

Generative models

Supervised vs Unsupervised

Supervised learning

- Take (x,y) pairs

Unsupervised learning

- Take x alone

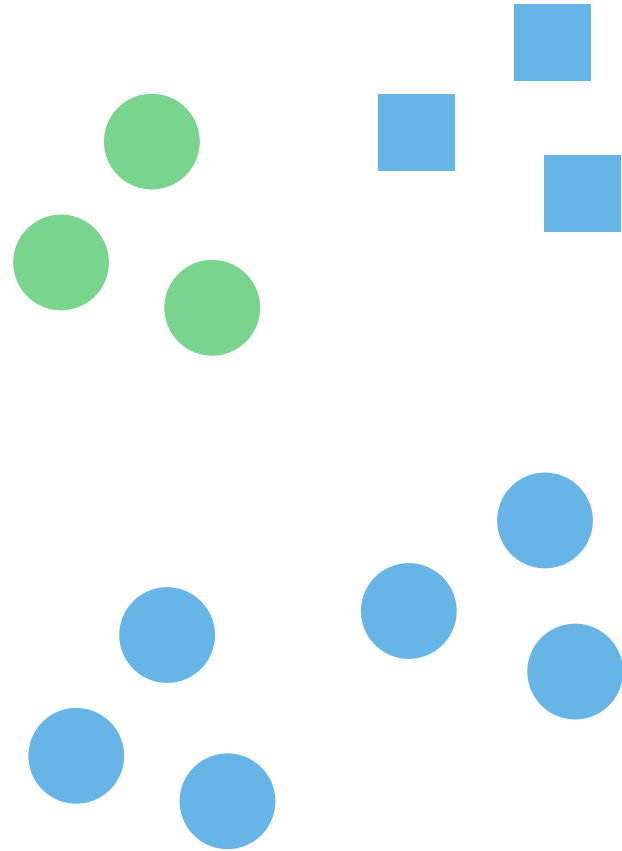
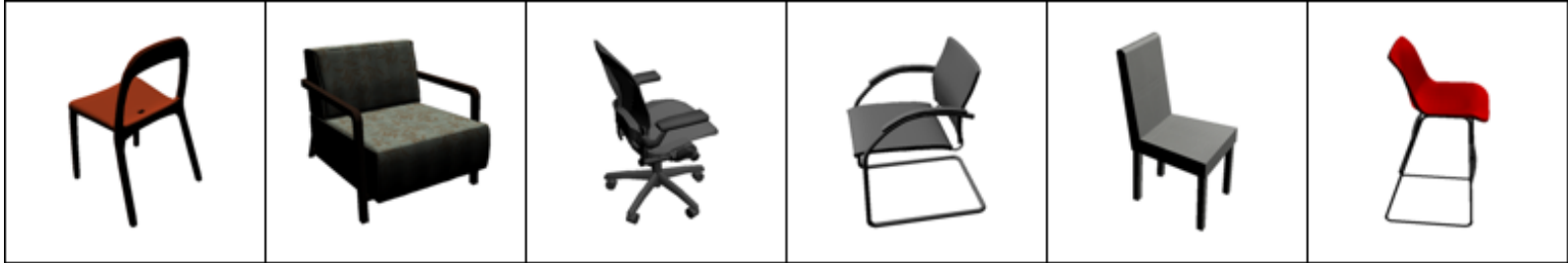


Image generation

Chairs (type, view, orientation)



Classifier

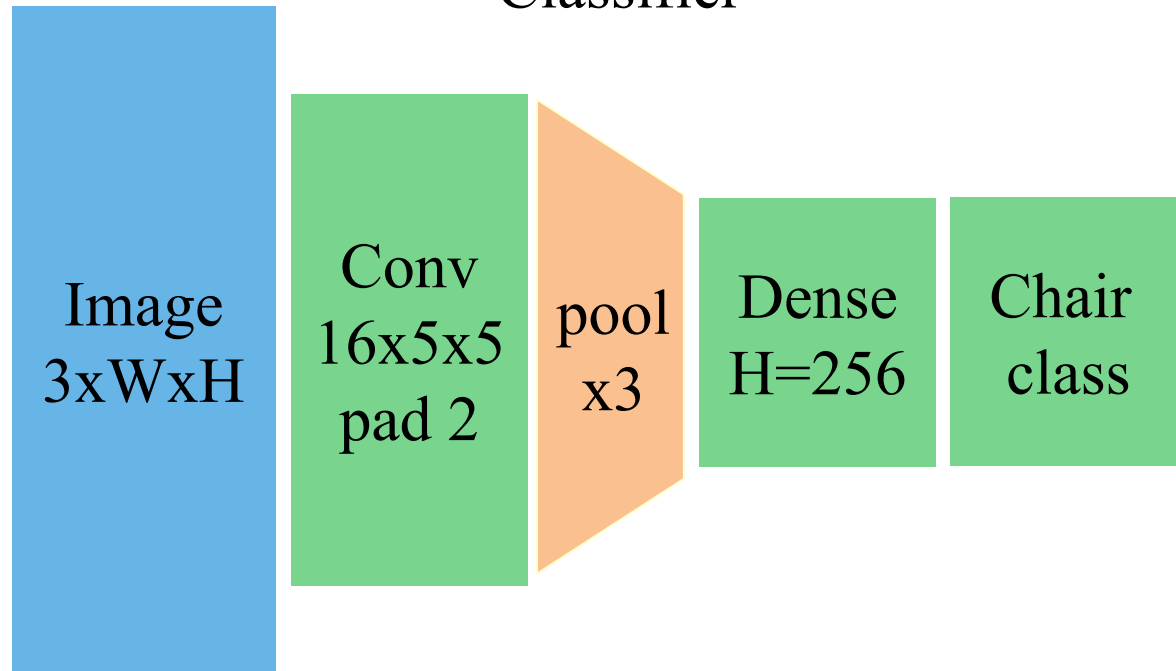
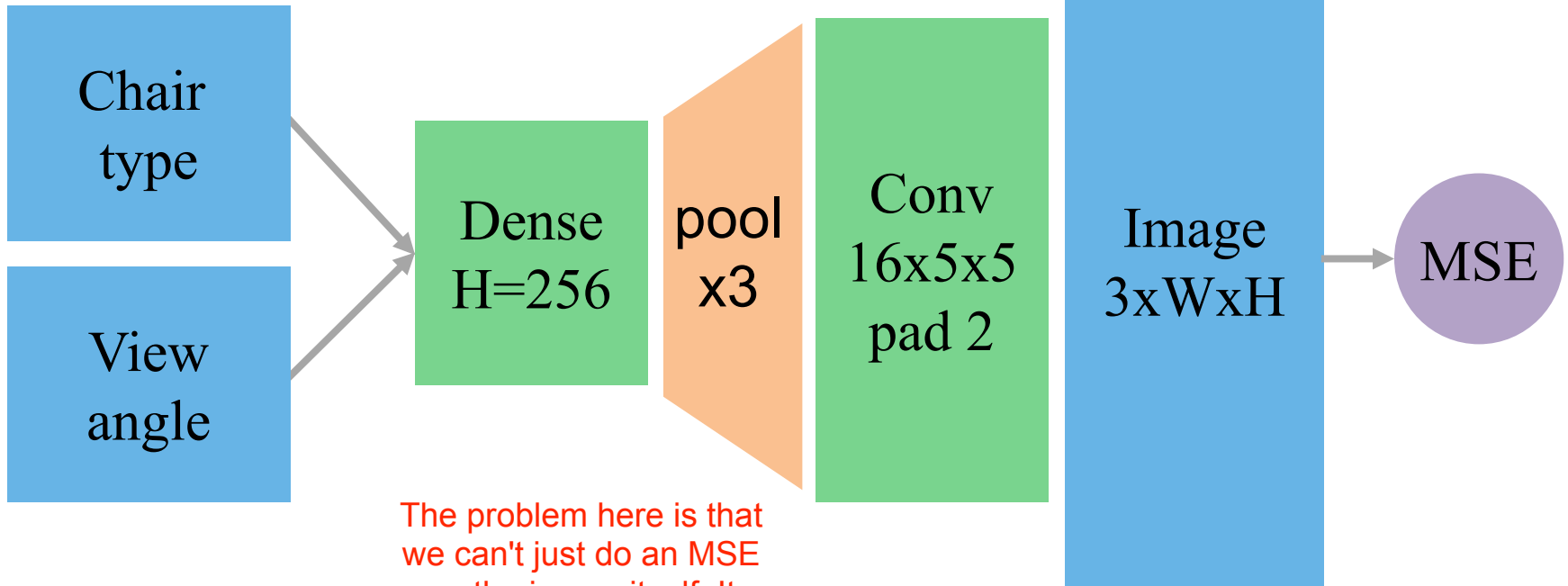


Image generation

Generator

Given a chair type, and a viewing angle, can we generate a chair image?



The problem here is that we can't just do an MSE on the image itself. It would be too 'literal' and low level, and when we optimise for gradient descent, we can't get good results.

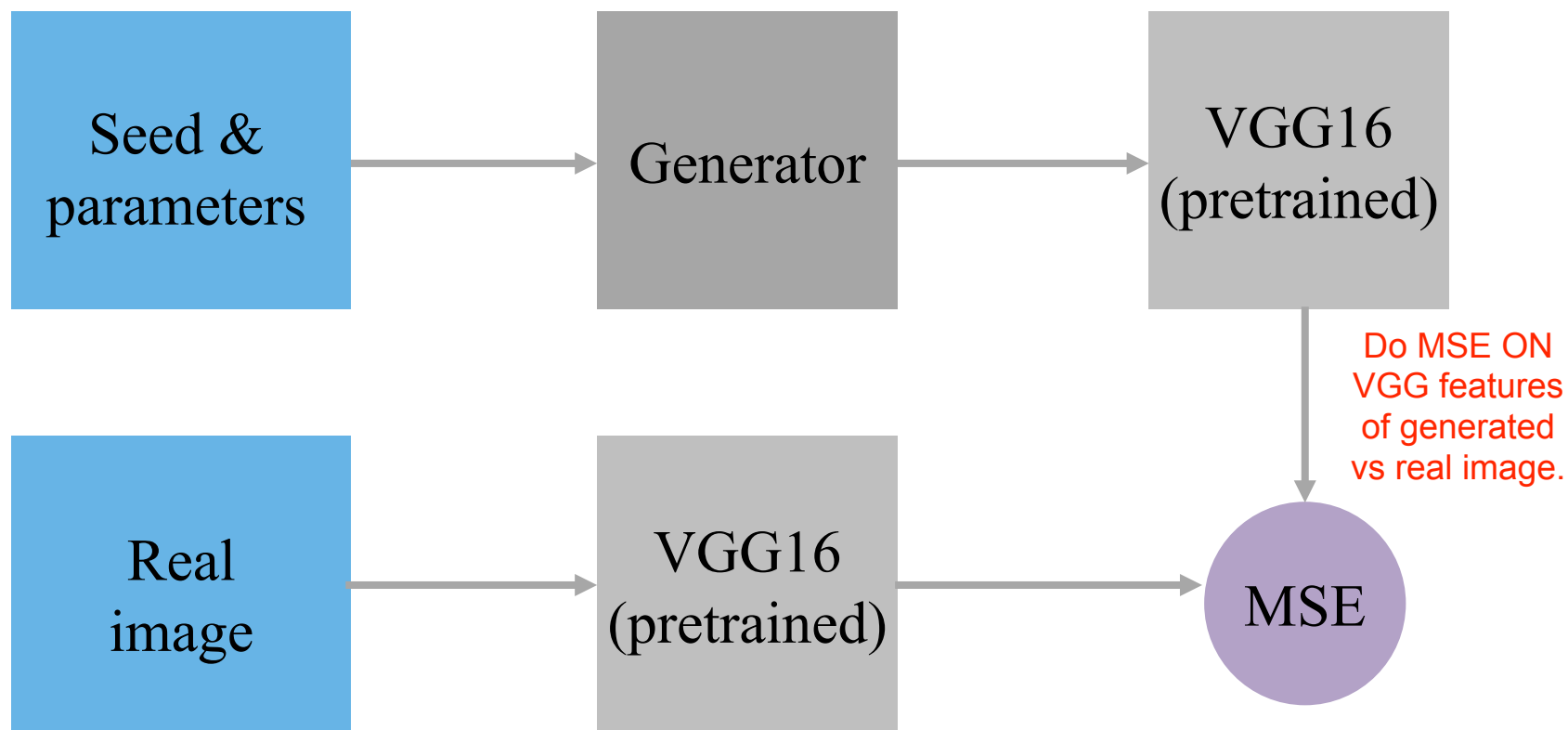
Mean Squared Error

Pixelwise MSE:

- A «**cat on the left**» is closer to «**dog on the left**» than to «**cat on the right**»
- We may want to avoid that effect
- Can we obtain image representation that is less sensitive to small shifts?

Problem: MSE sucks at this task.
Ideas?

Sketch: using pre-trained nets



We chop off the final layers (perhaps before the FC / dense layers) of VGG, and then flatten these to act as the 'embedding' of the image.

This is the $f()$ function below:

$$L = \|f(img) - f(Gen(seed))\|$$

our loss function would be to compute the difference between VGG sliced output of a real image, vs a VGG sliced output of a generated one.

We can't just compute the pixel-wise difference. This is because we want translational invariance in our Loss function.