Weakly Supervised Photo Enhancer Proposal

CSI 5139[Q]

September 28, 2018

Professor: Dr. Lang

Student Name & Number: Felix Singerman - 7970742

Background:

The paper I have chosen to present, and implement, was presented at the Computer Vision and Pattern Recognition (CVPR) conference in 2018 by A. Ignatov et al. from ETH Zurich, titled "WESPE: Weakly Supervised Photo Enhancer for Digital Cameras". The paper presents a novel image-to-image Generative Adversarial Network (GAN) based architecture trained under weak supervision in order to mitigate the limitations of poor photo quality from mobile cameras and old photos.

Image enhancement can be performed by graphical artists or by using a specialized software that enhances the photo's sharpness, contrast adjustments, and much more. However, this requires a set of skills that the day-to-day user might not have. Additionally, the process can be lengthy, and it is not possible to compute in a reasonable amount of time in the case of large-scale data processing. Past approaches to this problem have been through a supervised learning method which requires matched before/after learning pairs.

The paper proposes a deep learning architecture (see Appendix A) that "can be trained to enhance images by mapping them from the domain of a given source camera into the domain of a high-quality photos while not requiring any correspondence or relation between the images from these domains" (Ignatov et al., 2018). The only data required to train the network are two separate photo collections representing both domains.

Proposed Work:

I plan on implementing all the contents of the paper, except for the Flickr Faves Score and the User study. I will not be conducting comparison tests with a similar supervised learning method nor the commercial software baseline (the Apple Photos enhancement software).

The contents of the paper will be implemented in Python using TensorFlow and other data science and machine learning libraries such as pandas, matplotlib, NumPy etc. I will follow the architecture proposed in the paper and implement each part of the algorithm: content consistency loss, adversarial color loss, adversarial texture loss, total variation (TV) loss, and the sum of losses, which is a linear combination of the other four losses (see Appendix B). I will use the KITTI (A. Geiger, 2012), Cityscapes (M. Cordts, 2016), and DIV2K (Agustsson, 2017) datasets for training.

Appendix A

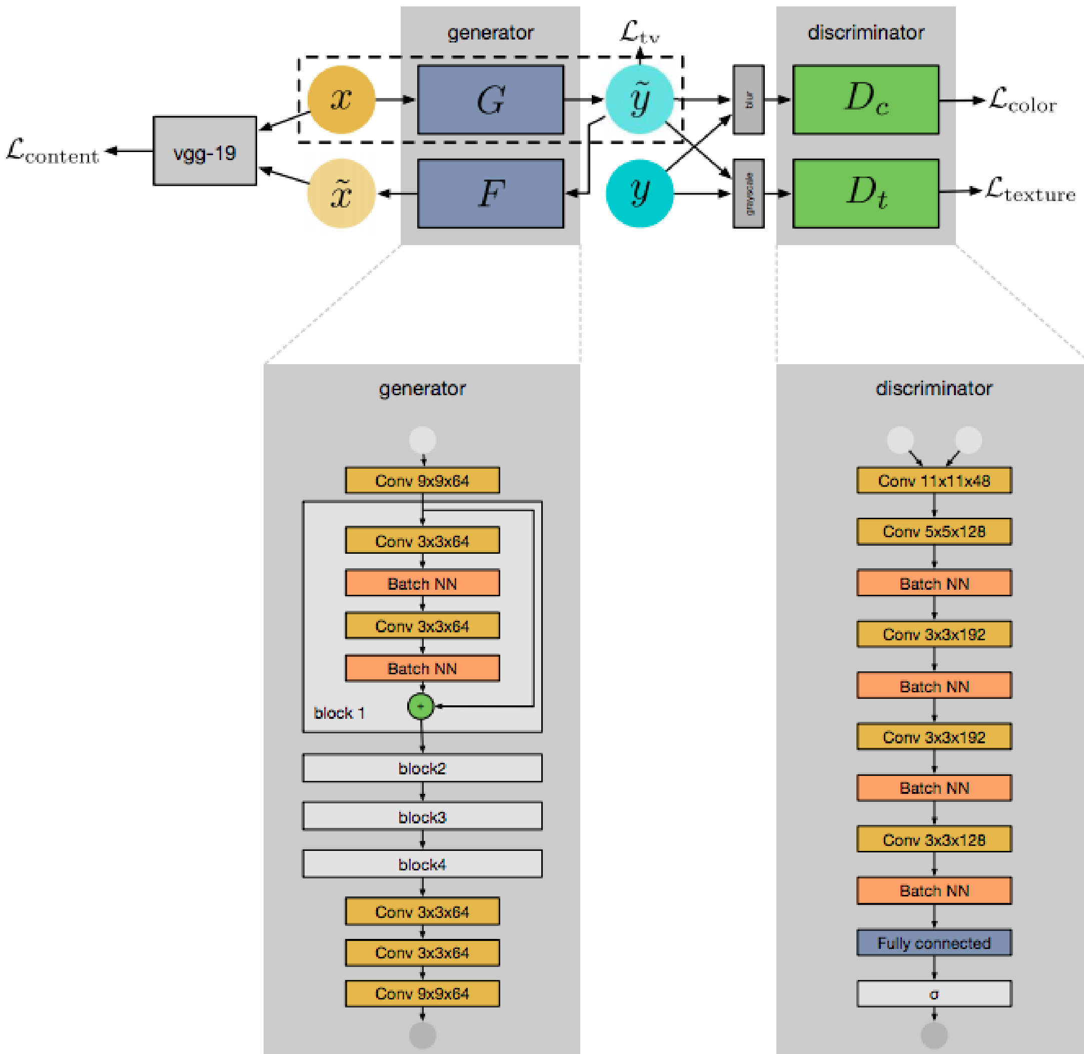

Figure 1: WESPE Architecture

Appendix B

$$\mathcal{L}_{\text{content}} = \frac{1}{C_j H_j W_j} \|\psi_j(x) - \psi_j(\bar{x})\|, \qquad (1)$$

Figure 2: Content Consistency Loss

$$y_b(i,j) = \sum_{k,l} y(i+k, j+l) \cdot G_{k,l}, \qquad (2)$$

Figure 3: Adversarial discriminator to differenciat between the blurred versions of enhanced and high quality images

$$\mathcal{L}_{\text{color}} = -\sum_i \log D_c(G(x)_b). \qquad (3)$$

Figure 4: Adversarial color loss

$$\mathcal{L}_{\text{texture}} = -\sum_i \log D_t(G(x)_g). \qquad (4)$$

Figure 5: Adversarial texture loss

$$\mathcal{L}_{\text{tv}} = \frac{1}{CHW} \|\nabla_x G(x) + \nabla_y G(x)\|, \qquad (5)$$

Figure 6: TV loss

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{content}} + 5 \cdot 10^{-3} \left(\mathcal{L}_{\text{color}} + \mathcal{L}_{\text{texture}}\right) + 10\, \mathcal{L}_{\text{tv}}. \quad (6)$$

Figure 7: Sum of losses

References

M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The Cityscapes Dataset for Semantic Urban Scene Understanding," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016

A. Geiger, P. Lenz, and R. Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In Conference on Computer Vision and Pattern Recognition (CVPR), 2012.

A. Ignatov, N. Kobyshev, R. Timofte, K. Vanhoey, and L. Van Gool. WESPE: Weakly Supervised Photo Enhancer for Digital Cameras. *arXiv:1709.01118 [Cs]*. Retrieved from http://arxiv.org/abs/1709.01118, 2017

A. Ignatov, N. Kobyshev, R. Timofte, K. Vanhoey, and L. Van Gool. DSLR-quality photos on mobile devices with deep convolutional networks. In Proceedings of the IEEE International Conference on Computer Vision, 2017

Agustsson, Eirikur & Timofte, Radu. NTIRE 2017 Challenge on Single Image Super-Resolution: Dataset and Study. 1122-1131. 10.1109/CVPRW.2017.150, 2017