YZV416E Computer Vision Project Progress Report The Effect of Super Resolution Method on Classification Performance

Ömer Faruk Aydın
Artificial Intelligence and Data Engineering
Istanbul Technical University
aydinome21@itu.edu.tr
150210726

Hasan Taha Bağcı
Artificial Intelligence and Data Engineering
Istanbul Technical University
bagcih21@itu.edu.tr
150210338

I. Introduction

Image super-resolution (ISR) is an image processing technique that aims to transform low-resolution images into high-resolution ones. This technique seeks to increase the resolution of an image, typically by a factor of 4 or more, while preserving its content and details as much as possible.
Image classification is the process of assigning an image to one of the predefined categories. This process aims to identify objects, patterns, or features in the image.
In this project, we aim to investigate the impact of different super-resolution methods applied to low-resolution images on image classification accuracy. We will examine how classification accuracy changes as image quality increases. Additionally, we will conduct a comparative analysis to determine which super-resolution method improves classification accuracy the most.
In this project, we will first make the images in the subset we will obtain from the ImageNet dataset low resolution by applying various low resolution methods. Then we will obtain super-resolution images using pre-trained models of different super-resolution methods. Then, we will compare the high resolution images using various metrics (PSNR, SSIM). Finally, we will put them into a classification model and observe the results.

II. Dataset

The ImageNet dataset, which contains over 14 million annotated images across more than 21 thousand classes, is utilized to evaluate the impact of super-resolution (SR) techniques on image classification accuracy. The dataset is divided into training, validation, and testing sets.

The classification performance of these models is assessed using the validation set of the ImageNet dataset, with top-1 accuracy as the primary evaluation metric. This setup allows for a controlled comparison to determine the effectiveness of SR techniques in improving the accuracy of image classification models, providing insights into the potential benefits of image resolution enhancement in practical applications.

III. Data Preparation

To determine the necessity of applying super-resolution to an image, the Laplacian function from the OpenCV library can be utilized as a sharpness metric. This function calculates the Laplacian of the image, which is a measure of the second spatial derivative and highlights regions of rapid intensity change, often corresponding to edges within the image. A lower variance of the Laplacian implies that the image is blurrier and could benefit more from super-resolution techniques to enhance its details. This approach serves as a preliminary step to identify images that lack sharpness and detail, making them suitable candidates for the super-resolution process to improve image quality.

IV. Methodology

Real-ESRGAN

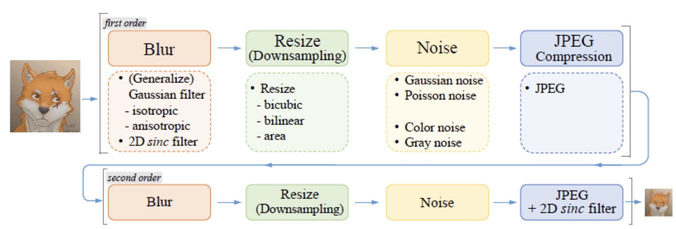

Fig. 1. Overview of the pure synthetic data generation adopted in Real-ESRGAN

It utilizes a second-order degradation process to model more practical degradations, where each degradation process adopts the classical degradation model. The detailed choices for blur, resize, noise and JPEG compression are listed. It also employ sinc filter to synthesize common ringing and overshoot artifacts [1].

BSRGAN

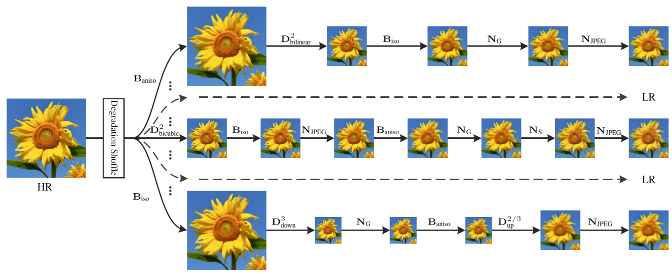

Fig. 2. BSRGAN Degradation Model

The main idea of the BSRGAN is to propose a degradation model instead of using the commonly-used blur/downsampling/noise-addition pipeline, it performs randomly shuffled degradations to synthesize LR images [2].

- It introduces more practical blur, downsampling, and noise degradations compared to traditional methods. This includes isotropic/anisotropic Gaussian blurs, various downsampling techniques (nearest, bilinear, bicubic), and different noise types (Gaussian, JPEG compression, camera sensor).
- Instead of a fixed degradation pipeline, it randomly shuffles the order of degradations to synthesize LR images.
- While it can produce some unrealistic degradations, this is expected to improve the generalization ability of the trained super-resolution model.
- Large capacity deep neural networks have shown the ability to handle different degradations via a single model, as validated by models like DnCNN.
- The degradation model can be conveniently modified by changing degradation parameters or adding new degradation types to improve practicality for specific applications.

Swin2SR

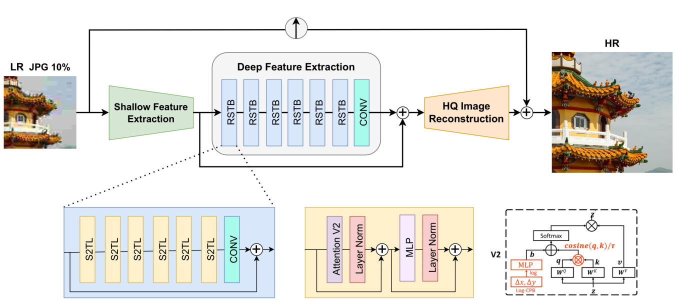

Fig. 3. The architecture of the proposed Swin2SR

The Swin2SR model, a variant of the Swin Transformer V2 tailored for image super-resolution tasks, comprises Swin Transformer blocks for multi-scale feature extraction, utilizing shifted window self-attention to capture long-range dependencies efficiently. Preceding these blocks, a shallow feature extraction module extracts low-level features from input images. Deep feature extraction is handled by stacked Swin Transformer blocks operating at various resolution levels. A reconstruction module then upsamples features for high-resolution image generation, incorporating residual connections for improved gradient flow and normalization layers for training stability. Leveraging Swin Transformer V2's enhancements, including improved model capacity scaling and training stability, Swin2SR excels in challenging scenarios like image super-resolution from compressed or low-quality input images [3].

V. Classification

The methodology for assessing the impact of super-resolution (SR) methods on image classification performance involves a systematic approach where low-resolution (LR) images are first enhanced using super-resolution techniques before being subjected to classification algorithms. The primary objective is to determine whether SR can improve the accuracy of image classification models when applied to LR images.

The extracted features are then fed into classification algorithm to categorize the images into predefined classes from ImageNet dataset. VGG19 algorithm will be used to evaluate to determine its effectiveness in classifying both the original and super-resolved images.

References

[1] X. Wang, L. Xie, C. Dong, and Y. Shan, "Real-esrgan: Training real-world blind super-resolution with pure synthetic data," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 1905–1914.

[2] K. Zhang, J. Liang, L. Van Gool, and R. Timofte, "Designing a practical degradation model for deep blind image super-resolution," in *IEEE International Conference on Computer Vision*, 2021, pp. 4791–4800.

[3] M. V. Conde, U.-J. Choi, M. Burchi, and R. Timofte, "Swin2sr: Swinv2 transformer for compressed image super-resolution and restoration," in *European Conference on Computer Vision*. Springer, 2022, pp. 669–687.

[4] A. Wahane, R. Gadade, A. Hundekari, A. Khochare, and C. Sukte, "Real-time sign language recognition using deep learning techniques," in *2022 IEEE 7th International conference for Convergence in Technology (I2CT)*. IEEE, 2022, pp. 1–5.

[5] R. Rastgoo, K. Kiani, and S. Escalera, "Real-time isolated hand sign language recognition using deep networks and svd," *Journal of Ambient Intelligence and Humanized Computing*, vol. 13, no. 1, pp. 591–611, 2022.

[6] S. Sharma, R. Sreemathy, M. Turuk, J. Jagdale, and S. Khurana, "Real-time word level sign language recognition using yolov4," in *2022 International Conference on Futuristic Technologies (INCOFT)*. IEEE, 2022, pp. 1–7.