Name: Mukund Dhar
UCF ID: 5499369

Report on "SDEDIT: GUIDED IMAGE SYNTHESIS AND EDITING WITH STOCHASTIC DIFFERENTIAL EQUATIONS"

The paper introduces a new method for image synthesis and editing that is based on a diffusion model generative prior. The image synthesis is done by iteratively denoising through a stochastic differential equation (SDE). Now, the main challenge in image synthesis is to find the right trade-off between the faithfulness of user inputs with the realism of the final output images. Existing GAN-based methods attempt to address this challenge using either conditional GANs or GAN inversions, but these methods are either too complex or often require additional training data or task-specific optimization algorithms or loss functions. SDEdit is, however, able to "hijack" the generative process of SDE-based generative models to produce realistic and faithful images. When an input image is given a user guidance input, such as a stroke painting or an image with stroke edits, SDEdit adds a suitable amount of noise to smooth out undesirable artifacts and distortions (e.g., unnatural details at stroke pixels), along with still preserving the overall structure of the input user guide. SDE can then be initialized with this noisy input image that is then iteratively denoised. SDEdit only uses a single pretrained SDE-based generative model trained on unlabeled data unlike the GAN-based methods. The right balance between realism and faithfulness can be achieved by the fact that when more Gaussian noise is added and the model is run for longer, the images generated are more realistic but less faithful. SDEdit has been demonstrated on three applications: stroke-based image synthesis, stroke-based image editing, and image compositing. SDEdit outperforms state-of-the-art GAN-based methods on realism score as well as on overall satisfaction score in stroke-based image synthesis experiments, according to human judgments. On image compositing experiments, SDEdit achieves a better faithfulness score and outperforms the baselines on overall satisfaction score in user studies.

Strengths:

- SDEdit does not necessitate the collection of new datasets for the "guide" images or the retraining of models unlike conditional GANs.
- It is computationally more efficient than GAN-based methods in terms of time and resources.
- It allows everyday users, regardless of their level of artistic expertise, to create and edit photo-realistic images with minimum effort.

Weaknesses:

- The highly "realistic" images generated by this model could be used in malicious ways to deceive humans and spread misinformation. The current forensic methods for detecting fake images struggle on the SDE-based models. Hence, it is critical to develop ways to prevent this and make this model more prevalent.
- There is a high possibility of bias in the human evaluation studies done to track the progress of image synthesis and editing tasks.

Questions:

- Are there any advanced techniques present for controlling the balance between realism and faithfulness in the generated images, such as incorporating user feedback or dynamically adjusting noise levels during the synthesis process?
- Are there any other ways to incorporate user guidance, beyond strokes?

Possible ideas:

- Exploring the use of SDE-based models for other types of image manipulation tasks, such as style transfer or image super-resolution.
- Extending SDEdit model to other modalities such as video or audio.
- The speed of SDEdit could be improved by recent works on faster SDE sampling.