Image Super-Resolution

Lasya Manthripragada
NUID : 002130377
Faiz Khwaja
NUID : 002133269

Abstract

The goal of this project is to create an image super resolution model using the current deep learning techniques that can generate high quality images. This project involves going through the literature survey of the previous methods/models used for the same, and training of CNN based models on a combination of high-resolution and low-resolution images and evaluation of the same models using various metrics such as peak signal-to-noise (PSNR) and similarity index measure (SSIM).

1. Overview

.

Image super resolution is a task to enhance low resolution images to high resolution by using deep learning techniques. It is mainly used to detect faces in a low resolution set up, like a security surveillance camera. [1]. To put it in simple terms, a low-resolution image is simply an image with a fewer number of pixels per inch (PPI). For example, if there is an image of size (60,60) and we want plot it in a canvas of size (6, 10) inches- the number of PPI would be (10,6) whereas if we want to plot the same image but of size (1400,931) in the canvas of same size, the PPI would be (233 , 248) thus giving a better quality image. There is a good explanation given by the author in [2] about the same. Now, our goal is to improve the lower quality image to make it as close as possible to the higher quality image. We are planning to build a working model by the end of the project's timeline, where an image of low quality can be uploaded, and the output, ideally, would be a high(er)-quality image. We will be referring to the architecture of the models built in the past to start building our model, the different architectures proposed and the metrics for evaluation used in previous work will be discussed in detail in section 2. The training of the models will be done on pairs of lower resolution and higher resolution images where the latter is the target of the model. The dataset being used in this project is taken from Kaggle [3] and will be further explained in section 3. We will first build a simple CNN model to get a baseline. Now to improve the model, we are going to use state-of-the-art architectures and techniques, one of them is using the concatenation operation [4] which is a "stacking" technique. So, if a layer has the input size (H, W) with 28 filters and some other layer has the same size (H, W) with 28 filters, if we combine these layers the model will learn new features because the feature space is increasing i.e., 56 filters. Please refer [4], it has a very insightful explanation on why this operation works. In addition to this, if time permits, we are going to use the U-net architecture, where we will tweak the input and output layers so that it can be trained with the spatial resolution of the dataset we are using. The results expected would be better than traditional resolution enhancing techniques like nearest neighbors or bilinear. The metrics we are considering evaluating these models are MSE (lower the better), PSNR - which is the log ratio of the maximum pixel (in our case, it 1 since we are normalizing the image between 0-1) and the MSE, the higher the better. One more metric we are considering is SSIM which gives the measure of how similar two images are in terms of luminance, contrast and structure, the metric is explained in further detail below.

$$SSIM(x, y) = (l(x, y))^\alpha * (c(x, y))^\beta * (s(x, y))^\gamma ,$$

where x,y are the predicted o/p and the ground truth respectively. l(x,y) gives the luminance, c(x,y) gives the contrast and s(x,y) gives the structure. Calculations on how to calculate each term is explained in this very insightful article [5].

Overall, we are taking a dataset with both low-resolution and high-resolution images and training it on models with different architectures and evaluating these models' using metrics like MSE, PSNR and SSIM.

2. Literature Survey

Deep learning has improved over the years, and it has many working models for image super resolution. We are presenting here 4 papers that are relevant to our project. Wenming Yang et al [6] discussed in their paper about the current models and their challenges. They divided the paper into reviewing the architectures and the optimizing parameters for the same. One of the best architectures was DBPN – deep back projection network which has the highest number of parameters (around 10 M) – in each block of this architecture the input is first down sampled and then the down samples images go through the convolutional layers the output of which is combined with the input image (which gives the residual image) and this process is called "back projection". This happens in every block of the DBPN architecture. Please refer the paper for more extensive details. From the paper, we inferred that CNN with more number of parameters yield better results. Chao Dong et al [7] proposed a SRCNN – super resolution convolutional neural network. The SRCNN's PSNR value is 27.95 db., so we can infer that the model performed good. The authors also mentioned that their model is way faster. They tested SRCNN model on Set5 data [8] by using different filter sizes, they came to a conclusion that more number of filters give good results. Also, larger filter sizes give the better result, but the computational speed decreases. They also tried adding one layer to the architecture to see if it will give better results, it took longer to compute, and eventually converged to the same result, indicating that deeper network did not help here. Xiaole Zhoa et al [9] proposed architecture without the residual connections, instead the skip connections are just the concatenation channels from different layers, hence they named the model FC^2-CN. It has much fewer parameters and gives better results. We are planning to heavily refer from this paper while we are building put concatenation architecture. Jin Yamanaka [10] et al discussed a CNN architecture with residual net. The architecture used a combination of deep convolutional layers and skip connections which gave them 10 times faster than a conventional CNN. They also noted that usage of ensemble learning gives better results , so they suggested that a deep learning model should be combined with ensemble learning for complex problems. Overall, we inferred from the above papers that the concatenation works well, and it is important to re-use lower feature space to higher feature by either using skip connections or concatenation. and PSNR is a good measure to check the goodness of the image.

3. Dataset

We are taking data from two sources and combining them. [11] has 100 images in low resolution and 100 images in high resolution. [3] has 855 images in each low resolution and high-resolution image sets. We are going to combine these data and split into train, validation, and test sets. [11] is a subset of the set5 [8] dataset. We will first train on these images, and if the model doesn't perform well will use the ImageNet [12] dataset and convert the HR images into LR. Figure 1 is a sample of our dataset.

4. References

1. M. Yang and J. Qi, "Reference-based Image Super-Resolution by Dual-Variational AutoEncoder," 2021 International Conference on Communications, Computing, Cybersecurity, and Informatics (CCCI), Beijing, China, 2021, pp. 1-5, doi: 10.1109/CCCI52664.2021.9583193.
2. https://www.linkedin.com/pulse/what-difference-between-low-high-resolution-image-vala-vincent/
3. https://www.kaggle.com/datasets/adityachandrasekhar/image-super-resolution
4. https://www.researchgate.net/post/What-are-the-operations-of-concatenation-or-combination-exist-in-the-literature-for-CNN-architecture-of-Deep-Learning
5. https://medium.com/srm-mic/all-about-structural-similarity-index-ssim-theory-code-in-pytorch-6551b455541e
6. Wenming Yang, Xuechen Zhang, Yapeng Tian, Wei Wang, Jing-Hao Xue, and Qingmin Liao. 2019. Deep Learning for Single Image Super-Resolution: A Brief Review. Trans. Multi. 21, 12 (Dec. 2019), 3106–3121. https://doi.org/10.1109/TMM.2019.2919431
7. Dong, C., Loy, C.C., He, K., Tang, X. (2014). Learning a Deep Convolutional Network for Image Super-Resolution. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds) Computer Vision – ECCV 2014. ECCV 2014. Lecture Notes in Computer Science, vol 8692. Springer, Cham. https://doi.org/10.1007/978-3-319-10593-2_13
8. https://huggingface.co/datasets/eugenesiow/Set5
9. Zhao, X., Liao, Y., He, T., Zhang, Y., Wu, Y., & Zhang, T. (2021). FCN: Fully Channel-Concatenated Network for Single Image Super-Resolution.

10. Yamanaka, Jin & Kuwashima, Shigesumi & Kurita, Takio. (2017). Fast and Accurate Image Super Resolution by Deep CNN with Skip Connection and Network in Network. 217-225. 10.1007/978-3-319-70096-0_23.
11. https://www.kaggle.com/datasets/akhileshdkapse/super-image-resolution.
12. https://image-net.org/download-images.php

5. Appendix

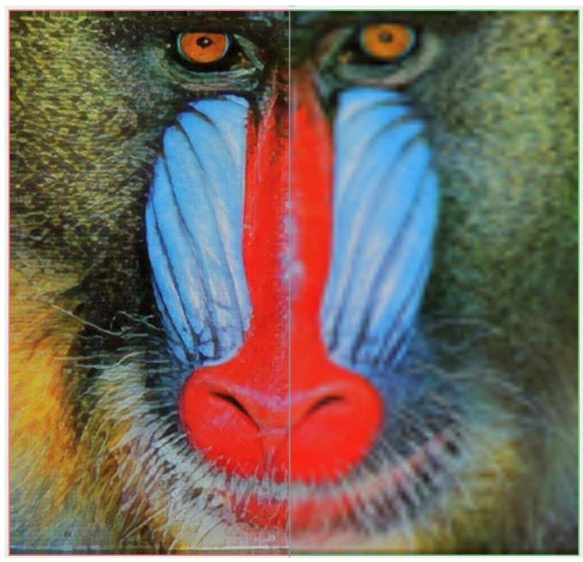

Figure 1 High resolution v/s Low resolution im