# Exploring Image Style Transfer and Combination Using Unified GANs

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

## Motivation

- This project is about Artistic Style Transfer & Combination among multiple categories
- Use an unified GAN is more cost-efficient than learning one model for each pair
- StarGAN is such a model originally developed for facial images, with conditional input controlling the labeling of output
- We are exploring StarGAN for this task
- For style combination, goal is image with **double** styles and challenge is lack of ground truth, we will proceed with the assumption that a combined image should be both style 1 & 2 at the same time

## Data Preparation

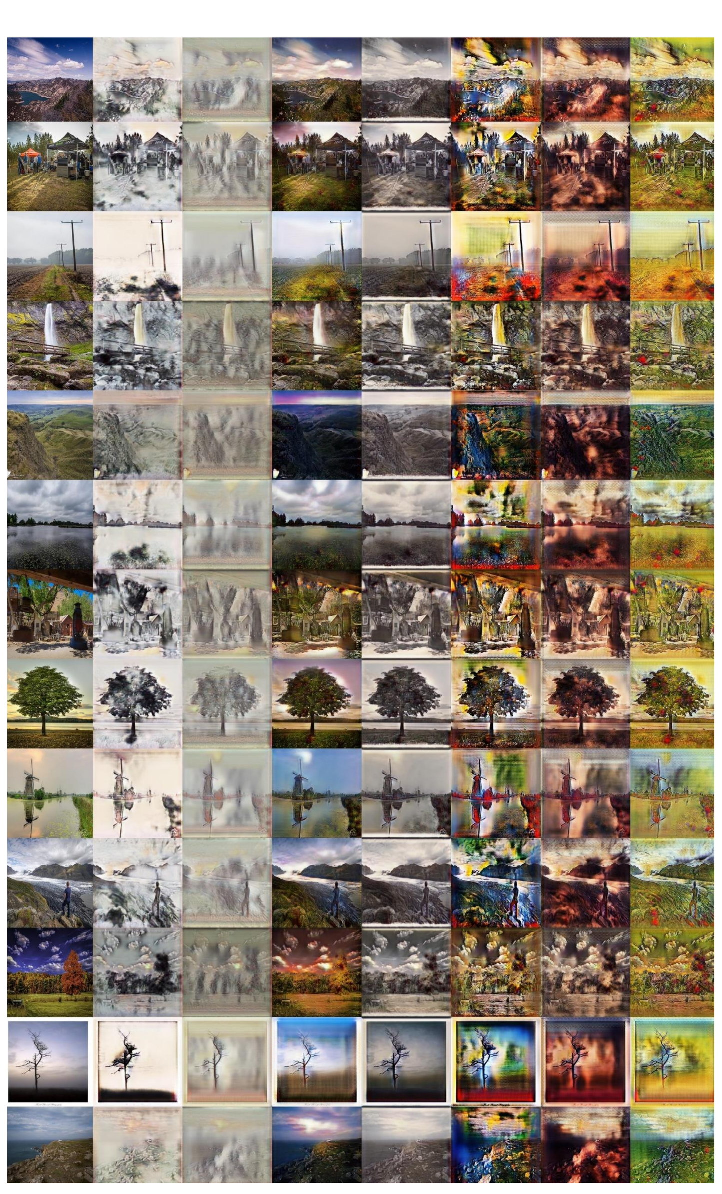
- 7 categories of images collected from image datasets and Google Image search:
  - Chinese Ink Paintings (564 – number of training images)
  - Paintings by Morandi (100)
  - Natural Colored Photos (6287)
  - Black and White Photos from Last Century (124)
  - Paintings by Picasso (348)
  - Paintings by Raphael (186)
  - Paintings by Van Gogh (400)
- Images are gathered as much as possible then selected for style consistency
- Class imbalance exists, but due to cycle consistency mechanism, for every image, two transfer A -> B & B -> A conducted, so the actual training on Generator with each category much flatter than it seems

## Image Style Transfer Using StarGAN

- StarGAN[1] basics
  - Convolutional based Generator with Residual Blocks & Instance Normalization
  - PatchGAN Discriminator
  - Wasserstein GAN objective function with penalty on gradient
- Train with default configurations
  - Adam with initial learning rate 0.0001
  - Batch size 16
  - Run In total 200,000 steps (batches)
  - Learning rate decay starting from 100,000
- Outputs shown from 146,000 steps
- Model learns a recoloring pattern
- Caveat
  - For unpaired style transfer, images selected in the training set has a large impact on the performance

    e.g. Transfer between similar genres (landscaping, portrait, still objects…) has better performance

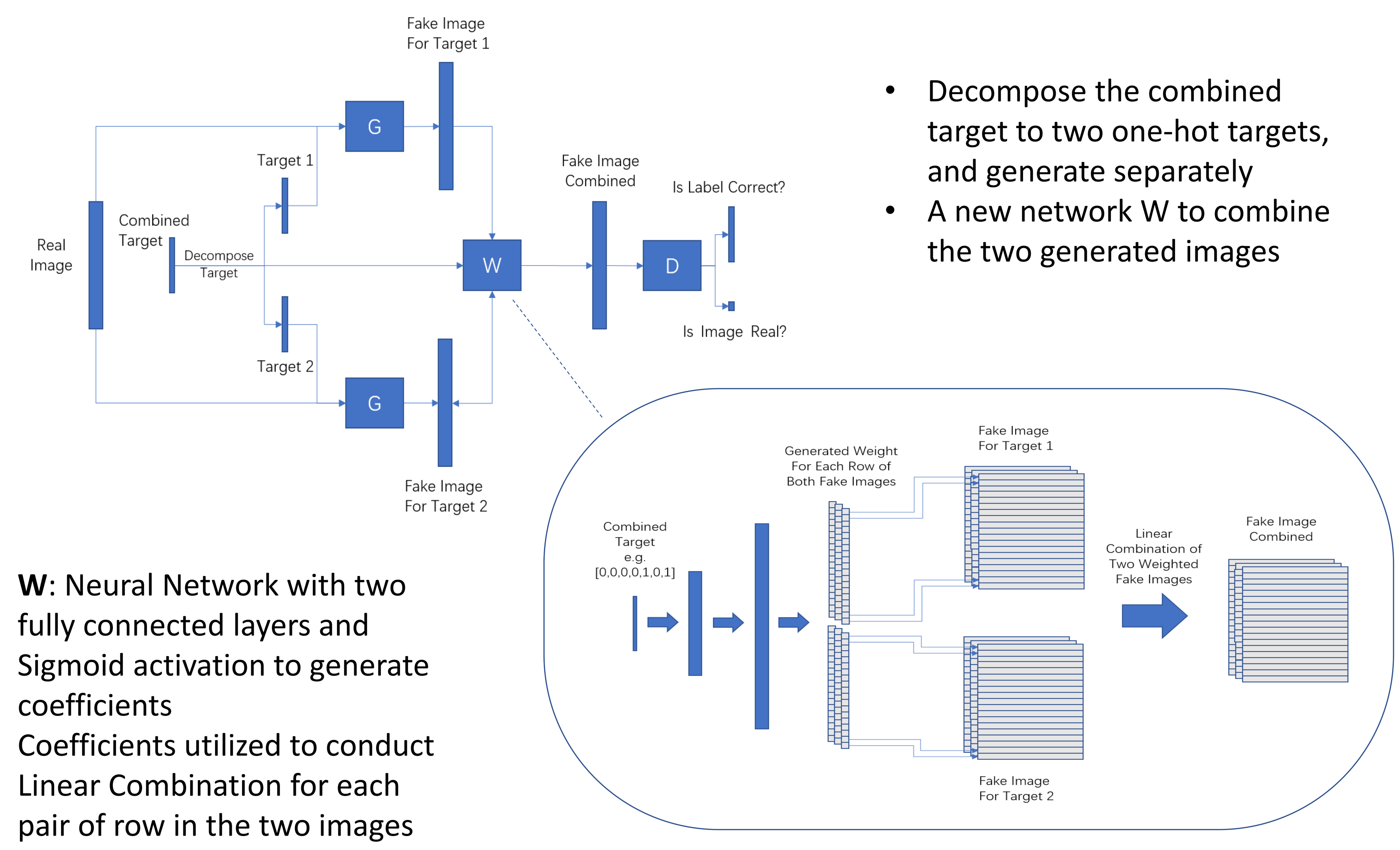
    e.g. Transfer between certain categories is easier




## Reference

[1] 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.

## Image Style Combination via Linear Combination



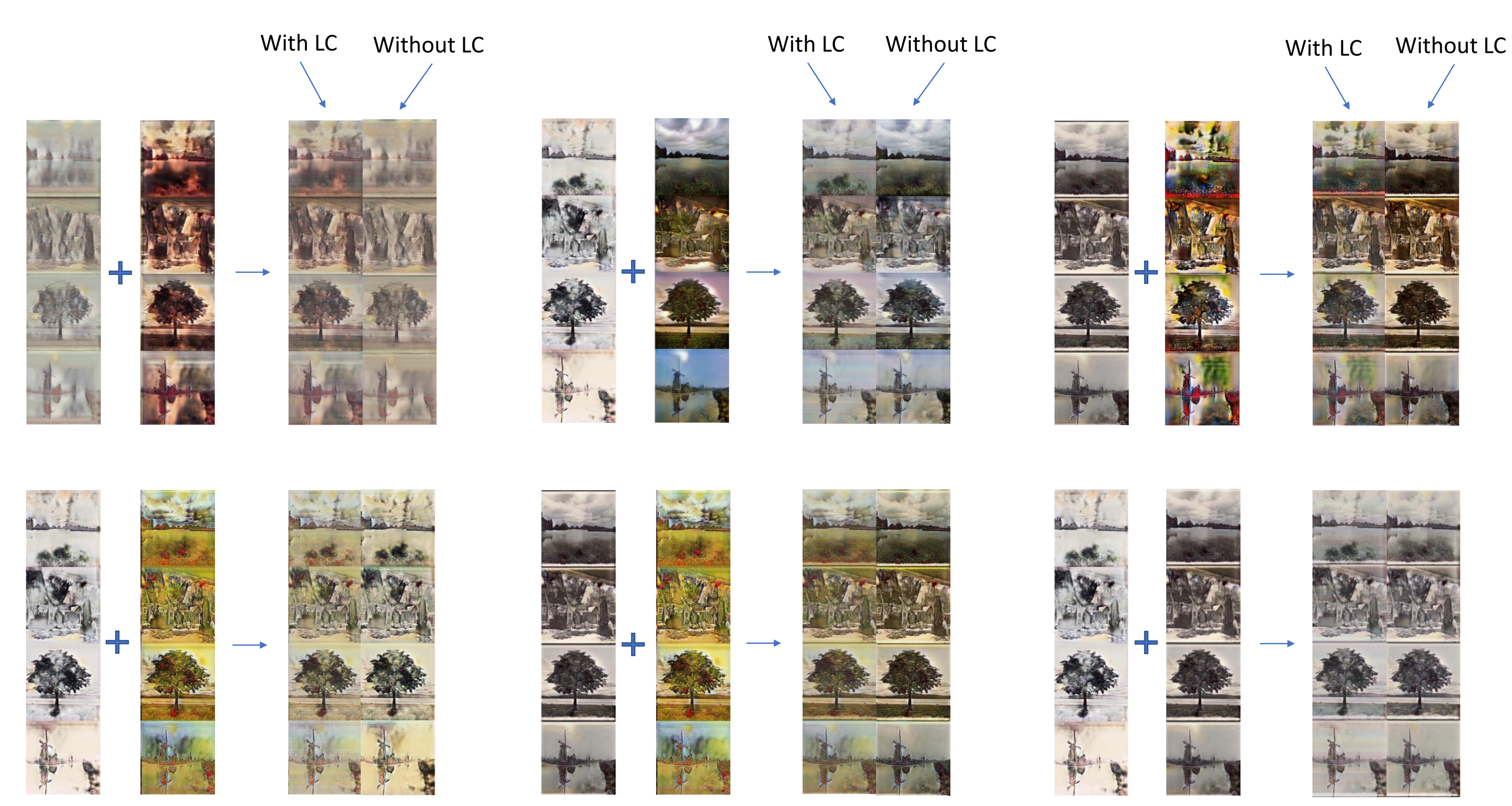

- Decompose the combined target to two one-hot targets, and generate separately
- A new network W to combine the two generated images

- **W**: Neural Network with two fully connected layers and Sigmoid activation to generate coefficients
- Coefficients utilized to conduct Linear Combination for each pair of row in the two images

$$Img_{Combined}[i,j,:] = Img_{Style1}[i,j,:] * c_1^{ij} + Img_{Style2}[i,j,:] * c_2^{ij}$$

- Fix Generator & Discriminator to the model of 146,000, retrain only W using Adam with initial learning rate 0.0001 and batch size 16, with loss on denoting as real and producing the right label

- Result of 150,000 steps shown below, as well as the combined images without W
- Personally speaking, more reasonable generations in terms of **combination** than original model

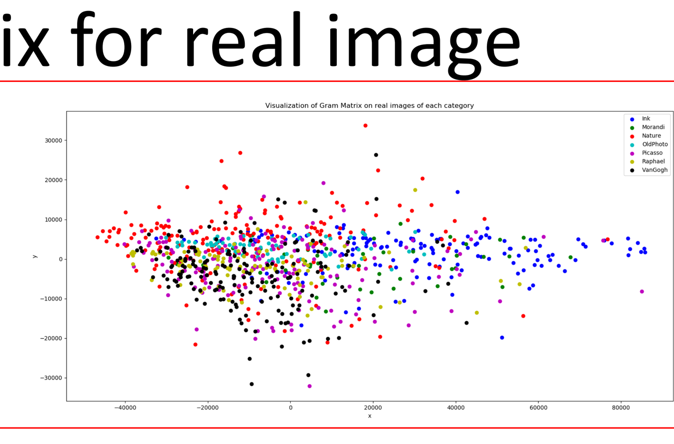



## Analysis on Gram Matrix for real image

- Calculate Gram Matrix on real images with vectorized channel map, resulting 3x3 matrix

$$G_{ij} = \langle X[i,:,:]^{Vec}, X[j,:,:]^{Vec} \rangle$$

- Visualize by applying PCA for 2D coordinates



## Conclusion

- Unified model like StarGAN can be used as a cost-efficient way in Multi-class artistic style transfer, but the performance relates with data set and similarity between styles
- Add a separate Neural Network to learn the coefficient of Linear Combination between individual images could yield perceptually better combination results
- The style defined by Gram Matrix is specific to each image rather than to each category

## Future Works

- Work with finer tuned dataset, for example, selecting all landscape images, training with more balanced class, and choosing painter with more consistent styles
- Is there way to find the reasonable "ground truth"? Try density estimation approaches and generates the pixel in combined image based on distribution of both styles