DeOldify

Authors: Katerina Fortova (xforto00), Tomas Venkrbec (xvenkr01), Jiri Dvorak (xdvora2u)
Date: 08/04/2021
Inst.: FIT BUT

Introduction

- Open-source deep learning model by Jason Antic

Original Approach

- Colorization to grayscale images
- Use of Self-Attention Generative Adversarial Network (SAGAN)
- Inspired by Progressive Growing GANs - number of layers are constant but size of the input and learning rates are progressively changed -> more stable, faster training
- Generator replaced by pretrained U-Net
- Generator loss - Feature Loss based on VGG16 and loss score from the critic

- Latest version of DeOldify - more focused on processing videos and uses approach of NoGAN

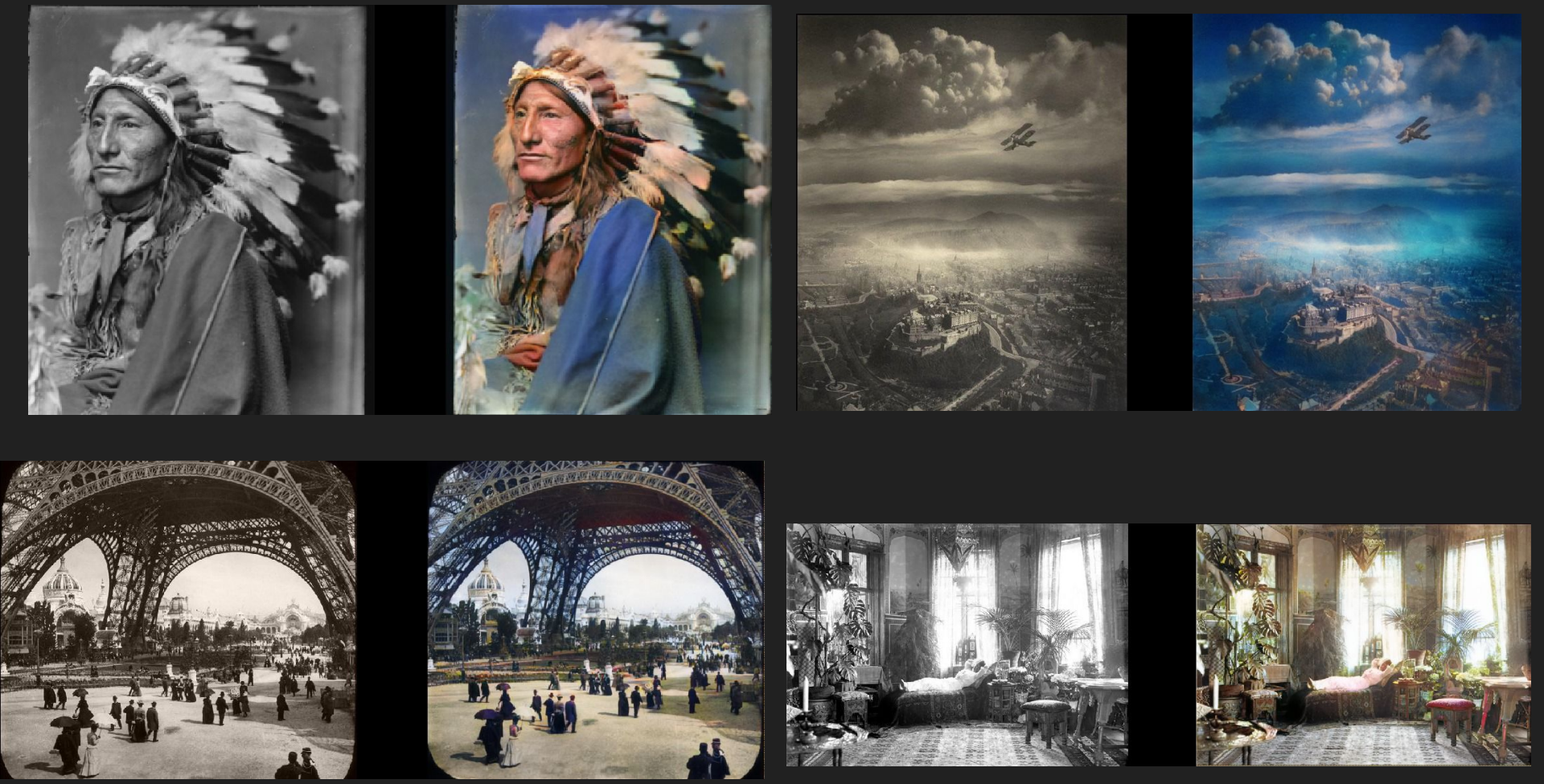

Figure 1: Comparison of ground truth images with the results from DeOldify. Taken from [1].

Self-Attention Generative Adversarial Network (SAGAN)

Traditional Convolutional GANs

- Generate high-resolution details as a function of only spatially local points in lower-resolution feature maps
- Difficulty in learning the image distributions of diverse multi-class datasets like Imagenet
- Easy generation of images with a simpler geometry like ocean or sky but images with some specific geometry like dogs, horses are problem
- Convolution is a local operation whose receptive field depends on the spatial size of the kernel

SAGAN Approach

- Details can be generated using cues (signals) from all feature locations
- Generator and discriminator efficiently model relationships between widely separated spatial regions
- Network first relies on the cues in the local neighborhood
- Gradually learning to assign more weight to the non-local evidence
- Progressive growth of the complexity of the task

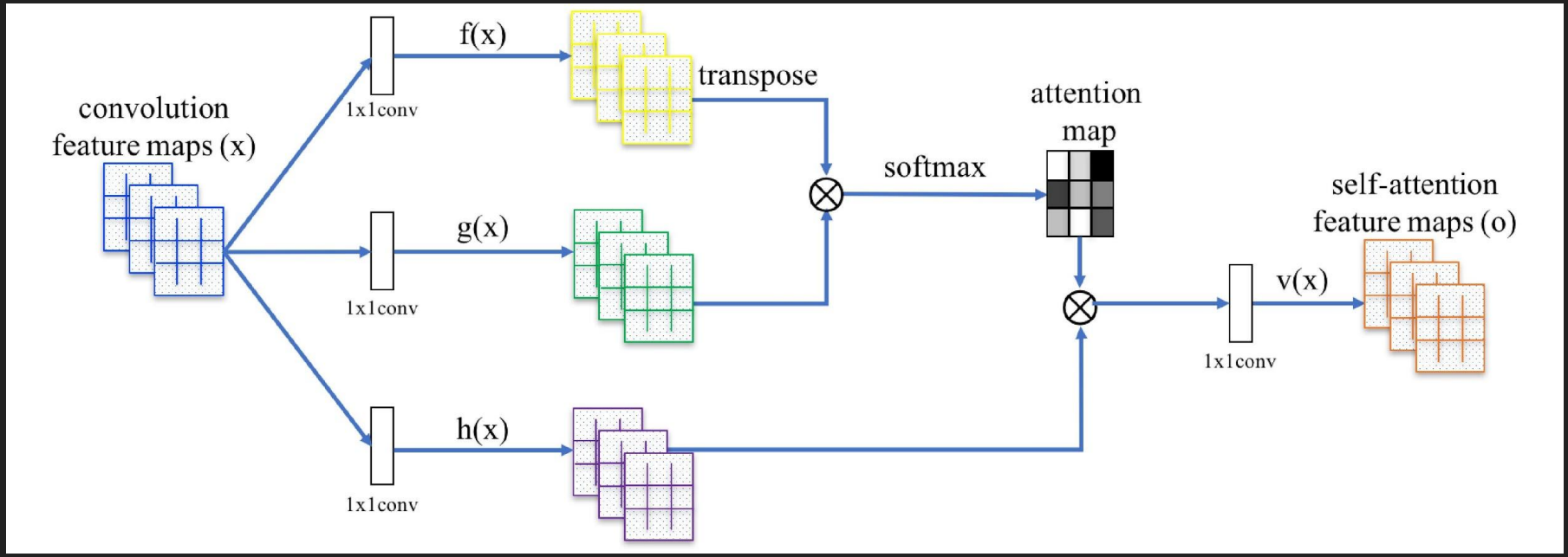

Figure 2: The self-attention module of the SAGAN proposed by [2].

- Use of spectral normalization in generator and discriminator
- Prevention of escalation of parameter magnitudes and unusual gradients for generator
- More stable training behaviour

U-Net

- Type of convolutional neural network
- Originally developed for biomedical image segmentation
- Resolution is gradually lower in contracting part and bigger in expansive part of network
- **Downsampling step** - number of feature channels doubled
- **Expansive step** - upsampling of feature map, concatenation with cropped feature map from contracting part, two 3x3 convolution with ReLU

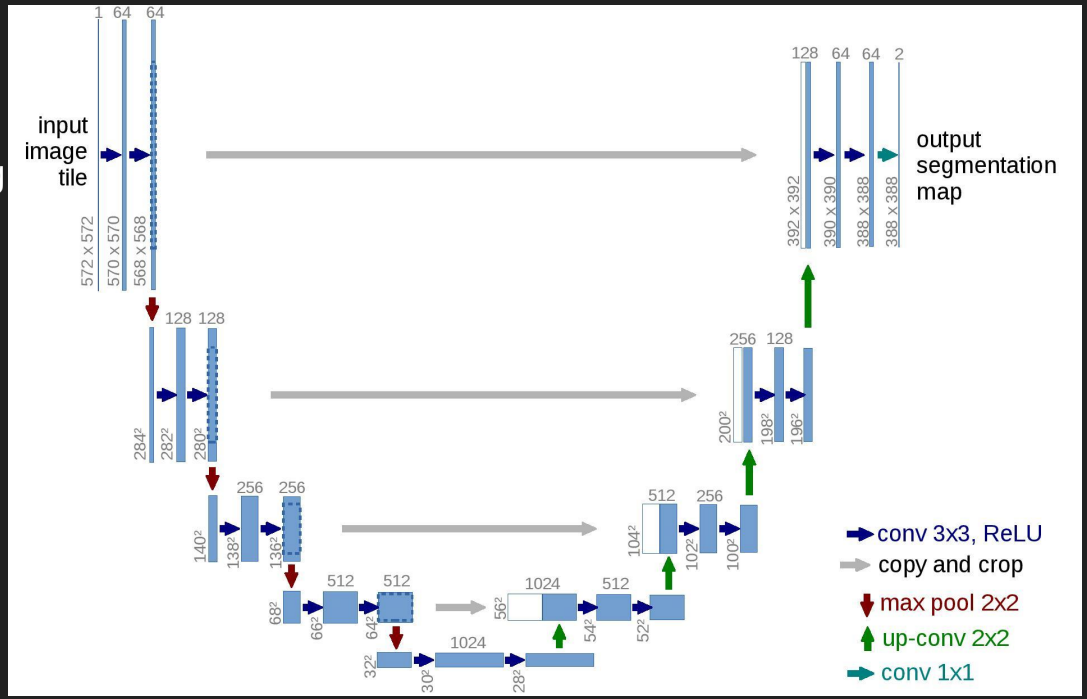

Figure 3: Example of U-Net architecture proposed by [3].

Bibliography

[1] Antic J., DeOldify, (2019), GitHub repository, https://github.com/dana-kelley/DeOldify

[2] Zhang, H., Goodfellow, I., Metaxas D., and Odena, A. Self-Attention Generative Adversarial Networks. arXiv:1805.08318v2, 2019.

[3] Ronneberger, O., Fischer, P. and Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. arXiv:1505.04597v1, 2015.