PixelGAN Autoencoders

Abstraction of representational behavior is a pivotal aspect from the generational viewpoint. The work proposes the PixelGAN autoencoder which abstracts global and local representations by utilizing a convolutional autoregressive neural network in conjunction with a Generative Adversarial Recognition (GAN). The generative path consists of a PixelCNN with the recognition path comprising of the GAN. PixelGAN, when combined with different prior distributions, results in different abstractions of latent code and autoregressive decoder.

The PixelGAN autoencoder combines autoregressive convolutional neural networks (the PixelCNN) with GANs in the conventional encoder-decoder style. Images are encoded using the posterior $q(z|x)$ which is utilized in the generative and recognition pathways. At the generative end, the PixelCNN makes use of the encoded representation as adaptive bias to reconstruct the original image by approximating the psoterior $p(x|z)$. This gives rise to two variants which utilize latent codes to construct vectors and add them as adaptive bias. (1) The *location-invariant* bias adds the adaptive bias within each layer whereas (2) *location-dependent* bias adds the term only to the first layer. The recognition pathway, on the other hand, utilizes a GAN to adversarially match the latents $p(z)$ to $q(z)$. PixelGAN autoencoder is jointly optimized using a lower bound on the log likelihood $\log p(x)$. However, the lower bound excludes the mutual information term $I(z;x)$ between latents and inputs. This step prevents z to become independent of x and provisions the generative pathway to be closely informative of the discriminative pathway.

PixelGAN autoencoders, when combined with suitable prior distributions, result in apt decompositions of global and local statistics. In the case of Gaussian priors, the hidden code learns the global information (the *what* in the image) while the autoregressive decoder successfully captures the local information (the *where* in the image). In the case of categorical priors, the latent codes are learned to capture the global class label information whereas the local information about the class distribution is captured by the decoder counterpart. Additionally, the PixelGAN autoencoder demonstrates suitability towards completely unsupervised settings such as clustering and presents improved performance in case of semi-supervised learning on MNIST, SVHN and NORB datasets. However, the approach could be made moer concrete with regards to its formulation. The lower bound objective does not demonstrate an apt justification for the removal of mutual information term. In the case of deterministic decoder, the lower bound is equivalent to maximizing $I(z;x)$ but it also motivates the reconstructed output to be different from the input sample, hence reducing the need for manual noising. In the case of stochastic decoder, the same scenario repeats and is more evident in light of an efficient recognition pathway. A more suitable way to evaluate the efficacy of $I(z;x)$ term would be to qualitatively compare sample quality and likelihood with and without mutual information.