AWS Machine Learning Engineer Nanodegree

Capstone Proposal for Image Super Resolution

Mohamed Hafez
August 6, 2024

Domain Background

Image Super Resolution is a problem in computer vision that aims to upscale a low-resolution image to high resolution while preserving important details. The work of Dong et al. (2014) that introduced CNNs for image super reslution laid the groundwork for subsequent research. The use of GANs, as demonstrated by Ledig et al. (2017), has further elevated the quality of generated images. GAN's were very popular in the field of image super resolution and many works adopted them such as Wang et al.'s works in (2018) and (2021), which is the model I am going to use in this project. Recently, some works used diffusion-based models for the image super reslution task. Examples include Stability AI's Stable Diffusion Upscaler. Although diffusion models show great performance in terms of generated image quality, they are computationally expensive in both training and inference. Some works tried to reduce inference time like Luo et al. (2023), which reduces the number of required steps for inference in diffusion models.

Problem Statement

Traditional image upscaling methods like bicubic interpolation often result in blurry and low-quality upscaled images, failing to capture fine details and textures present in high-resolution images. The challenge is to develop an image upscaling application using deep learning that can produce sharp and realistic high-resolution images from low-resolution inputs. Moreover, the application should be able to run in real-time scenarios which requires scalability and workflow optimization to minimize latency.

Datasets and Inputs

For this project, I will utilize the publicly available dataset DIV2K, which contain a diverse range of high-resolution images for training our deep learning model. The input to the model will consist of low-resolution images (e.g., 64x64 pixels) that need to be upscaled to higher resolutions. DIV2K is a popular single-image super-resolution dataset which contains 1,000 images with different scenes and is splitted to 800 for training, 100 for validation and 100 for testing. It was collected for NTIRE2017 and NTIRE2018 Super-Resolution Challenges in order to encourage research on image super-resolution with more realistic degradation. This dataset contains low resolution images with different types of degradations. Apart from the standard bicubic downsampling, several types of degradations are considered in synthesizing low resolution images for different tracks of the challenges. Track 2 of NTIRE 2017 contains low resolution images with unknown x4 downscaling. Track 2 and track 4 of NTIRE 2018 correspond to realistic mild $\times 4$ and realistic wild $\times 4$ adverse conditions, respectively. Low-resolution images under realistic mild x4 setting suffer from motion blur, Poisson noise and pixel shifting. Degradations under realistic wild x4 setting are further extended to be of different levels from image to image. source

Solution Statement

I am going to implement the Real-ESRGAN model for generating high resolution images. This model is the extension to the SRGAN and the ESRGAN models published before. I am going to train the model on AWS and prepare and endpoint to invoke the model and get predections.

Benchmark Model

As a benchmark model, I will compare the performance of my GAN-based image upscaling application against traditional methods like bicubic interpolation and state-of-the-art deep learning approaches such as SRCNN (Super-Resolution Convolutional Neural Network). The benchmark model will help us evaluate the effectiveness of the proposed solution in terms of image quality and computational efficiency. Note that in this project, I am going to focus on the implementation and deployment, rather than the model archeticture.

Evaluation Metrics

I will evaluate the performance of the image upscaling application using metrics such as Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM). These metrics will provide quantitative measures of image quality, fidelity, and similarity to ground truth high-resolution images. Moreover, I going to asses the latency and accessibility of the model.

Project Design

I am going to first prepare the training data in S3 bucket and push the code for the model to a Github repo. After that, I will launch a proper EC2 instance for training where I am going to clone my repo and start training immediately, minimizing the cost as much as possible by developing the code locally and then only training on EC2. The next step will be to write an inference code and prepare an endpoint to invoke the model.