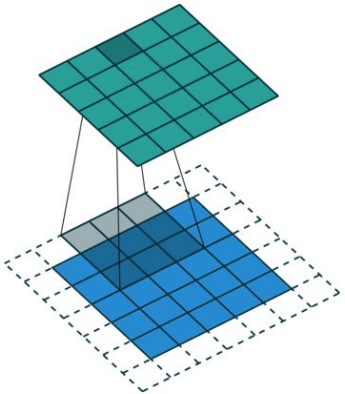
Dataset & Encoder

[1], [2]

Dataset

- Animal Faces from Kaggle/StarGAN v2
- Originally cats, dogs and wildlife → now focus on only cats and dogs
- Image resolution: 512x512 → 64x64
- Data split into
  - validation set of 1000 images (500 cats, 500 dogs)
  - training set of 8743 images
  - test set of roughly 1000 images



[3]

Encoder

- Three convolutional-convolutional-pooling blocks (Conv2D - Conv2D - MaxPool2D)
- Batch normalization in between layers for training stability
- Number of filters doubles after each block (32 → 64 → 128)
- Input dimensions (HxW): 64x64 → 32x32 → 16x16 → 8x8

[1] https://www.kaggle.com/datasets/andrewmvd/animal-faces
[2] https://github.com/clovaai/stargan-v2
[3] https://stackoverflow.com/questions/62166719/padding-same-conversion-to-pytorch-padding

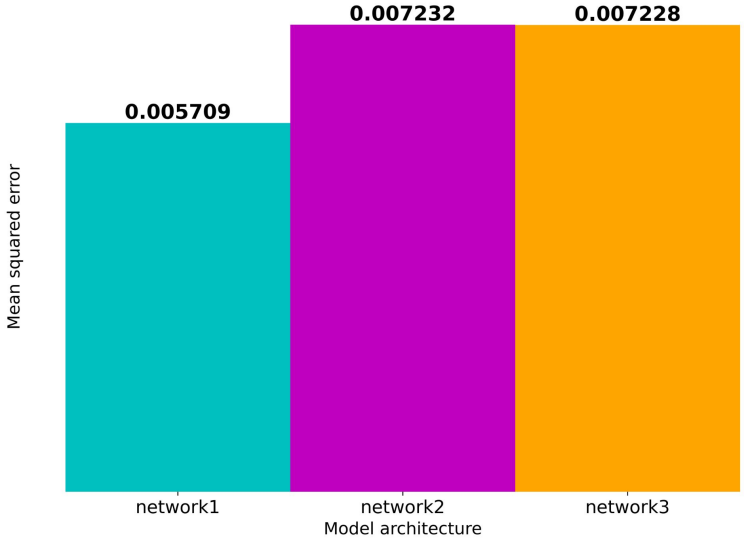
Decoder and bottleneck selection

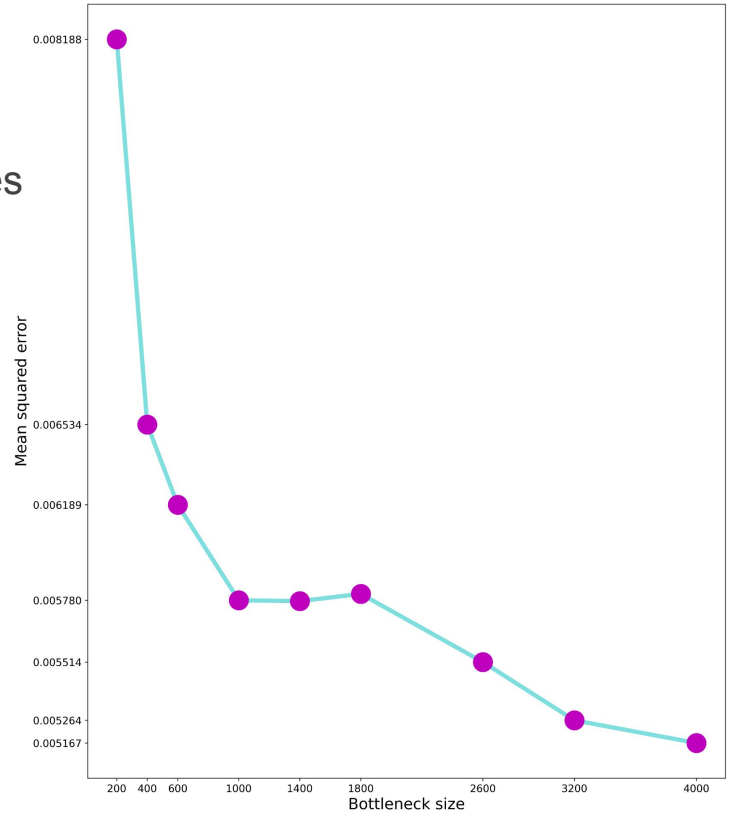
Three decoders with large latent dimension (4096), 3 to 1 compression, ~ 17 M trainable parameters

- Differ in transposed convolutions, upsampling and strides

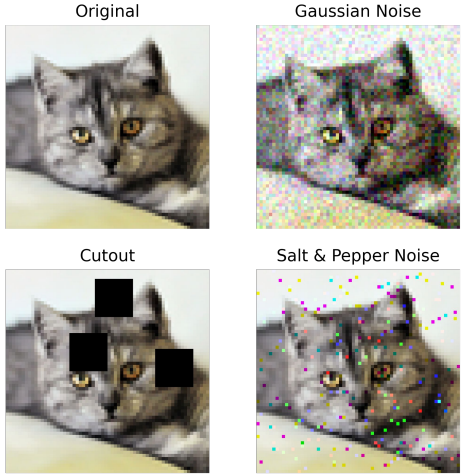
Bottleneck (latent dimension) sizes from 200 to 4000 tested

- Bottleneck size 1000 selected, ~ 12.2 to 1 compression



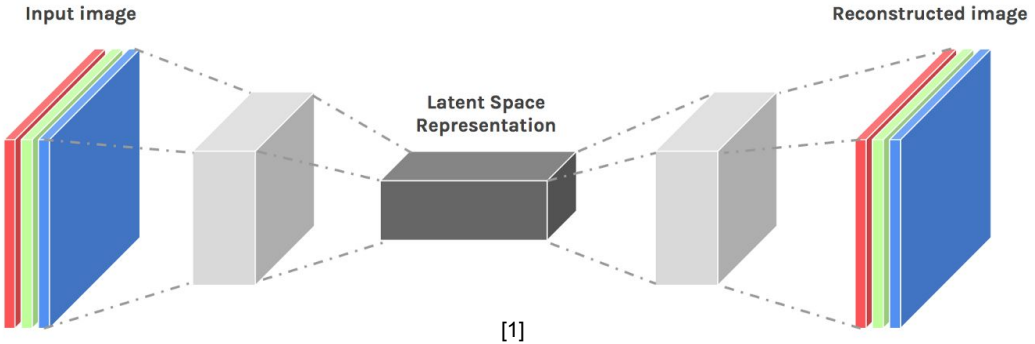


Perturbation application & Results

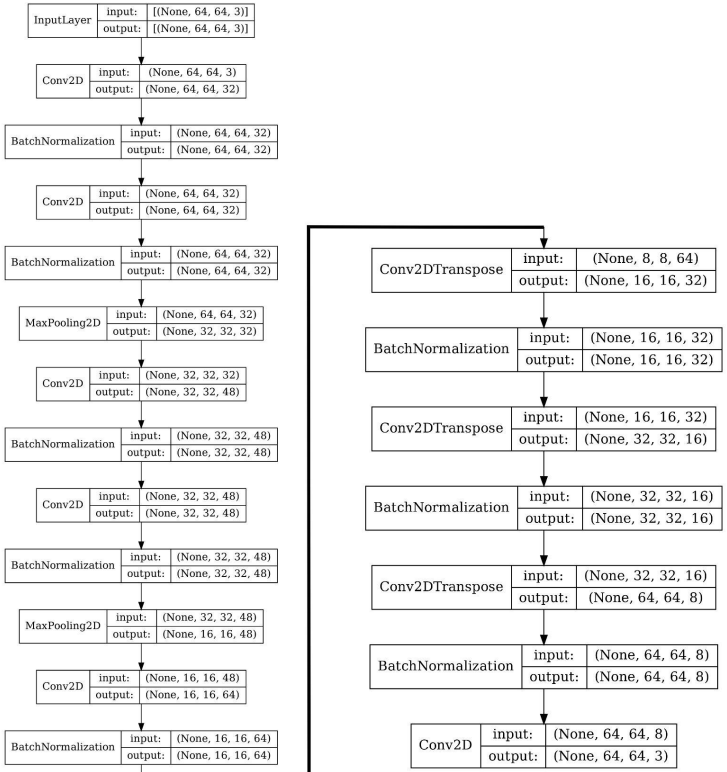


Distortion	Test error
Coarse	0.00765
Gaussian	0.00673
Salt and Pepper	0.00676
Coarse + Gaussian	0.00738
Coarse + Salt and Pepper	0.00729
Gaussian + Salt and Pepper	0.00626
Coarse + Gaussian + Salt and Pepper	0.00668

Fully-Convolutional Autoencoder



[1]



[1] https://towardsdatascience.com/autoencoders-introduction-and-implementation-3f40483b0a85