UnfilterMe – The “anti-filter” project

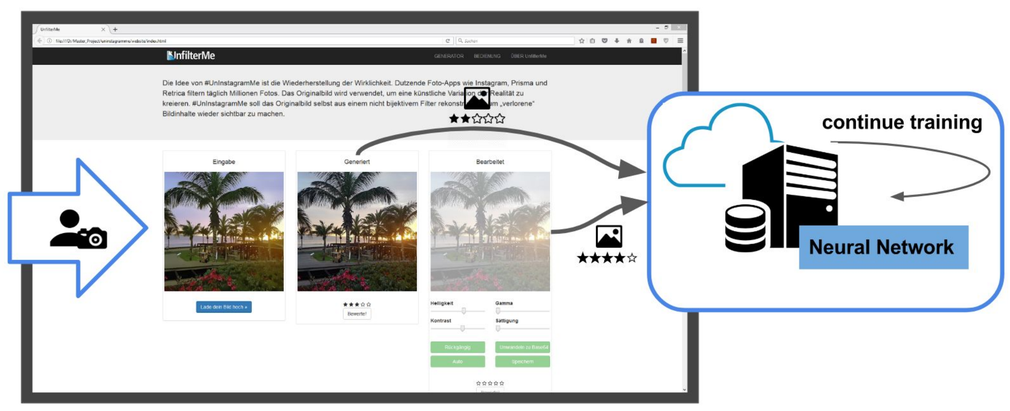
A. Abdeljaoued, M. Herzog, Th. Ilg, H. Lüdders, A. Morgner, K. Nejdl
HTW Berlin – Master Internationaler Studiengang Medieninformatik – July 2017

Objective

The web has become visual and dozens of apps offer the possibility to augment images with filters to obtain certain visual effects. This process, though, leads to loss of information. The goal is to restore this kind of information, to recover the original from an artificial variation of an image.

The area of Deep Learning and Neural Networks offers some algorithms for image-to-image translation. The objective of this project was to develop an API which allows continuous improvement of such an algorithm via user feedback.

Feedback via User Input

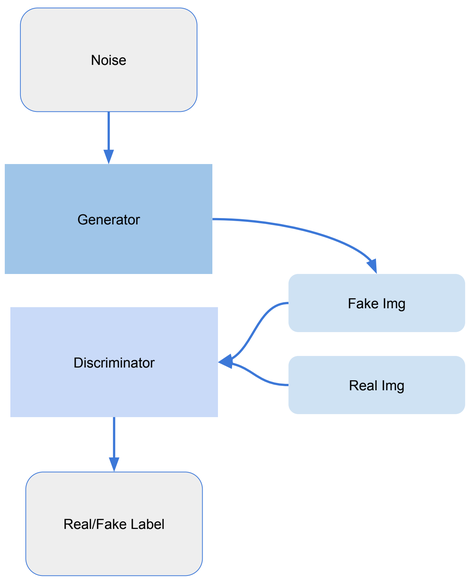


Through user input, the goal is to generate image pairs that improve the network training.
The webpage allows the user to upload their own image from which they want to remove a filter.
This picture will be processed by a trained GAN model. The server returns the generated image, and asks the user to rate the translation. Additionally, the user may edit the generated image to better fit the original, filter-less picture. This edited and the input image are then returned to the training, so that the algorithm learns to better map images it didn’t perform well on before.

Generative Adversarial Networks

The idea of Artificial Neural Networks is based on human neural cells. They take an input (in this case a pixel vector), multiply it with weights, and if the sum reaches a threshold, the neuron fires to activate other neurons. The neurons are arranged in layers, with an input layer, multiple hidden layers, and an output layer (here the generated image). The training of the Neural Network consists in automatically finding the correct weights to attain the desired output.

Generative Adversarial Networks (GANs) are a special kind of Neural Networks. They are commonly used for image-to-image translations to generate photo-realistic images. One network is trained to map a certain data set, and another tries to discriminate between these generated and the original images. The goal of the generative network is to fool the discriminator.



For this project, an implementation of Isola et al.’s Pix2Pix Neural Network [1] was used. Their network architecture has shown to be very effective at image-to-image translation.

Methods

The training of the Pix2Pix Neural Network was executed on an Nvidia Tesla GPU on an Amazon Web Server.
The trainings differed in dataset size, image size, used filter, and scenes depicted in the images. Beside a dataset of 5000 Instagram images, 15k images from the research dataset ImageNet from Berkeley University [4] were used. The Instagram dataset inherently contained more images of people, while Imagenet offers a wider range of image contents. If not otherwise specified, the models were trained with images of 128x128 pixel size. As test input, images of a size of 640x640 pixel were used, as a trained model can be applied to any image size, and the higher resolution produce considerably better results.
By default, Pix2Pix trains for 200 epochs, meaning that the whole dataset is passed through the network 200 times. The results get better with more epochs, as shown in the next section. As training with 15k images takes up to six days, some models were trained only for 100 epochs, which still produced good results.
Most common filters used in applications like Instagram etc. mainly consist of changes in color, saturation, or sharpness, resulting in filters such as a glow or sepia hue. As these do not illustrate the networks’ capabilities boldly enough, two different, hand-crafted comic filters were used.
Pix2Pix needs two paired images as input, so it can learn the mapping function from one to the other. In this case, these are images with and without a comic filter. The network then generates a “fake” unfiltered image.

Results

This section shows a range of results of differently trained models. The unfilterme.de website uses the model trained with the Imagenet dataset, 128x128px input image size, strong comic filter to generate an unfiltered image from the user’s input image.

Soft comic filter, comparison of input image size and epochs, Instagram-Dataset

Original | Filtered | 64px | 5 Epochs | 200 Epochs

Soft comic filter, 128x128, 200 epochs, comparison of datasets

Original | Filtered | Generated Instagram | Generated Imagenet

Strong comic filter, 128x128, 200 epochs

Filtered Strong | Generated Instagram | Filtered Strong | Generated Imagenet

References

[1] Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks
Jun-Yan Zhu*, Taesung Park*, Phillip Isola, Alexei A. Efros. In arxiv, 2017.
[2] Image-to-Image Translation with Conditional Adversarial Networks
Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, Alexei A. Efros. In CVPR 2017.
[3] Real-Time User-Guided Image Colorization with Learned Deep Priors.
Richard Zhang*, Jun-Yan Zhu*, Phillip Isola, Xinyang Geng, Angela S. Lin, Tianhe Yu, and Alexei A. Efros. In ACM Transactions on Graphics (SIGGRAPH 2017). (*equal contribution)
[4] http://www.image-net.org