

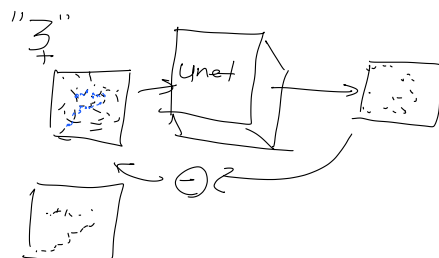
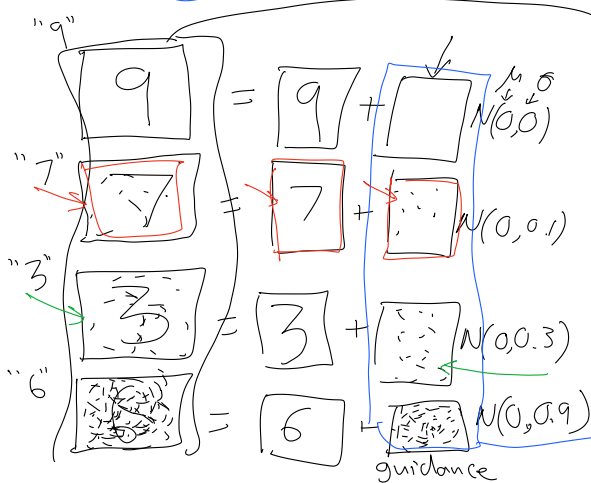
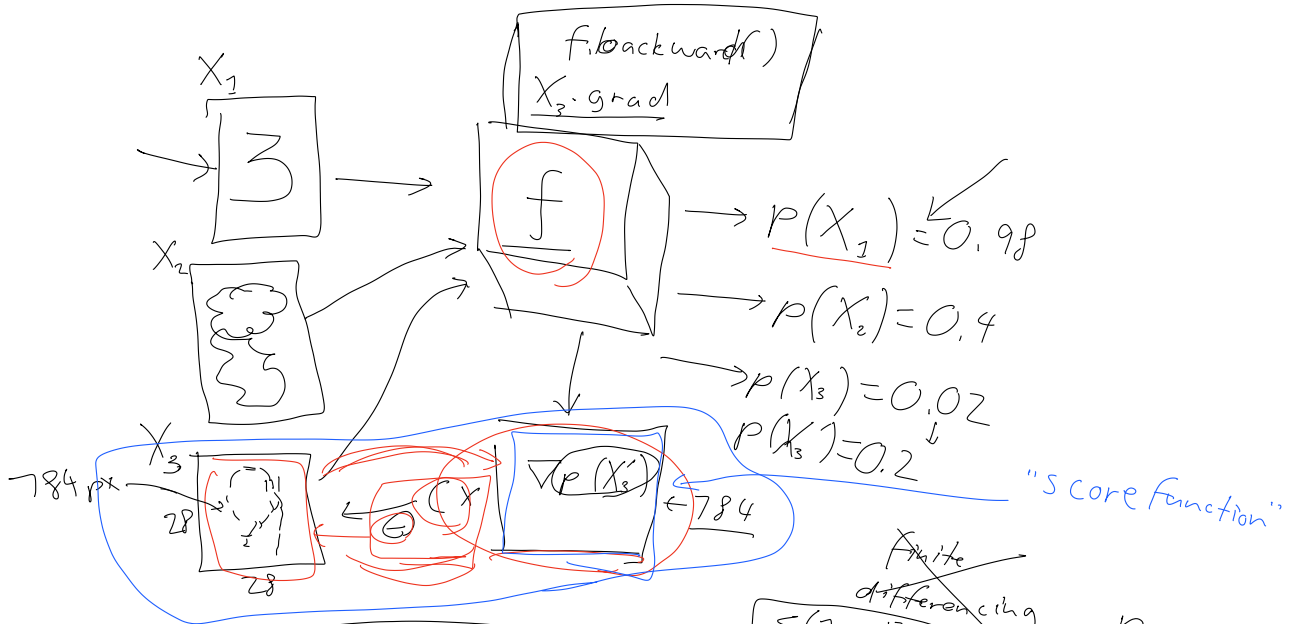
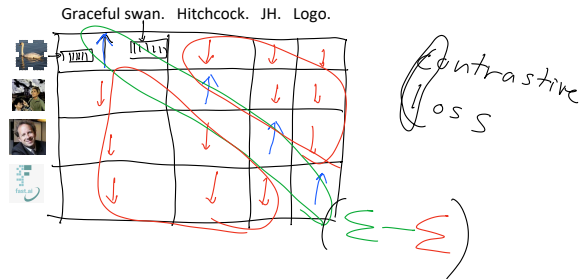
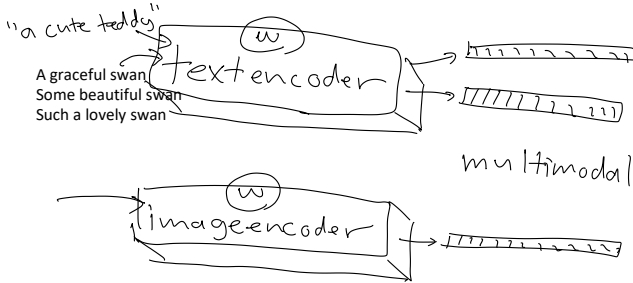
 scene from Hitchcock

 Jeremy Howard

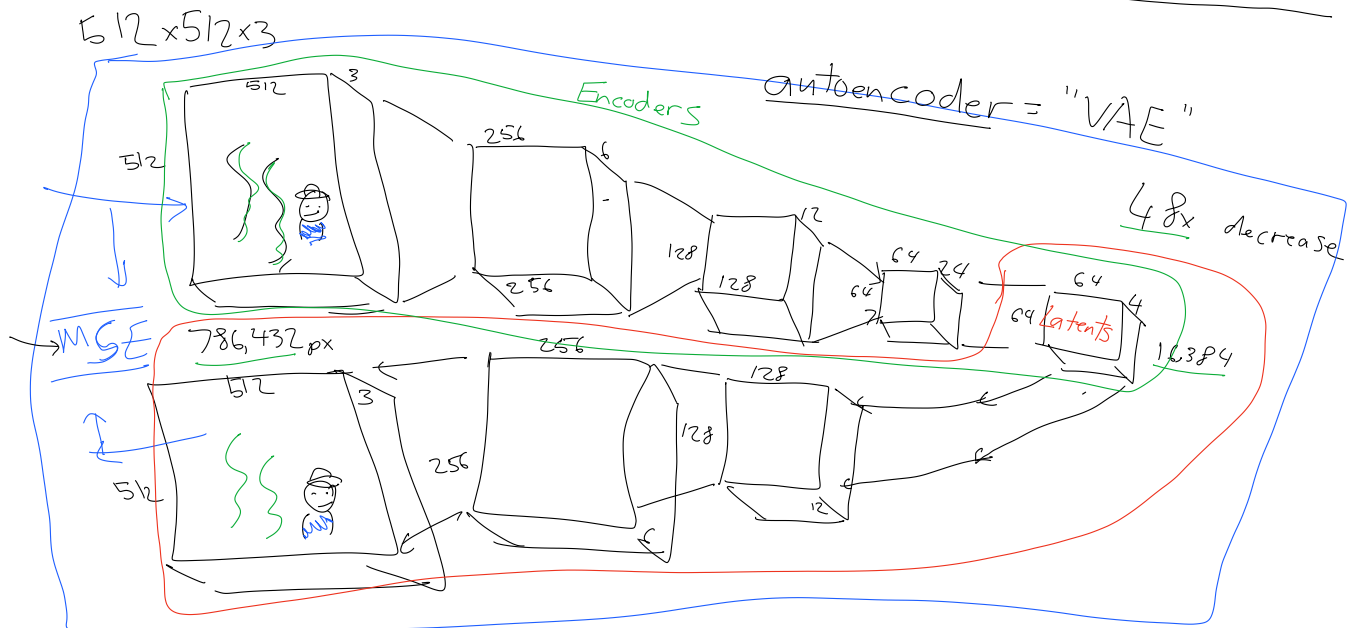
 fast.ai logo

$$\sum \left(\begin{array}{c} \text{image embedding} \\ \text{text embedding} \end{array} \right) \leftarrow \text{dot product}$$

CLIP



inp	out
VAE's decoder	small latents tensor
CLIP Text Encoder	text embedding
UNet	large image
VAE's encoder	small latents tensor
CLIP Image Encoder	image embedding



"a cute teddy"

"time steps"

