PML&DL Project Deliverable 1

🤖Neural Face Editor (NFE)🎨

Team

Andrey Palaev, a.palaev@innopolis.university, B19-DS

Mikhail Rudakov, m.rudakov@innopolis.university, B19-DS

Anna Startseva, a.startseva@innopolis.university, B19-DS

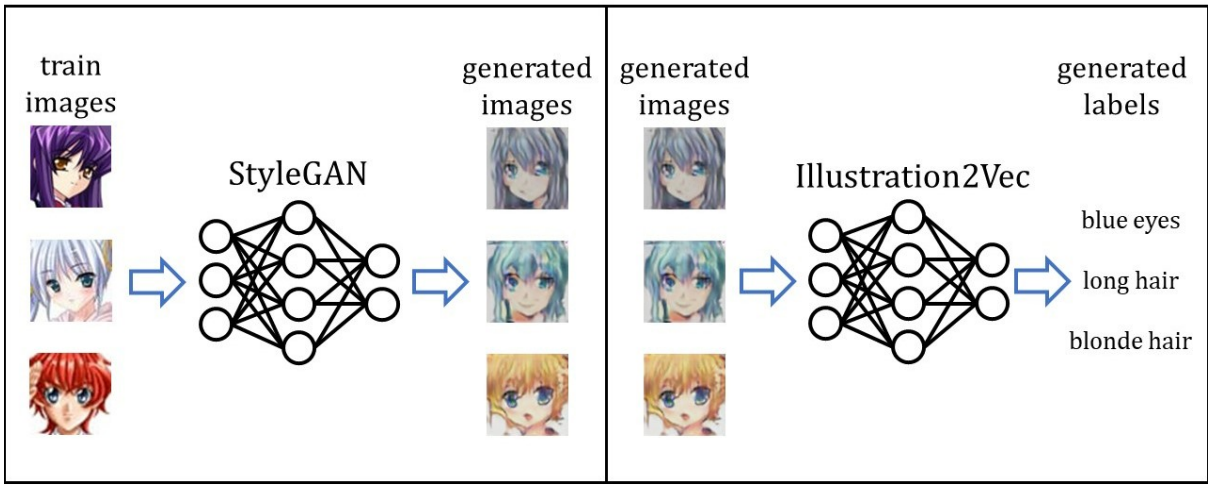
---

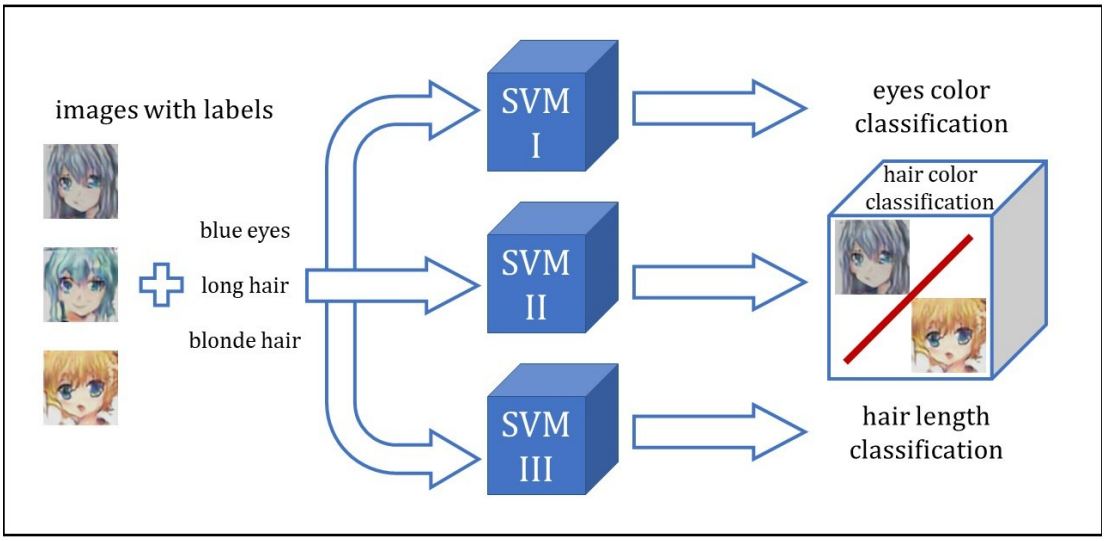
https://github.com/Palandr1234/NFE

Phase 1 Progress Overview

- Used dataset https://www.kaggle.com/datasets/splcher/animefacedataset, trained StyleGAN to generate anime faces
- Read papers about possible approaches
- Search for datasets suitable for training
- The attributes for the 10K generated images have been created

Project Architecture





3. Latent vector classification

Mainly adopted from Shen et al. [1], our project architecture consists of several models.

Stage 1 is image generation. StyleGAN is used to produce images of anime faces.

Stage 2 is generated image annotation. We use illustration2Vec proposed by Saito and Matsui [2] to tag images with tags we need.

Stage 3 is image classification. For the each attribute we want to be able to control further, we train a SVM for separating latent codes based on some feature. This would let us know the vector in which this feature changes in the latent space.

Currently, Stages 1 & 2 are implemented. See the results in Current Results section.

Papers & Possible Approaches

We explored several papers to decide how to approach the project main problem. Below are our findings on possible solutions with reference to the papers.

Main paper that our current architecture is based on is work by Shen et al. [1]. It uses GAN-generated images, classify them on every feature with SVM, and then manipulate semantic space of real images (produced by GAN inversion) to edit photos.

Similar approach of estimating feature variation trajectories in latent space is described by Plumerault et al. [3]. This approach required to manually apply several transformation and detect vector change in the latent space. This approach could be useful for general features of the image: recoloring hair/skin, rotating image, changing object position.

We also find work of Jin et al. [4] very useful for the project. They describe dataset collection process based on anime character website parsing, and propose a GAN architecture that is able to learn tag classification too, so user can specify the image tags at the generation stage. We consider using this architecture to produce styled version of the input image as an experiment.

Data

On this sprint, we found several datasets that may be useful for our project. Here we list datasets links, their advantages and disadvantages.

Anime Face Dataset

Mckinsey666 dataset scraped from www.getchu.com

https://www.kaggle.com/datasets/splcher/animefacedataset

This dataset contains 63k anime faces images. It can be useful for training GAN, no image preprocessing needed. We use this dataset on this phase to train StyleGAN. It can be simply downloaded from kaggle.

Tagged Anime Illustrations

Explore more than 300,000 pieces of fan art

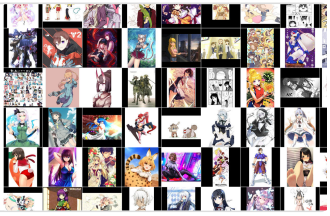
https://www.kaggle.com/datasets/mylesoneill/tagged-anime-illustrations

Dataset contains 300k anime images of general kind with labels. Additionally, there is a subset of anime faces with skin color, hair color attributes. At the current stage, we do not utilize this dataset as image preprocessing is needed for most of the images. Moreover, labels are chaotic, hardly usable. For faces subset, there are too few features tagged, so we can make no advantage out of it.

Danbooru2021: A Large-Scale Crowdsourced and Tagged Anime Illustration Dataset

Deep learning for computer revision relies on large annotated datasets. Classification/ categorization has benefited from the creation of ImageNet, which classifies 1m photos into 1000 categories. But classification/categorization is a coarse description of an image which

https://www.gwern.net/Danbooru2021



Danbooru is a large public dataset (5m images) of anime images. Second dataset if suitable for work (SFW) subset of this dataset. A lot of tag (162m) are present. This dataset could be used for big, high-quality GAN creation, but requires a lot of preprocessing work to be done (extract images, select proper tags). Currently we do not utilize this dataset.

We use first (anime-faces-dataset) dataset to train StyleGAN, which details are described further.

We apply automated tagging tool (illustration2vec) for labelling 10K images generated by our trained StyleGAN. Dataset of images that we labelled (vectors.csv with image latent code and attributes.csv with corresponding tags) is available here: https://drive.google.com/drive/folders/15NjOyO478uu3eZlCwyL7AZoZP4G1UnYQ?usp=sharing

Code & Models

Firstly, we create a StyleGAN following architecture proposed by Karras et al. [5], showed on figure 1. With a dataset of 63k images, we train StyleGAN using Kaggle computational service (with GPU's). Source code for training is available on github. Weights for trained GAN are on google drive.

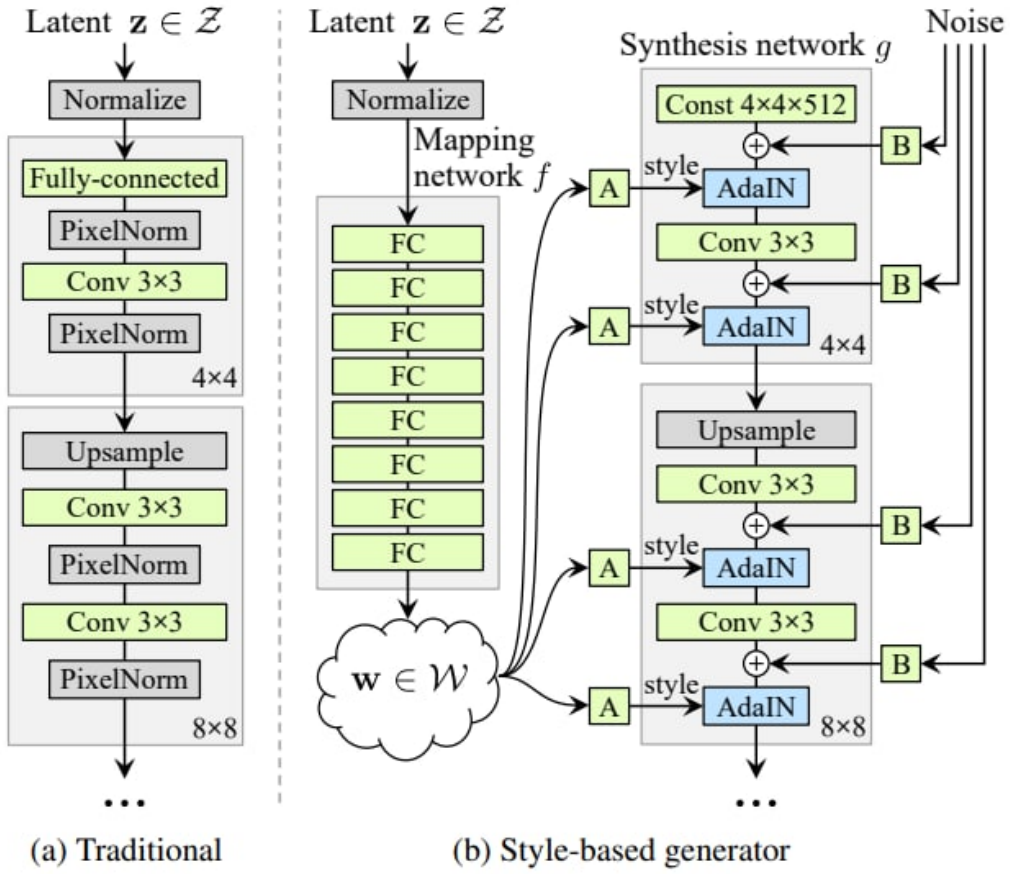


Figure 1. Traditional and StyleGAN architectures

Also, we use illustration2vec [2] to label images. This model is pretrained, so we only apply it on GAN-generated images: Notebook for images tagging.

Current Results

Sample generated images, produced by GAN:

Current results are far from perfect, but we will try to fix it in the future. Possible solutions are to try different, bigger dataset or to change model architecture.

As we stated before, we generated and labelled 10K samples. The dataset statistics is the following (the rest of values is NaN):

- Hair length

short	666
long	140

- Hair colour

blonde	2363
pink	559
purple	219
red	61
green	49
blue	33
silver	4
brown	4
black	1
aqua	1

- Eye colour

blue	4977
green	542
red	23
purple	19
pink	1

As we can see, there are a lot of NaN values. This is because the results are far from realistic ones and tagging tool is not 100% accurate (as it is a neural network)

In the future work , we may consider adding other attributes such as “Smile” and “Blush” or even “Quality” (to fix the artifacts) made by GAN.

Team Work Distribution

Andrey - Trained StyleGAN for faces generation, applied illustration2vec to label images

Anna - Searched appropriate datasets for the project, explored their features and applicability to our project, studied StyleGAN

Mikhail - Read papers about to formulate possible approaches for project implementation, managed team & deliverable creation

References

[1] Shen, Y., Gu, J., Tang, X., & Zhou, B. (2020). Interpreting the latent space of gans for semantic face editing. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 9243-9252) paper link

[2] Saito, M., & Matsui, Y. (2015). Illustration2vec: a semantic vector representation of illustrations. In *SIGGRAPH Asia 2015 Technical Briefs* (pp. 1-4). paper link

[3] Plumerault, A., Borgne, H. L., & Hudelot, C. (2020). Controlling generative models with continuous factors of variations. *arXiv preprint arXiv:2001.10238.* paper link

[4] Jin, Y., Zhang, J., Li, M., Tian, Y., Zhu, H., & Fang, Z. (2017). Towards the automatic anime characters creation with generative adversarial networks. *arXiv preprint arXiv:1708.05509.* paper link

[5] Karras, T., Laine, S., & Aila, T. (2018). A Style-Based Generator Architecture for Generative Adversarial Networks. *arXiv*. paper link