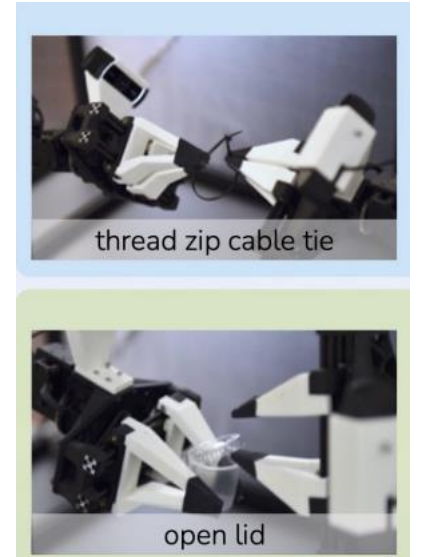
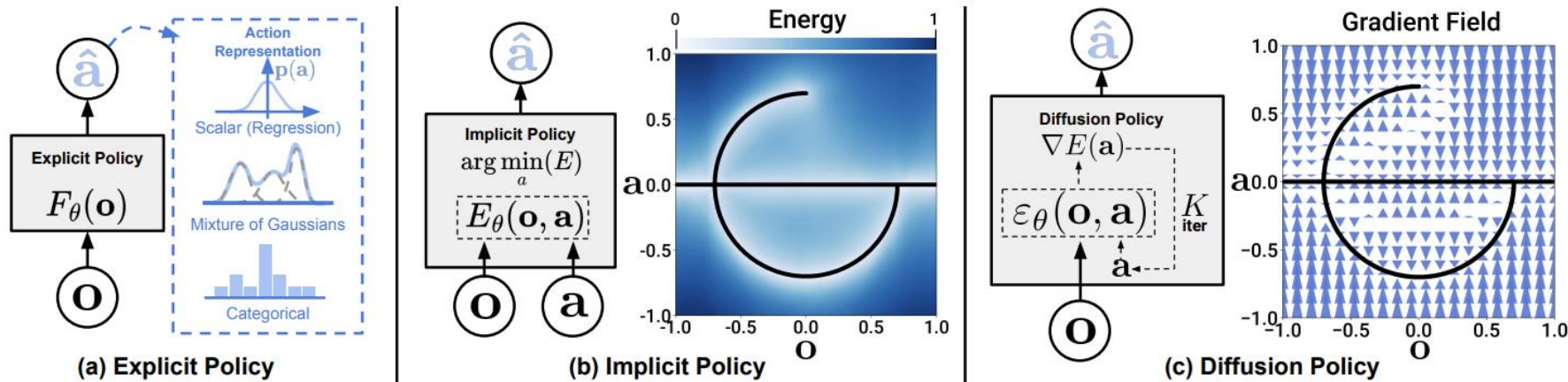


Advanced Behavioral Cloning



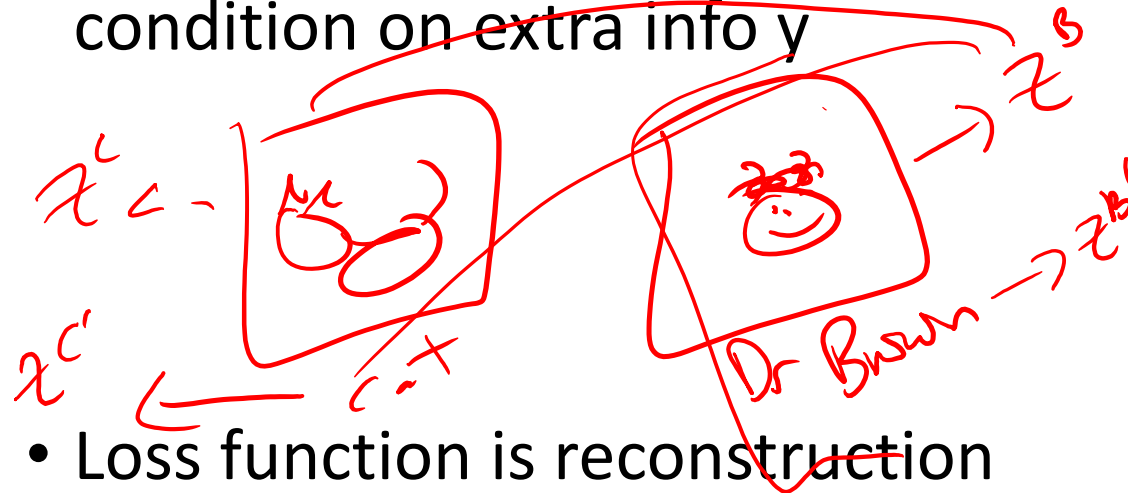
Instructor: Daniel Brown

Action Chunking with Transformers (ACT)

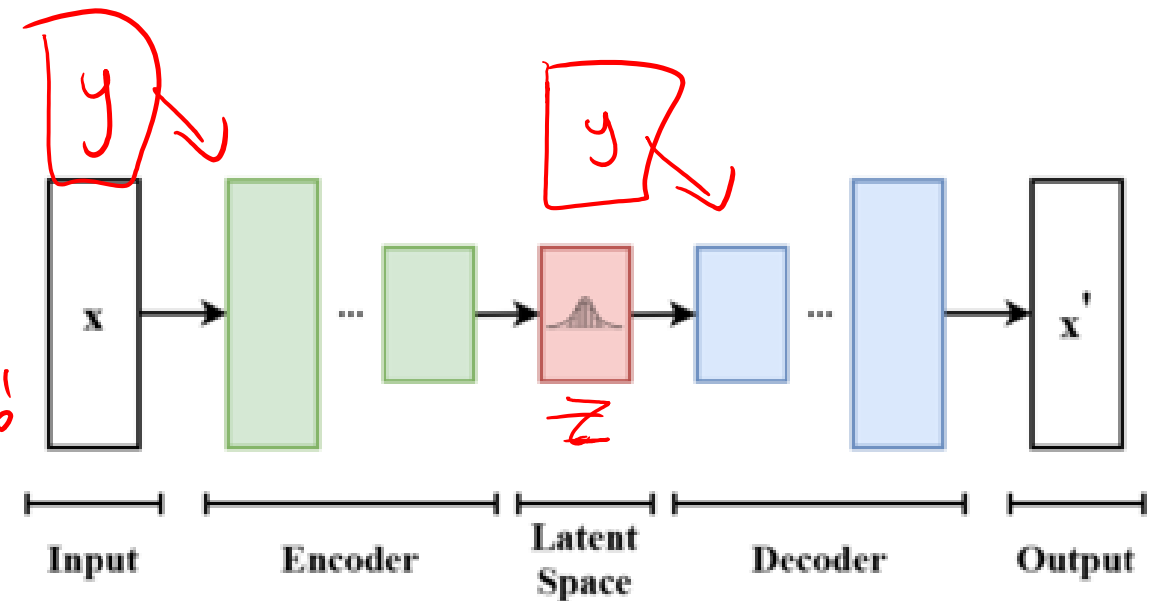
- Paper: <https://arxiv.org/pdf/2304.13705>
- Videos: <https://tonyzhaozh.github.io/aloha/>

Conditional Variational Autoencoders (CVAEs)

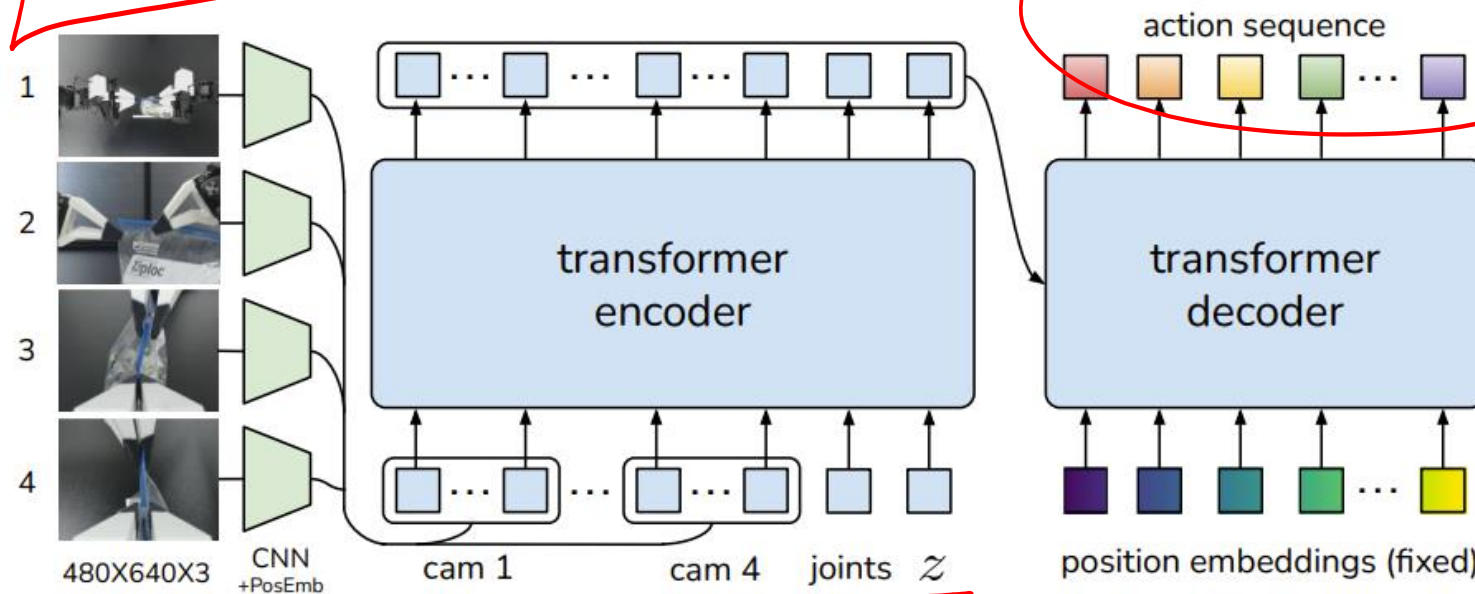
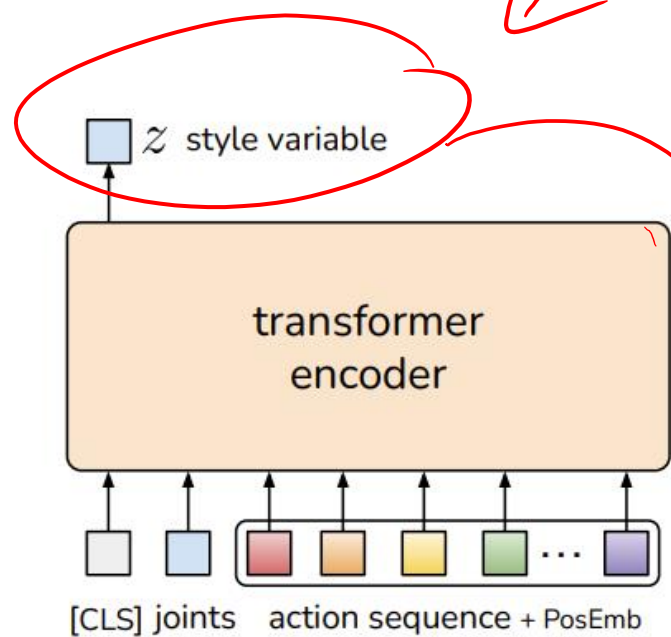
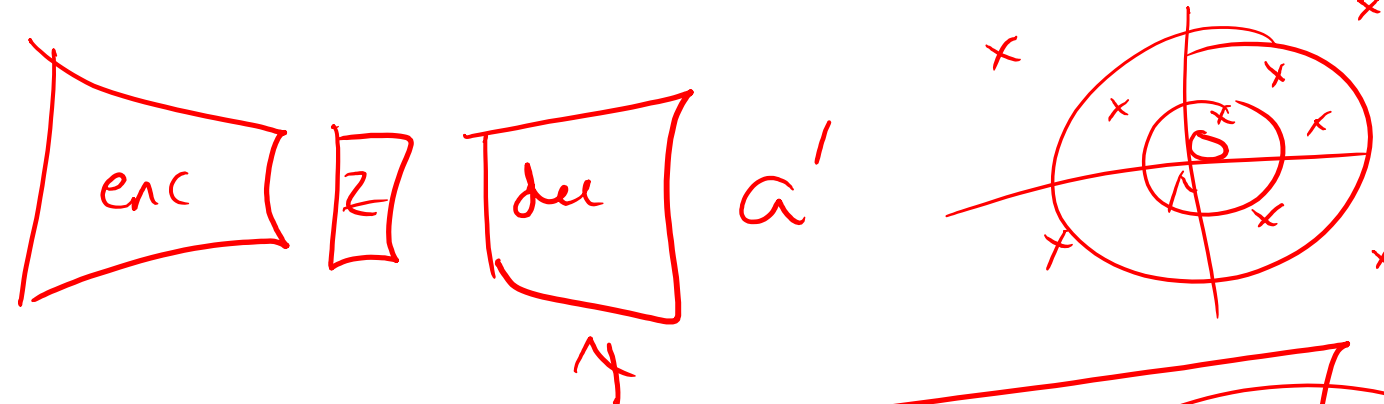
- Encoder and decoder both condition on extra info y



- Loss function is reconstruction plus KL divergence



$$\mathcal{L} = \mathbb{E}_{q(z|x,y)} [\log p(x|z,y)] - D_{\text{KL}}(q(z|x,y) || p(z|y))$$



MSE loss

What is the purpose of the latent variable z ?

- Model human noise
- Multi-modal behavior
- See Figure 8 (c)

Diffusion Policy

- Paper: <https://arxiv.org/pdf/2303.04137v4>
- Videos: <https://diffusion-policy.cs.columbia.edu/>

Denoising Diffusion (high-level)

$x_0 \rightarrow x_1$

$$q(x_t | x_{t-1}) = \mathcal{N}(x_t | \alpha x_{t-1}, \beta I)$$

Fixed forward diffusion process

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

Data



Noise

Generative reverse denoising process

apply 3 steps of noise

ϵ

$\| \hat{\text{noise}} - \text{noise} \|^2$

$t=3$