DeeperFour: Modernize Old Movies through Colorization and Animation

Zhuo Han, Muxin Liang, Xi Jin, Bo Sun

Department of Computer Science, University of Southern California

USC University of Southern California

Goal: Modernize Old Movies

It is a pity to watch old classic movies are forgotten by people due to limited shooting technology and the relatively boring black-and-white presentations. Therefore, we decided to make those films more enjoyable by filling them with vivid color by virtue of existing colorization networks, and creating an animated cartoon version thanks to style-transfer techniques. In order to make the final output video smooth and well displayed, our job is to combine those networks for colorization and style-transfer that originally work independently, and figure out a way to integrate another network to stabilize and increase consistency of frames through temporal loss.

Model Description

Input

Consecutive video frames from B&W films.

Colorization

We integrated the existing work of DeOldify[1], which composed of a generator with Unet34 architecture[4], and a discriminator with DCcritic architecture[5] which perform colorization, into stabilization as described beneath.

Stabilization

Learn video consistency by wrapping a short term temporal loss and a long term temporal loss[3], we modified the model so that it accommodate our goal of colorization, trained with 140 ~10s-long grayscale video frames.

Animation

The stabilized output of colored images were directly fed into CartoonGAN[2], which trained with 5418 real-world images and 4308 cartoon images clipped from Shinkai's movies and cropped into 256 × 256.

Output

Animated Video: stream together the output images and add on audio to form a video output.

Model Architecture

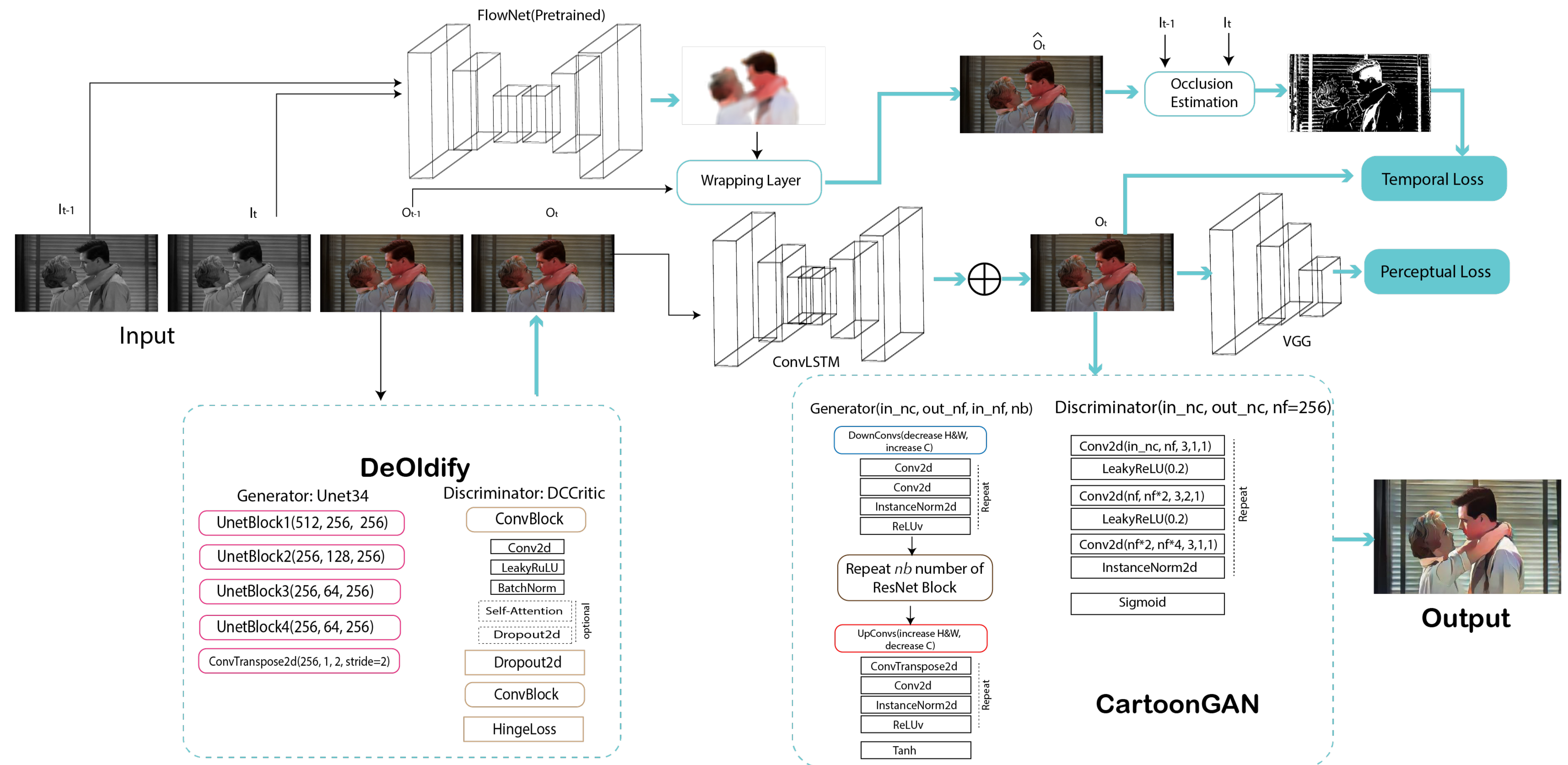

Fig. 1: Model Architecture

Loss Function

VGG perception loss

$$\mathcal{L}_p = \sum_{t=2}^{T}\sum_{i=1}^{N}\sum_{l}\|\phi_l(O_t^{(i)}) - \phi_l(P_t^{(i)})\|_1 \quad (1)$$

Short term loss

$$\mathcal{L}_{st} = \sum_{t=2}^{T}\sum_{i=1}^{N} M_{t\Rightarrow t-1}^{(i)}\|O_t^{(i)} - \hat{O}_{t-1}^{(i)}\|_1 \quad (2)$$

Long term loss

$$\mathcal{L}_{lt} = \sum_{t=2}^{T}\sum_{i=1}^{N} M_{t\Rightarrow t-1}^{(i)}\|O_t^{(i)} - \hat{O}_1^{(i)}\|_1 \quad (3)$$

Overall loss

$$\mathcal{L} = \lambda_p\mathcal{L}_p + \lambda_{st}\mathcal{L}_{st} + \lambda_{lt}\mathcal{L}_{lt} \quad (4)$$

Results

Fig. 2: Input sequence from Hitchcock's *Psycho*

Fig. 3: Colorization result and directly cartoonized result

Fig. 4: Colorization result with stabilization and its cartoonized result

References

[1] Jason Antic. *DeOldify: A Deep Learning based project for colorizing and restoring old images*. `https://github.com/jantic/DeOldify`. 2018.

[2] Yang Chen, Yu-Kun Lai, and Yong-Jin Liu. "CartoonGAN: Generative Adversarial Networks for Photo Cartoonization". In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2018), pp. 9465–9474.

[3] Wei-Sheng Lai et al. "Learning Blind Video Temporal Consistency". In: *CoRR* abs/1808.00449 (2018). arXiv: `1808.00449`. URL: `http://arxiv.org/abs/1808.00449`.

[4] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. "U-Net: Convolutional Networks for Biomedical Image Segmentation". In: *CoRR* abs/1505.04597 (2015). arXiv: `1505.04597`. URL: `http://arxiv.org/abs/1505.04597`.

[5] Han Zhang et al. "Self-Attention Generative Adversarial Networks". In: *arXiv e-prints*, arXiv:1805.08318 (May 2018), arXiv:1805.08318. arXiv: `1805.08318 [stat.ML]`.