Exploring Image Style Transfer and Combination Using Unified GANs

Ximeng Mao
260770984
School of Computer Science, McGill University
ximeng.mao@mail.mcgill.ca

***Abstract*—Image style transfer model such as Cycle-GAN provides a valid mapping of an image from one style category to the other. However it needs a model for each transfer pair, which makes it sub-optimal in the context of multi-category style transfer. On the other hand, an unified model can be much more cost-efficient in this task as it is capable of generating images of various styles conditioning on the target label. In this project, one such model (StarGAN), who is originally developed for the facial expression transfer and synthesis task, is explored on transfer and combination of images of different artistic styles. After several experiments, we found that: firstly this model is capable of generating reasonable images among multiple categories; secondly with slight alternations on the architectural design, we can further improve its generated results in style combination. It is in our conclusion that although the quality of generated images is not optimal, this model is still a viable choice for multi-category artistic style transfer and combination.**

I. Introduction

Image style transfer is about generating images that appear with a specific style, for example, transferring a modern natural landscape photo to a painting supposedly done by Van Gogh. When it is firstly introduced in [1], the goal is to generate image to match the style of a style target image and content of a content target image, where a style transfer loss, constructed by Gram matrix of hidden layer representation, is estimated to evaluate how good it does in term of style matching. Later in 2016, a conditional Generative Adversarial Network (GAN) [18] approach named CycleGAN [2] is proposed for translation between unpaired group of images, where the mapping is not from the input to one specific image but to another category of images as a whole. The model produces impressive results in terms of artistic style transfer. Unlike the earlier works, CycleGAN doesnt adopt style transfer loss, instead they used the combination of GAN loss, reconstruction loss and cycle consistency loss to achieve the goal.

But when it comes to style transfer among multiple categories, models like CycleGAN are not optimial as the they are learning a one-to-one relationship between two categories and the number of training needed scales heavily with the number of categories involved in the task. On the contrary, an unified model can gain us significantly because of the cost-efficiency aspect of the translation.

As an unified model, StarGAN [3] becomes a potentially competitive candidate. However, it is originally developed to perform with face and facial expression dataset, therefore in this project, we are going to investigate its performance on the topic of artistic image style transfer among multiple categories. Furthermore, because StarGAN is using binary target label to condition the translation domain, it is inherently convenient for us to experiment on style combination, where the goal is to generate the images that exhibits the characteristics of multiple styles instead of one. The challenge of style combination is the lack of real, or should I say valid, images with multiple styles, so we need to proceed it with caution so as to not generate arbitrary results. The basic idea behind our methodology is that the image with multiple styles should be at least be regarded as all the individual style at the same time. In the scope of the project, we are attempting to generate the image with double styles.

The rest of this report is constructed as follows: We will start with describing the images of various categories we collected for this project in Section 2; Then we are going to give a general introduction on the StarGAN model and the change we applies onto its network architecture in order to accommodate the task of style combination, in the following two Sections; we will provide the discussion on the results in Section 5 and an analysis on Gram Matrix of each input images in Section 6; In the last two sections, we will conclude our project and point out some of the potential future extension of this work.

II. Data Preparation

To comprehensively investigate the performance of StarGAN model on artistic style transfer, 7 categories of images are collected as transfer candidates, which are, in the labelling order: Chinese ink paintings (later referred as Ink), paintings by Morandi (referred as Morandi), natural photos (referred as Nature), black and white photos from last century (referred as OldPhoto), paintings by Picasso (referred as Picasso), paintings by Raphael (referred as Raphael), and paintings by Van Gogh (referred as VanGogh). These images are collected from various online resources as described below: Nature and VanGogh are from the online repository provided by the authors of CycleGAN [4]; Morandi, Picasso, and Raphael are from the combination of datasets of Kaggle "Paintings by Numbers" competition [5] and images downloaded from Google Images search by method introduced in this blog [6]; Ink is from the combination of the same Kaggle competition, Google Images search, and the online repository provided by the authors of DualGAN [7], [8]; OldPhoto is from the

combination of Google Images search and digital collections of New York Public Library [9].

Due to data limitations, the number of training images in each genre is imbalanced, with Nature the largest (around 6200); Ink, Picasso and Van Gogh in the middle (around 400 - 600 each); Morandi, Raphael, and OldPhoto the smallest (around 100 - 200 each). Note that since StarGAN includes the procedure to compute cycle consistency, there are two training per each input image: One forward transfer where the target category is chosen randomly and the other backward transfer to translate the image back to its original category. So although the dataset is largely imbalanced, the actual times of training each category are much more flatter than it seems. Other reasons we didn't downsampling the images in Nature is that we are more interesting in the style transfer and combination coming from the natural photos, for it is in general easier for us to evaluate the generated single, and double styles, results by perception from some domain we are familiar with, such as natural photos. However, it must be admitted that the imbalance is stilla disadvantage as it might result in the under-training of Discriminator for categories other than Nature. But luckily, as we will discuss in the result section, it opens up some observation on relatedness between styles.

As the size of the images is originally different, pre-processing steps such as random flip, random crop, resize and normalize are applied to ensure the all the inputs are of size 256*256.

III. Related Background: StarGAN

StarGAN is proposed as an unified GAN solution for Multi-domiain image-to-image translation, with a outstanding performance on face dataset CelebA and facial expression dataset RaFD. StarGAN incorporates a target label into the input of the Generator so that it can reuse the same model for generating images of multiple domains therefore improve significantly in terms of translation cost in multiple domains, compared with previous successful model such as CycleGAN. The Generator of StarGAN is a convolutional Neural Network, associated with several Residual blocks [14] and Instance Normalization [15]. The images will firstly going through a few convolutional layers with strides for downsampling and finally some transpose convolutional layers for upsampling and reconstructing the same dimension as the input. The Discriminator of the StarGAN is derived from PatchGAN [2], [16], where it output patch by patch decisions on whether or not this is a real image. The Discriminator is also responsible of classifying each image into one specific domain as a multi-class classification problem. Its loss function includes GAN loss, classification loss, and a cycle-consistency loss to penalize deviation on the reconstructed input image from the output image. Its final version adopts Wasserstein GAN objective function with gradient penalty [17]. The composition of loss of both Generator and Discriminator is shown as follows:

$$L_{Disc} = -L_{GAN} + \lambda_{cls} L_{cls}^{real} \tag{1}$$

$$L_{Gene} = L_{GAN} + \lambda_{cls} L_{cls}^{fake} + \lambda_{rec} L_{rec} \tag{2}$$

Other than being an unified model, StarGAN has features of incorporating data from different dataset by utilizing a mask vector, but since it is not within the goal of our project, we didn't involve it in our experiment.

IV. Architectural Change Upon for Style Combination

The problem with style combination is that there are no ground truth about how, for example, a combined Raphael and Morandi could be. However, the characteristic of the combined images should be partially deducible from that of each of the styles in the combinations. For example, it is natural to think that a combined image of Raphael and Morandi should be determined as a Raphael and a Morandi at the same time. Moreover, it is natural to base the model for combination on the well trained one from transfer task (referred later as Base model). In the attempts to explore style combination, three ways are tried for extending the base model to generate image with double styles:

- Without any training, run the trained generator in test mode with labels indicating double styles (e.g. [0,0,0,0,1,0,1]).
- Without any alternation, retrain the whole model directly conditioning on a random mix of single and double style target labels, with slower learning rate (10 times slower).
- Fix both the Generator and Discriminator networks and train a Linear Combination network in between to learn the weights to perform a linear combination from different single style image that supposed to be combine together.

The results from the first two will be discussed in the next sections, and this section will be focusing on describing the architectural changes on the StarGAN model for the third attempts.

The motivation behind adding a separate layer between Generator and Discriminator originates from the summary of the previous two attempts: Although the first one does exhibit a level of variation, it is not quite explainable in the sense of style combination, hence this naive way will be treated like a baseline performance; Meanwhile, the generated results of the second one indicates that even with reduced learning rate, the weight update in the network (specially the Generator) is detrimental, as even the single style image generations are degenerating. This could relate to the lack of ground truth, thus avoided afterwards.

Then for the third attempt, aside from detaching both the Generator and Discriminator from training, a Linear Combination layer (LC-layer) is added, in order to linearly and pixel-wisely (row-wisely to be precise) combine two generated images. The weight utilized in the linear combination is the outputs of the combined target labels through a Feedforward Neural Network (FNN), and its parameters are trained by backpropagating the Discriminator's loss with respect to the combined images and the combined target labels. It is designed

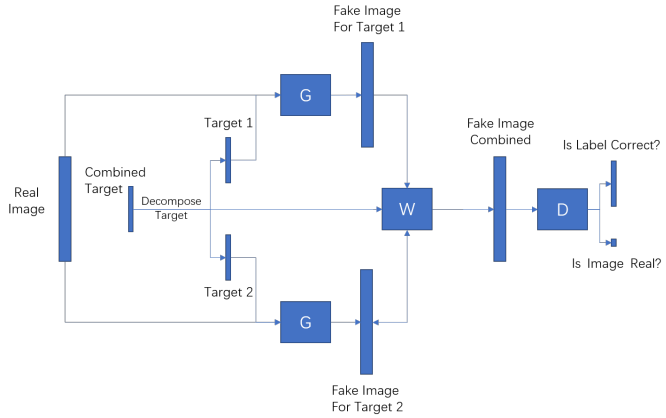

Fig. 1. Network Architecture in the 3rd Attempt

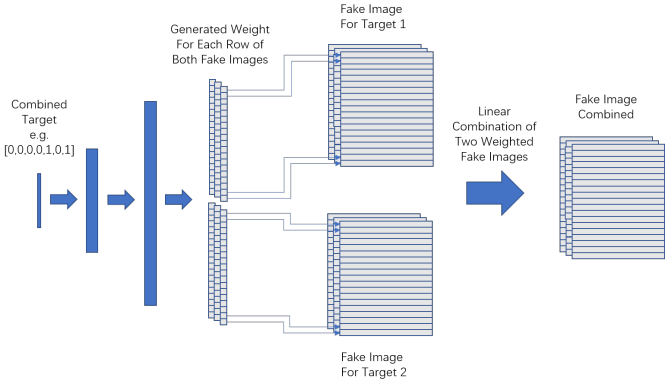

Fig. 2. W Network that Generates Coefficients for Linear Combination

under the principle of both fixing the results in single style transfer and introducing reasonable variations in double style combination. The idea behind this approach is coming from the original paper of style transfer [1], where VGG [10] is used for extracting features and training is done on the pixel of the input image until the input is matching style and content images in their respective way.

After adding LC-layer and FNN, the whole network architecture is shown in Fig. 1, Note that network denoted as W includes both of the two parts, and will be referred to as W in the following. For each combined target label, it will be decomposed into two one-hot encodings, which then served as conditional inputs for two separate style transfer to generate two outputs. In the following, W will take the two generated images along with the double-style target label, and generate the combined image. A more detailed demonstration of W is shown in Fig. 2. Note that in the implementation, the target label is a 7-dimensional binary vector with exactly two 1s; The FNN is comprised of two full-connected linear layers of size (7, 512) and (512, 1536) respectively, where 1536 equals to 2*(number of channels)*(image size in one axis),with non-linear activation leak ReLU and Sigmoid after each layer ; The 1536-dimensional vector is then rewritten as two [3, 256] 2-dimensional matrices, and utilized as the coefficients associated to every row of the two generated images in the LC-layer to form one final combined image. The loss to train the parameters of W is a subset of the loss originally in StarGAN when training the Generator:

$$L_{LC} = L_{GAN} + \lambda_{cls} L_{cls}^{fake} \tag{3}$$

For the training, Adam [11] is chosen as the optimizer with initial learning rate as 0.00001, and all the implementations on the additional features are done in PyTorch [12]. Code is made publicly available in https://github.com/ximmao/MLFinalPrj

V. Results and Discussion

A. Style Transfer

The training is done with the default configuration of StarGAN, including the learning rate initially as 0.0001 for

Fig. 3. Test Samples of Style Transfer

Generator and Discriminator, 64 convolutional filters in the first layer of the both network, and batch size of 16. As mentioned above, input images are cropped and resized as 256 * 256 and the dimension of conditional input is 7, corresponding to 7 candidate categories. The training steps in StarGAN implementation are the number of batches trained in total, and we are observing the sample generation outputs jointly with model loss to determine when to stop the training. In addition, two more flags are added to accommodate our style combination exploration: ”label combine list” has nothing to do with the training itself, it creates fixed double style target label to show the style combination results when generating the samples during training and generating the test results;

Fig. 4. Good Transfer Results

"train combined label" is a boolean flag to alternate between style transfer and style combination mode.

For style transfer task, the model is trained with "train combined label" as "False". In Fig. 3, we show the testing results of the model at 146000 steps. In the figure, the leftmost column shows the input images in the test set, and consecutive 7 columns are the generations for each of the following category: Ink, Morandi, Nature, OldPhoto, Picasso, Raphael, and VanGogh from left to right. Note that both the loss of Generator and Discriminator stabilized starting from around 95000 steps, with the classification loss of both the real and fake images mostly 0. The total step ran in this part is 200000 steps, and the step chosen above is intuitive, by finding the sample output perceptually best to the eyes.

Before moving on to discussion of results, it is necessary to talk about the evaluation metric for the task. As with other generative applications in computer vision domain, due to the lack of ground truth, it is inherently difficult to evaluate the quality of the generated images. As mentioned by author of CycleGAN [13], loss curve is not a good indicator in the GAN model, and the simplest way is actually by observing it and evaluating with common knowledge. But since it is fair to say that personal opinion is very much biased, papers in the field also resort to extrinsic evaluation such as conducting an online survey and computing the classification loss via separately trained Neural Network. However, due to the time constraint, we are going to analyze the generated images purely via perception.

Back to the testing results, one straightforward observation in Fig. 3 is that they mixes good generations with bad ones. Fig. 4 shows some of the good transferring from the results. While it is common knowledge in GAN based generative model, it demonstrates from another perspective that none of the models is not truly intelligently knowing the logic to generate the images, and maybe it is more suitable to say that the model just happens to be right than to be wrong. In our case specifically, it seems that the model is learning a recoloring pattern on top of each input image. For example, since most of the training images of Morandi is still-object painting with the signature beige-like color in the background, the model seems to learn both the beige like mask and a frame-like artifact on the generated image, while in the case of OldPhoto, it becomes a threshold mask to put everything extremely black and white. But since the task is unpaired style transfer between various categories rather than various images, the result is probably reasonable in the sense as the Generator needs to learn a universal characteristic that appears in each and every image in the category to more easily fool the Discriminator.

We believe that this observation can lead to a potential caveat in doing similar task in the future, as not only the count of images matters, the uniformity in the genre (landscape, portrait or still-object) of images may contribute to the overall performance as well. When it comes to optimize the performance and lower the failure rate, maybe it is beneficial, for example, to keep all the images in the dataset as landscape photos or paintings. Moreover, it is likely that there is similarity between various categories, for example, between Nature and OldPhoto, since the structure is basically the same with just the color difference, the transfer between these two becomes a easy task, even if there are smallest number of images in OldPhoto. Finally, we should be more careful about the choice of categories involving in the training, as certain category can be essentially harder to learn in the current setup, such as the style of Picasso, who is more likely a concept difference rather than color difference, as well as the style of Dali, who I thought about putting in but later avoided. On the contrary, the style of VanGogh (or Monet and other impressionist) is more resembling to reality hence could result in better generations from the model.

B. Style Combination

For style combination task, training is based on the trained model from transfer task (referred later as original trained model), and the "train combined label" will be set to "True". The "label combine list" is chosen arbitrarily during training.

1) With Original Trained Model: The combination results using just the original trained model are shown in Fig. 5 denoting as "without LC". By definition, with the only difference

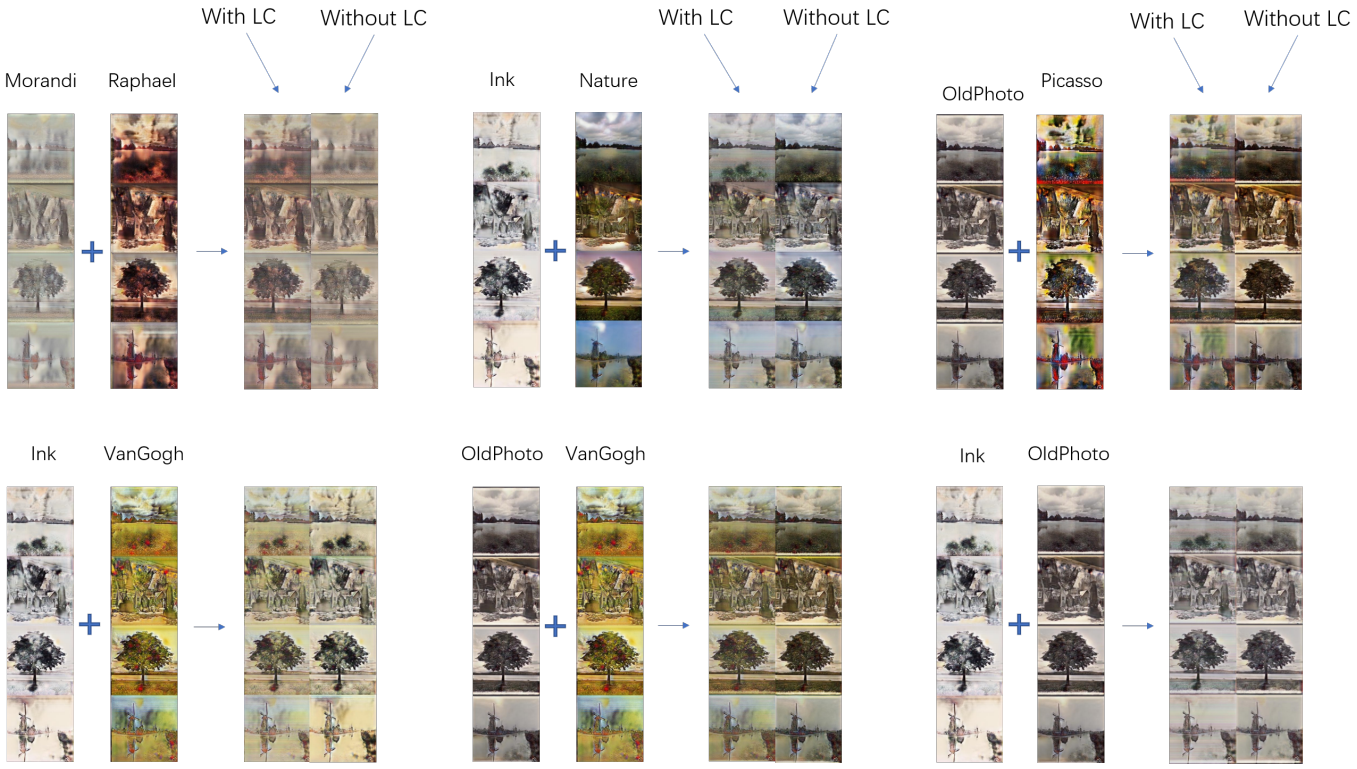

Fig. 5. Comparison of Linear Combination and Original Model

being in the conditional inputs from one-hot vector to a vector with two 1s, the final combined results should be essentially equal to the pixel-wise addition of two generated single style images, each corresponds to one of the 7 categories in the dataset, then applying squashing, shifting and cropping to end up between 0 and 1, so as to form a valid image. This would be more reasonable in the case of facial dataset like CelebA in StarGAN, where label usually denotes more locally affecting features such as yellow hairs or brown eyes. While the changes in transferring artistic style occur more globally in the image, that we need to alter value of almost every pixel so that it can appear as a whole a different category. In this case, saturation effect of nonlinear squashing could make the combined images ending up with one style dominating over the other.

Some of the results could back up this claim: For example, all the combined images from Morandi and Raphael exhibit more Morandi (with the beige color over the entire image) between the two, and Picasso is nearly unnoticeable through its combination with OldPhoto. Nevertheless, the generations are still perceptually valid and acceptable, so in the following this will be regarded as the baseline when we search for potentially better solution, and it will be referred as ”original model”.

2) With Directly Retrained Model: It is clear that some retraining will be involved if we want to further improve the above results, but the reasonable thing to do is to keep the deviations within a small range, as no ground truth for these combined style images can support the more radical changes. Naturally for the retraining, we will follow the minimum change principle and prefer re-balancing how the two images combine rather than recreating a brand new image.

The most simple and straightforward way to bring in the variation would be to retrain the entire network, but this time with combined target label. Under the principle of minimum change, the idea is to slow down the retraining process and to obtain conservative and controllable variation each step, hence the 10 times reduced learning rate. But it still turns out to be a failed attempt, as shown in Fig. 6, where the leftmost column is the real image, consecutive 7 columns as 7 single styles transfer and last 5 columns as some of the style combination (from Left to right, combined Morandi and Raphael, combined Ink and Nature, combined Ink and VanGogh, combined OldPhoto and VanGogh, and combined Ink and OldPhoto). Although a more observable changes are brought onto the combined image, it comes with the price of performance degeneration even on the single style generations. For example, the generated Ink and Morandi images becomes more abstract and losing a great amount of details, the Picasso and Raphael images is totally corrupted with the various round color stains, and the rest also suffers from observable degradation to various extents. This phenomenon is, in turn, extended onto the generated combined images.

This is a predictable failure, as we impose too much on Discriminator’s capability of correctly inferring a unknown distribution, and without the real distribution for the combined image, the training becomes a procedure where instead of trying match the true distribution, Generator just actively looks

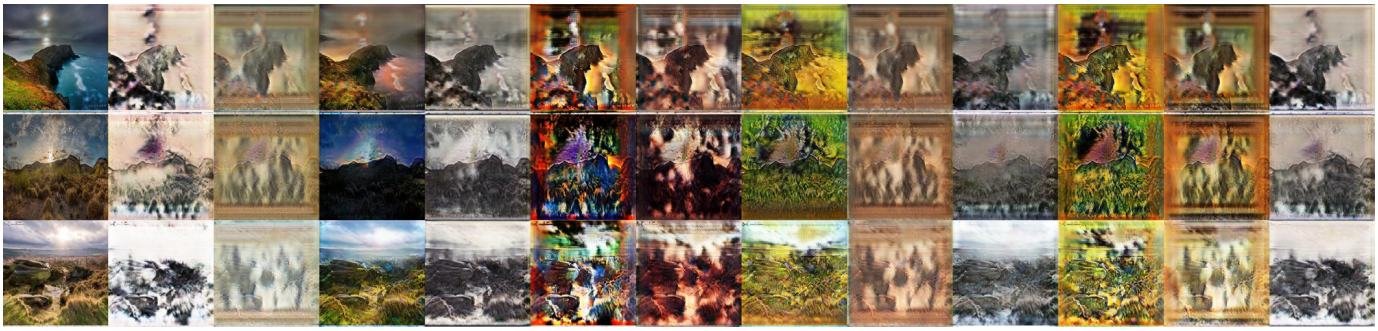

Fig. 6. Compromised Generation from Second Attempt

for the loophole of Discriminator's decision procedure so that it can trick the Discriminator into thinking the wrong way. This is easy for Generator to do that as it has access to its adversary's information through gradient, but it is by no mean a good generation. This would be a form of model collapse where Generator is eventually doing the task poorly.

3) With Trained Additional Linear Combination Layers: So in the third attempt, we are going to jump out of the GAN settings, and try to deal with the task in different perspective: with W network and LC-layer. Note that in this approach, we still rely on the fact that the trained Discriminator has some knowledge of what a combined image would be by inferring from single style generations, but since the generation of single style image is completely fixed (by fixing the Generator), it is more fitting in supporting the claim that a double style image is both the two styles at the same time. Moreover, by concatenating a Sigmoid layer at the output, we prevent the image from changing arbitrarily and guarantee that the generated image is visually acceptable.

W network is trained in total for 40000 steps, and in Fig. 5 denoting as "with LC" is results using the model from 150000 steps. Fig. 7 plots the loss for each steps throughout the training, and it shows that even with constraint on the coefficient, the training can finally reduce the loss to a lower level. However, smaller loss doesn't necessarily mean that the combined image is getting better. In fact, as the samples of training results from 170000 steps shown in Fig. 8, the generated images has obvious horizontal color stripes, so it is very likely that the network W just find a way to fine-tuning its coefficients so that the Discriminator is fooled into thinking it is getting a better results. Moreover, even the one in 150000 steps has similar phenomenon though much less obvious. Therefore, it becomes an issue similar to overfitting in supervised learning, and we can early-stop the training to fight against it. In our case, the model from 150000 steps is chosen perceptually.

We can observe the difference between the results from this attempt and the one from original model. By explicitly modeling the linear combination, it is possible to restore the balance and make the result image more convincing to be considered as in between the two styles. Take style combination of Morandi and Raphael as an example, the image generated

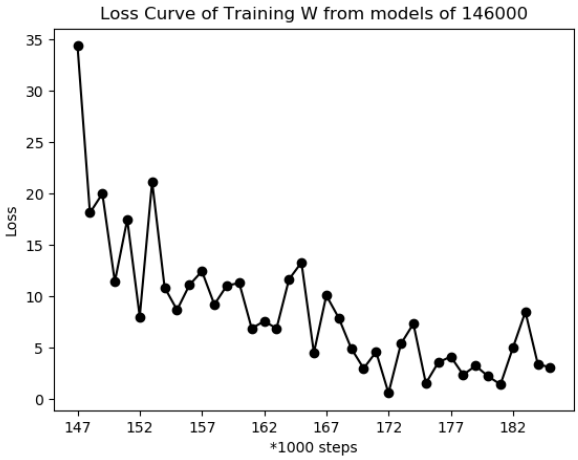

Fig. 7. Procedure of Training Linear Combination Coefficients

Fig. 8. Failed Combination from 170000 steps

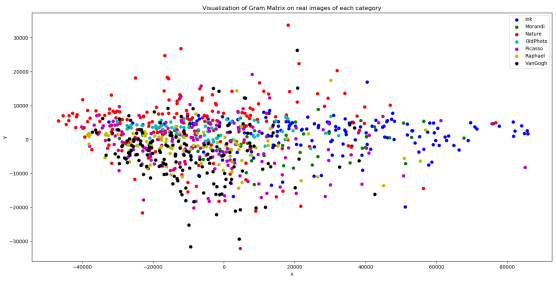

Fig. 9. Gram Matrix projected 2-D space for real images

from the linear combination has not only the beige color from Morandi, but also the darker brown from Raphael; Another example is the combination of Picasso and OldPhoto, where we can see some bright colors on the whole black and white setup. Note that there is still the case where the difference between the two approaches is almost unnoticeable, such as the case combining OldPhoto and Ink. That is mainly because both the case has similar color components. In conclusion, although fairly subjective, in terms of keeping the balance of style combination, it is in our personal opinion that linear combination is doing a better job than original model.

VI. Analysis of Gram Matrix on Real Image of Each Category

The last part is an analysis on the Gram Matrix of each input image. Gram Matrix loss is an essential part in original style transfer paper [1], serving as style transfer loss and indicating how close is the generated image to the target image in terms of style, and it is computed via calculating the Gram Matrix of feature maps of various layer from a well-trained deep neural networks. However, it is not included in the later GAN based approaches in the task of unpaired style transfer, and we are curious of to why it is the case. Here we are just going to do a simplified version: the Gram Matrix of directly the real image, with the intention of to investigate its potential effectiveness in our problem. In detail, we reshaped each channel (R, G, B) to be a 256*256 dimensional row vector, respectively, and calculated their inner product. It results in a 3*3 Gram Matrix, where each element of the matrix can be defined as

$$G_{ij} =< X[i,:,:]^{Vec}, X[j,:,:]^{Vec} > \tag{4}$$

Due to the symmetric characteristic of this matrix, out of the 9 elements there are 6 unique variables. We then apply Principle Component Analysis (PCA) (implementation of sklearn) to visualize it on a 2-D plane. Due to memory and speed consideration, the process is done only on a random subset of all the training images of each category, and the visualization result is shown in Fig. 9.

It can be seen from the figure that although we wished to see a clear distinction between each category, the region of images in each category actually overlaps largely, therefore, it is likely that the style defined by the Gram Matrix is more specific to each individual image than to each category. Hence, it would be hard to incorporate it effectively into our task. In addition, while the distribution of VanGogh and Raphael is more compact, that of Picasso is more sparse, and it may due to the fruitful variations within Picasso's paintings. Another thing worth noticing is that Ink seems to be forming around x-axis and more extensively distributed compared with other categories, it may also imply some uniqueness of Chinese ink painting as a whole.

VII. Conclusion

In this project, we investigate the capability of an unified model StarGAN for the task of artistic style transfer and combination among multiple categories. For style transfer, it proves its advantage as a cost-efficient model to generate reasonable results, and its performance seems to relate with genres of images in the data set and similarity between different category. We need to be cautious about what images should be involved in the training set instead of feeding any images that can be found, in the attempts of optimizing the translation performance. For style combination, adding a separate Neural Network to learn the coefficient of Linear Combination between images could yield perceptually better combination results than directly feeding the model with combined target label. By explicitly modeling the weights of each image in the combined picture, we can obtain a more balanced combination and not let one style dominating the whole. At last, we found that the style defined by Gram Matrix is more likely to be specific to each image rather than to each category, so it won't be much useful in the case of transferring unpaired group of images.

VIII. Future Works

One of the future Work would be trying with better selected dataset, and to see if there is boost onto the generative performance. For example, we can train the model with more balanced dataset; we can also be more cautious about the genre of the images in the dataset, such as selecting all landscape images instead of portrait or still object paintings; we may as well optimize the choice of category, like by choosing painter with more consistent styles. Another potential direction is with the style combination, in order to impose more constraints onto the training of linear combination, it may be beneficial if we can find better images with combined style to serve as reasonable ground truth. We could try density estimation based approaches, by finding the probability distribution associated to each style and generating the next pixel by sampling from the two distributions. It may be costly, but it is a favorable trade-off as it would be very helpful in training a better linear combination coefficients.

IX. Acknowledgments

The author would like to thank the authors of StarGAN by making their complete implementations available online and all the providers of the images involved in this project, and thank our two instructors for the wonderful courses and all the

valuable opinions and discussions in the poster representation session in better forming the report. Special thank to Manxi Guo for the help in both the discussion on the painter style and the effort of sticking the poster together.

REFERENCES

[1] L. A. Gatys, A. S. Ecker, M. Bethge; Image Style Transfer Using Convolutional Neural Networks. The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 2414-2423.
[2] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros. Unpaired image to-image translation using cycle-consistent adversarial networks. IEEE International Conference on Computer Vision (ICCV), 2017.
[3] Y. Choi, M. Choi, M. Kim, J. W. Ha, S. Kim, J. Choo; Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018.
[4] https://github.com/junyanz/pytorch-CycleGAN-and-pix2pix
[5] https://www.kaggle.com/c/painter-by-numbers/data
[6] https://32hertz.blogspot.com/2015/03/download-all-images-from-google-search.html
[7] Z. L. Yi, H. Zhang, P. Tan, M. L. Gong; DualGAN: Unsupervised Dual Learning for Image-to-Image Translation. IEEE International Conference on Computer Vision (ICCV), 2017, pp. 2868-2876.
[8] https://github.com/duxingren14/DualGAN
[9] https://digitalcollections.nypl.org
[10] K. Simonyan, A. Zisserman; Very Deep Convolutional Networks for Large-Scale Image Recognition. CoRR, 2014, abs/1409.1556.
[11] D. Kingma, J. Ba; Adam: A Method for Stochastic Optimization. International Conference for Learning Representations (ICLR), 2014.
[12] https://pytorch.org/
[13] https://github.com/junyanz/pytorch-CycleGAN-and-pix2pix/blob/master/docs/tips.md
[14] K. He, X. Zhang, S. Ren, and J. Sun; Deep residual learning for image recognition. The IEEE conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770-778.
[15] D. Ulyanov, A. Vedaldi, and V. Lempitsky; Instance normalization: The missing ingredient for fast stylization. arXiv preprint arXiv:1607.08022, 2016.
[16] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros; Image-to-image translation with conditional adversarial networks. The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.
[17] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville; Improved training of wasserstein gans. NIPS, 2017
[18] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio; Generative Adversarial Nets. Advances in Neural Information Processing Systems, 2014, pp. 2672–2680.

APPENDIX A

More style transfer and combination results from input images of various categories are provided. Note that the each column represents, from left to right: input, Ink, Morandi, Nature, OldPhoto, Picasso, Raphael, and VanGogh, combined Morandi and Raphael, combined Ink and Nature, combined Ink and VanGogh, combined OldPhoto and VanGogh, combined Ink and OldPhoto, and combined OldPhoto and Picasso.

Fig. 10. Generation from Chinese Ink Paintings

Fig. 11. Generation from Morandi Paintings

Fig. 12. Generation from Old Photos

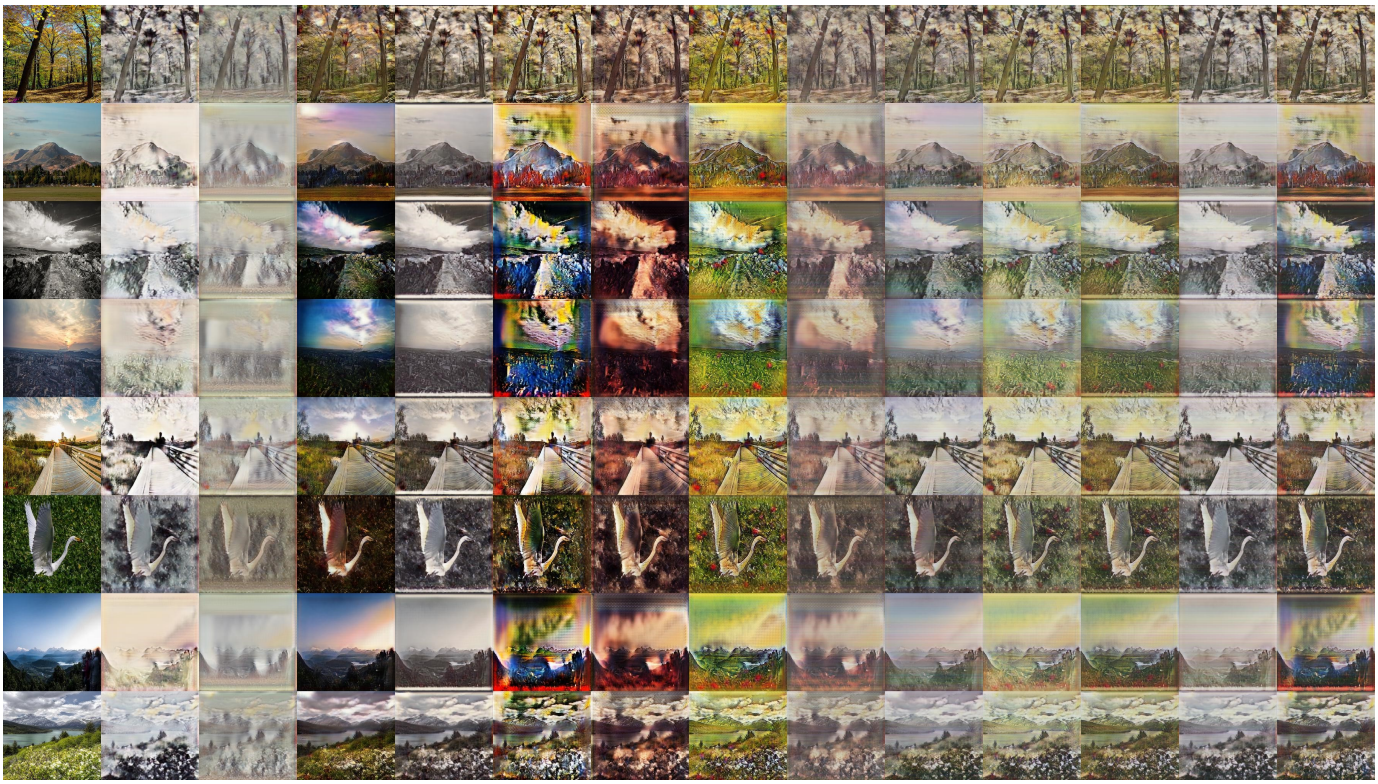

Fig. 13. Generation from Nature Photos

Fig. 14. Failed Case from Nature Photos

Fig. 15. Generation from Raphael Paintings

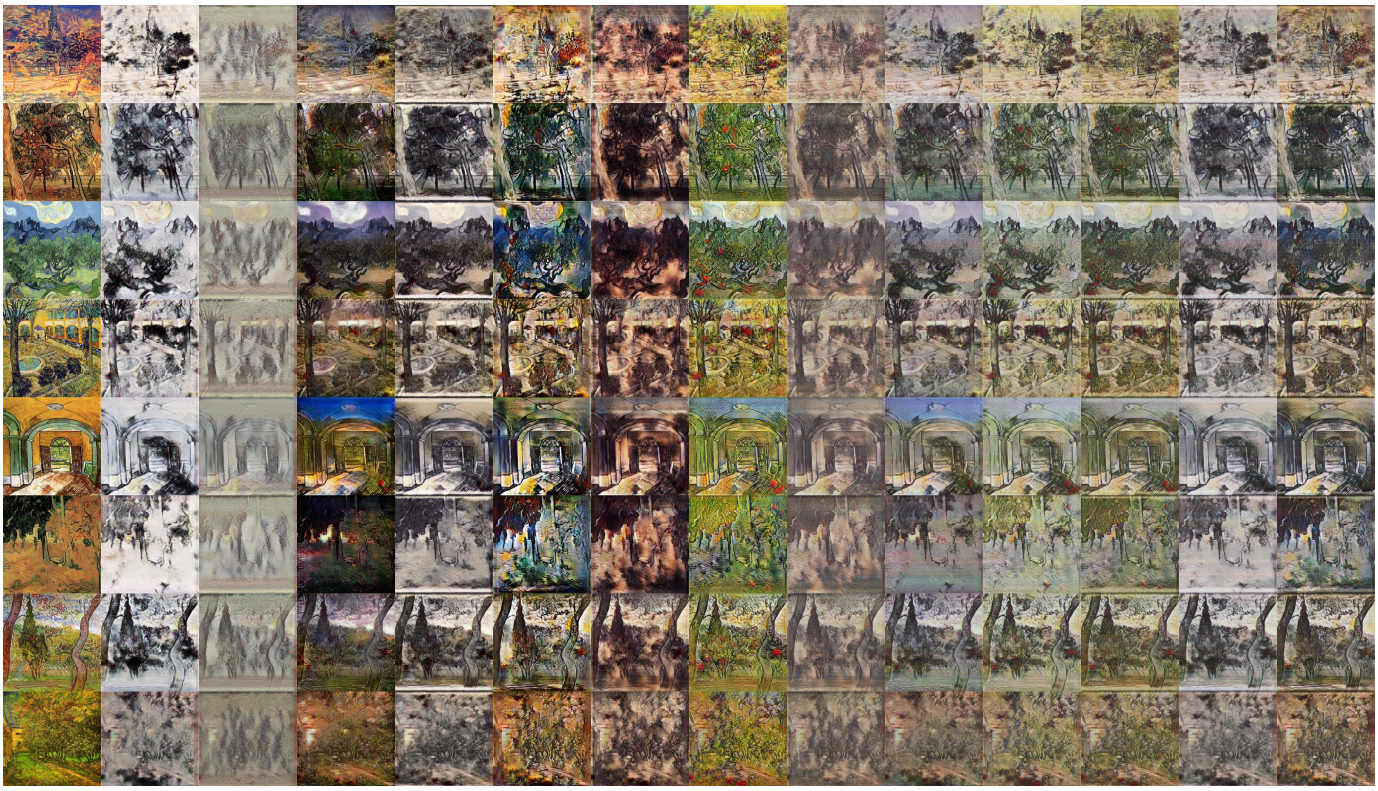

Fig. 16. Generation from Van Gogh Paintings (Landscape)

Fig. 17. Generation from Van Gogh Paintings (Portrait)

Fig. 18. Generation from Picasso Paintings