Neural Style Transfer for Videos

Degree programme: BSc in Computer Science | Specialisation: Computer Perception and Virtual Reality
Thesis advisor: Prof. Dr. Jürgen Eckerle
Expert: Han van der Kleij (SBB AG)

Art has always been an essential part of human culture. With algorithms taking over more tasks of our daily lives, the question arises if they can create art. Artistic Neural Style Transfer is a tool that lets a computer reimagine a picture in a different style, for example, the style of a famous painting. That‘s impressing, but could a computer do this for a video?

Bernard Jaquet

Introduction

Artistic Neural Style Transfer is an AI-based concept that allows transferring style features like colours and textures from one image to another. Taking a famous painting as style input can transform any picture into something that the painter may have drawn. Since computers work faster than humans, this technology could be used to style entire videos. Unlike images, videos must be consistent in the time dimension. Since there are many ways of applying a style to an image, two consecutive frames may look substantially different, which leads to flickering in the video.

Image styling

Neural Style Transfer originates in a paper proposing a solution to a big problem in AI-based image processing. Learning algorithms need feedback to improve, which initially was provided by functions comparing the pixels of a result to a true-ground image. This approach has a big drawback. If a result gets shifted by a single pixel, the error value calculated is significant despite barely being visible to the human eye. This paper proposes to use a neural network trained for image classification to extract features from the images and compare these instead. This strategy allows results to differ if the impact on perception is small. Neural Style Transfer uses two of these perceptual error functions, one for style and one for content accuracy. The content error function compares the original picture to the styled image. The style error function scrambles the features to lose location information before comparing the result to the style reference.

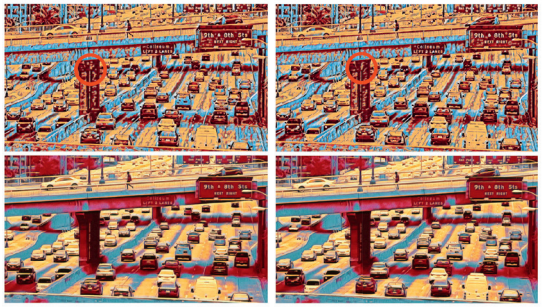
Video styling

Styling videos works similar to the method described previously but adds a new error function using the optical flow of the input video. Optical flow describes the movement of depicted objects between two frames. An image warping algorithm can use this optical flow to approximate the next frame of the video based on the previous frame. The temporal error function uses this algorithm on the previous styled frame to create a temporally coherent reference. Comparing this reference to the generated image detects temporal incoherences.

Results

Image styling requires finding a balance between the two error functions so that colours and textures get exchanged without changing the content. The results for styling videos are less clear cut. While some styles show little to no flickering, other styles still contain a lot. Increasing the weights on the temporal error function leads to side effects like a darker video or fewer colours in the video. These can be counteracted to a certain degree, but even then the results contain some flickering. In conclusion, the methods implemented for styling videos, while promising for some styles, don’t seem to be universally applicable.

A demonstration of Artistic Neural Style Transfer using Kandinsky’s Composition VII.



Comparison of a method with and without measures to enforce temporal coherence.