So, I had this general question of how we can use AI to generate art and I came across style transfer, which allows you to generate new art based on reference images. I thought this was an interesting idea and wanted to research this problem more deeply. Style transfer is a way in which two images are supplied: the content image and the style image. The content image is the image that will be modified, and the style image is the image that will have its features extracted and imprinted onto the content image to generate a new image.

Style transfer has been around since 2017, so it isn’t a new concept. But traditional style transfer uses convolutional neural networks which face many problems. The first one is content leak: where over multiple iterations of style transfer the original shape of the image will be completely lost. This is obviously not the desired outcome as we want some traces of the original content image to remain.

The second issue is lack of semantics awareness. This means that the convolutional neural network architecture does not understand the contents of the image. So, this means in repetitive images such as a single photo duplicated four times or grass, the style transfer algorithm will just transfer the shape of the style image instead of comprehending what the objects in the image are.

The final issue is the lack of scale invariability. With CNNs, if the style image and the content image are different resolutions, there will be a lot of random visual artifacts and the image will not be stylized properly. As shown in the image, the non-invariant images all have these strange vertical bars of color that show they haven’t been stylized properly.

For implementation, I first started by rerunning the experiments. I took the paper StyTR-2: Image Style Transfer with transformers as my code base and tested the claims that they solved the three problems with convolutional neural networks. First, all my images were of different resolution so that confirms that the algorithm is scale invariant. Next, I ran an image through the style transfer algorithm 20 times to see if the content was still there and sure enough it was still there, meaning the content leak issue was solved. Finally, I duplicated an image four times and ran it through style transfer and confirmed that each image had the exact same style applied to it, meaning that the style transfer algorithm is semantics aware.

For my extension, I felt there could be other ways to generate interesting images, so I decided to add the option to do style transfer with multiple style images and select the weighting of each image. This will have extracted the features from multiple style images and blend everything together to create a more interesting image. This was done by modifying the encoding and decoding methods in the transformers and changing the way the style transfer model stylized the images by having individual weights of each style image decide the way the features would be extracted.

From this, I noticed that using multiple styles generates an interesting blend of images. From visual inspection, all of the images turned out looking interesting-looking. Something I noticed was that when using a bunch of styles, it’s hard to pick out the specific details of each style image, which is why weighting is very important in picking which features you want to stand out the most. Also, having weightings over 1.0 can generate some interesting looking images, but this comes at the cost of content loss, with the original content image losing a lot of its original features.

Some take home messages from this project is that style transfer is a great way to generate artwork using neural networks, and some interesting images can be generated. While doing my extension, I noticed that there’s a lot of challenges in balancing the content and style and how much content loss I’m willing to put up with. There’s a bunch of potential usages for style transfer, such as colorizing black and white images, converting people into cartoon styles, and visualizing how certain aesthetics would look in interior design. There’s also potential ethical issues. Like with all neural network projects trained on large datasets, there could problems with artists not getting credit for their contributions to the images generated from style transfer, and potential hidden plagiarism if images are taken without consent.