Re-Imagen

- □□□□□RE-IMAGEN: RETRIEVAL-AUGMENTED TEXT-TO-IMAGE GENERATOR
- □□□□□https://arxiv.org/abs/2209.14491
- ICLR 2023

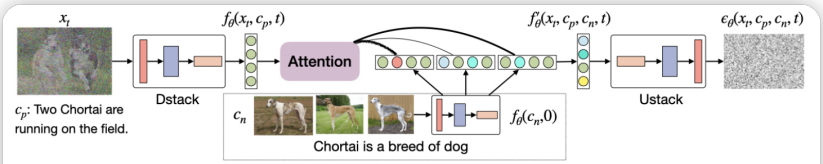

Figure 3: The detailed architecture of our model. The retrieved neighbors are first encoded using the DStack encoder and then used to augment the intermediate representation of the denoising image (via cross-attention). The augmented representation is fed to the UStack to predict the noise.

□□□□