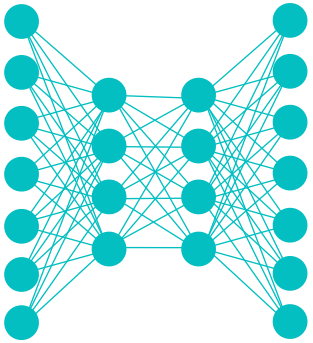
Lecture Notes for

Neural Networks and Machine Learning

Stable Diffusion



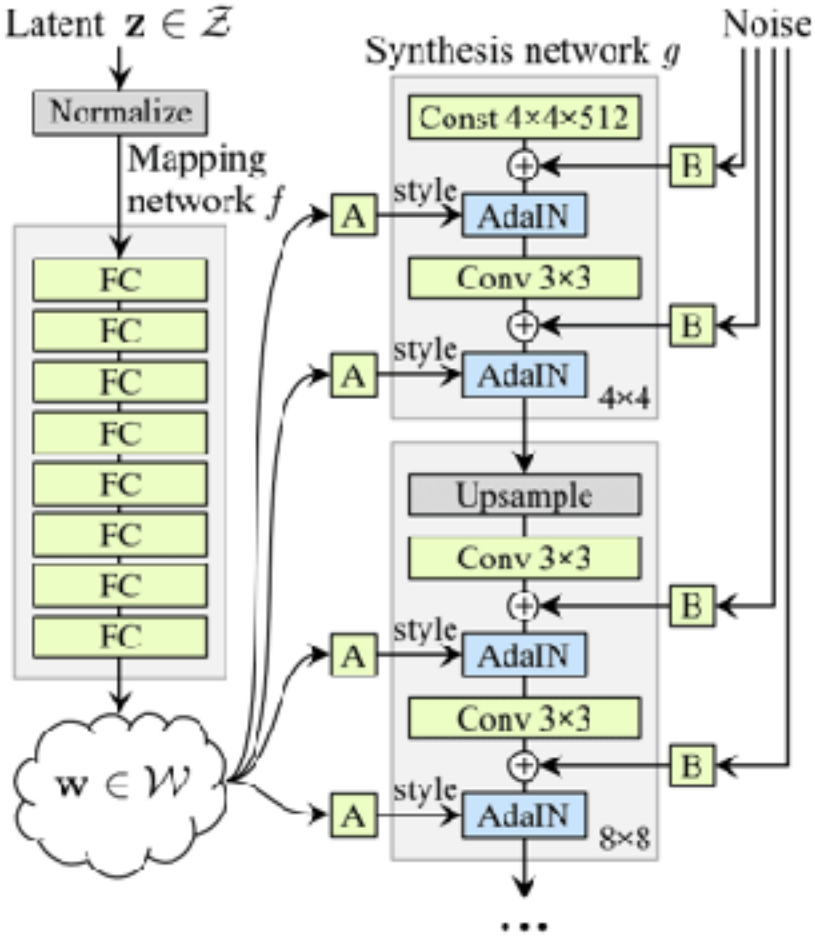
Logistics and Agenda

- Logistics
  - None
- Agenda
  - Stable Diffusion
  - Final Project Town Hall

Stable Diffusion

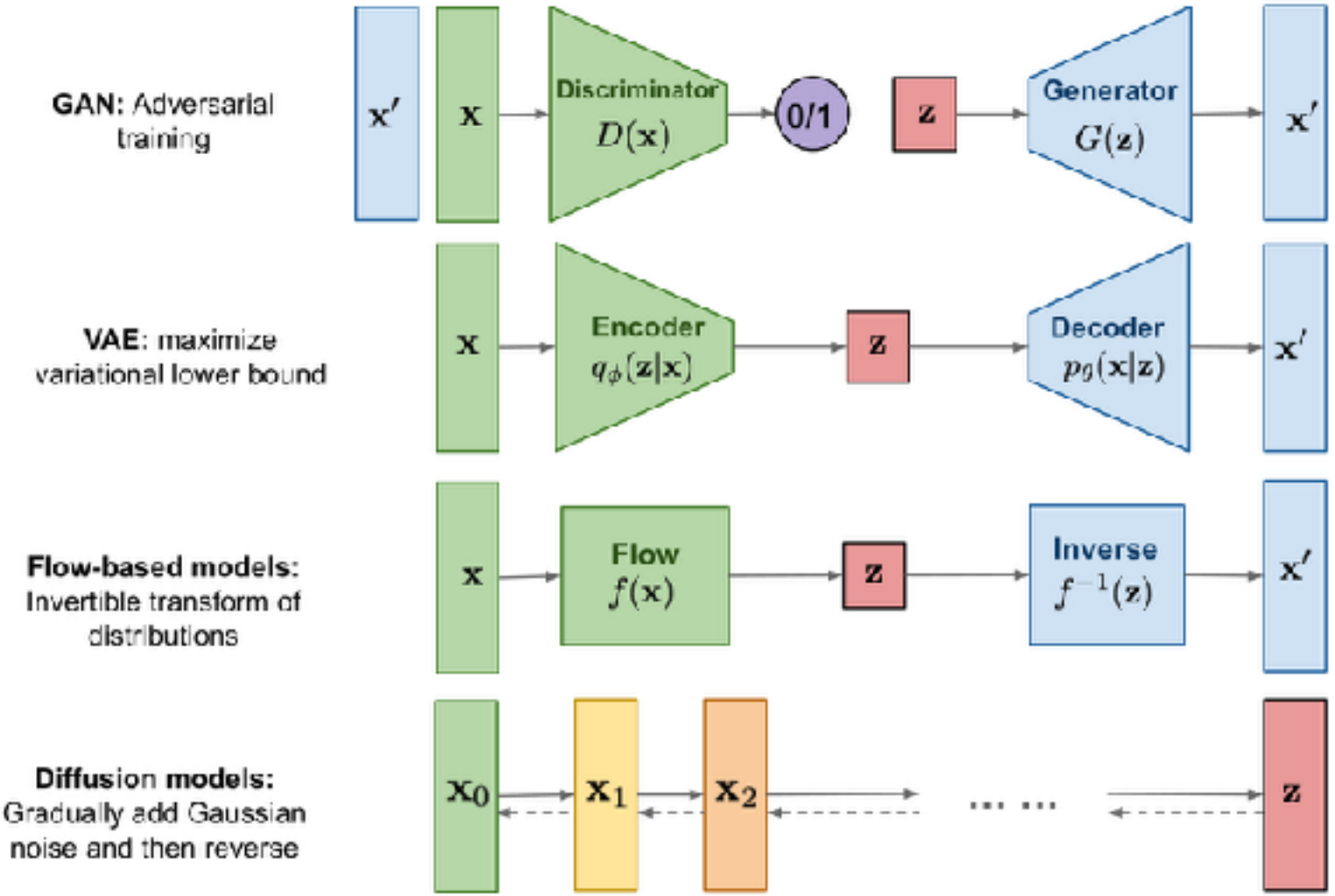
Recall StyleGAN

- A latent sample was chosen that initialized a “mapping network”
- Processing of this vector took place and representations “A” were added to a synthesis network
- Noise was systematically added to the network activations
  - Why?



Comparative Overview

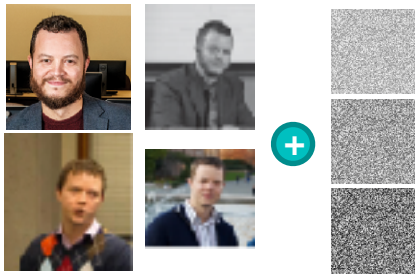
https://learnopencv.com/image-generation-using-diffusion-models/



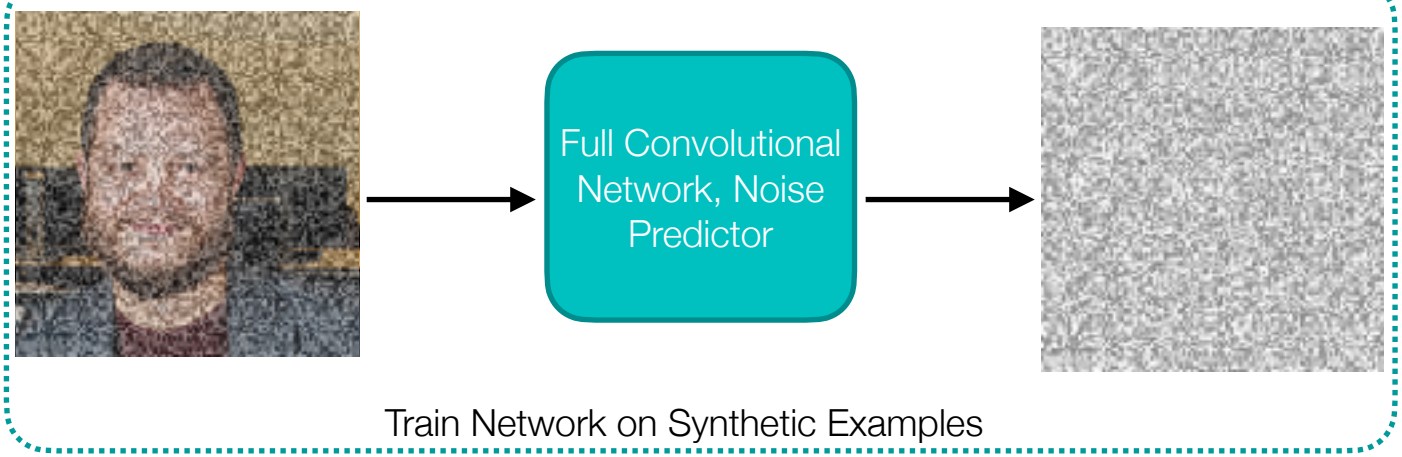
A great resource for understanding at high level: https://jalammar.github.io/illustrated-stable-diffusion/

The Diffusion Process, Simplified

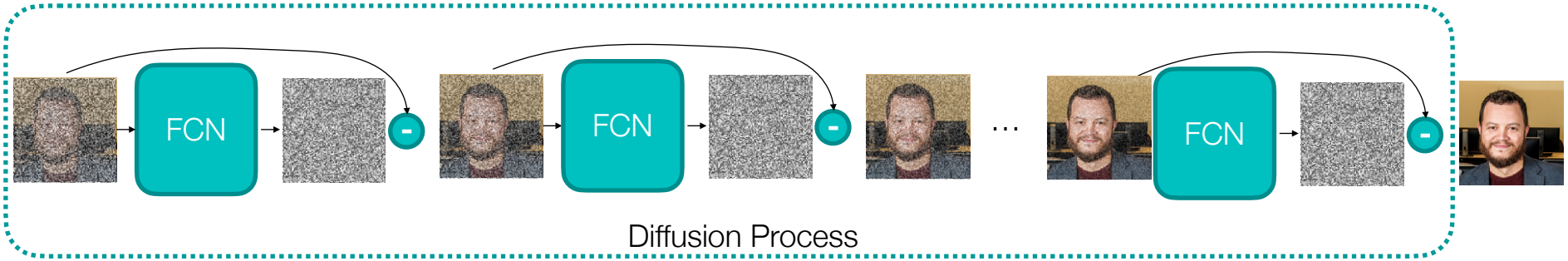
- Guiding Example: Predict noise sample in an image



Create Dataset



Train Network on Synthetic Examples



Diffusion Process

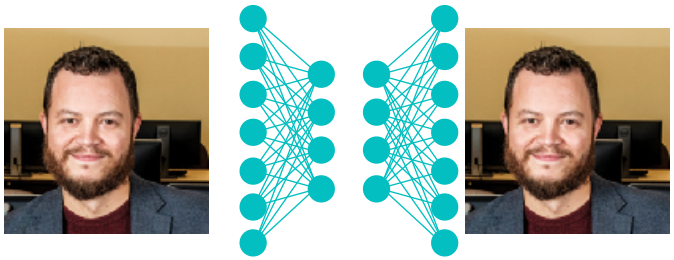
- Now we could generate great looking images from noise!!

Departure to Latent Space

Departure to Latent Space Our approach starts with the analysis of already trained diffusion models in pixel space: Fig. 2 shows the rate-distortion trade-off of a trained

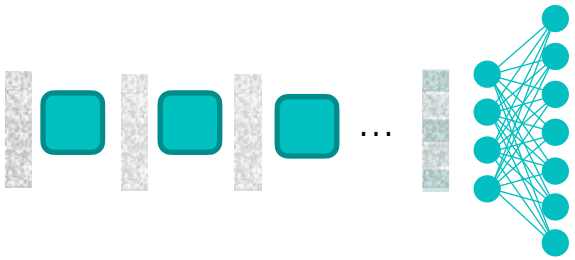
Rombach et al., 2022, https://arxiv.org/pdf/2112.10752.pdf

- Start with a nice auto encoder
- Train noise prediction in latent space
- Perform diffusion in Latent Space
- Generate new images…



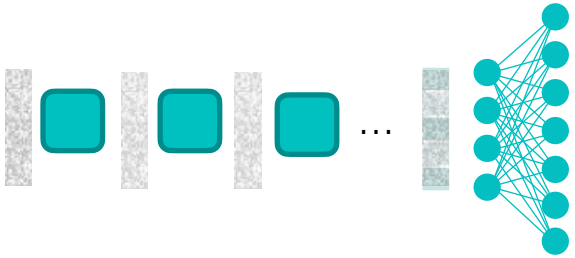
Random Noise

$z \sim \mathcal{N}(0,I)$



Eric Larson, Brown University

$z \sim \mathcal{N}(0,I)$



Eric Larson, Disney Animator

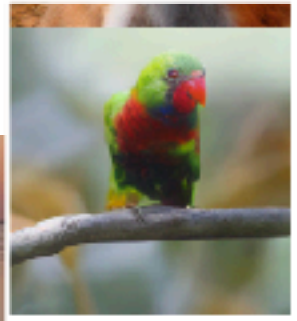
Examples of Diffusion in Latent Space

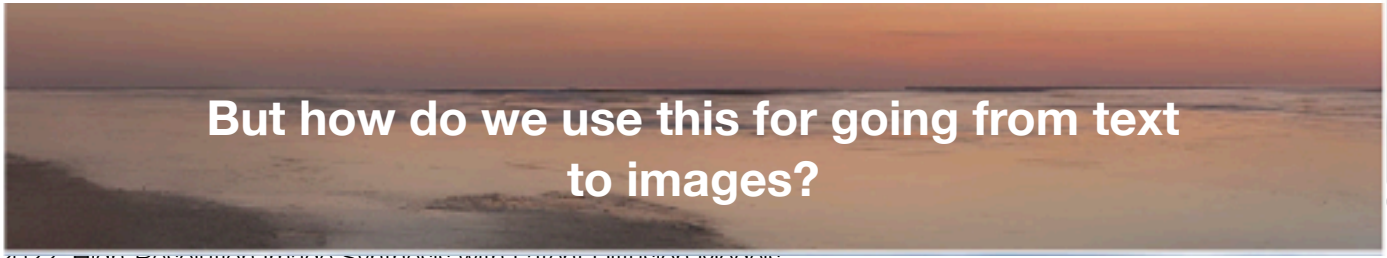
CelebA-HQ 256 × 256				FFHQ 256 × 256			
Method	FID ↓	Prec. ↑	Recall ↑	Method	FID ↓	Prec. ↑	Recall ↑
DC-VAE [63]	15.8	-	-	ImageBART [21]	9.57	-	-
VQGAN+T. [23] (k=400)	10.2	-	-	U-Net GAN (+aug) [77]	10.9 (7.6)	-	-
PGGAN [39]	8.0	-	-	UDM [43]	5.54	-	-
LSGM [93]	7.22	-	-	StyleGAN [41]	4.16	0.71	0.46
UDM [43]	7.16	-	-	ProjectedGAN [76]	**3.08**	0.65	0.46

Method	FID↓	IS↑	Precision↑	Recall↑	N_{params}	
BigGan-deep [3]	6.95	203.6±2.6	**0.87**	0.28	340M	-
ADM [15]	10.94	100.98	0.69	**0.63**	554M	250 DDIM steps
ADM-G [15]	4.59	186.7	0.82	0.52	608M	250 DDIM steps
LDM-4 (ours)	10.56	103.49±1.24	0.71	0.62	400M	250 DDIM steps
LDM-4-G (ours)	**3.60**	**247.67**±5.59	**0.87**	0.48	400M	250 steps, c.f.g [32], $s = 1.5$

LDM-8 (ours, 200 s) 4.02 0.64 0.52 *LDM-4* (ours, 200 s) 2.95 0.66 0.46

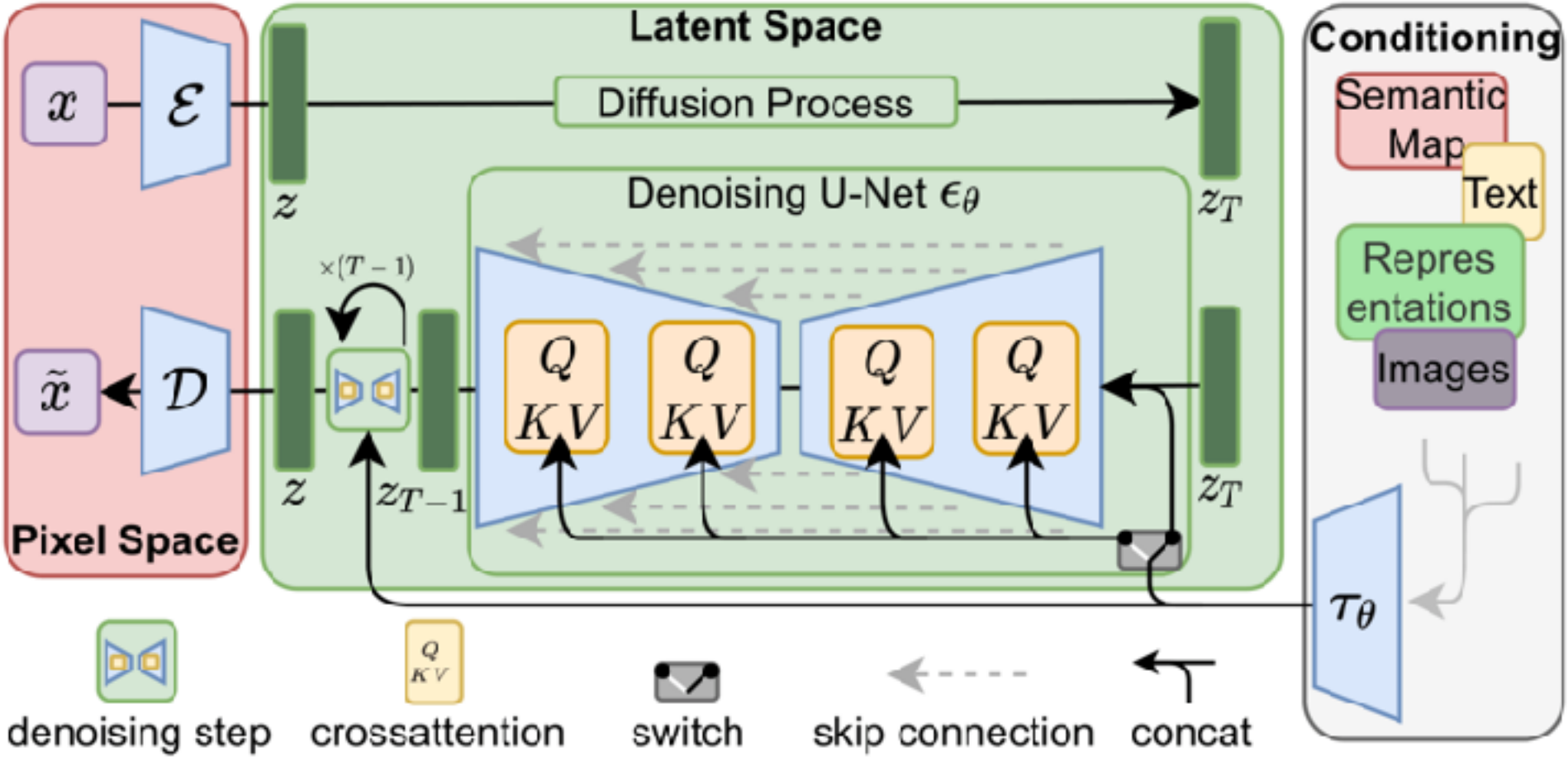
Table 1. Evaluation metrics for unconditional image synthesis.





Rambach et al., 2022, High-Resolution Image Synthesis with Latent Diffusion Models https://arxiv.org/pdf/2112.10752.pdf

Conditioning for Denoising, Overview

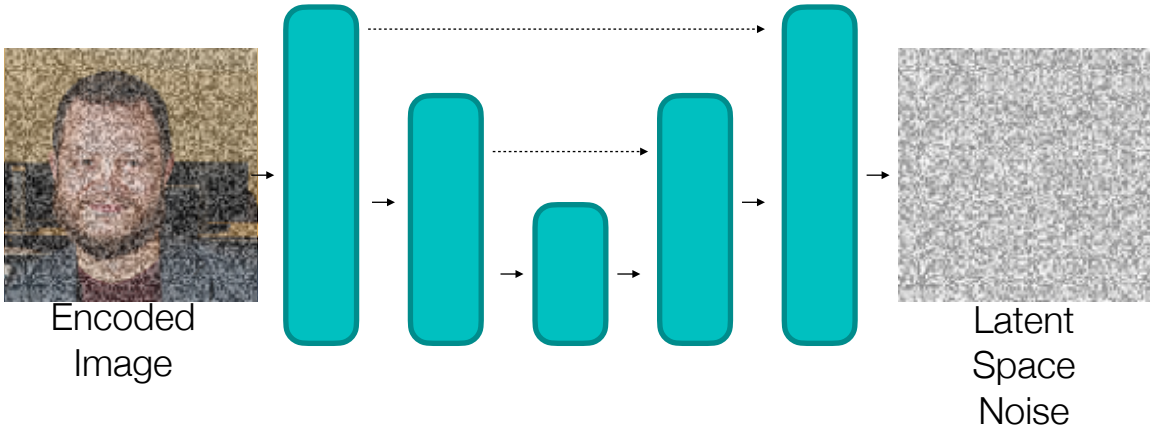


Rambach et al., 2022, High-Resolution Image Synthesis with Latent Diffusion Models https://arxiv.org/pdf/2112.10752.pdf

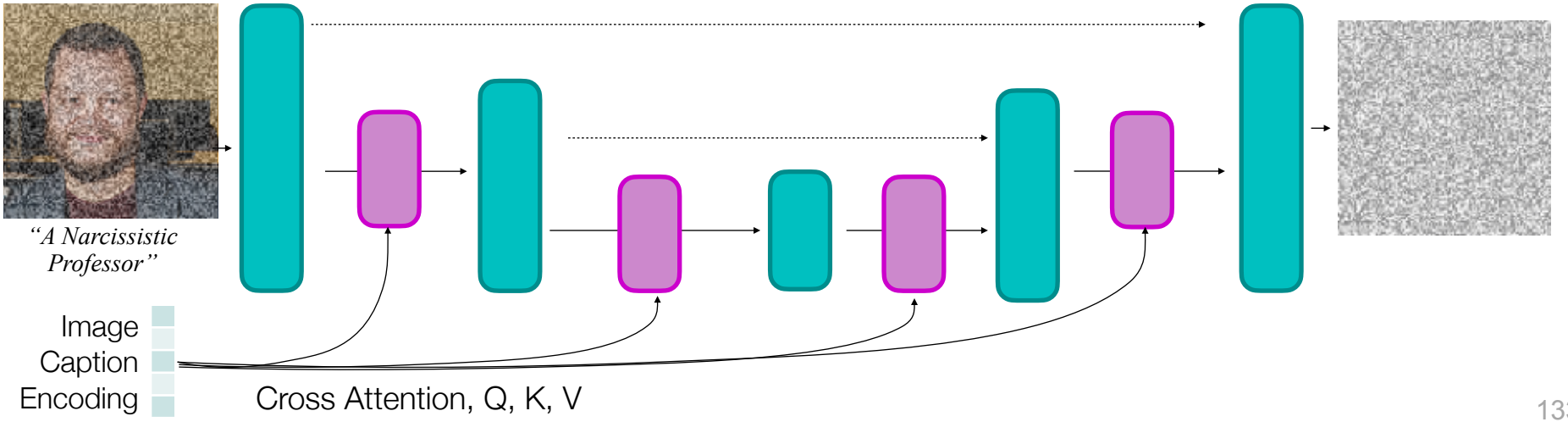
Unpacking "Text Conditioning"

- Employ text embeddings in model denoising

Full Convolutional Network, Noise Predictor



Now model denoising is sensitive to the content within it!



Are all text embeddings created equal?

- No! We need embeddings that are good for images…

Image Encoder Pre-trained

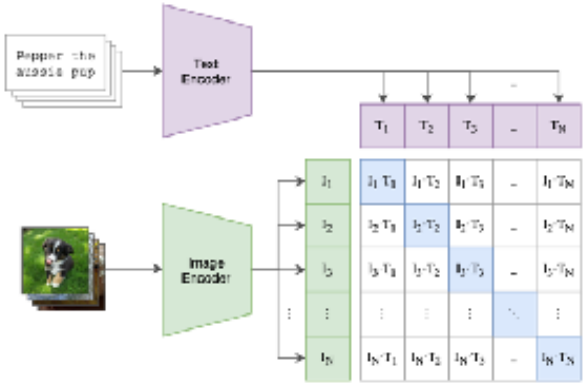
W_i

"A Narcissistic Professor"

Text Encoder Pre-trained

W_t

Positive Examples: Make similar
Negative Examples: Make dissimilar

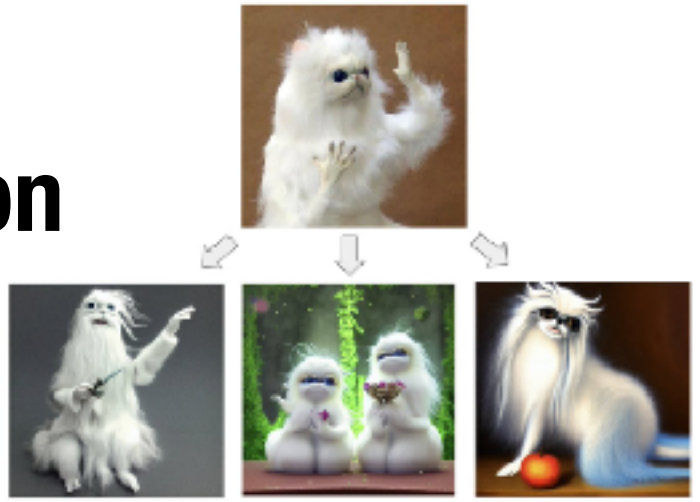


Contrastive Language-Image Pre-training

CLIP

Radford et al. 2021, Learning Transferable Visual Models From Natural Language Supervision, https://arxiv.org/pdf/2103.00020.pdf

Stable Diffusion



High-performance image generation using Stable Diffusion in KerasCV

Authors: fchollet, lukewood, divamgupta
Date created: 2022/09/25
Last modified: 2022/09/25
Description: Generate new images using KerasCV's StableDiffusion model.

View in Colab • GitHub source

Sponsor Follow

LukeWood Luke Wood

KerasCV Author, Full Time Keras team member & Machine Learning researcher @ Google, Part Time UCSD Ph.D student

PRO

Final Project

One Idea from Dr. Stephanie Langin-Hooper
SMU Meadows Professor