Name: Mukund Dhar
UCF ID: 5499369

Report on **"Image Super-Resolution via Iterative Refinement"**

The paper introduces SR3 (Super-Resolution via Repeated Refinement), an approach that uses stochastic iterative denoising process to achieve image Super-Resolution based on Denoising diffusion probabilistic models (DDPM). The reverse process starts with pure Gaussian noise and iteratively refines the noisy output using a U-Net model trained on denoising at various noise levels. This process generates high-quality, super-resolved images from low-resolution input images. SR3 works by learning to transform a standard normal distribution into an empirical data distribution through a sequence of refinement steps. The main important factor is the U-Net architecture trained with a denoising objective. The DDPM is adapted to conditional image generation by modifying the U-Net architecture. SR3 is effective on face and natural image super resolution at different magnification factors. It is demonstrated that by cascading a 64x64 image synthesis model with SR3 models to progressively generate 1024x1024 unconditional faces in 3 stages, and 256x256 class-conditional ImageNet samples in 2 stages. The SR3 architecture is similar to the U-Net found in DDPM, with modifications to replace the original DDPM residual blocks with residual blocks from BigGAN and re-scaling the skip connections. The number of residual blocks and the channel multipliers at different resolutions have also been increased. SR3 exhibits strong performance on super-resolution tasks on faces and natural images. SR3 achieves a fool rate close to 50% and yields a competitive FID score of 11.3 on ImageNet in cascaded image generation. This suggests the effectiveness of the SR3 in generating photo-realistic outputs.

Strengths:

- SR3 is a new approach to conditional image generation that proves effective on face and natural image super-resolution at different magnification factors.
- It achieves a significantly higher fool rate than state-of-the-art GAN methods and a strong regression baseline, as demonstrated in human evaluations.
- These models can be cascaded with another image synthesis model, allowing for more efficient inference and the generation of high-fidelity images.

Weaknesses:

- The author notes that bias is an important problem in all generative models, SR3 being no different.
- There is evidence of mode dropping by these diffusion-based models. The models generate the same images during sampling when conditioned on the same input.
- The model also generates human face images with unrealistic continuous skin textures with no moles, pimples, or piercings.

Questions:

- Can the SR3 model be applied to other image-to-image translation tasks, such as colorization or deblurring?
- How can Automated image quality scores be improved for scenarios, such as, when the input resolution is low and the magnification ratio is large, so that they resemble human preference?

Possible ideas:

The SR3 model can be further improved by incorporating additional prior knowledge. For example, if the input image contains a face, the model could take in further information about the face structure and texture to generate more realistic and detailed outputs. Self-supervised learning could be used to train the model on tasks such as image colorization and inpainting. We could also explore incorporating attention mechanisms to help the model focus on relevant parts of the image leading to better performance.