Final Project 吳宜靜、黃芷柔、蔡長頤

1. Method:

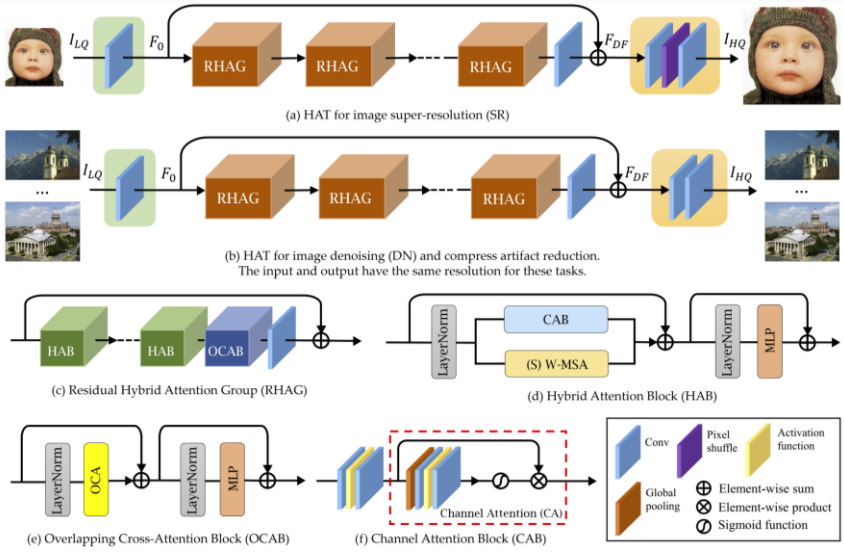
HAT is the super resolution method in our project. Besides, we integrated filtering and Codeformer to further improve the result.

(1) Super resolution—HAT

- HAT :

  Traditional Transformer-based methods are limited in their spatial range of input information utilization. Therefore, HAT is designed to activate more input pixels for better image reconstruction, integrating both channel attention and self-attention mechanisms.

  The Hybrid Attention Transformer (HAT) Design includes 2 main features, Hybrid Attention Block (HAB) and Overlapping Cross-Attention Module (OCAB). HAB combines channel attention with window-based multi-head self-attention (W-MSA), enhancing the ability to utilize global statistics and local details effectively. OCAB facilitates enhanced interaction between neighboring window features by allowing overlaps in the attention mechanism, improving the integration of cross-window information.



- HAT-GAN:
  HAT-GAN, a GAN incorporated with the HAT architecture, is the model we used in the project. It is a model for real-world image super-resolution.
  The paper showed that employing a "same-task" pre-training strategy using large-scale datasets can improve the effectiveness HAT. Therefore, the HAT-GAN is trained by using the pretrained MSE-based model and then adopt DIV2k, Flick2K and OST datasets for training. First, the MSE-based model is trained, and then introduce the generative adversarial training to fine-tune the GAN-based model.

(2) Filtering

By observation, we can see that there is noise in the pictures provided by the TA. For example, the white dots in the fig1 above the left eye and below the head of the cat. Another obvious example is the table tennis racket, featuring abnormal black dots. Therefore, we decided to perform filtering for the purpose of denoising.

Noteworthily, we performed filtering after super resolution. The reason is that, if we filter before super resolution, the resulting image would be extremely blurred, loosing an extensive amount of detail. This may be since blurring worsens the problem of insufficient information in super resolution problems. In contrast, if we filter the image after super resolution, we can keep more details, while smoothing the image and reducing the noise.

We tried many kinds of filters, for example, mean filter, max filter, min filter, median filter, and contraharmonic filter. We found that, for different pictures, they have different bast-performed filters.

Take the cat photo as an example:

HAT+ No filter + Codedeformer	HAT+ Median filter + Codeformer	HAT+ Contraharmonic filter + Codeformer
		

(-) Obvious noise above the left eye and below the head of the cat. (+) Cat fur has delicate detail.	(-) A certain amount of the detail at cat fur is lost. Blurred. (+) Attenuated noise.	(-) Very Blurred. (+) The white dot noise is al modt gone.

(3) Codeformer:

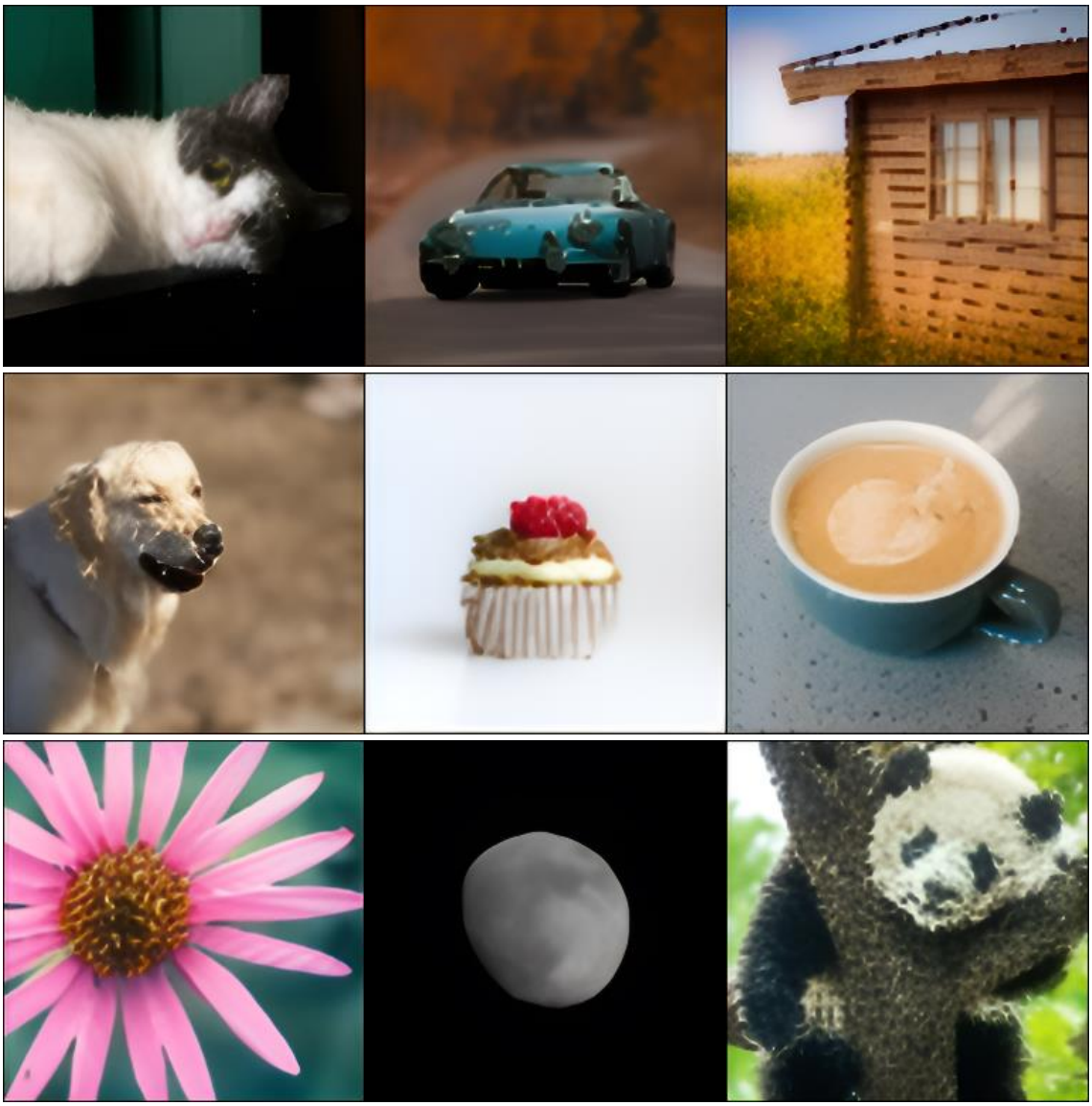
The CodeFormer is a Transformer-based face restoration network designed to address the challenges of blind face restoration by utilizing a learned discrete codebook prior. This codebook prior reduces the uncertainty in mapping degraded inputs to high-quality outputs and enriches the process with detailed visual elements necessary for generating clear and natural faces. The CodeFormer models the global composition and context of low-quality face images for accurate code prediction, which aids in reconstructing faces that closely resemble the target, even from severely degraded inputs. Additionally, it includes a controllable feature transformation module that facilitates a flexible balance between fidelity and image quality.

There are 2 reasons that motivated us to use Codeformer. First, obviously, after applying filtering, the image quality is degraded and blurred, so we want to restore it. Second, we observed the difficulties of super resolution is mainly on more complicated images such as human face, dog, and panda. Therefore, to focus on these human-face associated images, we decided that codeformer would be helpful.

HAT	HAT + Codeformer	HAT+ Contraharmonic filter + Codeformer
		

2. Results:

HAT + Contraharmonic filter + Codeformer:



Observations:

(1) Our image processing structure performed ideally on human face, dog, and a great improvement on the bird image compared to only using HAT.
(2) We can see that some images are still poorly restored. For example, the panda image is not smooth, and in the fruit image, the margins of the fruit and the moon are rugged and unnatural.

3. Reference:

- HAT:
  Paper: https://arxiv.org/pdf/2205.04437 / https://arxiv.org/pdf/2309.05239
  Github: https://github.com/XPixelGroup/HAT?tab=readme-ov-file
- Codeformer:
  https://github.com/sczhou/CodeFormer/tree/master