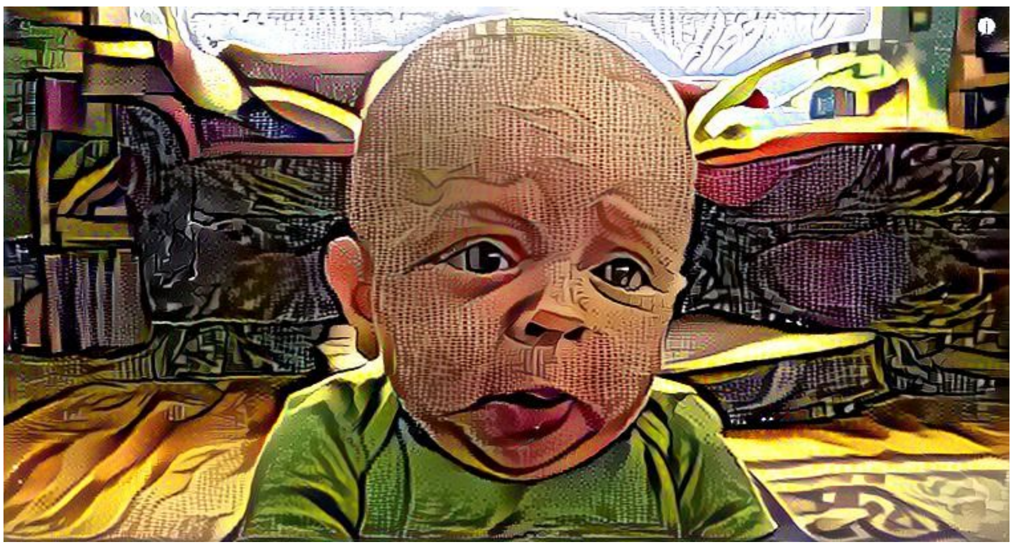
ART AND MACHINE LEARNING
CMU 2019 SPRING
PROJECT 1

Stabilization of Neural Style Transfer for Video



Piyusha Sane
Manisha Chaurasia
Simral Chaudhary

Concept:

Animation and graphic design elements are essential to a variety of media formats from magazines and books to television and movies to video games. In the cases of videos, which are composed of numerous frames, the animation is a more challenging task as a flow between the frames must be maintained. A key point is that most times, these videos obtain inspiration from living people or existing scenes. What if we could use human actors or real-world scenes to generate artistic graphics in animation? We took this motivation to style images of real life people with the aim to 'cartoonize daily lives' which could be used for comics characters.

In an era where Machine Learning has come to intersect Art, neural networks have proven to be effective in style transfer which ensures a myriad of applications in the graphic design and movie production industries. In this project, we explored the use of style transfer in images and videos, experimenting with existing methods and ways to improve the quality of the media produced.

While exploring the generation of videos through style transfer techniques, we observed that everytime an image was styled it would result into significantly different style output. This lead to the conclusion that when the frames of a video are styled and combined to form a video using the same technique, they would turn out to be quite unstable w.r.t the relatively constant background. It was evident that the challenge of ensuring flow in the background scenes would arise given the temporal instability. Extending the previously outlined neural style transfer technique to videos from images, we aspired to enhance the stability in the videos produced by incorporating the difference of change from previous image frame in the loss function. This was similar to Manuel Ruder et al. work on Artistic Style Transfer which incorporates the idea of optical flow [1].

Technique:

Style transfer of video frames with optical flow:

Leon A. Gatys's *A Neural Algorithm of Artistic Style* [2] outlines an algorithm known as neural style transfer in which a pretrained VGG19 model inputs a style image and a base content image which is reconstructed in the texture or theme of the style image. In the traditional method, three types of losses (content, style, and total variation) are minimized at evaluation time. Content loss compares the variation between the base image and the reconstructed image while the style loss compares the style representation of the style image and reconstructed image. Total variation loss augments the coherency of an image as it encourages minimal changes within the image. In addition to the existing loss terms, we added a fourth loss term, stabilization loss. This loss was computed as the mean squared error between the previous frame and current frame pixel values, and minimized over iterations in an effect to reduce the variability and transition noise between frames.

Process:

We explored several techniques before finalizing our resulting artwork. The experiments are highlighted below:

1. Generation of a stylized image with default parameters using the repository 'style-transfer' on various style images
2. Conversion of a video to frames using ffmpeg:
    a. Video was broken down into 202 frames with a frame rate of 15.
3. Stylizing each image frame of the video

4. Hyperparameters tuning:
    a. In this we modified the weights of each loss parameter i.e total_variation_weight, style_weight and content_weight.
5. New loss function:
    a. Added the component of stabilization loss to the final loss used to optimize the image. This was a mean squared error between images of consecutive frames: mean_over_pixels(square(frame1_iteration5 - frame0_iteration_0))
    b. We also experimented with computing stabilization loss w.r.t corresponding iteration of consecutive frames: mean_over_pixels(square(frame1_iteration3 - frame0_iteration_3))
6. Moving new loss function to training phase:
    a. The effective outcome of the new loss function could be judged better only if it was incorporated in the training phase of any network. Since style-transfer does not train the network based on styles, it uses a pretrained vgg19 (on Imagenet), so it doesn't learn to minimise the loss apart from the eval phase which uses fmin_l_bfgs_b to do that.
    b. Hence we planned to incorporate this new loss to the given 'fast-style transfer' approach. However, training on this network was very resource-intensive and we could not complete it hence we do not include those results in our report.

Result:

1. Simple generation of stylized images on various styles:

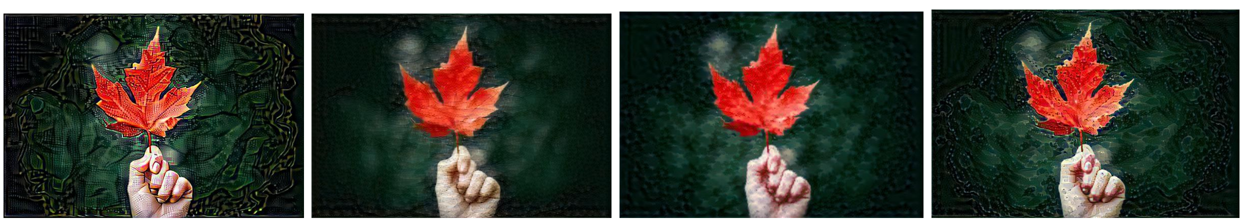
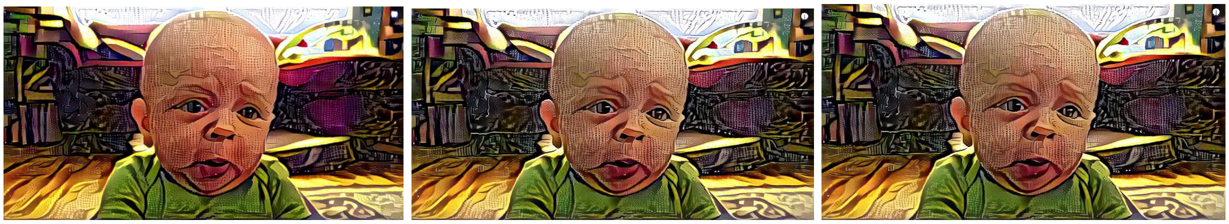


Image styles from left to right: Clouds, Liechtenstein, Picasso, Wave

2. Stylizing each frame of the video:

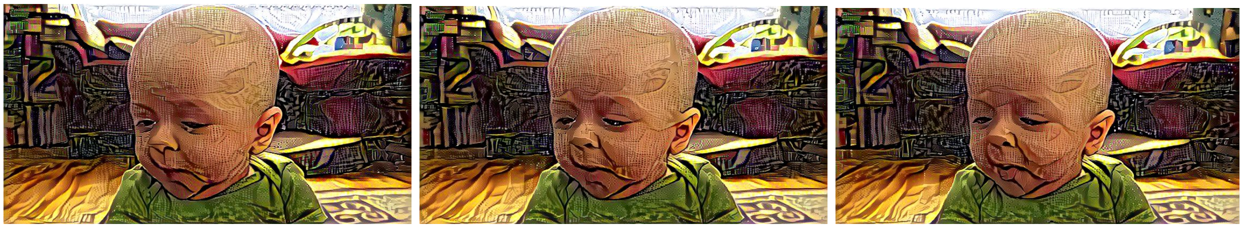
These are consecutive frames of the decomposed video which were stylized on Liechtenstein style

3. Effect of tuning total variation loss weights:



Images styled with total variation loss weights of 1, 1e-2 and 8.5e-5. The details of the variation are distinguishable. Amount of the distortion or change cause over the initial image is different.

4. Incorporating Stabilization Loss



Three consecutive frames of the final stylized and stabilized video. These frames show that the background details do not vary significantly across frames.

The final result is depicted in final_output_video.mov on the github page.

Reflection:

Reviewing our experiments, we found that this final video resulted in the most temporal stability as compared to the videos generated with the original hyperparameters and three loss terms only. Nevertheless, we feel that the optical flow in the video can still be optimized for example by using a weight to control the amount of stabilization loss being introduced. Recent studies [3] have recognized this problem as "popping" and have resorted to minimize this frame-to-frame noise at training time. In the future, we would like to train the model of fast-style transfer with this new loss function. Given additional time and resources, integration of the loss function during training would minimize the stylized versions of consecutive frames, potentially leading to more stabilization in the videos generated.

Performing the different experimentations in style transfer, we learned how various parameters and loss functions affect the stylized output in neural style transfer. We also discerned the importance of stabilization in videos and improved upon our results while experimenting with various techniques of optical flow. In a more holistic perspective, we have learned how style transfer can be applicable to so many use cases and medias.

References:

[1] Ruder et al. 2016. *Artistic Style Transfer for Videos.* https://github.com/manuelruder/artistic-videos
[2] Gatys et al. *2015. A Neural Algorithm of Artistic Style*
[3] Stabilizing neural style-transfer for video. Jeffrey Rainy and Archy de Berker. https://medium.com/element-ai-research-lab/stabilizing-neural-style-transfer-for-video-62675e203e42
[4] https://github.com/kangeunsu/ArtML/tree/master/fast-style-transfer
[5] https://github.com/titu1994/Neural-Style-Transfer

[6] https://github.com/jcjohnson/neural-style
[7] https://github.com/ElementAI/chainer-fast-neuralstyle/blob/stable-style/train.py#L177

CODE: https://github.com/simralc/style_transfer

RESULT: final_output_video.mov