Generating High Fidelity Images with Sub-scale Pixel Networks and Multidimensional Up-scaling.

Abraham Jose
ID :5068109, CAP6614
abraham@knights.ucf.edu

1 Summary

The author proposes a new network for generating an image as a sequence of sub-images of equal size and and learning the distribution of the image for generating high fidelity images. Proposed conditional decoder network is called Sub-scale Pixel Network(SPN). They used the Auto-regressive neural models and extended on their SPN network to create large high fidelity images. They achieves state-of-the-art results in the large footprints of image set from ImageNet and CelebAHQ data-set.

2 Strengths of the proposal

1. Likelihood based models which intrinsically captures data distribution and uses the direct negative log-likelihood towards loss.
2. They captures the entire data distribution and the training will be stable and delivers best results.
3. Using Sub-scale Pixel Networking, slicing reduces the required computation for the images for generating high fidelity images and the memory required will be reduced.

3 Weaknesses of the proposal

1. Performs badly with low resolution images and they are not able to reduce the loss any further due to the slicing operation they have in the SPN network. However, other techniques performs much better in low resolution images.
2. The loss is based on the negative log-likelihood. There are many other metrics that hasn't been considered like VAE loss,-ELBO or geometric metric etc...

4 Results

They proposed network with Sub-scale Pixel Network involving size up-scaling and depth up-scaling techniques performs really good in bigger footprint images at 128x128(ImageNet data-set) or 256x256(CelebAHQ data-set). They also could establish benchmarks for these data-set for bigger footprints of images and hypothesis that the slicing is the reason for poor performance in small footprints of image.

5 Discussion

How these learned distributions of the images can be used for describing a particular image w.r.t. the pattern, background clutter etc..?
The model was able to learn the distribution of complex nature of the image to attain high sample fidelity for the generated images. However why it is not able to achieve the proposed performance in smaller footprint images? And how much performance gain can occur in a very large image frame(>500x500 size)?.