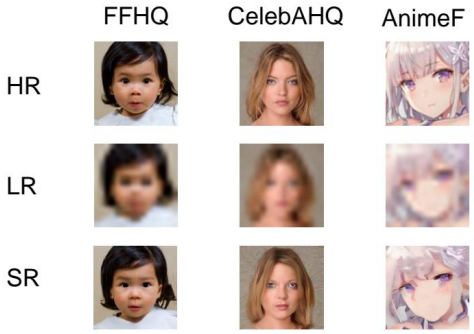
Can Diffusion Model Generalize Well in Image Super Resolution with Limited Fine-tuning

Kejia Yin
MScAC Student, Department of Computer Science, University of Toronto

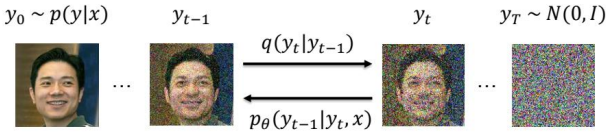
Motivation

- Image Super Resolution models perform well on original in-domain data but fails to generalize to new out-of-domain data.
- Related research is almost absent!
- Try to find out whether fine-tuning can mitigate this problem.



Method

- This project is based on existing Image Super Resolution method SR3[1].



Algorithm 1 Optimizing Process of SR3

1: **repeat**
2: Sample (x, y_0) from Dataset
3: Sample time step $t \sim U(1, MAX\_TIMESTEP)$
4: Sample noise $\epsilon \sim N(0, I)$
5: Take gradient step on

$$\nabla\theta \| f_\theta(x, \sqrt{\beta_t} y_0 + \sqrt{1-\beta_t}\epsilon, \beta_t) - \epsilon \|_p^p$$

6: **until** converged

Algorithm 2 Inference Process of SR3

1: Sample $y_T \sim N(0, I)$
2: **for** $t = T, ..., 1$ **do**
3: Sample noise $\epsilon \sim N(0, I)$ if $t > 1$, else $\epsilon = 0$
4:

$$y_{t-1} \leftarrow \frac{1}{\sqrt{\alpha_t}}(y_t - \frac{1-\alpha_t}{\sqrt{1-\beta_t}} f_\theta(x, y_t, \beta_t)) + \sqrt{1-\alpha_t}\epsilon$$

5: **end for**
6: **return** y_0

Where $q(y_t|y_{t-1}) = N(y_t|\sqrt{\alpha_t}y_{t-1}, (1-\alpha_t)I)$, and $\beta_t = \prod_{t=1}^{t} \alpha_i$

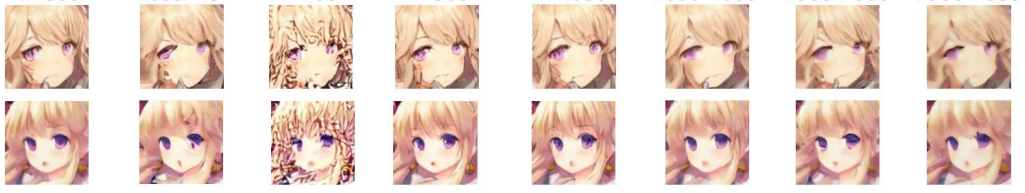
- We fine-tuned the pre-trained SR3 model with limited steps, amount of new domain data and range of time step.

Qualitative Results

Steps

Amount of New Domain Data

Range of Time Step



Quantitative Results

	PSNR↑	SSIM↑	LPIPS↓ [2]
FFHQ	22.56	0.673	**0.2331**
CelebAHQ	**23.27**	**0.676**	0.2355
Baseline	20.01	0.495	0.3770

Steps	PSNR↑	SSIM↑	LPIPS↓
400	20.81	0.499	0.3314
1200	20.98	0.515	0.3192
1600	20.71	0.508	0.3111
2000	21.16	0.524	0.3136
4000	20.69	0.504	0.3088
12000	**21.57**	**0.540**	**0.3062**

Amount	PSNR↑	SSIM↑	LPIPS↓
128	21.16	0.524	0.3136
64	21.35	0.527	**0.3063**
32	21.56	0.536	0.3208
16	21.99	0.552	0.3417
8	**22.07**	**0.555**	0.3465

Timestep	PSNR↑	SSIM↑	LPIPS↓
1-2000	21.16	0.524	0.3136
1-100	15.62	0.302	0.4692
1-500	20.71	0.514	**0.3082**
1-1000	20.91	0.513	0.3096
1000-2000	**21.72**	**0.541**	0.3349
1500-2000	21.29	0.519	0.3678
1900-2000	20.93	0.527	0.3890

References

[1] Saharia, Chitwan, et al. "Image super-resolution via iterative refinement." IEEE Transactions on Pattern Analysis and Machine Intelligence. 2022
[2] Zhang, Richard, et al. "The unreasonable effectiveness of deep features as a perceptual metric." Proceedings of the IEEE conference on computer vision and pattern recognition. 2018