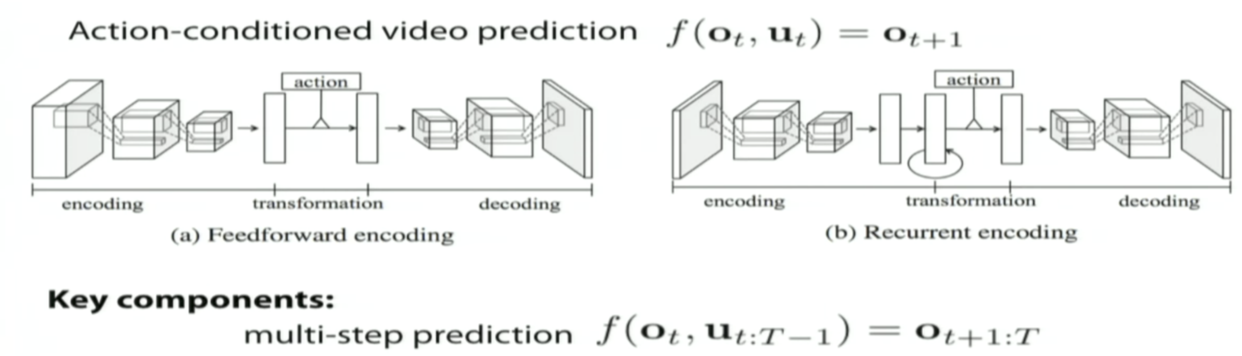
advanced model learning

- we now jump back to model based RL → model here refers to dynamics model
- remember that in practice, we don't get to see the state, we only get to see the observations which could be quite different from state, especially in partially observable environments (POMDPs)
- the state itself could be quite low dimensional (like position of something) but the observation could be high dimensional, like the image of it.
- learning models in latent space:
    - one approach is to learn a model in a much lower dimensional latent space/embedding space. We can use, for example, an autoencoder to do this. we can train an autoencoder to map an input to a latent space (autoencoder training is usually done by minimizing reconstruction loss in some way that forces it to learn useful latent space representations) once trained, we can learn the dynamics in this latent space.
    - the advantage of learning in this much lower dimensional latent space is that we can learn from much lesser data. this is all true only if we are able to get the autoencoder to capture the right information though. its possible that we can't get the autoencoder to learn the right representation.
    - the middle layer of an autoencoder (hidden layer) is sometimes called the bottleneck.
    - since we are representing state in latent space, we would have learn the rewards and the dynamics model in the latent space too. for example, by assigning rewards to latent representations of different states.
    - the trick about learning a good autoencoder is shaping the loss function used for autoencoder training in such a way that it picks out the right information needed → gotta be careful with trivial/degenerate solutions
- learning models in image space:
    - another approach is to do video prediction using either a feedforward architecture or a recurrent architecture (for multi step prediction as actions and images are fed in)



    - we can use this to do model based RL using the prediction as the dynamics model.
- another approach is to make up some alternate metric that we will predict. instead of predicting the full state given a state and action, we can predict if a particular action would get an end effector closer to a goal. this is more like a hand tuned rule, and going away from the "just learn it" approach.