Report: Generating High Fidelity Images with Subscale pixel networks and Multidimensional Upscaling

Kesar TN*
University of Central Florida
kesar@Knights.ucf.edu

The goal of the paper is to build auto-regressive models to unconditionally generate high fidelity images. The motivation being the capacity to encode vast previous context used to generate images and to learn a distribution that preserves both global semantic coherence and exactness of details, using Subscale pixel networks. The author further use dimensional upscaling to generate high fidelity images on ImageNet and CelebA-HQ datasets.

0.1 Strengths of the paper

The capacity to generate high fidelity images is made possible by slicing the image into S slices, which implicitly captures the size upscaling required to generate high resolution images.

Setting the benchmarks for Negative Log-likelihood on 256x256 ImageNet images which hasn't been possible before.

0.2 Weaknesses of the paper

The negative log-likelihood loss may not be the correct metric as it just describes how confident the model is on it's successful predictions. Hence, by calibrating the scale of the softmax output, NLL might vary drastically.

Although this method works exceptionally on high-resolution images, it struggles to beat other auto-regressive models with 32x32, 64x64 pixels.

0.3 Main Results

(1) Unconditionally generate images up to size 256x256 pixels and assess the fidelity of generated images. (2) Improve state-of-the-art negative log-likelihoods on ImageNet images at 128x128 pixels by a large margin (3.08 bits/dim). Down from 3.55 bits/dim (3) First negative log-likelihood benchmark on 256x256 pixel ImageNet images (2.97 bits/dim without depth-upscaling and 3.01 bits/dim with depth-upscaling)

0.4 Discussion

The SPN and Multidimensional upscaling approach accomplishes state-of-the-art MLE scores on large-scale images from complex datasets like ImageNet-128 and CelebAHQ-256. It is able to generate full 8-bit images from the learnt distribution in turn being able to show unprecendented amount of semantic coherence and exactness of details.

*Graduate student at University of Central Florida (https://github.com/kesaroid)

Preprint. Under review.