1 Summary

InfoGAN learns interpretable and disentangled latent representations in a completely unsupervised fashion. The paper accomplishes this by decomposing the input GAN noise into two parts, incompressible noise z and latent code c. An information maximization term is added to the usual GAN loss, which maximizes the information between latent codes and the generated samples guided by latent codes. However, this objective function is intractable since we don't know the true posterior distribution, which is why a variational lower bound is proposed which becomes tractable with the re-parameterization trick. Experiments reveal that infoGAN can learn interpretable latent codes corresponding to visual factors like azimuth, elevation and lighting.

2 Strengths

- The paper is supported by a clear theoretical motivation instead of just throwing things together and presents convincing experimental evaluations which back their claims.
- Training infoGAN only adds negligible computation costs over simple GAN and is easy to train.
- Since the method does not use any domain specific knowledge, it can be applied to a wide range of generative tasks.

3 Weaknesses

- The paper does not exactly mention why the incompressible noise z is needed. What would happen if we completely remove z and increase the dimension of c to be equal to z. Will some portion of c act as pure noise to GAN with the other part acting as latent code ?
- There is no mention whether all latent codes are always representable. Intuitively it makes sense that complex tasks should have a larger dimension of c, but what happens if a large c is used for a simple task like MNIST. Will the latent code still be disentangled?

4 Critique of Experiments

- To verify that infoGAN is able to learn disentangled representation, authors train it on MNIST dataset. Latent code is a vector of dimension 3. The first dimension $c1$ is uniformly sampled from categorical distribution with 10 classes while $c2$ and $c3$ are sampled from uniform distribution. It is observed that $c1$ captures digit style while $c2$ and $c3$ capture continuous variations like rotation and width of digits.
- Experiments on 3D faces dataset reveal that infoGAN is able to learn interpretable representations like azimuth, elevation and lighting. The latent code is of dimension 5 and all numbers are sampled from uniform distribution. The same representations are learne by DC-IGN method but under supervised setting. On chairs dataset, infoGAN is able to learn concept of rotation using a single continuous code.

5 Follow Ups/Extensions

- The experiments are performed varying only one latent code while all other latent codes are fixed. It would be interesting to see what happens if more than one latent codes are varied together. For example, in MNIST experiment, if $c1$ and $c2$ are varied together, will the digit type and rotation angle both change together ? To be precise, are the transformation linear in terms of variations in latent codes ?
- In MNIST experiment, $c1$ is uniformly sampled from categorical distribution of $K = 10$ classes. Ironically, MNIST has exactly 10 classes and $c1$ basically corresponds to class type. It would be interesting to see what happens if we change K.