Image Upscaling by code prediction (CodeFormer)

April 24, 2023

Agenda

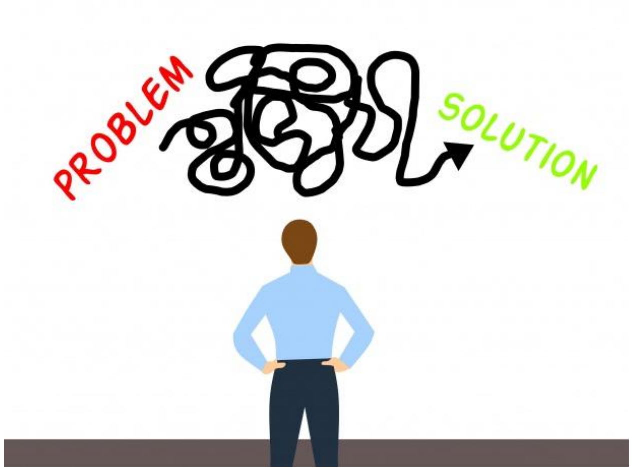
- Team intro
- Problem description
- Approaches
- Evaluation metrics
- Experiments
- Results
- Lesson learned
- Future work

Team Intro

- Justin Wu, syuanyingwu@vt.edu, MEng CS @ NVC
- Yi-Han Chen, yihanchen@vt.edu MEng CS @ NVC
- Diego Romero, rdiego@vt.edu, MEng CS @ NVC

Problem description

1. Restoring face images often suffer from various degradation, such as compression, blur, and noise.
2. Details lost in the inputs (degraded image) needs to be complimented.
3. CodeFormer offers limited superiority to other methods. Sometimes cannot produce good results
   ➪ *our experiment*



Approaches: CodeFormer Methodology

- Exploit **Discrete** codebook prior
- See blind face restoration as a **code prediction task**
- Proposed a **Transformer-based code** prediction network
- Proposed a **controllable feature transformation** module with an adjustable coefficient -> a trade-off between restoration quality and fidelity

Codebook Learning (Stage I)

Codebook Lookup Transformer Learning (Stage II)

Controllable Feature Transformation (Stage III)

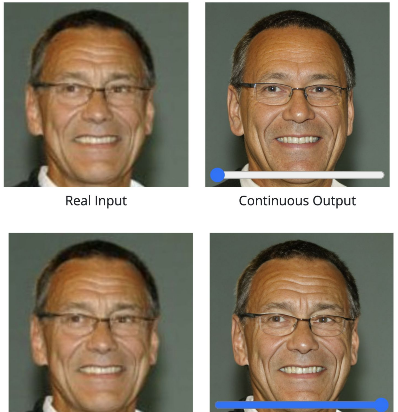
What?
Pre-train the quantized autoencoder to learn a context-rich codebook

Objective
Reduce uncertainty of the LQ-HQ mapping and complement high-quality details for restoration

What?
Employ a Transformer to model global interrelations to make the prediction

Objective
Improve the accuracy of code prediction for face restoration

What?
Create a flexible tradeoff between quality and fidelity of face restoration.



quality

fidelity

Evaluation metrics

- SSIM
- PSNR

SSIM (Structural Similarity Index Measure)

- Measure the similarity between two images
- It measures three factors: luminance, contrast, and structural similarity.
  - Luminance: the brightness of the images
  - Contrast: the difference between the images
  - Structure: how similar the patterns in the images are.

$$SSIM = l(x,y) + c(x,y) + s(x,y)$$

SSIM (Structural Similarity Index Measure)

$$l(x, y) = \frac{2\mu_x \mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1}$$

$$c(x, y) = \frac{2\sigma_x \sigma_y + c_2}{\sigma_x^2 + \sigma_y^2 + c_2}$$

$$s(x, y) = \frac{\sigma_{xy} + c_3}{\sigma_x \sigma_y + c_3}$$

SSIM (Structural Similarity Index Measure)

$$\text{SSIM}(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}$$

with:

- μ_x the pixel sample mean of x;
- μ_y the pixel sample mean of y;
- σ_x^2 the variance of x;
- σ_y^2 the variance of y;
- σ_{xy} the covariance of x and y;
- $c_1 = (k_1L)^2$, $c_2 = (k_2L)^2$ two variables to stabilize the division with weak denominator;
- L the dynamic range of the pixel-values (typically this is $2^{\#bits\ per\ pixel} - 1$);
- $k_1 = 0.01$ and $k_2 = 0.03$ by default.

SSIM (Structural Similarity Index Measure)

MAPPING SSIM TO MEAN OPINION SCORE SCALE

SSIM	MOS	Quality	Impairment
≥ 0.99	5	Excellent	Imperceptible
$[0.95, 0.99)$	4	Good	Perceptible but not annoying
$[0.88, 0.95)$	3	Fair	Slightly annoying
$[0.5, 0.88)$	2	Poor	Annoying
< 0.5	1	Bad	Very annoying

Source: https://ieeexplore.ieee.org/document/6850361

PSNR (Peak Signal-to-Noise Ratio)

- Evaluate the quality of compressed or reconstructed images or videos
- It measures the ratio between the maximum possible power of a signal and the power of corrupting noise that affects the fidelity of its representation.
- PSNR is expressed in decibels (dB)

$$PSNR = 20 \cdot \log_{10}\left(\frac{MAX_I}{\sqrt{MSE}}\right)$$

$$MSE = \frac{1}{m\,n}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[I(i,j) - K(i,j)]^2$$

PSNR (Peak Signal-to-Noise Ratio)

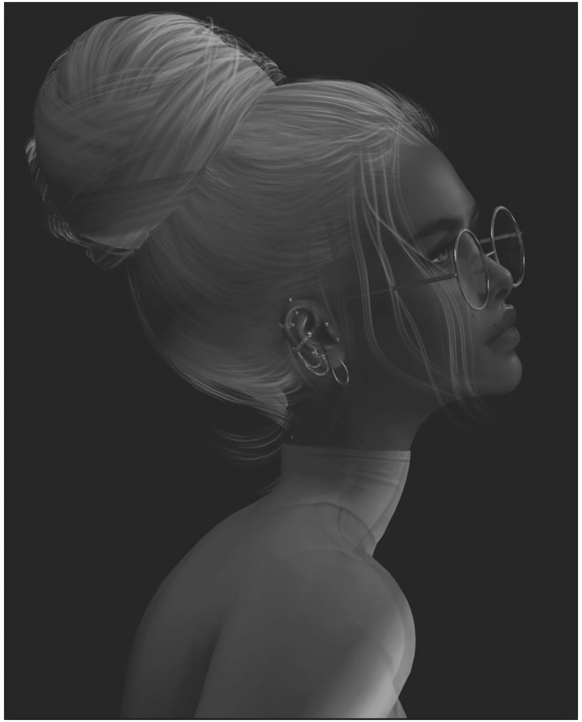
PSNR Value	Quality	MOS
PSNR > 33 dB	Excellent Quality	5
33 dB > PSNR > 30 dB	Fair Quality	2
PSNR < 30 dB	Poor Quality	1

Source: https://ieeexplore.ieee.org/document/6141267

Experiment

1. Collect side face images
2. Generate LQ images
3. Upscale LQ images
4. Evaluate with ground truth
    a. Visual Comparison
    b. Evaluation Metrics

Collect side face images

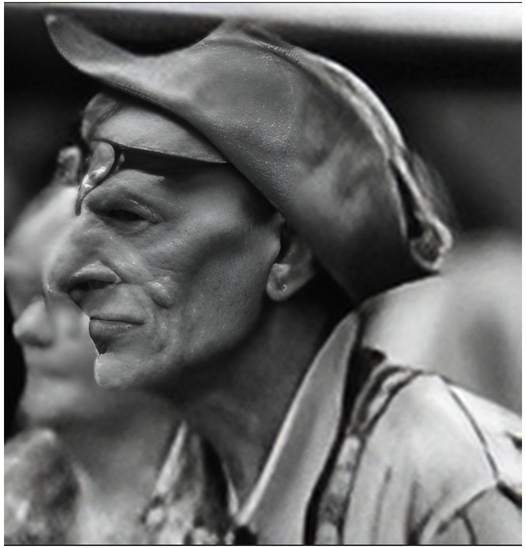


Generate LQ images

Upscale LQ images

Evaluate with ground truth - Visual Comparison

Ground Truth



Upscaled image

Evaluate with ground truth - Visual Comparison

Ground Truth

Upscaled image

Evaluate with ground truth - Evaluation Metrics

Methods	SSIM	PSNR
Score	0.66	23.52
Quality	Poor	Poor

Ground Truth

Upscaled image

Evaluate with ground truth - Evaluation Metrics

Methods	SSIM	PSNR
Score	0.89	32.29
Quality	Poor	Fair

Ground Truth

Upscaled image

Evaluate with ground truth - Evaluation Metrics

Methods	**Side Face (Our experiment)**	**CodeFormer**
SSIM	0.75	0.61
PSNR	26.74	22.18

Results

- While the experiment yields better evaluation metrics than the paper, CodeFormer's output for the side face image still lacks naturalness, and has significantly lost detail such as accessories.
- The reason behind the better evaluation metrics could be the differences in the experimental setup, such as utilizing different degradation model parameters.

Future work

- Find side-face dataset to train the data for better code prediction.
- Same technique to restore old movies?

References

1. Shangchen Zhou Kelvin C.K. Chan Chongyi Li Chen Change Loy, “Towards Robust Blind Face Restoration with Codebook Lookup Transformer”, 2022
2. Chitwan Saharia, Jonathan Ho, William Chan, Tim Salimans, David J. Fleet, Mohammad Norouzi, “Image Super-Resolution via Iterative Refinement”, 2021
3. Yoanda Alim Syahbana; Herman; Azizah Abdul Rahman; Kamalrulnizam Abu Bakar, “Aligned-PSNR (APSNR) for Objective Video Quality Measurement (VQM) in video stream over wireless and mobile network”, 2011
4. Marco Zanforlin; Daniele Munaretto; Andrea Zanella; Michele Zorzi, “SSIM-based video admission control and resource allocation algorithms”, 2014