Undoing Instagram Filters

Padraig Boulton
P.Boulton@bath.ac.uk
Department of Computer Science,
University of Bath

Christian Richardt
christian@richardt.name
Department of Computer Science,
University of Bath

With 600 million users and counting, Instagram is among the most popular apps around today. It provides a photo-sharing platform as well as tools that allow everyone to create artistic photos. However, Chen et al. [2] found that state-of-the-art image recognition (CNNs) tends to fail on Instagram photos. Following this, Bianco et al. [1] showed AlexNet can classify Instagram-like filters with around 99% when trained from scratch. Here, the next logical step is taken: undoing Instagram filters.

We propose two approaches, both using deep architectures. For the first approach, referred to as targeted filter removal, we implemented and trained AlexNet from scratch to classify which filter was applied, as Bianco et al. [1] did. Instead of their approximated filters, we used 10k Instagram images with classes assigned based on 'hashtagged' filter names. After identifying the filter, we explicitly remove it by passing the image through an autoencoder trained to undo that specific filter. Rather than typical autoencoders, we use U-nets as Isola et al. [3] found these preferable for image-to-image translation; the skip-connections directly pass high level content between the encoder and decoder.

Starting from the pix2pix's [3] U-net, the architecture was optimised for undoing Instagram filters, to give:

- Encoder: $C64_{4,1} - C64_{4,1} - C64_{4,1} - C128_{4,2} - C256_{4,2} - C512_{4,2} - C512_{4,2} - C512_{4,2}$
- Decoder: $CD512_{4,2} - CD1024_{4,2} - CD1024_{4,2} - C1024_{4,2} - C1024_{4,2} - C512_{4,2} - C256_{4,2} - C128_{4,1}$

Where $Ck_{m,n}$ denotes a Convolution-BatchNorm-ReLU layer with k filters, kernel size m and stride n. $CDk_{m,n}$ denotes a Convolution-BatchNorm-Dropout-ReLU layer with a dropout rate of 50%. The optimised architecture uses a stride of 1 rather than 2 for the first 3 layers; without the downsampling of pix2pix, the pixel-to-pixel correspondence between image pairs may be exploited across multiple layers. Additionally, we appended coordinate channels to the image (i.e. images are RGBXY) to be able to learn spatial varying effects, such as vignettes.

Data consists of 100 unfiltered and Mayfair-filtered image pairs, split into 75 training and 25 test images. The optimised architecture reduces average MSE on the test set from 0.0026, with pix2pix's standalone U-net, to 0.0009. Furthermore, it did so without noticeably increasing the computational complexity. Ultimately, the downside is that so many networks are required: a U-net for each filter and one filter-classifying AlexNet.

For the second approach, we investigated undoing any filter with a single GAN using the full pix2pix framework as the starting point. First, for the optimised conditional GAN (CGAN), we replace pix2pix's generator with our optimised U-net architecture. Also, we investigate the unconditional variant, like Isola et al. [3] did, in addition to our "No filter vs Instagram" unconditional variant in which the discriminator compares filtered versus unfiltered images rather than real versus fake unfiltered.

Our dataset contains 100 images with 7 filtered-unfiltered pairs each, split into 75 training and 25 test images. Images are split so a unique image set (all 7 pairs) only appear in either the training or test set. Thus, at test time, the images presented have not been seen in any form. Another test dataset was generated featuring 10 images with 13 Instagram filters the models were not trained on, to asses how the networks extend to undo new filters. The results in Table 1 show that integrating the optimised U-net into pix2pix reduces average MSE significantly across all tests. Moreover, Figure 1 illustrates that pix2pix's outputs remain quite Instagram-like; when undoing 'Clarendon' the output remains excessively blue, and with 'Hefe' the vignette is still very visible. Conversely, the optimised CGAN's outputs are near-identical to the ground truth.

Table 1: Performance of the models on various test sets.

Model	pix2pix	Optimised CGAN	Unconditional	No filter vs Instagram
Test MSE	0.0062	**0.0048**	0.0071	0.0054
Test MSE (No Moon)	0.0047	**0.0034**	0.0054	0.0042
New filters MSE	0.0078	**0.0065**	0.0091	0.0083

Of the two methods, targeted filter more accurately removes filters, but one network per filter and an AlexNet must be trained to create the full pipeline. Using the GAN approach only requires one network to remove many filters, albeit at the cost of some accuracy. Furthermore, our GAN approaches appear to learn to undo Instagram-like features, so the networks can also perform zero-shot filter removal of new filters.

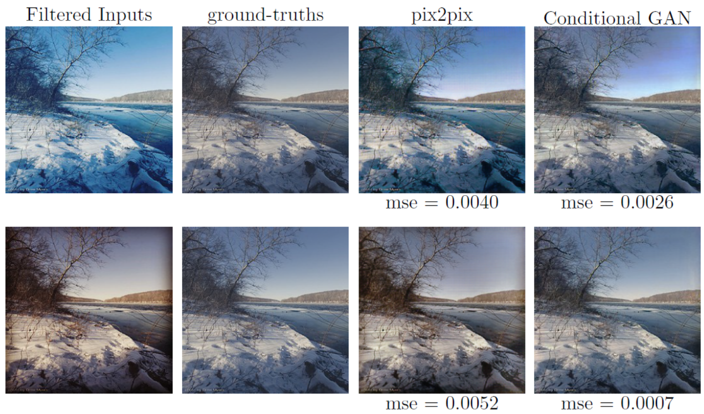

Figure 1: Example output images of pix2pix and our optimised CGAN (labeled Conditional GAN) and the images' MSE for 'Clarendon' (top) and 'Hefe' (bottom) filtered images.

The GANs' outputs given 'Moon'-filtered images, in Figure 2, highlight there is some subjectiveness to the results. The GANs are not trained to learn image colourisation, but clearly do learn some inherent colour-scene relationships. With a CGAN, the objective assesses the output given the input, so some implicit filter recognition occurs to infer colours with and without the filter. With unconditional variants, then the output is assessed on its own and the colouring of the image (apart from removing Instagram features) is inferred based on implicit scene/object recognition. The outputs confirm this, as the unconditional model ties colours to semantic objects far better than the conditional one. Interestingly, the unconditional outputs are different to the ground-truth, but are still a plausible. Thus, future work could investigate the subjective quality of the unconditional GAN-generated outputs.

(a) Input (b) ground-truth (c) Conditional GAN (d) No filter vs Instagram GAN

Figure 2: Example output images of the optimised CGAN (labeled Conditional GAN) and "No filter vs Instagram" unconditional GAN for a 'Moon' filtered image.

Finally, as Chen et al. [2] found image recognition performed poorly on Instagram images, future work should investigate whether state-of-the-art image recognition works as expected on Instagram images after undoing the filters.

[1] Simone Bianco, Claudio Cusano, and Raimondo Schettini. Artistic photo filtering recognition using CNNs. *International Workshop on Computational Color Imaging, CCIW 2017, Milan, Italy, March 29-31, 2017*, pages 249–258, 2017.

[2] Yu-Hsiu Chen, Ting-Hsuan Chao, Sheng-Yi Bai, Yen-Liang Lin, Wen-Chin Chen, and Winston H. Hsu. Filter-invariant image classification on social media photos. *International Conference on Multimedia*, pages 855–858, 2015.

[3] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-image translation with conditional adversarial networks. *CVPR*, 2016.