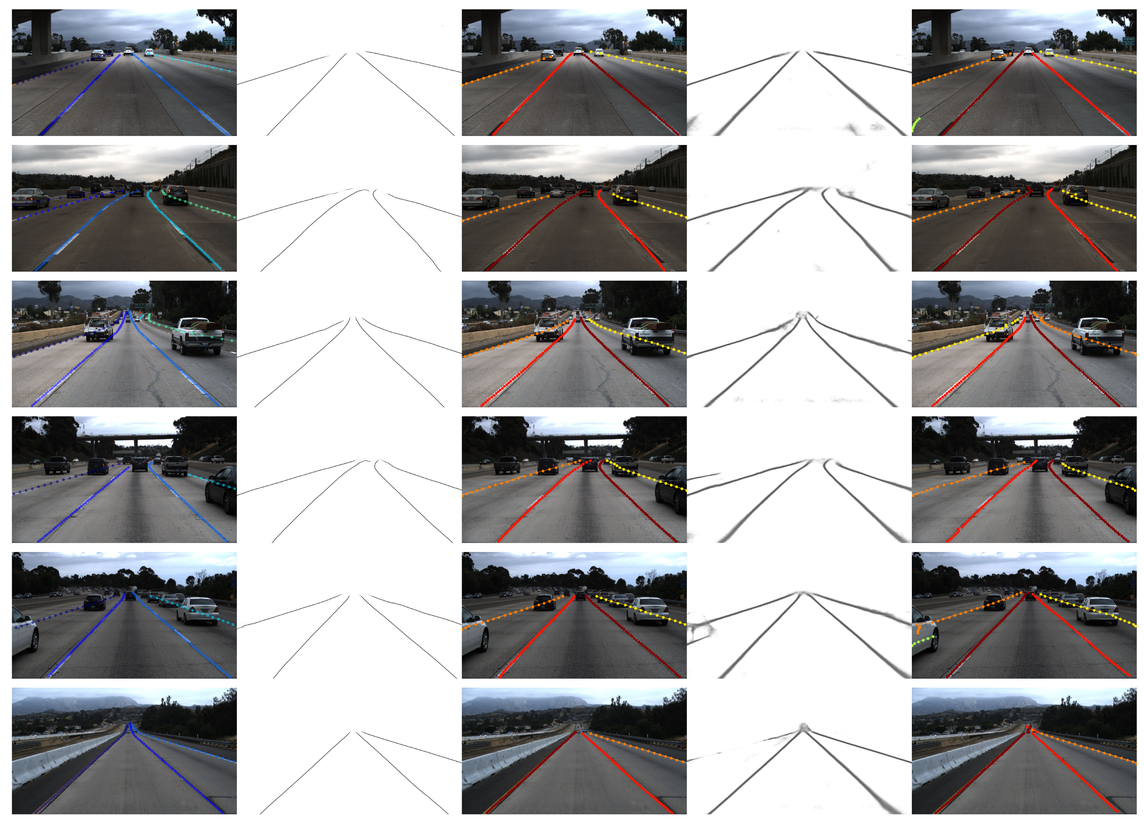


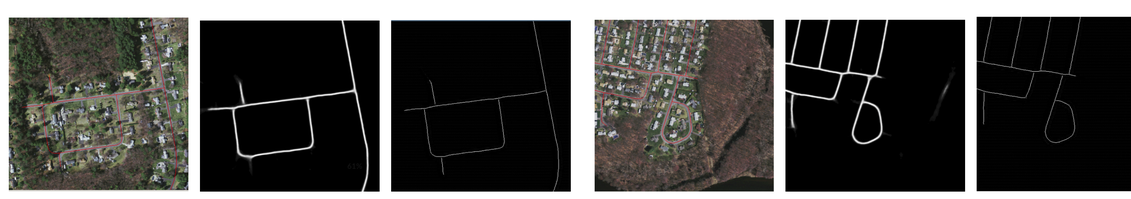
Figure: Schematic of EL-GAN architecture

Experimental Results



label on data prediction post-processed prediction post-processed
EL-GAN regular CNN

Figure: Qualitative comparison of EL-GAN to regular CNN.



GT on img CNN EL-GAN GT on img CNN EL-GAN

Figure: EL-GAN tested on satellite imagery road extraction.

Table: Results on TuSimple lane marking *validation* set

Method	Post-processing	Accuracy (%)	FP	FN
Baseline (no GAN)	basic	86.2	0.089	0.213
Baseline (no GAN)	basic++	94.3	0.084	0.070
EL-GAN	basic	93.3	0.061	0.104
EL-GAN	basic++	**94.9**	**0.059**	**0.067**

Table: TuSimple lane marking challenge leaderboard (test set) as of March 14, 2018

Rank	Method	Extra data	Acc.	FP	FN
#1	leonardoli	?	**96.87**	0.0442	0.0197
#2	Pan et al. [3]	Yes	96.53	0.0617	**0.0180**
#3	aslarry	?	96.50	0.0851	0.0269
#5	Neven et al. [4]	No	96.38	0.0780	0.0244
#6	li	?	96.15	0.1888	0.0365
#14	Baseline (no GAN)	No	94.54	0.0733	0.0476
#4	EL-GAN	No	96.39	**0.0412**	0.0336

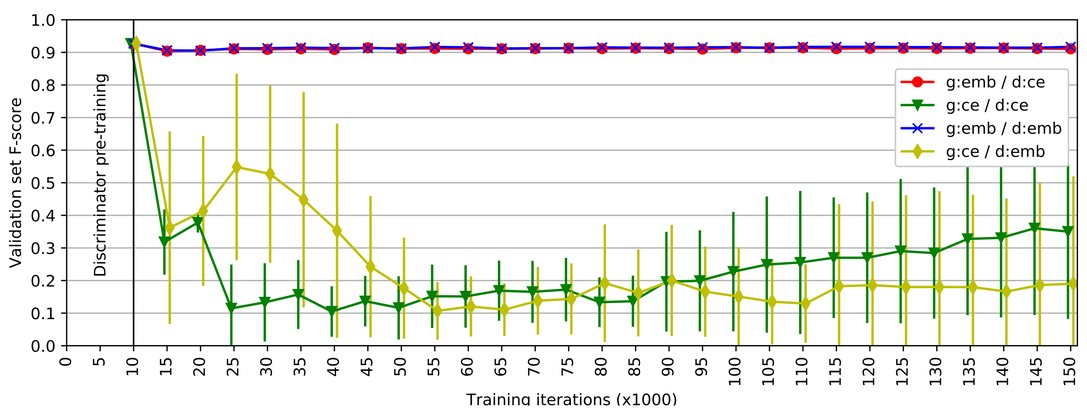


Figure: Ablation study on different adversarial loss terms. d: discriminator, g: generator, emb: embedding loss, ce: binary cross entropy loss.

References

[1] P. Luc *et al.*, "Semantic Segmentation using Adversarial Networks," in *NIPS-W*, 2016.

[2] P. Isola *et al.*, "Image-to-image Translation with Conditional Adversarial Networks," in *CVPR*, 2017.

[3] X. Pan *et al.*, "Spatial As Deep: Spatial CNN for Traffic Scene Understanding," in *AAAI*, 2018.

[4] D. Neven *et al.*, "Towards End-to-End Lane Detection: an Instance Segmentation Approach," 2018.

Our Paper: https://arxiv.org/pdf/1806.05525.pdf