Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks
Radford, Metz, and Chintala, 2015
https://arxiv.org/abs/1511.06434

Architecture

- Uses strided convolution instead of maxpooling, allowing the model to learn its own spatial downsampling.
- Uses an all-convolutional net with only 2 fully connected layers (from z to first convolution in generator, and from final convolution of discriminator to output).[1]
- Last layer of discriminator is flattened to fed into a single sigmoid output.
- Uses batch norm on all layers except generator output layer and discriminator input layer.[2]
- The generator uses ReLU activation everywhere with the exception of the output layer, which uses Tanh.[3]
- The discriminator uses LeakyReLU activation everywhere.

Training

- Images were scaled to [-1, 1], the range of the Tanh activation function. Otherwise no pre-processing was done.
- Trained with minibatch SGD with a mini-batch size of 128.
- Weights were initialized from a Normal distrbution with mean 0 and variance 0.02.
- The slope of the leak for LeakyReLU was set to 0.2 in all models.
- Uses Adam optimizer with $\alpha = 0.0002$ and $\beta_1 = 0.5$.

[1] Using global averaging pooling with fully connected layers resulted in increased model stability but hurt convergence speed.
[2] Using batchnorm everywhere resulted in sample oscillation and model instability.
[3] Using a bounded activation allowed the model to learn more quickly to saturate and cover the color space of the training distribution.