Machine Learning for the Arts
UCSD fall 2019
**FINAL PROJECT**

# Loving Kar-Wai

(experimental drama film)


(Hypnotic images of love and loss finally wear down your resistance as seemingly discordant slow motion, pixilation, rushing backgrounds and frozen foregrounds into a magic story. This is one of the most classic scenes from the film Chungking Express )

(Femme a Marguerite, pianted by Alphonse Mucha

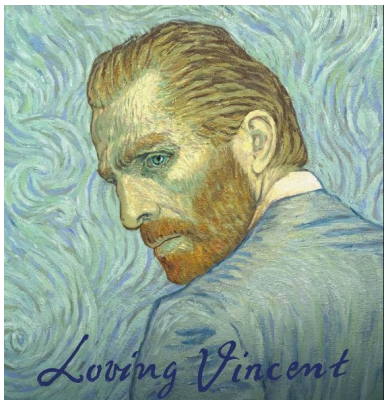
One of representative work of Art Nouveau)

（Zizhen Wang）

**Description**

**Concept:**

My idea is from a movie called “Loving Vincent” which is an experimental animated biographical drama film. Each of the film's 65,000 frames is an oil painting on canvas, using the same technique as Van Gogh, created by a team of over 100 painters.[1]




I started to think I could generated a film by using style transfer to get the same effort. There are two reasons I choose Art Nouveau as my style transfer. Based on my project 3, I generated 14 different art style of modern arts, and Art Nouveau style generated result is my favorite one and also I think is one of successful results. Modern art includes artistic works produced during the period extending roughly from the 1860s to the 1970s, and denotes the style and philosophy of the art produced during that era. The term is usually associated with art in which the traditions of the past have been thrown aside in a spirit of experimentation. Modern artists experimented with new ways of seeing and with fresh ideas about the nature of materials and functions of art. A tendency away from the narrative, which was characteristic for the traditional arts, toward abstraction is characteristic of much modern art.[2] Modern artists experimented with new ways of seeing and with fresh ideas about the nature of materials and functions of art. This ideology is the same as that this course wants us to learn which is we need to use new things (machine learning) to generate new art works to insist on this spirit. Art Nouveau is one of representative styles in modern arts which were a sense of dynamism and movement, often given by asymmetry or "whiplash" curves, and the use of modern materials to create unusual forms and larger open spaces. It was often inspired by natural forms such as the sinuous curves of plants and flowers. The second reason is the storyline of this film Chungking Express. Every day, Cop 223 (Takeshi Kaneshiro) buys a can of pineapple with an expiration date of May 1, symbolizing the day he'll get over his lost love.[3] By his encounter with an Asian woman wearing a trench coat and a blond wig(Brigitte Lin),  he is instantly attracted, oblivious of the fact she's a drug dealer. Originally thinking that everything in "love" had a shelf life, he unexpectedly ushered in a short warmth of soul. The scene I picked showed they were relying on each other in the bar at the first

night even though they are strangers.  I wanted to use the sense of dynamism and movement from Art Nouveau to show the warmth and the shortness.

**Technique:**

Style transfer is an optimization technique used to take three images, a content image, a style reference image such as an artwork, and the input image you want to style. And blend them together such that the input image is transformed to look like the content image, but in the style of the style image.[4] There 5 steps to perform style transfer: 1. Visualize data 2. Basic Preprocessing/preparing our data 3. Set up loss functions 4. Create model 5. Optimize for loss function. We use model VGG19 (VGG19 model, with weights pre-trained on ImageNet)[5].to train the set. Here is the sample for our model.

Deep dream is an algorithm by Google that magnifies the visual features that a CNN detects in an image, producing images where the recognized patterns are amplified. The method is to learn visual features from paired visual and textual data in a self-supervised way.[6]
Here is one example:

There are some problems when I use style transfer to generate video. The first one is that there are so many frames in video, so it will take too long time. The second one is that when used frame-by-frame on movies, the resulting stylized animations are of low quality which is inconsistent stylization from frame to frame. The stylized features (lines, strokes, colours) are present one frame but gone the next frame.

1, The first approach is to produce a fast style-transfer algorithm which significantly reduces the effect of popping on the learned style. The stabilization is done at training time, allowing for an unruffled style transfer of videos in real-time. [7] The changes in pixel values from frame-to-frame are largely noise. Therefore, they manually add a small amount of noise to the images during training and minimize the difference between the stylized versions of the original and noisy images to impose a specific loss at training time.
The original video: https://youtu.be/gp98OojLxfs
The modified video: https://youtu.be/VYbXYfjBBSM

Also, the stability difference is readily visible from the images which the environment is much more stable.
However, there is still about 1-2 minutes for each style frame generation. For my video, I have 266 images for 11 seconds video. Therefore, I tried another method.

2. They made the image stylization procedure real-time by training a neural network for this optimization problem instead of optimizing each image separately. The method keeps the temporal consistency but works about 10 times as fast as the method of regular style transfer. The method makes it possible to stylize movies with reasonable time costs.[8] They are use 2 ways to accelerate the training time and model compression: transfer learning and knowledge distillation.[9] Although neural networks often benefit from an excess of trainable parameters during the training process to learn an approximation function, the actual computation of the function learned doesn't necessarily need all of the network parameters. Transfer learning is that instead of having to learn all of the weights from scratch, the network can benefit from some general knowledge learned from the first task. Knowledge distillation is that the smaller network is trained to mimic the output of the larger pre-trained network for each training example. Since the larger neural network has already done much of the work in learning high level features of the function, the smaller network can benefit from its knowledge without having to go through nearly as many training examples. The results are shown below:

Resolution: 640*480

| Label (As in Video) | Method | Time |
|---|---|---|
| Ruder et al. (30 iterations) | Ruder's method, 30 iterations | 0.80 hour |
| Johnson et al. (Real-time) | Simply run Johnson's method for each frame | 77.8 seconds |
| Ruder et al. (1000 iterations) | Ruder's method, 1000 iterations | 7.56 hours |
| Our method (30 iterations), no pixel loss | Our method, with pixel-loss weight 0 | 0.80 hour |
| Our method (30 iterations), with pixel loss | Our method, with pixel-loss weight 1.5e-3 | 0.80 hour |

Result: https://youtu.be/61nrG1fCfz0

Since the method is based on the real-time algorithm, the result showed that the background and the character sometimes will be mixed which is not stable. Therefore, I tried another method.

3. The method gets as input one or more keyframes then automatically propagates the stylization to the rest of the sequence. The approach for interpolating style between keyframes could preserves texture coherence, contrast and high frequency details to facilitate less training time while preserving visual quality.[10] The method based on the module called Ebsynth. Ebsynth is a versatile tool for by-example synthesis of images. It can be used for a variety of image synthesis tasks, including guided texture synthesis, artistic style transfer, content-aware inpainting and super-resolution. The focus of Ebsynth is on preserving the fidelity of the source material. Unlike other recent approaches, Ebsynth doesn't rely on neural networks. Instead, it uses a state-of-the-art implementation of non-parametric texture synthesis algorithms. Ebsynth produces crisp results, which preserve all the fine detail present in the original image.[11]

Result: https://youtu.be/XslAbz9jj5I


If we provide (a) (b) (d) (f), they will generate (c) and (e). I picked up this method since it could keep accurate details while short process time.

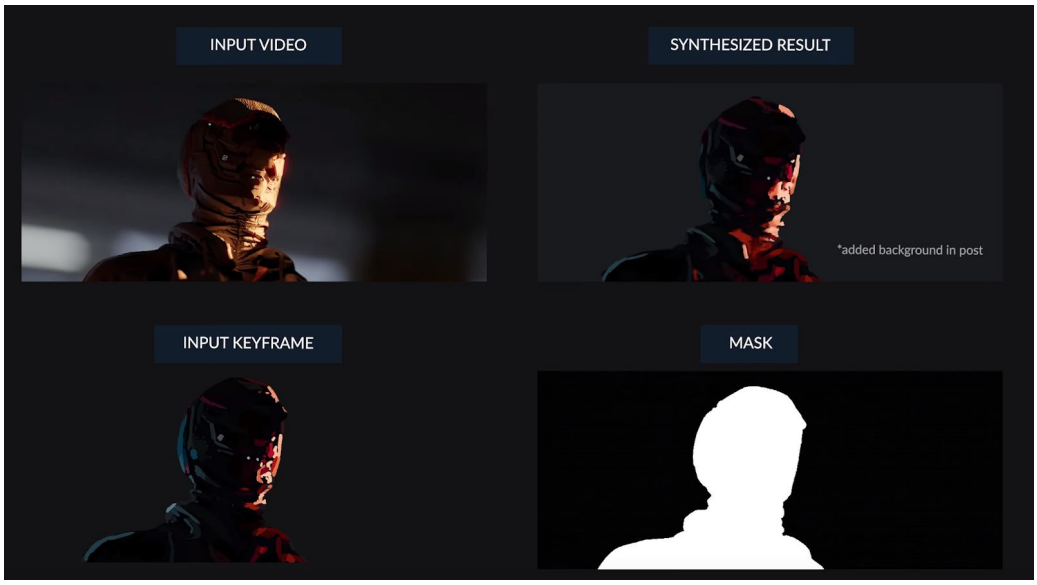
Process:
Firstly, I used movie cutting tool to cut the video include the scene I used before and also in project 3. I got 11-second video which includes two scenes: one is revolving disc to show temporal Hongkong was fast-paced city to express the short love; and the other is I showed before to show the warmth I talked about before. Firstly, I tried to use python FFmpeg to transfer video to frame, then I found the resolution is bad. The divided frames is 128. Then I used Premerie to divide video to frames. I set to 24fps then I totally got the 266 frames and resolution is also good. I also saved the audio from the original video.
First test: I randomly picked up 2 of 266 frames of two different scenes as keyframes. I put them in Deep Dream Transfer first and put them into Ebsynth. Ebsynth need the frame and keyframe have exactly the same resolution which should have the same length and width. I used photoshop to change each keyframe to the same resolution.  After Ebsynth finished all frame, I put them into Premerie, combined with the audio and produced the caption.
The final video is
https://drive.google.com/file/d/1wCo2f0B2hYhIOOppnTutDa4gLcWkOSaS/view?usp=sharing
The result is not good that too many details lost that the character's face disappeared since he was holding glass before.

Test 2:
This time I put them in style transfer and increase the keyframes to keep the detail. I pick up one keyframe for each 20 frames and so the total keyframes is 13.
The result is
https://drive.google.com/open?id=18HCcVmSTT8VY9zdY4PUOsCvLjPrIILbT
I found the result is also not good. The video looks like not consistent due to too many frames. The generated frames are very different for each keyframes, so when they combine together, the video would change  a lot.
Test3 and Test 4:
I picked up 3 frames after I checked all frames. One is shown all 3 discs, one is shown the man holding the glass and one is shown the whole face of the man. The result looks much better.

Result:
Original video:
https://drive.google.com/file/d/1gyJaZ2K49fw2qia3kkyq94DYDlxWLV0v/view?usp=sharing
Test 3 Deep Dream:
https://drive.google.com/file/d/1vAeVPPLNpkl-RMFHOhaj3KR0Ik7b3HzW/view?usp=sharing
Test 4 Style transfer:
https://drive.google.com/file/d/1UK9WICYDjbUX7H7aq1QlQTxmidVZhj5k/view?usp=sharing

We will see the deep dream video: the whole color is brighter and background is more clear. Also characters' face details are more clear. However, the color difference and shadow is not good especially when the face moving, the color looks like a little bit strange.
The style transfer: the color difference is more intense and also character motion is much more smooth than deep dream. However, it lost more details such as the shape and emotions, but I like this one because it would be more close to the Art Nouveau which showed natural forms such as the sinuous curves of plants and flowers to express the warmth. Therefore, I picked up Test 4 as my final result.

Reflection:
For this project, we learnt two generated approach of machine learning and 3 improvement for style transfer video generation. I also learnt art tools Premerie which is very useful in my future video cutting.
For the future improvement, there is also one more thing which could use in Ebsynth: mask.




Since the background of my video is very complex, I don't have enough time to edit each frame to the mask like black and white. I think if we do so, the environment will be more stale and the motion of character will be more clear.

Code: See https://github.com/ucsd-ml-arts/ml-art-final-zizhen_wang

Result:
https://drive.google.com/file/d/1UK9WICYDjbUX7H7aq1QlQTxmidVZhj5k/view?usp=sharing

Reference

[1]https://en.wikipedia.org/wiki/Loving_Vincent
[2]https://en.wikipedia.org/wiki/Art_Nouveau
[3]https://www.imdb.com/title/tt0109424/plotsummary
[4]https://medium.com/tensorflow/neural-style-transfer-creating-art-with-deep-learning-using-tf-keras-and-eager-execution-7d541ac31398
[5]https://keras.io/applications/#vgg19
[6]https://gombru.github.io/2019/01/14/miro_styletransfer_deepdream/
[7]https://medium.com/element-ai-research-lab/stabilizing-neural-style-transfer-for-video-62675e203e42
[8]https://zeruniverse.github.io/fast-artistic-videos/
[9]https://towardsdatascience.com/real-time-video-neural-style-transfer-9f6f84590832
[10]https://dcgi.fel.cvut.cz/home/sykorad/ebsynth.html
[11]https://github.com/jamriska/ebsynth