UGC/FDS11/E06/24

Title of Project:	
(English)	Image Generative Model breaking away from Denoising in Diffusion-Like Environment
(Chinese)	圖像生成 - 突破擺脫噪音的類擴散模式

Abstract of Research Comprehensible to a Non-specialist

The impact of Generative AI with deep learning on machine learning is huge, which brings tremendous advancement in machine learning and its applications. It can be used to generate “fake” images for image processing. Many people misunderstand that “generating fake images for chaotic uses” be its major applications. Actually, there are many useful applications with good academic, industrial and commercial uses. In an ideal situation, we can train a latent space with a latent vectors z which can map to a desired photo-realistic image (latent) space. This **will turn an image processing problem into a searching problem**, and be able to generate new images which are not available in the training dataset. Its applications include image super-resolution, image inpainting with its generative power, image/video compression by transmitting LR image with HR quality display at the receiving end; augmentation of small datasets, mixed-reality for virtual try-on, etc. etc.

Recently, diffusion models have been considered as the most effective way to make image generation. It depends on noise adding for training and denoising for testing. Noise is a means to bring and to form the required statistical distributions. Hence noise properties and distribution matching are the major concern of diffusion models. The **major novelty of this proposal is to break away from the requirement of using explicit noise operations as a means for processing** in our proposed model which is structurally close to Diffusion Models (DM). **The concept is new. Without any noise adding and denoising process,** results of our initial test show that the super-resoluted images and images generated match or even better than (for x16) those from Diffusion Models. This is extremely encouraging.

Proposing an image Super-Resolution with Domain Transfer: Note that there is a certain data distribution in the LR image domain. The super-resolution process is **to transfer the LR data distribution into the HR data distribution domain**; hence **this is a process of domain transfer**. If we make the transfer abruptly from LR domain to HR domain, it may be too drastic to achieve high quality. Hence, we suggest to make the transfer progressively, even say for example from 511 x511 to 512x512 to one extreme. We suggest that **the transfer does not need to make use of noise as the mechanism**, but it can be achieved by making use of a time-step U-net (or any network to be studied in details), to make the domain transfer with distributions closer and closer to the final photo-realistic domain. Our initial test shows that this does not work well, but the approach sounds. The problem could come from errors accumulation during domain transfer, which may make

the statistics of the input images differ too much from the training situation. To get away from this problem, we suggest to guide the tour to the right direction with reference to the source LR images. **This idea appears promising**. However, is the U-net good? What is the interval for the transfer? What is the right correction step? …

Commented [HHAC1]: To solve this problem?
-- Already used "get away from" using Gaussian Noise in Objectives section.

<u>Creating a Generative Model</u> from our Domain Transfer in Latent Space structure: How does the domain transfer approach relate to image generation? We can consider the LR image (say a 32x32 image) be a latent vector. Apparently, if we can randomly assign pixel values to this latent vector, it may generate a face or even new faces if the structure has been trained by a face dataset. It does not work, since the statistic of the randomly generated latent vector does not match the properties of the training dataset. To make this up, we can start with a general latent vector, z (say 1x100) and construct a deep circuit to convert z domain into a 32x32 LR image space with the required statistical distribution. There are many possible designs of the circuit, we may simply use GAN or StyleGAN. This forms a new direction for our research.

The major novelty of this proposal is to start using a **Diffusion like model without denoising** to effect image super-resolution with the concept of **domain transfer**, and subsequently **to generalize it to form an image generative model** with deep circuit for distribution matching.