ART AND MACHINE LEARNING
CMU 2022 SPRING
PROJECT 2

Connecting Different Perspectives

Group 25

Winfred Wang
Undergraduate in BXA Program (Computer Science and Music Performance)
No ML background, First project involving ML

DESCRIPTION

Used Style Transfer to generate opposite perspectives of music concerts in order to help people understand the views of those that are different from them especially during controversial events or news.

Concept

My project was created on my curiosity to understand a controversial event that occured around the end of last year. The event revolved around a prominent rapper named Travis Scott who made headlines for organizing a concert in his hometown of Houston for his touring 'Astroworld' album. On November 5, due to a lack of concert security and structure, tons of fans rushed the concert floor area, knocking down gates and trampling security guards as well as each other in order to get into the concert for free. As a result, according to the Wall Street Journal, at least 10 people died and dozens of others were severely injured primarily from the crowd surge, including trampling and stepping on each other. The artist and his organizers were heavily scrutinized for their lack of foresight in improving security as well as responding to the audience yelling to stop the show in order to help people being literally buried by others. Instead, Travis just continued performing. Despite his apologies, this event caused the public to reconsider the atmosphere of concerts and festivals where it is a strategy to create very dense crowds in order to have the best experience. The news also mentioned controversial theories that Travis Scott meant for the incident to occur that way or that his concerts are designed to look like hell or some kind of dangerous environment. This view was heavily biased between those who knew Travis Scott's music and aesthetic and those who didn't. Thus, my project attempts to use style transfer to help each side understand the other by drawing similarities of the real event to having more apparent dark and ominous textures/style as well as the complete opposite, where the concerts don't seem to pose any danger at all.

Technique

In order to carry out this concept, my technique was primarily focused on finding the correct type of image pair and then figuring out which one of the machine learning algorithms we've learned would give me the best results. For one side of this pair, I needed photos of specific concerts that mainly captured what made that specific event unique like the ambience, color, and vibe or how the concert stages and audience areas are laid out. This way we could also see if different concerts that host different niches of people or genres of music have different balances of both dark and light styles. For the second side, I needed to find reference photos that would contain the styles of the two sides of people that I am trying to connect, including textures, colors, and other details of the image that could possibly be seen as opposites. For example, what I eventually used as my two reference images was one that was only orange and black with rusty textures, rough terrain, and depicted jagged mountains. The other was a photo of nature with lots of blue and green with a smoother texture and containing a big field of grass and flowers. These ideas could only be created by making a custom dataset where I searched and picked out certain concert photos and reference images and tested many combinations of them to see which one would be best. One strategy I had for the reference or style images was to look for more open areas like landscape photos or also photos of abstract art that elicited certain emotions.

Process

In retrospect, one of the hardest obstacles of this project was starting out. The process of trial and error in creating a viable dataset with image pairs was very tedious. I tried searching for datasets online of music concerts but nothing really had what I was looking for.

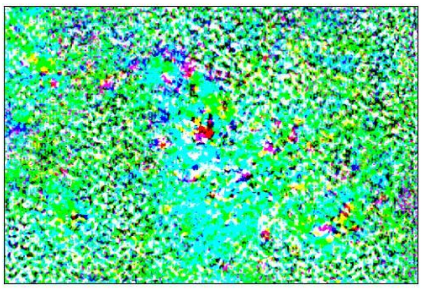
I also started out believing that CycleGAN would be the best algorithm as I was very inspired by the example in the lecture slides of converting human faces into ramen that my image pairs would work under the algorithm. But I was unexpecting how much preprocessing I needed in order for my custom dataset to work. I eventually was able to test a couple pairs but they all did not turn out correctly at all. The result just looked like a blurrier version of the input image. I tried looking for more details about the 'person2ramen' project to help me understand how to adjust the algorithm but it was not successful. The realization was that my image pairs were simply way too different from each other and that CycleGAN primarily worked well on pairs that share a lot of features, so it was best to try out other algorithm options.

I settled for Style Transfer as the few result photos I tested were the closest. Before I settled on the nature image to act as the 'opposite' of the dark image, I tried testing more ethereal looking photos such as a sprawl of clouds or ocean waves thinking it would also potentially serve as a good opposing texture. However, the results didn't really come as clear as I wanted them to and it was too difficult to pinpoint the components of the original image as shown here:

I then began testing one control concert image with both the dark and nature references with different weight values. I saved them under 'results/test_content_weights' (folder also includes more tests with the cloud style image) where I kept different weight-valued results to understand the difference and see which integers would work the best. I realized from my tests that the higher the integer used as the weights value, the less texture and components of the reference style image was transferred to the target content image. Additionally, any value I tested below 0.5 felt like too much of the textures were being brought through. I eventually realized that I needed two different weights for my two different references. The orange-colored dark style image had fewer complex textures to transfer, so a value of 0.85 worked best. For the nature image, however, there were more details and different textures from the grass to the sky and mountain in the background, so I had to dial the translation down (increasing the weight) so that it wouldn't overwhelm the original. So with a 1.35 for nature, I believed these were the best values at balancing the content and style images together so that most of the outlines and people in the concert images could be told apart from the features added from the style image. The

larger the integer for weights also decreased the ability for the algorithm to apply the style image to emptier spaces in the content images. This included skies and the crowds of audiences in the pictures as they sort of blend in to become a big 2D plane of consistent color and texture.

At this point, the results were already looking pretty good. But to fine tune it a bit more, I looked at the STROTSS method's code more in-depth to see which additional parameters or code I could experiment with. I eventually tested different values of the learning rate used. The default was a value of '2e-3' and I also created a folder at 'results/test_code_adjustments' to test different values to examine its behavior. There was no detailed change in lowering to '1e-3' but I observed that raising to '5e-3' or '8e-3' really brought out more details in the textures from the style images. Raising these values even more drastically to '8e-2' or '8e-1' did not help at all as the algorithm eventually wasn't able to learn properly:



Because I wanted more of the grassy texture in the results, I was determined to use a learning rate of '11e-3' for the nature style image and keep the default learning rate for the dark style image as bringing more texture on that style made it blurrier.

After finishing all of these trial and errors, I wrote a script file, main.py, to handle processing multiple image urls at a time since the default source repository from https://github.com/futscdav/strotss could only handle them one by one. So I stored my dataset as a txt file of image urls rather than downloaded images when I was testing the CycleGAN and I separated the urls by using some python string methods.

Reflection

After style transferring all of my picked concert images using the dark and nature style images, I am pretty satisfied with the results. Most of the images had a great balance between incorporating the textures, colors, and components of the style images while retaining the notable shapes and structures of the original concert image. Many examples, as I will include in the next section, had not only very contrasting depictions of these concerts but it was also very convincing. These were determined to be my final versions (stored in 'results/final_dark' and 'results/final_nature') because they offered a perspective onto that specific concert and how someone could have a complete 180 degree view of the same context. Some of the concert's final depictions were only convincing on the dark or the nature side and I believe this is purposeful in understanding that a part of the event design itself contributed to this. It raises some questions about whether some concerts purposefully design their events to look very ominous and fiery like hell to encourage wild and crazy behavior or design to look very friendly, colorful, and positive to encourage more civilized or arguably enjoyable experiences with other event-goers. It may almost be interpreted that these resulting images reveal the hidden motive behind some concerts when it is difficult to uncover them through the original image. Perhaps more trial and error could've solved these one-sided convincing results or trying to style transfer a video of a concert would help reinforce my goals.

RESULT

1. This is from an original image of the Travis Scott concert which caused the horrible events talked about in the concept section. You can see with the dark image the concert looks varyingly evil bearing similarities to hell while the nature image makes the show look more peaceful and friendly.
(https://ca-times.brightspotcdn.com/dims4/default/6e4f32f/2147483647/strip/true/crop/5568x3712+0+0/resize/1486x991!/quality/90/?url=https%3A%2F%2Fcalifornia-times-brightspot.s3.amazonaws.com%2Faa%2F18%2Fe54c32e242b684c943ea0cf6c222%2Fap21310170175809.jpg)

2. This is an image of the entrance of Travis Scott's astroworld festival with his iconic head figure entrance gate. The dark styled image also looks very creepy and like hell while the nature styled one makes the crowd look like a field of grass and flowers. (https://www.billboard.com/wp-content/uploads/media/travis-scott-astroworld-2019-cr-Julian-Bajsel-billboard-1548.jpg?w=1024)

3. This is extracted from a country music festival that contains different colors and concert layouts. (https://www.thenashnews.com/wp-content/uploads/2021/12/Country-Music-Concert.jpg)

4. These were derived from an electronic dance music festival or rave called Tomorrowland. (https://images.prismic.io/tomorrowland/8cc9c6bb-ff08-46eb-ba33-332db6278695_190720-194141-TML2019-KV-5679-HR.jpg?auto=compress,format&rect=0,0,5266,3511&w=480&h=320)

CODE

https://github.com/winfwang/ArtMLProject2

- Contains all project code
- Project's required media files are also here under path 'strotss/results/final_dark' and 'strotss/results/final_nature'
- Project's custom dataset is under path 'strotss/dataset'

REFERENCE

[1] Anthony DeRosa, Neil Shah, December 9, 2021
https://www.wsj.com/articles/what-happened-travis-scott-astroworld-tragedy-11636478633#:~:text=At%20least%2010%20people%20died%20and%20two%20dozen%20others%20were,Houston%20Police%20Department%20is%20investigating