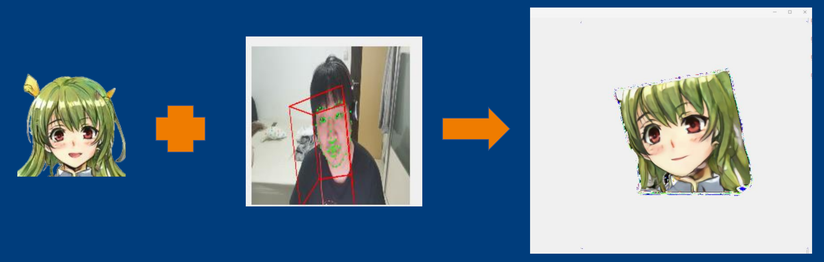
Enhanced Talking Head Anime from a Single Image towards a User-friendly Application

Keywords: Image processing, face detection, GAN, CNN, facial landmark detection, super-resolution
Authors: Huang Liu, Sun Weiran, Tan Zhenwei, Wang Yuxuan



CURRENT DEVELOPMENT AND LIMITATIONS TO BE RESOLVED

The content of our project is to input a facial picture of an animated character, combined with the camera to capture the expressions and actions made by a real person, so that the animated character can make expressions and actions consistent with the real person.

Below is part of the structure of our project:

We developed our model based on an open-source project, talking-head-anime, which provides with fundamental structure of network. We aim to improve on the existing work by solving the following limitations:

1. Simplify input requirements, the framework has a strict restriction on the size of the image. This restriction means that is impossible for a user to choose a random image, which hasn't been edited deliberately, as the input, causing a great inconvenience.
2. We find that the facial landmark detection supported by the original framework is not accurate enough to recognize the features of a human face, which may lead to the failure of generating the corresponding changes to the image.
3. We also find that some images are in low resolution, which will greatly impact the user's experience.

ENHANCEMENT 1: BETTER INPUT IMAGE PROCESSING AND GENERATION

Detect anime character head in the input image and pre-process the head to fit input requirements

Detect anime character face with lbpcascade

Traditional face detection models, e.g., MTCNN, do not work well with anime characters.

Process the cropped head

1. Extract head, resize to 128*128
2. Expand the picture to 256*256 with transparent pixels
3. Place the head at the centre

Generate anime characters using pre-trained GAN models and pre-process accordingly

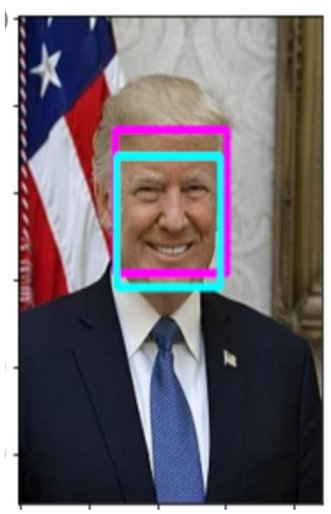
WAIFU LABS

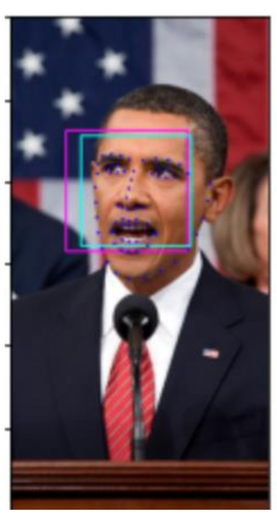
WaifuLabs is a world-class, state-of-the art GAN model for anime character generation.

The model is integrated into the application to randomly generate anime characters. Generated images are processed in the same way as manual inputs automatically.

ENHANCEMENT 2: MORE ACCURATE FACIAL LANDMARK DETECTION

Use pretrained CNN model (mmod_human_face_detector.dat) to better locate the 3D faces.





This is trained on this dataset: http://dlib.net/files/data/dlib_face_detection_dataset-2016-09-30.tar.gz. which is created by finding face images in many publicly available image datasets (excluding the FDDB dataset). In particular, there are images from ImageNet, AFLW, Pascal VOC, the VGG dataset, WIDER, and face scrub.

CNN model improve the accuracy of detecting faces compared to the original model based on Histogram of oriented gradient and linear classifier.

CNN Dlib

ENHANCEMENT 3: ANIMATED CHARACTERS WITH HIGHER RESOLUTION

We offer a post-processing super-resolution option for advanced users with better hardware wanting higher resolution.

For the backend, we use state-of-the-art SR model DCSCN, which is known for efficiency. The pretrained model is based on the anime character dataset and achieved excellent performance with real-time processing for video streaming on GTX 2060.

All post-processing is transparent to the user and switch-on can easily be done with a single button.