DOI: 10.54254/2755-2721/81/20241188

Enhancing Image Classification with Super-Resolution Techniques: A Comparative Study of SRResNet and SRGAN

Siyu Deng

College of letters and science, University of California, Santa Barbara, California, USA

siyudeng@ucsb.edu

Abstract. This paper explores the use of super-resolution (SR) approaches to enhance the performance of image classification, with particular attention to the effects of various SR models on classification accuracy. Using the ImageNet Dogs dataset for assessment, the paper examines two SR techniques: Super Resolution Generative Adversarial Networks (SRGAN) and Super Resolution Residual Networks (SRResNet). The research demonstrates that deep learning-based SR methods can enhance classification accuracy. Notably, SRResNet is identified as the superior method for improving classification performance, despite generating less visually appealing images compared to SRGAN. These finding highlights that while Generative Adversarial Networks (GANs) and perceptual loss functions can enhance image quality, their impact on classification accuracy may not always be substantial. The study offers important new perspectives on the relative merits of different SR approaches, highlighting the necessity of choosing the right SR methods in accordance with the particular demands of picture classification tasks. The results suggest that SRResNet, with its focus on accuracy over visual appeal, is more effective for boosting classification model performance, offering guidance for optimizing SR methods in practical image classification applications.

Keywords: Super-Resolution, Image Classification, SRResNet, SRGAN.

1. Introduction

Reconstructing high-resolution (HR) features from their low-resolution (LR) counterparts is a crucial job in image super-resolution (SR) in computer vision. An image's resolution is increased using this procedure. This procedure is especially important in fields like object detection, facial recognition, and medical imaging, where the caliber of visual data directly affects how well tasks are performed in the future [1]. In the context of dog breed classification, high-quality images are crucial for accurate model predictions, as subtle differences between breeds may be difficult to discern in LR images. The significance of this study lies in its exploration of whether employing a Super Resolution Generative Adversarial Networks (SRGAN) during data preprocessing can improve the accuracy of dog breed classification models, particularly when using a deep convolutional network such as a Residual Network (ResNet). By enhancing the resolution of training images, the study aims to determine if the improved image quality leads to better feature extraction and, consequently, more accurate classification.

In the field of computer vision, SR has received a lot of attention, and many approaches have been developed to address the problem of reconstructing HR pictures from LR inputs. A common strategy used in traditional SR techniques was to minimize the mean squared error (MSE) between the ground truth and the produced HR picture [2], as the peak signal-to-noise ratio (PSNR), a commonly used parameter for assessing SR performance, is likewise maximized by this optimization goal [3]. However, these methods often resulted in images that, while high in PSNR, lacked fine details and appeared overly smooth. A major development in SR approaches was the advent of deep learning techniques, specifically convolutional neural networks (CNNs). Early models such as Super Resolution Convolutional Network (SRCNN) [4] and Faster RCNN (FSRCNN) [5] demonstrated considerable improvements in SR performance by leveraging deep networks to learn hierarchical representations of images. More recent developments, such as very deep SR (VDSR) [6] and enhanced deep residual networks (EDSR) [7], further pushed the boundaries by employing deeper architectures and advanced techniques like residual learning to achieve superior results.

SRGAN is one of these innovations that stands out as being especially promising. SRGAN makes use of GANs' capabilities to generate HR pictures that are both perceptually and numerically more similar to natural photos. It achieves this by employing an adversarial loss, which encourages the generated images to reside on the manifold of natural images, and a perceptual loss, which emphasizes perceptual similarity over pixel-wise accuracy [2]. In addition to these technical advancements, SR techniques have been applied across a range of practical scenarios, including medical imaging [8], surveillance [9], and security [10]. Nevertheless, there has not been much research done on their possible effects on classification job performance, especially when it comes to dog breed categorization. This study aims to fill this gap by investigating whether SRGAN can be effectively utilized in the data preprocessing stage to enhance image quality and, consequently, classification accuracy, especially when using state-of-the-art classification models like Residual Networks (ResNet), which are deep learning architectures designed to solve the vanishing gradient problem by using shortcut connections to skip layers and enable the training of very deep networks for tasks like image classification.

This study examines how using SRGAN, especially in conjunction with ResNet, affects data preparation for dog breed categorization. The research reviews key concepts and advancements in SR techniques, focusing on SRGAN's ability to enhance image quality. It evaluates the performance of ResNet models trained on super-resolved images compared to those trained on original low-resolution images. Experimental results demonstrate that SRGAN significantly improves classification accuracy by providing higher-resolution images, which enhance ResNet's ability to detect subtle breed characteristics. The paper highlights SRGAN's strengths and limitations in this context and discusses implications for future research. It addresses whether SRGAN's image enhancement leads to better model performance in practical applications. The study contributes to the understanding of integrating SR techniques into machine learning workflows to boost overall performance. The paper is structured as follows: an introduction to SRGAN and related SR methods, a presentation of experimental results with ResNet, a discussion of the findings, and a summary with recommendations for future research.

2. Methodology

2.1. Dataset description and preprocessing

The ImageNet Dogs dataset, a subset of the broader ImageNet dataset, is the one that is currently being used the most [11]. It contains images of 120 different dog breeds and is widely used for fine-grained image classification in machine learning, particularly to train and evaluate CNNs. The dataset is originally from the ImageNet project. Preprocessing steps like resizing, normalizing, and data augmentation are often necessary to prepare the images for model training, and breed labels need to be encoded for compatibility. As a pre-processing phase, this work uses data augmentation, which uses horizontal flipping, cropping, and random rotations to increase the variety of the training data and strengthen the model's resistance to changes in input pictures [12].

2.2. Proposed approach

The research investigates the impact of integrating SRGAN in the preprocessing stage of dog breed classification using the ResNet model. The focus is on determining whether enhancing image resolution with SRGAN improves ResNet's classification accuracy. ResNet, renowned for its deep learning capabilities and proficiency in learning complex features from high-quality images, is central to this study. The methodology includes a review of SR techniques, particularly SRGAN, analyzing its principles and performance in generating high-resolution images. Experimental evaluations compare ResNet models trained on SRGAN-enhanced images with those trained on original low-resolution images. The findings will reveal SRGAN's effectiveness in improving classification accuracy and address its strengths, limitations, and future research directions. Figure 1 illustrates how SRGAN is integrated into the preprocessing workflow, emphasizing its role in enhancing image resolution before inputting data into the ResNet model. Initially, a pre-trained SR model is used to perform 4x SR on the images in the dataset. The SR models applied include Super Resolution Residual Networks (SRResNet), which leverages traditional deep learning techniques, and SRGAN, which incorporates adversarial mechanisms.

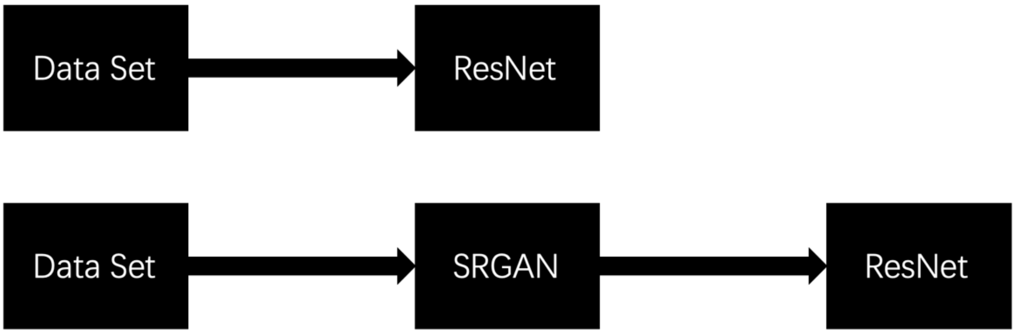


Figure 1. Pipeline of the experiment.

2.2.1. SRResNet. The SRResNet is a fully convolutional network engineered for 4x SR. As its name suggests, it integrates residual blocks with skip connections, which enhance the network's optimization capability, even with its considerable depth [13].

2.2.2. SRGAN. SRGAN is a SR network architecture introduced by Christian Ledig, which significantly advances the performance of SR techniques. SRGAN is built upon the GAN framework, comprising two main components: a discriminator and a generator. The discriminator employs the Visual Geometry Group-19 (VGG-19) network structure [14], which consists of eight convolutional layers. The function utilized for activating the hidden layers is Leaky Rectified Linear Unit (LeakyReLU). Lastly, the discriminator calculates the likelihood that the input image is a produced image or a genuine HR image using a sigmoid activation function and a fully connected layer. The generator utilizes a ResNet architecture [14], in which many residual blocks are present in the network's first portion. Each residual block contains two 3 × 3 convolutional layers, followed by batch normalization. Later on, two subpixel network modules are added to upscale the image, and this is enabled via the ReLU function. This design enables the generator to focus on learning HR image details in the early layers during training and to enhance image resolution in the subsequent layers, thereby reducing computational demands.

2.2.3. ResNet. ResNet, introduced by Kaiming He et al. in their pioneering 2015 paper [15], was designed to overcome the difficulties associated with training very deep neural networks. The impetus for ResNet came from the observation that increasing the depth of a network led to a point where performance would plateau and eventually decline. This issue, known as the vanishing gradient problem, occurs when gradients become too small, impeding the network's ability to learn effectively.

ResNet used a novel architectural strategy known as residual learning to address the vanishing gradient issue. The key idea was to employ skip connections, sometimes called shortcut connections, which made it possible for data to go straight from earlier layers to later ones, avoiding specific levels in the process. By allowing the network to learn residual mappings rather than direct mappings, these skip connections made optimization easier and made it easier for deep representations to be learned. The idea behind residual learning is that the network learns the residual mapping, or the difference between the input and the desired output, instead of learning the whole mapping from input to output. By focusing on this residual information, the network can more easily capture the subtle details or modifications needed to reach the desired output. A residual connection typically involves a direct link from the input of a block to its output, which sometimes includes additional layers, though often it does not. ResNet models are categorized by their layer count, such as ResNet18, ResNet34, ResNet50, ResNet101, and ResNet152.

2.2.4. Perceptual loss function. Mean squared error (MSE) is the loss function used for optimization in the majority of image SR techniques., often producing images with a high signal-to-noise ratio but lacking in high-frequency details [13].

$$l_{MSE}^{SR} = \frac{1}{r^2 WH} \sum_{x=1}^{rW} \sum_{y=1}^{rH} (l_{x,y}^{HR} - G_{\theta_G}(I^{LR})_{x,y})^2 \tag{1}$$

SRGAN [13] incorporates the VGG loss, which relies on the Rectified Linear Unit (ReLU) activation layer within the pre-trained 19-layer Visual Geometry Group (VGG) network, as detailed by Simonyan and Zisserman [14]. In particular, it uses as a guide the feature map produced by the VGG19 network's i^{th} convolutional layer (post-activation), which comes before the i^{th} max-pooling layer. It computes the VGG loss as the Euclidean distance between the reconstructed images $G_{\theta_G}(I^{LR})_{x,y}$ and the feature representation of the reference image $l_{x,y}^{HR}$. In order to preserve high-frequency information in the reconstructed pictures, this loss function is essential.

2.2.5. Top-k accuracy. When dealing with a large number of classes, top-k accuracy is a statistic used to evaluate the performance of the classification model. It calculates the percentage of cases where the real label, as determined by the model's confidence ratings, is among the top k projected labels. Unlike Top-1 accuracy, which only considers the highest-ranked prediction, Top-k accuracy offers a more lenient evaluation by including the top k predictions. This metric is particularly valuable in contexts like image classification and recommendation systems, where multiple relevant options may be possible. In this study, with over 120 dog breeds, Top-k accuracy is employed to provide a more flexible evaluation criterion, accommodating the complexity of distinguishing among many classes.

3. Result and Discussion

3.1. Performance comparison of SR models

The assessment criteria used in this work were the Structural Similarity Index Measure (SSIM) and the Peak Signal-to-Noise Ratio (PSNR). Tables 1 and 2 display the results. All tests involved down-sampling and SR magnification with a factor of four. The chart displays the PSNR and SSIM metrics of images of four types of dogs under two SR models. The PSNR calculates the difference between the original picture's maximum potential power and the power of corrupting noise, which is the difference between the original and reconstructed image. Since there may be less of a disparity between the original and reconstructed pictures, higher PSNR values are indicative of higher quality images. SSIM uses three visual criteria to evaluate properties: structure, contrast, and brightness. The range of SSIM values is -1 to 1, where a value of 1 denotes complete structural similarity.

Table 1. PSNR of testing set.

	Japanese spaniel	Maltese	Pekinese	Toy terrier
SRResNet	28.761	29.172	29.708	28.885
SRGAN	26.132	26.953	26.577	26.260

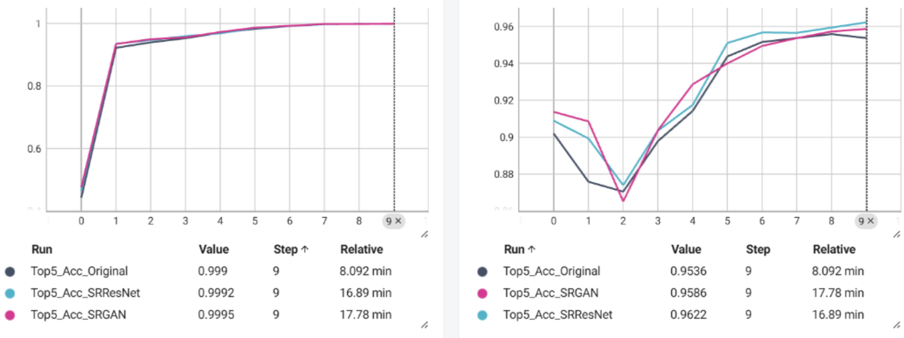
Table 2. SSIM of testing set.

	Japanese spaniel	Maltese	Pekinese	Toy terrier
SRResNet	0.806	0.818	0.791	0.797
SRGAN	0.744	0.773	0.736	0.735

The performance metrics of SRResNet surpass those of SRGAN. However, the SRGAN paper [13] emphasizes that PSNR and SSIM might not be able to convey the clarity of super-resolved images. Although SRResNet produces images that tend to be less realistic and overly smooth, it achieves higher scores on these metrics compared to SRGAN. Nevertheless, from a visual standpoint, SRGAN generates results that are more visually convincing and discernible to the human eye than those produced by SRResNet.

3.2. Performance comparison of classification models

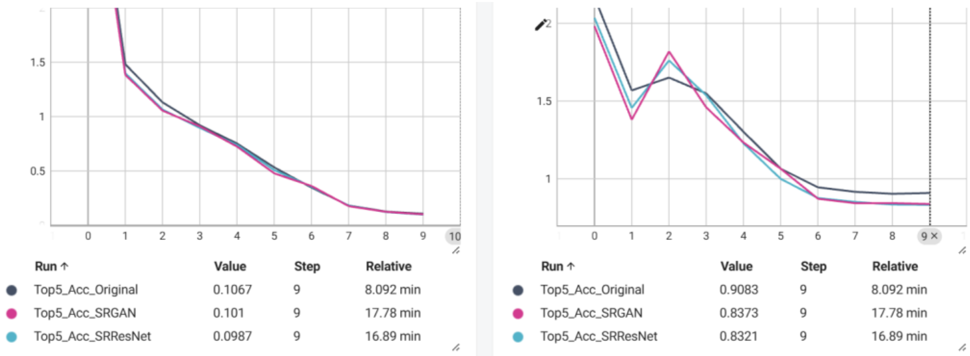
This article utilizes ResNet-34 as the foundational classification model and employs transfer learning to train it. Initially, the pre-trained model, based on ImageNet weights, is loaded, and hyperparameters are optimized for the specific dataset. The accuracy and loss during the training process are presented in Figure 2 and Figure 3.



	Run	Value	Step ↑	Relative
●	Top5_Acc_Original	0.999	9	8.092 min
●	Top5_Acc_SRResNet	0.9992	9	16.89 min
●	Top5_Acc_SRGAN	0.9995	9	17.78 min

	Run ↑	Value	Step	Relative
●	Top5_Acc_Original	0.9536	9	8.092 min
●	Top5_Acc_SRGAN	0.9586	9	17.78 min
●	Top5_Acc_SRResNet	0.9622	9	16.89 min

Figure 2. Accuracy during training.



	Run ↑	Value	Step	Relative
●	Top5_Acc_Original	0.1067	9	8.092 min
●	Top5_Acc_SRGAN	0.101	9	17.78 min
●	Top5_Acc_SRResNet	0.0987	9	16.89 min

	Run ↑	Value	Step	Relative
●	Top5_Acc_Original	0.9083	9	8.092 min
●	Top5_Acc_SRGAN	0.8373	9	17.78 min
●	Top5_Acc_SRResNet	0.8321	9	16.89 min

Figure 3. Loss during training.

This study investigates SR approaches and evaluates their effect on classification model performance. An investigation into the effectiveness of two deep learning-based approaches, SRResNet and SRGAN, demonstrates that both methods positively influence classification accuracy. Notably, classification models perform better and faster convergence when deep learning models are applied to improve datasets during training. Improved datasets for the validation set produce reduced losses and increased accuracy over the same number of epochs. However, despite the original SRGAN authors' claims that SRGAN achieves lower PSNR and SSIM scores but higher visual clarity compared to SRResNet [13], this study found that SRGAN is less effective than SRResNet in improving classification model performance. This suggests that the visual clarity perceived by the human eye may not be the primary factor influencing machine vision accuracy. It may be concluded that improving classification performance may not always be achieved by the use of perception-based loss functions and GANs. Consequently, when evaluating these two deep learning approaches, SRResNet emerges as the preferred option, maintaining superior classification performance and visual quality under similar training conditions and iterations.

4. Conclusion

In the rapidly evolving field of neural network technology, particularly deep learning, integrating advanced methods into various applications shows great promise. This study evaluates the impact of different SR models on image classification, specifically comparing SRResNet and SRGAN. The findings reveal that SRResNet achieves superior classification accuracy, despite SRGAN producing more visually appealing images. This suggests that while GANs and perception-based loss functions offer benefits for image enhancement, they do not always translate into better performance for classification tasks. In practical scenarios, SRResNet is the preferred choice for improving classification accuracy while upholding high visual quality standards. To get further insight into how classification performance differs with various SR approaches, future research should keep evaluating these models in a variety of datasets and application situations.

References

[1] Yang C Y Ma C Yang M H 2014 Single-image super-resolution: A benchmark Computer Vision–ECCV pp 372-386

[2] Agarwal A Chhotaray S Roul N K et al. 2023 A Review Super Resolution Using Generative Adversarial Network-Applications and Challenges J. Eng. Technol vol 3 no 1 pp 1-6

[3] Yang Q Yang R Davis J et al. 2007 Spatial-depth super resolution for range images IEEE conference on computer vision and pattern recognition pp 1-8

[4] Chen Y Phonevilay V Tao J et al. 2021 The face image super-resolution algorithm based on combined representation learning Multimedia Tools and Applications vol 80 pp 30839-30861

[5] Dong C Loy C C He K et al. 2014 Learning a deep convolutional network for image super-resolution Computer Vision–ECCV pp 184-199

[6] Dong C Loy C C Tang X Accelerating the super-resolution convolutional neural network Computer Vision–ECCV pp 391-407

[7] Kim J Lee J K Lee K M 2016 Accurate image super-resolution using very deep convolutional networks Proceedings of the IEEE conference on computer vision and pattern recognition pp 1646-1654

[8] Ren H Kheradmand A El-Khamy M et al. 2020 Real-world super-resolution using generative adversarial networks Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops pp 436-437

[9] Johnson J Alahi A Fei-Fei L 2016 Perceptual losses for real-time style transfer and super-resolution Computer Vision–ECCV pp 694-711

[10] Liu C Shum H Y Freeman W T 2007 Face hallucination: Theory and practice International Journal of Computer Vision vol 75 pp 115-134

[11] Aditya K Nityananda J Bangpeng Y Fei-Fei L 2020 Stanford Dogs Dataset Retrieved on 2024, Retrieved from: http://vision.stanford.edu/aditya86/ImageNetDogs/

[12] Shorten C Khoshgoftaar T M 2019 A survey on image data augmentation for deep learning Journal of big data vol 6 no 1 pp 1-48

[13] Ledig C Theis L Huszár F et al. 2017 Photo-realistic single image super-resolution using a generative adversarial network Proceedings of the IEEE conference on computer vision and pattern recognition pp 4681-4690

[14] Simonyan K Zisserman A 2014 Very deep convolutional networks for large-scale image recognition arxiv preprint 1409.1556

[15] He K Zhang X Ren S et al. 2016 Deep residual learning for image recognition Proceedings of the IEEE conference on computer vision and pattern recognition pp 770-778