PML&DL Project Deliverable 2

🤖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 2 Progress Overview

- Tried different GAN architectures and compared results for better quality images
- Trained SVM for 3 attributes
- Explored approaches for GAN inversion to enable passing our own anime images to the model

GAN Architecture Selection & Results

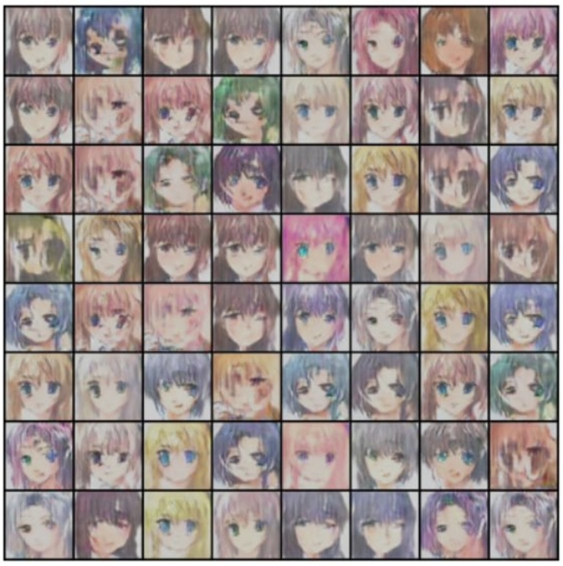


Fig. 1: Old StyleGAN results

Since the last iteration, we tried to change StyleGAN architecture to produce more detailed and diverse faces in terms of face shape, hair density, eyes and mouth detalization.

The first version of GAN-generated images, shown on Fig. 1, are often too corrupted. Eyes location and size, mouth and hair shape look terrifying sometimes.

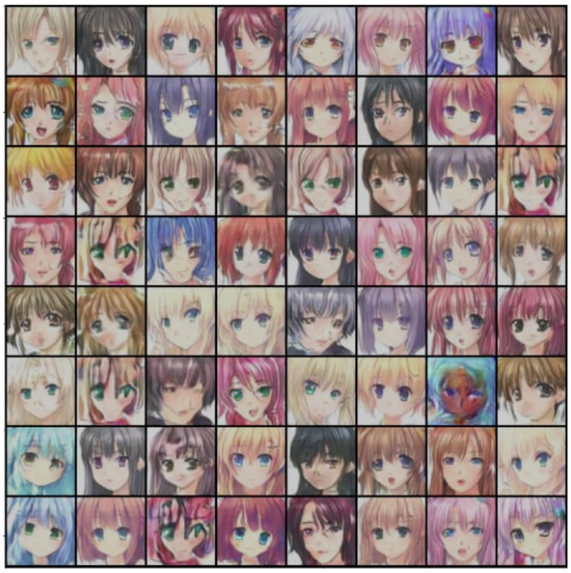


Fig. 2: New StyleGAN results (currently the best one)

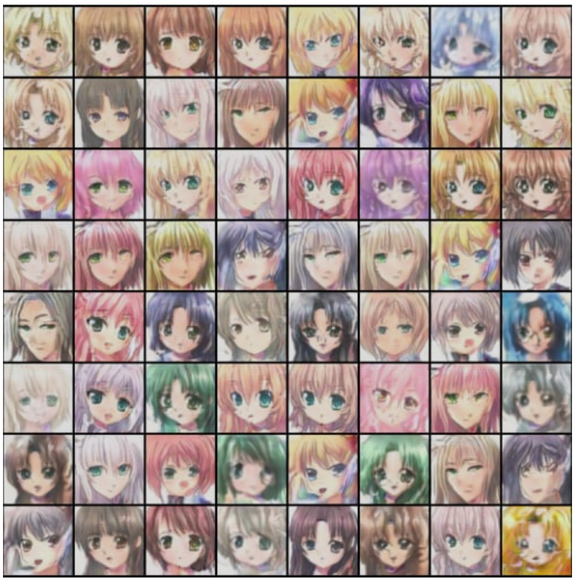


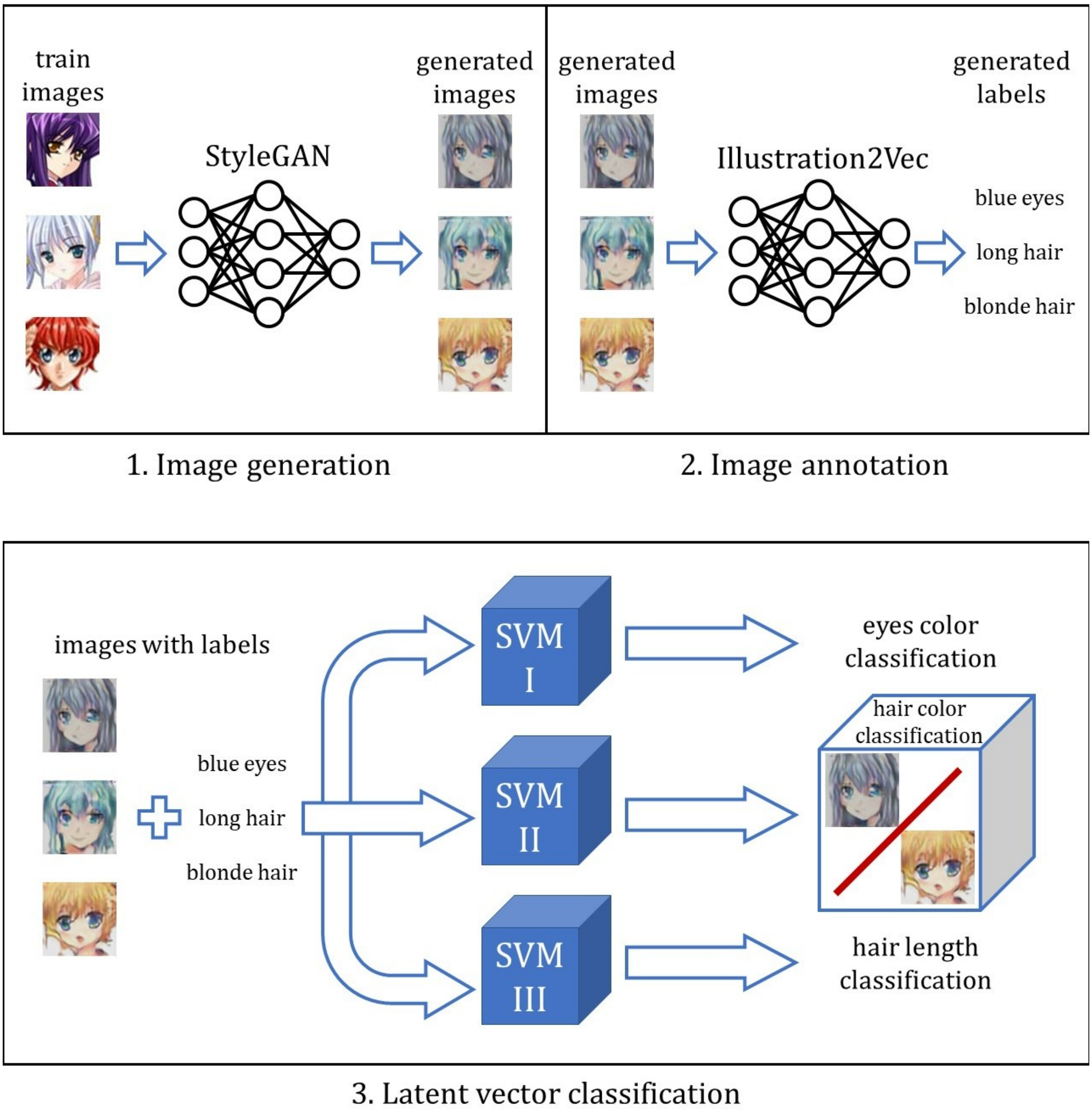
Fig. 3: New StyleGAN results (worse)

To fix these issues, we increased the number of weights in both Discriminator and Generator layers to 8.9M and 13.7M respectively. Also, the amount of epochs was increased from 50 to 150. The results can be seen in Fig. 2. Then, we tried to increase it even further (to 11M and 13.5M parameters respectively). However, due to Generator and Discriminator imbalance results became worse (Fig. 3). As we see, faces in Fig. 3 are less diverse in terms of face shape and hair length (for example, row 4, col 1-3 on Fig. 3). Also, eyes shape and color diversity is worse. In the future, we will try to use the latent space with decreased dimensionality (decrease from 512 to 128) to fix the imbalance between Disriminator and Generator layers.

In the line below, we show cherry-picked generated images from the current best version of StyleGAN (from Fig. 2). As we see, face angle, hair, eyes, and even blusing vary across images greatly. We believe this version would be enough for the project.

We also needed to re-label images with illustration2vec as done on the first phase.

SVM for Attributes Classification



According to step 3 of the NFE architecture, we trained several SVM classifiers on labelled GAN-generated images as proposed in [1]. We currently have 3 classifiers: eye colour, hair length, hair colour. SVM training details and result metrics are presented in Table 1. Training quality is far from perfect, though. Possible reason is tagging quality of Illustration2Vec: it generates too much tags that are often not usable for our purposes, and the prediction quality is also poor sometimes. Another reason is training set size, as some attributes contain too few examples in the training set. We clearly see that SVMs are overfitting as train accuracy is 100%.

Table 1. SVM Training Results

	Train Dataset Size	Train Accuracy	Test Accuracy
Eye colour	4258	0.9	0.67
Hair length	1711	1.0	0.73
Hair colour	3182	1.0	0.67

Next step regrading SVM is to get separating hyperplane from the trained classifier and use it to change facial attributes (one or several at a time).

GAN Inversion Approaches

For NFE to be actually used in the facial attributes editing, we need to convert user's input image to the latent vector representation that GAN produces. Then latent representation of the image could be changed to edit selected parameters and output the changed image. However, GAN does not provide such function for converting input image to latent vector. Hence, we need to use some other approach. We analyzed GAN inversion survey [2] to choose the best technique for our purposes.

One approach for GAN inversion is learning-based method. It involves training an encoder network that produces approximated latent vector on the encoder output. The training objective of the encoder model is to reproduce an input image as the decoder output.

An opposite approach for GAN inversion is optimization-based technique. It optimizes the latent vector to produce the desired input image. Altough this approach does not involve training a complex and large neural network, it suffers from local minima in the GAN latent space and highly depends on the initialization.

A hybrid approach based on two disscussed ones exist. It firstly constructs approximate latent vector in the encoder part, and then optimizes obtained vector to match input image. This way, we take best of both worlds and assumed to have better inversion quality, as confirmed by the survey.

We wil next try to implement hybrid approach for GAN inversion. This would be a step towards using NFE on user-provided images.

Team Work Distribution

Andrey - Finetuned GAN architecture, trained SVM for attributes

Anna - Evaluated GAN's and SVM's results and drawbacks, experimented with real images

Mikhail - Explored GAN inversion approaches to be used in the NFE pipeline

Sprint 3 Tasks:

- Use SVM hyperplanes and approach from [1] to change facial attributes; evaluate results
- Implement GAN inversion
- Decide on including more facial attributes to GAN and SVM pipeline if possible
- Start telegram bot implementation

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] Xia, W., Zhang, Y., Yang, Y., Xue, J., Zhou, B., & Yang, M. (2021). GAN Inversion: A Survey. *arXiv*. paper link