

Selected Topics in Visual Recognition using Deep Learning Homework 4

Github link of the codes

Project Repository: [VRDL_21au](#)

Reference

Github repository [KAIR:Image Restoration Toolbox](#) is used and modified in this project.

Introduction

The project is an experiment of implementing deep learning model to deal with image super resolution task. The dataset used here is composed of 305 images with different resolution, with 291 high-resolution images as training data and 14 low-resolution images as testing data. In this project, we analyze the capability of image restoration using currently state-of-the-art model SwinIR on our dataset.

Data Preprocessing

Compared to other common computer vision problems like image classification or object detection task, the preprocessing method of image resolution focus more on upscaling/downscaling the image size than normalization or augmentation. Here we follow the default settings that KAIR provides, using bicubic interpolation to downscale the target image, then train the model to restore it. For more information, [the corresponding source code](#) could be the reference.

Basic Deep Learning Techniques

We believe that applying transfer learning or directly using pretrained data could provide better performance in this case, transfer learning is not used, however, in this project to follow the homework requirement. For learning rate scheduler part, `StepLR` is used so that smaller learning rate will be used in the last few epochs.

Thing should be noticed is that in our case, the training dataset consists of only 291 images, which is relatively small dataset. The training epochs (also, learning rate) hence should be carefully chosen to avoid overfitting issue. We've tried a range from 100,000 to 500,000 iterations and choose the one with highest PSNR in validation set.

SwinIR: Image Restoration Using Swin Transformer

Based on the Swin Transformer, SwinIR consists of three parts: shallow feature extraction, deep feature extraction and high-quality image reconstruction. In particular, the deep feature extraction module is composed of several residual Swin Transformer blocks (RSTB), each of which has several Swin Transformer layers together with a residual connection. According to the original paper, SwinIR outperforms other methods by up to 0.14~0.45dB, while the total number of parameters can be reduced by up to 67% on three different tasks (image super-resolution, image denoising and JPEG compression artifact reduction).

Experiment Result

Model	PSNR
SinwIR	28,0461

Summary

As we could see above, the basic implementation of SwinIR model could achieve a decent PSNR with around 28.06, which outperforms the baseline 27.41 based on VDSR implementation. However, there exist an obvious gap compared to the 27.56~32.93 PSNR performance shown in the original paper. This may be because of the different training/testing dataset, or not-the-best hyperparameter settings. More comparison should be done to further improve the performance.