Exposure correction with Adversarial Networks

Technical University of Denmark DTU

Pierre Le Jeune (s182169)

Abstract

Exposure is a fundamental component of a good picture. However, it can be quite challenging to set the camera parameters to get it right. Overexposed shots can be corrected, but this also demands some expertise. In this work, we try to show that this correction can be automated using deep learning. We use conditional adversarial networks in order to correct overexposed images. We mainly build our method on previous work from [1] that introduced a successful GAN architecture for image-to-image translation problems. On top of that we make use of more recent techniques to improve the quality of the reconstructions, such as spectral normalization, noise injection and perceptual loss.

Overexposure and trivial solution

Overexposure happens on a picture when some areas are totally white. This means that it is not possible to distinguish any shades of color within those areas, even though the actual scene had some. Overexposure can be simulated easily by directly cropping the pixel intensities of a picture to a certain threshold. However, this is not how it happens in real life. A camera takes pictures using a sensor that converts light intensity into pixel value and this sensor has a range of light intensities that it can detect. Overexposure happens when the sensor is exposed to light intensity outside of its range. Most of the time, pictures are then mapped to a non-linear light intensity space, in order to match what we see with our own eyes.

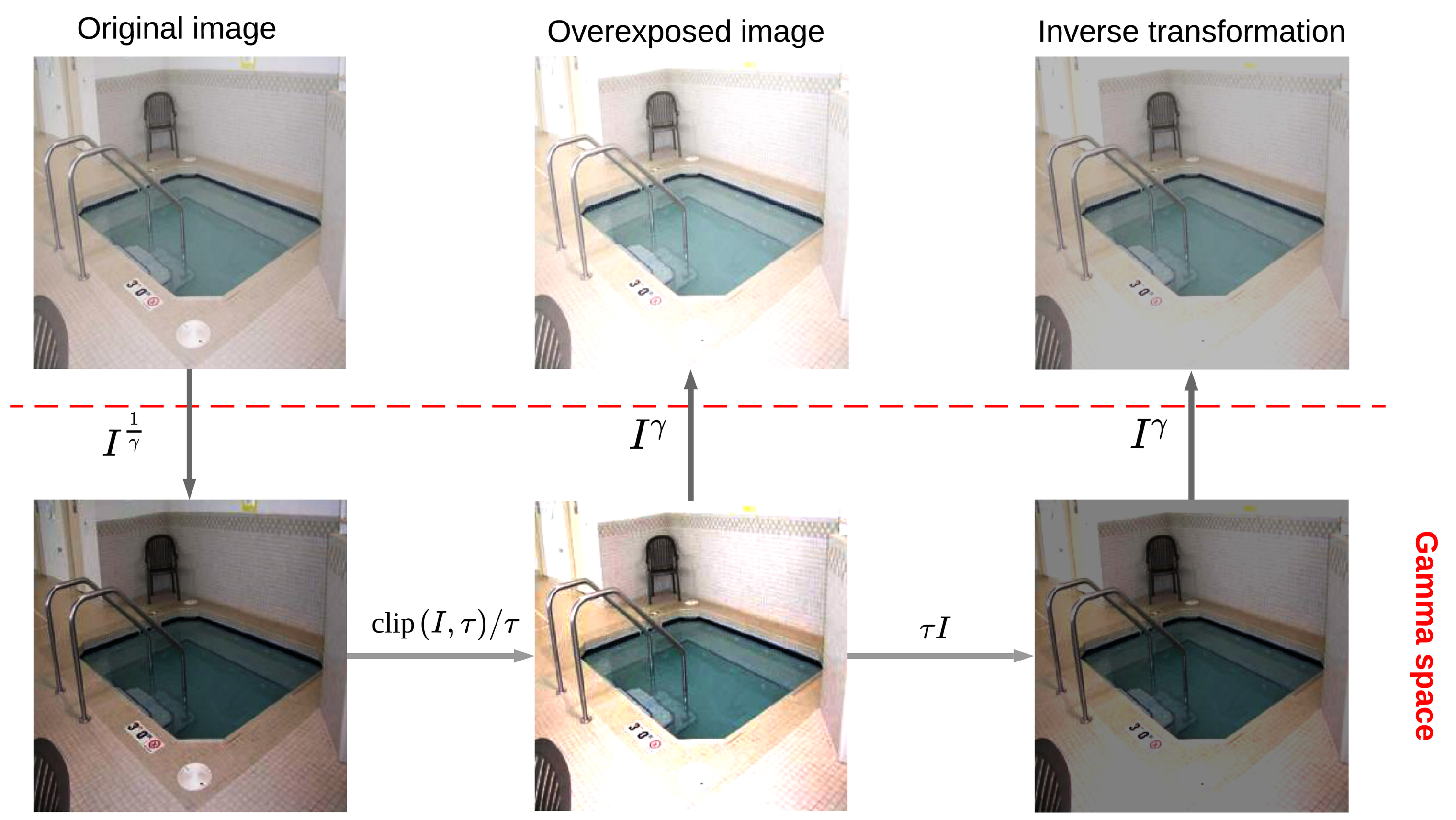

FIGURE 1: Overexposure simulation and trivial inverse operation. In order to generate a dataset with pair of images and overexposed images we used a subsample of the Places365 dataset. Overexposed images are simply generated by cropping pixel intensities in the gamma-space. In all the experiments: $\gamma = 1/2.2$ and $\tau = 0.5$

.

It is possible to correct the overexposure by applying the inverse operation but the resulting images do not look very natural.

Generative Adversarial Networks

Loss functions for the discriminator and the generator in the original GAN framework [2].

$$\mathcal{L}_D = \mathbb{E}_{x\sim p_{data}(x)} \log(D(x)) + \mathbb{E}_{z\sim p_z(z)} \log(1 - D(G(z))) \tag{1}$$

$$\mathcal{L}_G = \mathbb{E}_{z\sim p_z(z)} \log(D(G(z))) \tag{2}$$

Image-to-image translation and Pix2pix

This problem can be seen as an image-to-image translation problem: transforming overexposed images into correctly exposed ones. Therefore we chose to base our model on the Pix2pix architecture [1]. This architecture (see Figure 2) presents two advantages:

- Local discriminator: returns real/fake probability for multiple patches of the input image.
- Image size invariance: any image size can be fitted in the network.

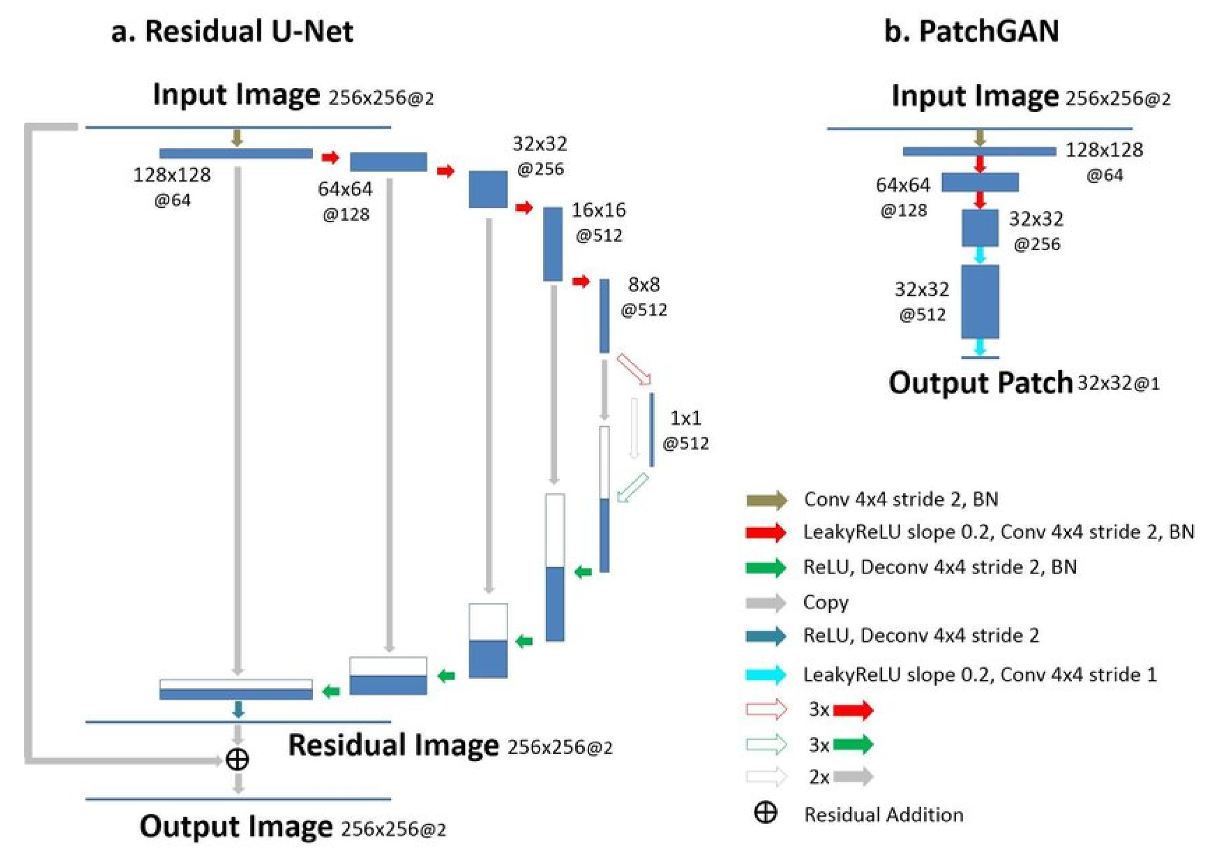

FIGURE 2: Example of U-net/PatchGAN architecture. Slightly different from the one used for our experiments.

Loss design

- Least squares adversarial losses [3] for improved stability.

$$\mathcal{L}_{LS} = \frac{1}{2}\mathbb{E}_{z\sim p_z(z)}\Big(D(G(z)^2)\Big) \tag{3}$$

- Supervised content loss: matching the reconstruction with the original image.

$$\mathcal{L}_{content} = \mathbb{E}_{x,y}\Big(\|y - G(x)\|_1\Big) \tag{4}$$

- Perceptual loss: matching features maps of VGG-19 for original and reconstructed image. Similar losses are widely used in image-to-image translation and improve a lot the reconstruction's quality.

$$\mathcal{L}_{perceptual} = \mathbb{E}_{x,y}\Big(\sum_{i\in\mathcal{C}} \frac{1}{W_i H_i}\|\phi_i(y) - \phi_i(G(x))\|_1\Big) \tag{5}$$

- Overall adversarial losses: a weighted combination of all of the above.

$$\mathcal{L}_G = \mathcal{L}_{LS} + \lambda_c \mathcal{L}_{content} + \lambda_p \mathcal{L}_{perceptual} \tag{6}$$

Experiments

- Adding noise with a learnable scaling factor into after every convolution to give 'inspiration' to the network, to fill areas where information was lost due to overexposure.
- Varying the field of view of the network. Each network has a field of view of 140 pixels. We tried multiple sizes and 140, besides having the best scores, produces the most natural reconstruction.

Results

	Pix2pix baseline	Noise	Perceptual + noise
Inception Score	18.36	19.26	**21.46**
Luminance Sim.	**90.19**	89.97	88.11
Contrast Sim.	88.29	88.74	**90.63**
Saturation Sim.	83.23	83.55	**84.73**

FIGURE 3: Example of reconstructed image (on the left) and ground truth (on the right).

Conclusions and future work

Overall this method produced relatively satisfying results. It learnt to correct small overexposure quite well, even if the whites in reconstructed images are still dimer. However, for highly overexposed images, the networks have trouble to fill the areas where all information was lost. This could be addressed using techniques from image inpainting such as [4].

References

[1] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134, 2017.

[2] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.

[3] Xudong Mao, Qing Li, Haoran Xie, Raymond YK Lau, Zhen Wang, and Stephen Paul Smolley. Least squares generative adversarial networks. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2794–2802, 2017.

[4] Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, and Thomas S Huang. Generative image inpainting with contextual attention. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5505–5514, 2018.