Photographic Image Synthesis with Cascaded Refinement Networks

Zejia LV
zejialv@zju.edu.cn

5th December 2018

Introduction

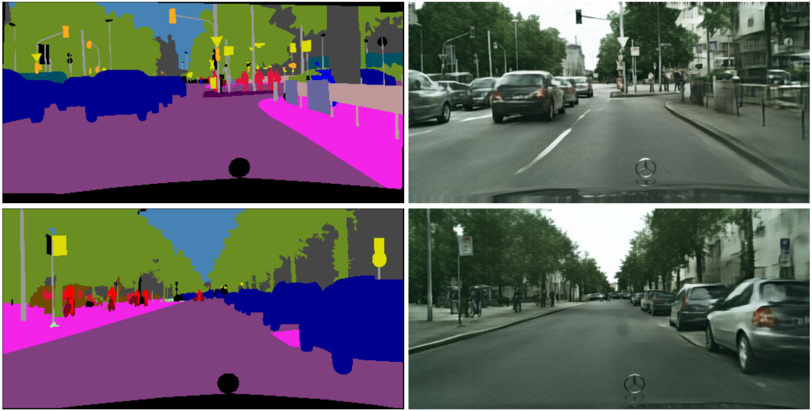
Background

- Given a semantic layout of a novel scene, can an artificial system synthesize an image that depicts this scene and looks like a photograph?
- Mental imagery is believed to play an important role in planning and decision making. Our second source of motivation is the role of mental imagery and simulation in human cognition

What we do

Our model is a convolutional network, trained in a supervised fashion on pairs of photographs and corresponding semantic layouts. Such pairs are provided with semantic segmentation datasets.

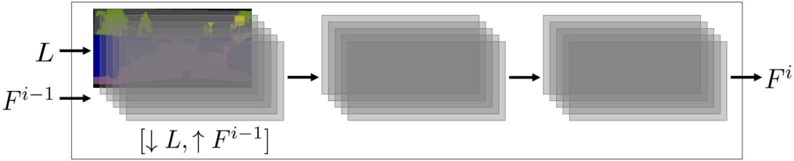
Introduction



Algorithms

Architecture

The Cascaded Refinement Network(CRN)



Algorithms

For a training pair $(I, L) \in \mathcal{D}$, our loss is

$$\mathcal{L}_{I,L}(\theta) = \sum_{l} \lambda_l \left\| \Phi_l(I) - \Phi_l(g(L;\theta)) \right\|_1 \tag{1}$$

Algorithms

Our first version of the modified loss is based on the hindsight loss developed for multiple choice learning

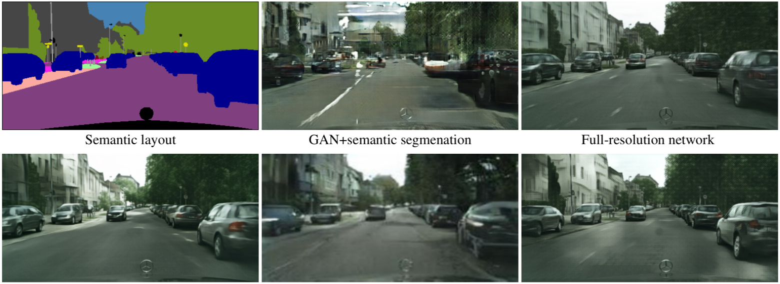
$$\min_u \sum_l \lambda_l \left\| \Phi_l(I) - \Phi_l\left(g_u(L;\theta)\right) \right\|_1 \tag{2}$$

We now define a more powerful diversity loss as

$$\sum_{p=1}^{c} \min_u \sum_l \lambda_l \sum_j \left\| L_p^l \odot \left(\Phi_l^j(I) - \Phi_l^j\left(g_u(L;\theta)\right) \right) \right\|_1 \tag{3}$$

Experiments

Qualitative comparison on the Cityscapes dataset



Semantic layout | GAN+semantic segmenation | Full-resolution network

Our result | Isola et al. [16] | Encoder-decoder

Experiments

Qualitative comparison on the NYU dataset

Further

Ongoing Optimization

- Encoder-decoder and convolutional network have good performance on image processing.
- This result, while significantly more realistic than the prior state of the art, are clearly not indistinguishabld from real HD images.

References

- Qifeng Chen, Vladlen Koltun. *Photographic Image Synthesis with Cascaded Refinement Networks*