Indonesian Journal of Electrical Engineering and Computer Science
Vol. 39, No. 1, July 2025, pp. 634~643
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v39.i1.pp634-643

Low-resolution image quality enhancement using enhanced super-resolution convolutional network and super-resolution residual network

Mohammad Faisal Riftiarrasyid[1], Rico Halim[1], Andien Dwi Novika[1], Amalia Zahra[2]
[1]Department of Computer Science, School of Computer Science, Bina Nusantara University, Jakarta, Indonesia
[2]Department of Computer Science, BINUS Graduate Program - Master of Computer Science, Bina Nusantara University, Jakarta, Indonesia

Article Info

Article history:

Received May 11, 2024
Revised Mar 12, 2025
Accepted Mar 26, 2025

Keywords:

ESPCN
Image processing
SRResNet
Super resolution
VGG16

ABSTRACT

This research explores the integration of enhanced super-resolution convolutional network (ESPCN) and super-resolution residual network (SRResNet) to enhance image quality captured by low-resolution (LR) cameras and in internet of things (IoT) devices. Focusing on face mask prediction models, the study achieves a substantial improvement, attaining a peak signal-to-noise ratio (PSNR) of 28.5142 dB and an execution time of 0.34704638 seconds. The integration of super-resolution techniques significantly boosts the visual geometry group-16 (VGG16) model's performance, elevating classification accuracy from 71.30% to 96.30%. These findings highlight the potential of super-resolution in optimizing image quality for low-performance devices and encourage further exploration across diverse applications in image processing and pattern recognition within IoT and beyond.

Corresponding Author:

Andien Dwi Novika
Department of Computer Science, BINUS Graduate Program - Master of Computer Science
Bina Nusantara University
Jakarta, 11480, Indonesia
Email: andien.novika@binus.ac.id

1. INTRODUCTION

In recent years, internet of things (IoT) has become a hot topic for research in industries, especially topics on integration of deep learning image classification or detection in IoT technologies [1]. Within those research, camera will always be present as the main sensor to gather input for the system. Modern camera which is being utilized in IoT system will capture light which is going towards it, then construct a digital picture from the light data received by the sensor. Camera will perceive data such as color information, saturation, and brightness from incoming light and then assign those data into each tiny part within a picture called pixel. Those set of pixels are going to be assembled into one and projected as a picture. The more pixel within a picture, the sharper that picture gets or it can also be said that the picture contains more information. The amount of pixel can be captured by camera means more information can be stored within a picture. It means, larger resolution image contains more information which can be processed further.

Before using camera as an input, ones must consider its limitation and capabilities. Each camera has their own limitations, which is the amount of pixel that could be captured in a single picture or the amount of power needed to operate it [2]. Camera capability is often shown in the amount of mega pixel of information can handled. Even though more pixel seems like the best option, it needs significantly larger space to store the image and it needs more computing power to further process it. In recent studies which utilize camera as the main source of input, researcher focuses on improving performance or efficiency of their prototype. In order to improve performance and efficiency, recent studies suggest lowering the image resolution for object detection [3]. By using minimizing resolution, researcher aims to allocate computing power to other processes like deep learning [4] or minimize battery consumption of remote controlled devices like drones [5]. Furthermore by using lower resolution (LR), minimizing chances of unwanted information to appear [6]. Other things to consider is the cost to buy camera which capture high resolution (HR) image is significantly more expensive, so people needs to consider which camera is best suited for its intended functionality.

Modern artificial intelligence breakthrough open its image generating abilities, in which opens more possibilities to train deep learning model to generate image with larger resolution from its original resolution [7]. Super resolution image enhancement aims to expand image resolution and enrich its information content based on the original image. Super resolution will expand the resolution size of the original image by an upscale factor then predict the expanded pixels based from information of the adjacent original pixel. Some studies indicates that by improving quality of images has a significant impact in deep learning prediction results [8], [9]. It means by utilizing super-resolution to pre-process image could improve performance of deep learning integrated system. This is very beneficial, especially with low performance machine from previous generation. By improving software without any needs to improve its hardware such as camera as the main sensors of its system which can be very expensive to upgrade. Nevertheless, there is still yet any research utilize enhanced super-resolution convolutional network (ESPCN) and super-resolution residual network (SRResNet) super resolution to pre-process image using low performance devices like IoT devices. This research tries to explore the idea of improving image captured by low performance devices by using low resolution camera with super resolution has significant impact in classifying mask prediction deep learning model (visual geometry group-16 (VGG16), MobileNetV2, and ResNet50). This research tries to explore researches which has related topics. This is done by searching the Google Scholar page with the keywords "super resolution deep learning", "SRResNet", "ESPCN" and "super resolution deep learning classification" which were published between 2019 and 2023. In order to maintain the relevance of the research topic and fill gaps in existing research. Study on single image super-resolution (SISR) was first conducted by [10]. The present study employs super-resolution generative adversarial network (SRGAN) and SRResNet to produce super-resolution images (SISR) for laser confocal images of the root cells of Solanum nigrum, a hyperaccumulator [11] in the year 2023. The evaluation methods employed in this study are peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), and mean opinion score (MOS). The findings of this work indicate that both the primary reconstruction and the subsequent reconstruction of SRGAN and SRResNet have demonstrated enhanced resolution capabilities for laser confocal pictures.

In order to enhance the intricate features of texture, the authors of [12] propose the employment of SRGAN, a framework that demonstrates the ability to generate photo-realistic natural images with a four-fold increase in resolution. The SRGAN model employs a framework based on generative adversarial networks (GANs), comprising a generator network and a discriminator network. The generator network is trained to produce HR images that exhibit visual similarity to the ground truth HR photos. Conversely, the discriminator network is trained to discern between the generated images and the authentic HR photographs [13]. Through the utilization of a perceptual loss function that amalgamates an adversarial loss and a content loss, SRGAN demonstrates the capability to generate super-resolved images that exhibit not only elevated PSNR, but also effectively capture intricate texture features, resulting in visually appealing outputs for human observers. The evaluation methods employed in this study are PSNR, SSIM, and MOS. The findings of [12] and [14] on the application of SRGAN in picture super-resolution demonstrated that SRGAN exhibited superior performance compared to other advanced techniques, as determined through the utilization of the MOS evaluation method, specifically in terms of perceptual quality. In terms of quantitative performance, SRResNet exhibits superior results in comparison to SRGAN, as measured by metrics such as PSNR and SSIM. Furthermore, ResNets have been shown to generate sharper super resolution image output [15].

ESPCN was proposed by [16], where the network primarily focuses on enhancing the resolution from LR to HR in the latter stages of the network. It achieves this by super-resolving HR data using LR feature maps. This obviates the necessity of doing the majority of the SR operation within the significantly HR. The present work employs the ESPCN technique for the purpose of real-time super-resolution of 1080p films. The findings of this study indicate that the ESPCN model outperforms the SRCNN model in terms of performance. ESPCN excel in running times with decent PSNR image output [17], this means that ESPCN is best suited to real-time super resolution processing that focuses on processing speed.

2. METHOD

The objective of this study is to determine the process of generating a HR image based on a provided LR image. The initial stage of this research involves pre-processing, specifically the reduction of the HR image to create the LR image. The training approach utilized two datasets, namely BSD500 and 600 webscrapped photos. The ESPCN-SRResNet model will utilize LR images as its input. The ESPCN-SRResNet model is trained to accurately predict HR images. The averaging of layers from ESPCN and SRResNet will be performed, followed by the calculation of the PSNR. Hyperparameter tuning is also conducted to ascertain the optimal model.

The testing procedure employed the face mask dataset supplied from Kaggle. The ESPCN-SRResNet model is utilized for the purpose of super resolution prediction. The dataset consists of super resolution images of face masks, along with their corresponding original photos. These images are subsequently utilized as input for three different CNN models: VGG16, MobileNet, and ResNet 50. There are three models that are commonly employed for the purpose of evaluating the performance of ESPCN-SRResNet. The study flowchart is depicted in Figure 1.

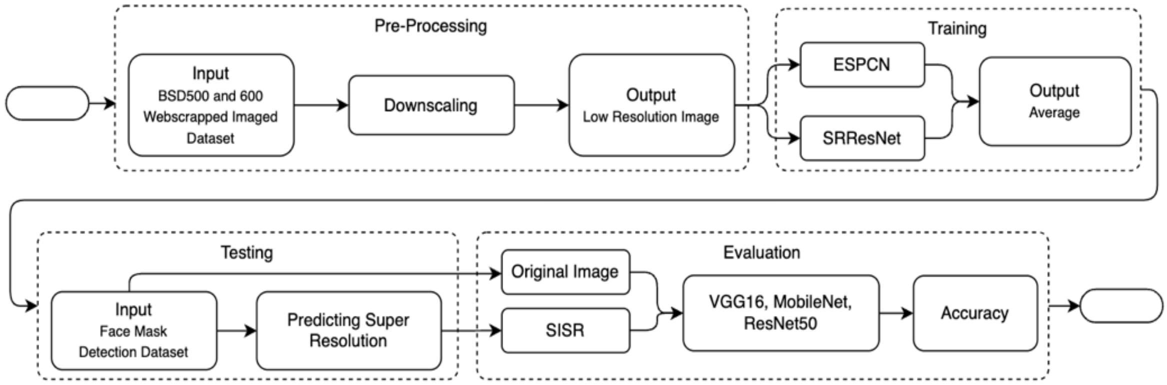


Figure 1. Flowchart of this study

2.1. ESPCN

The acronym ESPCN represents efficient sub-pixel convolutional neural network. Shi *et al.* [16] introduces a unique CNN architecture designed specifically for the real-time enhancement of 1080p videos through super-resolution techniques. The primary concept underlying ESPCN involves the extraction of feature maps within the LR space [18], as opposed to the typical strategy of utilizing the HR space. This facilitates enhanced computational efficiency and mitigates the complexity associated with the super-resolution procedure.

Furthermore, ESPCN presents a highly effective sub-pixel convolutional layer. The present layer is designed to acquire knowledge regarding an assortment of upscaling filters [19], which are subsequently utilized to upscale the ultimate LR feature maps into the HR output. By substituting the manually designed bicubic filter with these trained upscaling filters, the ESPCN model demonstrates the capability to produce images with enhanced sharpness and increased contrast. The proposed method exhibits superior performance in terms of both reconstruction accuracy and computational efficiency compared to earlier methods based on CNNs.

In general, ESPCN refers to an architectural design within the CNN framework that effectively utilizes LR feature maps and a sub-pixel convolution layer to accomplish real-time super-resolution of films with a resolution of 1080p. The enhancement of super-resolved image quality is achieved with a simultaneous reduction in computing complexity. The architectural design of ESPCN can be observed in Figure 2.

2.2. Super-resolution residual network

The SRResNet algorithm is employed for the purpose of enhancing the resolution of digital core images through a process known as super-resolution reconstruction. The proposed approach utilizes deep learning techniques in order to improve the resolution and overall quality of photos that have a poor level of detail. The network architecture of SRResNet comprises a generator network that is trained to establish the relationship between LR and HR pictures [20].

The difficulty of utilizing limited data from small samples for super-resolution reconstruction is effectively addressed by SRResNet with the implementation of transfer learning. Transfer learning is a widely employed approach in the field of machine learning, wherein the parameters of a pre-existing model are leveraged and repurposed for a different job [20]. In the context of SRResNet, the pre-existing model has acquired shared characteristics from an extensive dataset. Through the use of transfer learning, the SRResNet technique has the capability to effectively train distinctive features within limited sample data [21], hence diminishing the necessity for an extensive training dataset and alleviating the issues of overfitting that arise due to inadequate data.

In the domain of super-resolution reconstruction of digital core images, the utilization of SRResNet with transfer learning encompasses a two-step process [15]. The initial part of pre-training involves utilizing the plunger sample image and the LR image acquired from downsampling as the training dataset. This stage serves to illustrate the capability of SRResNet in achieving super-resolution reconstruction. Furthermore, the feature extraction layer of the pre-trained model is immobilized, and the model undergoes fine-tuning by utilizing real plunger samples and subsample images as the fresh training dataset. The utilization of transfer learning in the SRResNet method enables the reconstruction of finer details, including texture and pores, despite the presence of a limited number of training examples.

In brief, the SRResNet approach is designed for the purpose of super-resolution reconstruction [22]. It tackles the issue of limited data resulting from small sample sizes by employing transfer learning techniques. This involves leveraging pre-existing models and training specific features using tiny sample datasets. The architectural design of SRResNet can be observed in Figure 3.

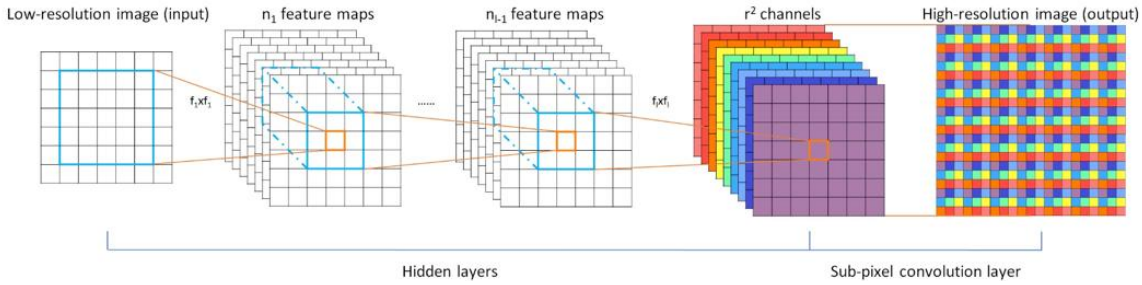


Figure 2. ESPCN architecture [16]

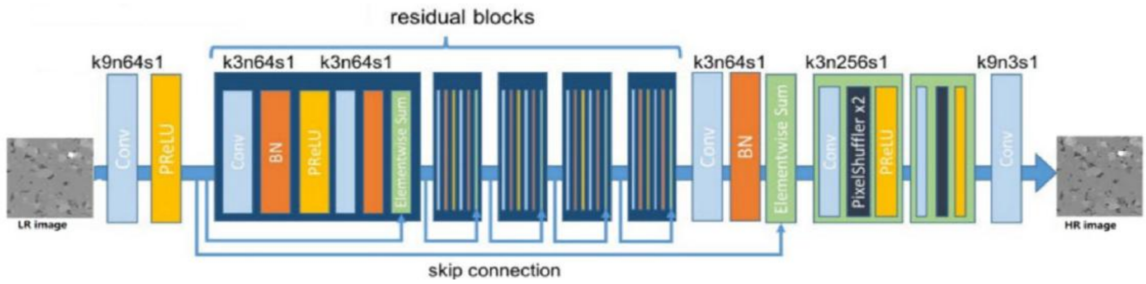


Figure 3. SRResNet architecture [20]

2.3. Signal-to-noise ratio

PSNR is a commonly used metric to evaluate the quality of reconstructed pictures in the field of super-resolution reconstruction. The relevance of this resides in its capacity to quantify the correlation between the signal's maximum potential power and the level of noise that impacts the accuracy of the signal. PSNR is crucial in evaluating the accuracy of super-resolution reconstruction by measuring the similarity between the reconstructed HR image and the actual HR image. The PSNR measure provides a numerical evaluation of the quality of the reconstructed image. A higher PSNR value indicates a closer match to the original image and hence fewer distortions. The primary advantage of PSNR is its dependability in catching subtle intricacies and general similarity between the reconstructed and original HR images, rendering it an indispensable tool for complete evaluation of image quality [23].

The PSNR is a widely used metric for evaluating the quality of compressed or distorted images [24]. It measures the discrepancy between the original and distorted images in terms of pixel values. PSNR, which is measured in decibels (dB), is determined by calculating the mean squared error (MSE) between corresponding pixels in two pictures and expressing it as the ratio of peak signal power (highest pixel value) to MSE. The PSNR metric is crucial in assessing image quality, since a greater value indicates superior image quality, while a lower value signals poorer quality. The primary advantage of PSNR is its dependability as a quantitative metric, offering a distinct indicator of the quality of the reconstructed or compressed image. Its widespread usage stems from its capacity to accurately measure the similarity in pixel values between the original and distorted images, providing a clear and understandable evaluation of image quality [25]. The formula for PSNR is elucidated in (1).

$$PSNR = 20 . \log 10 \left(\frac{MAX_i}{\sqrt{MSE}}\right) \tag{1}$$

2.4. Mask classification

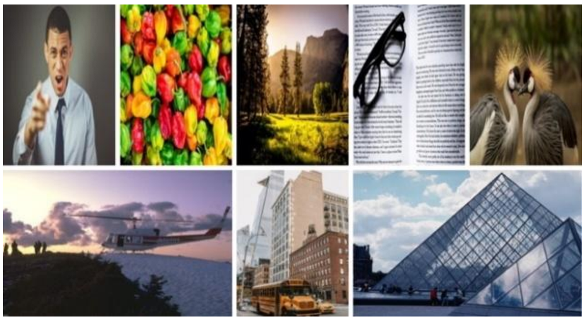
After image enhancement has been built and fine-tuned, this research will have to prove its impact on future research especially in computer science subjects. To do it, this research employ deep learning to measure image enhancement effect on image classification. To determine whether image enhancement improve or worsen deep learning performance, this research will focus to measure the accuracy of tested models. Accuracy is chosen as metrics of performance because it measure model performance on correctly classifying data. Accuracy is with (2).

$$Accuracy = \frac{CorrectlyGuessedImage}{AmountofImagesTested} x100\% \tag{2}$$

2.5. Dataset

The research utilized three datasets: 500 photos sourced from BSDS500, 600 images obtained from pexels.com, and 3,444 face mask images extracted from the face mask dataset. The BSDS500 and 600 datasets consist of a combined total of 1,100 photos. These datasets aim to encompass a wide range of subjects, including buildings, animals, humans, nature, and other relevant categories. The intention behind this comprehensive collection is to ensure that all aspects of imagery in society are well represented.

In the initial training phase, a total of 450 photographs were utilized for training purposes, while an additional 450 images were allocated for testing. Furthermore, an additional set of 200 images were designated for validation in order to obtain the most optimal model for super resolution. The second training session utilized a dataset consisting of 3,444 photos specifically designed for face mask detection. This dataset was further divided into subsets, with 2,754 images allocated for training purposes, 485 images for validation, and 205 images for testing. Illustrations depicting sample images for the initial and subsequent training sessions are presented in Figure 4. Figure 4(a) shows several examples of BSD500 dataset, also several examples of 600 scrapped images. Figure 4(b) shows several examples of face mask detection dataset.



(a)

(b)

Figure 4. Examples of dataset (a) BSD500 and 600 scrapped images and (b) face mask detection dataset

3. RESULTS AND DISCUSSION

The implementation of the super-resolution methodology required doing multiple iterations of tests that involved merging the ESPCN and SRResNet techniques. The conducted experiments encompassed the process of hyperparameter optimization, employing both manual and grid search methodologies. Experiments were performed using different hyperparameter configurations, including modifications in batch size, learning rate, ESPCN structure, and SRResNet structure. The manual approach encompassed the process of optimizing the quantity of layers and filters in both models. In the study conducted, four different structural variants were examined in the ESPCN model. These variations included (64, 64, 32), (64, 64, 64, 32), (128, 64, 64, 32), and (128, 128, 64, 32). Additionally, two different learning rate variations were tested, specifically 0.001, 0.0001, and 0.00001. Amount of SRResNet layers examined in this study are (64, [5*64], 64, 256) and (64, [7*64], 64, 256). That variations also used for grid search. The test results are presented in Table 1.

Table 1. Manual hyperparameter tuning

Number	ESPCN	SRResNet	Learning rate	PSNR predict (db)	Execution time (s)
1	64, 64, 32	64, [5*64], 64, 256	0.001	28.43	0.106
2	64, 64, 32	64, [7*64], 64, 256	0.001	28.4	0.381
3	64, 64, 64, 32	64, [5*64], 64, 256	0.001	28.51	0.272
4	64, 64, 64, 32	64, [7*64], 64, 256	0.001	28.38	0.127
5	128, 64, 64, 32	64, [5*64], 64, 256	0.001	28.39	0.377
6	128, 64, 64, 32	64, [7*64], 64, 256	0.001	28.35	0.452
7	**128, 128, 64, 32**	**64, [5*64], 64, 256**	**0.001**	**28.51**	**0.347**
8	128, 128, 64, 32	64, [7*64], 64, 256	0.001	28.52	0.92
9	64, 64, 32	64, [5*64], 64, 256	0.01	27.7	0.82
10	64, 64, 32	64, [5*64], 64, 256	0.0001	28.09	0.826
11	64, 64, 32	64, [7*64], 64, 256	0.0001	28.13	0.461
12	64, 64, 64, 32	64, [5*64], 64, 256	0.0001	28.05	0.427
13	64, 64, 64, 32	64, [7*64], 64, 256	0.0001	28.24	0.348
14	128, 64, 64, 32	64, [5*64], 64, 256	0.0001	28.05	0.487
15	128, 64, 64, 32	64, [7*64], 64, 256	0.0001	28.14	0.693
16	128, 128, 64, 32	64, [5*64], 64, 256	0.0001	28.12	0.347
17	128, 128, 64, 32	64, [7*64], 64, 256	0.0001	28.14	0.766

From Table 1, it can be seen that the highest PSNR obtained by the 8th experiment and the second highest PSNR obtained by the 7^{th} experiment. The PSNR values obtained by that two experiments are not much different, but there are quite significant differences in execution time. The 7^{th} experiment run 62.288% faster than the 8^{th} experiment, hence the 7^{th} trial in yielded the optimal result. The results of this experiment demonstrate that ESCPCN has exceptional time efficiency in computational tasks. There is only one experiment that use 0.001 learning rate and that is the 9th experiment. Its because 0.001 learning rate produce lower PSNR and longer execution time. Deeper ESPCN and SRResNet layers takes longer execution time, could be seen in experiment 8^{th} and 17^{th}.

Grid search is also used in this study to find the best hyperparameter for image enhancement. Result shown that compared to the best model from manual tuning (experiment 7), grid search has a little lower PSNR. The result could be seen in Table 2. As seen in Table 2, number of SRResNet layer and learning rate from grid search and manual tuning has the same value. However, model from grid search has less filter in the second layer than model from manual tuning. Execution time of grid search and manual tuning has no significant difference.

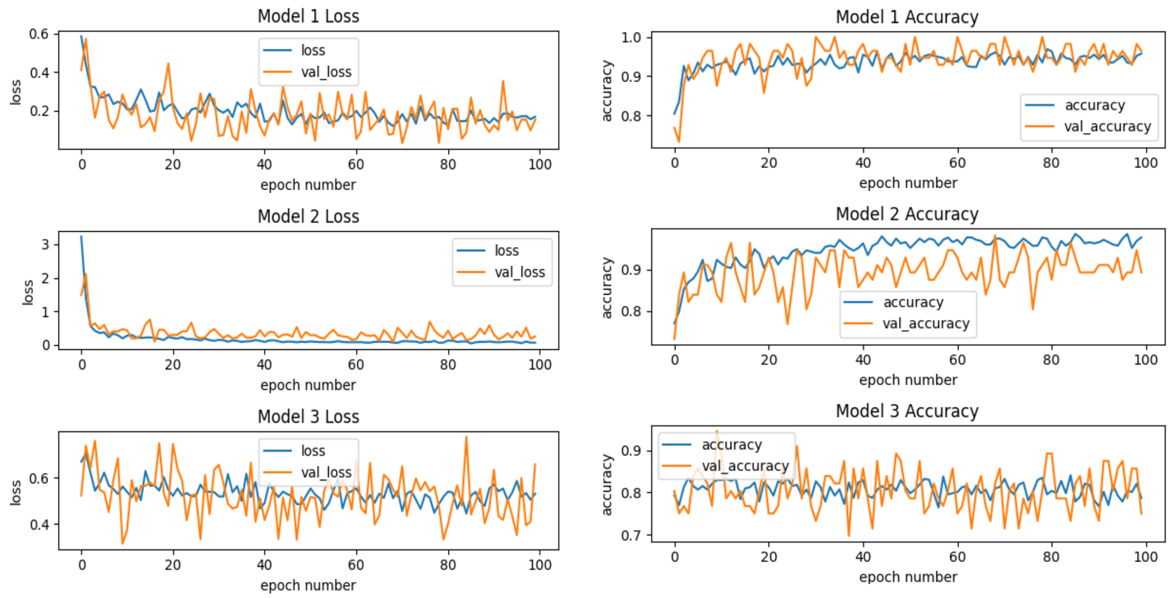
Table 2. Hyperparameter tuning with grid search

Experiment	ESPCN	SRResNet	Learning rate	PSNR predict (db)	Execution time (s)
Manual Tuning	**128, 128, 64, 32**	**64, [5*64], 64, 256**	**0.001**	**28.5142**	**0.347**
Grid Search	128, 64, 64, 32	64, [5*64], 64, 256	0.001	28.5125	0.387

As previously mentioned, there are three classification model used in this research. VGG16 as model number one, MobileNetV2 as model number two, and ResNet50 as model number three. Each models was trained using transfer learning method by freezing its layer and connecting 1 layer of fully-connected nodes and closed of with output layer which represents each classes in label. Figure 5 shows model training result. Model training loss could be seen in Figure 5(a) and model training accuracy could be seen in Figure 5(b).

Model 1 stands for VGG16, model 2 stands for MobileNetV2, model 3 stands for ResNet50. In the testing phase, input from the mask classification model is upscaled with super-resolution from the image enhancement model. Image enhancement model with the greatest performance with optimal PSNR and execution time result was chosen as the model which and integrated to the mask classification. Based on Table 1 the most optimal model with the greatest PSNR value and minimal execution time is model number 7. And based on the grid search method resulted in PSNR value of 28.5125 and execution time 0.3872871092 ms. Integration results is presented as comparison between three popular classification architecture which is

VGG16, MobilenetV, and ResNet50. Accuracy was used to measure performance impact of those three classification models, a comparison before and after super resolution will be presented.



(a) (b)

Figure 5. Model training (a) loss and (b) accuracy

Data from Table 3 shown an increase in accuracy by using super resolution to pre-process input images. VGG16 has the most significant improvement with 25% increase in accuracy, which is followed by MobileNetV2 with 5% increase in accuracy, and ResNet50 which has the least improvement in accuracy with less than 0.001% increase. Further analysis in deep learning classification model shows that VGG16 architecture has 16 layers of neural network [26], MobileNetV2 has 11 layers [27], and ResNet50 has 50 layers of neural network with residual network [26]. This indicates that increasing model complexity layer reduces the impact of super-resolution. Nevertheless, this finding also supports this research purposes which is improving results on deep learning model in a low performing device.

Data from Table 4 shown that fine-tuning image enhancement model with grid search and manual fine-tuning achieve similar results. VGG16 model test resulted in 96.30% accuracy using manual and grid search fine-tuning. followed by MobileNetV2 whose accuracy scored 76.30% in both manual and grid search fine-tuning. and ResNet50 achieved 71.30% accuracy in both manual and grid search. All the tested models showed a minuscule difference with less than 0.001% difference in accuracy between the two fine tuning methods.

Table 3. Classification accuracy comparison

Model	Classification accuracy	
	With super resolution	Without super resolution
VGG16	96.30%	71.30%
MobileNetV2	76.30%	71.30%
ResNet50	71.30%	71.30%

Table 4. Manual tuning and grid search impact on mask classification

Model	Hyperparameter tuning	
	Manual tuning	Grid search
VGG16	96.30%	96.30%
MobileNetV2	76.30%	76.30%
ResNet50	71.30%	71.30%

4. CONCLUSION

This research demonstrates the effective integration of ESPCN with SRResNet, resulting in a substantial enhancement of picture quality for face mask classifiers. The hand tweaking trials yielded an ideal configuration with a PSNR of 28.5142 dB and an execution time of 0.34704638 seconds. These findings showcase the supremacy of this technique in effectively producing exceptional outcomes through meticulous hand tweaking. The experimental results further emphasize the significant need of manually adjusting the hyperparameters to attain optimal outcomes. Additionally, the integration of super resolution greatly enhanced the performance of the VGG16 model, resulting in a notable increase in classification accuracy. Specifically, the accuracy improved from 71.30% without super resolution to 96.30% with super resolution.

Subsequently, this research presents the possibility of integrating ESPCN and SRResNet into alternative deep learning frameworks. Incorporating these two super-resolution approaches into a more comprehensive deep learning framework can significantly enhance picture quality in many application scenarios, extending beyond the scope of face mask classifiers. Hence, the forthcoming objectives involve investigating the prospective uses of ESPCN and SRResNet in additional deep learning frameworks, so expanding the beneficial influence of this study across many fields of image processing and pattern recognition.

ACKNOWLEDGEMENTS

The research presented in this paper was made possible through the financial support provided by Bina Nusantara University.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Mohammad Faisal Riftiarrasyid	✓	✓	✓	✓	✓	✓		✓	✓	✓			✓	
Rico Halim		✓				✓		✓	✓	✓	✓	✓		
Andien Dwi Novika	✓			✓			✓			✓			✓	✓
Amalia Zahra	✓		✓	✓			✓			✓	✓		✓	

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The datasets used in this study are publicly available and can be accessed as follows:

– Face mask detection dataset: the dataset for face mask detection, containing annotated images of three classes: individuals wearing mask, wearing it incorrectly and without wearing face masks, is available on Kaggle at https://www.kaggle.com/datasets/andrewmvd/face-mask-detection.
– Super-resolution dataset: the core dataset for image super-resolution is the Berkeley Segmentation Dataset 500 (BSD500 dataset) introduced by Martin et al. (2001), which can be accessed through the Berkeley Computer Vision Project website: https://www2.eecs.berkeley.edu/Research/Projects/CS/vision/bsds/.
– In addition, the study incorporates 600 supplementary images randomly sourced from the free stock photo website Pexels (https://www.pexels.com) to increase dataset diversity and robustness.

All data were used in accordance with the terms and conditions specified by the respective data sources.

REFERENCES

[1] K. Almezhghwi, S. Serte, and F. Al-Turjman, “Convolutional neural networks for the classification of chest X-rays in the IoT era,” *Multimedia Tools and Applications*, vol. 80, no. 19, pp. 29051–29065, Jun. 2021, doi: 10.1007/s11042-021-10907-y.

[2] M. Preti, F. Verheggen, and S. Angeli, “Insect pest monitoring with camera-equipped traps: strengths and limitations,” *Journal of Pest Science*, vol. 94, no. 2, pp. 203–217, Dec. 2021, doi: 10.1007/s10340-020-01309-4.

[3] L. Fu, F. Gao, J. Wu, R. Li, M. Karkee, and Q. Zhang, “Application of consumer RGB-D cameras for fruit detection and localization in field: a critical review,” *Computers and Electronics in Agriculture*, vol. 177, p. 105687, Oct. 2020, doi: 10.1016/j.compag.2020.105687.

[4] Y. Xi *et al.*, “DRL-GAN: dual-stream representation learning GAN for low-resolution image classification in UAV applications,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 1705–1716, 2021, doi: 10.1109/JSTARS.2020.3043109.

[5] V. Magoulianitis, D. Ataloglou, A. Dimou, D. Zarpalas, and P. Daras, “Does deep super-resolution enhance UAV detection?,” in *2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, Sep. 2019, pp. 1–6, doi: 10.1109/AVSS.2019.8909865.

[6] H. Yao, R. Qin, and X. Chen, “Unmanned aerial vehicle for remote sensing applications - a review,” *Remote Sensing*, vol. 11, no. 12, p. 1443, Jun. 2019, doi: 10.3390/rs11121443.

[7] Y. K. Ooi and H. Ibrahim, “Deep learning algorithms for single image super-resolution: a systematic review,” *Electronics (Switzerland)*, vol. 10, no. 7, p. 867, Apr. 2021, doi: 10.3390/electronics10070867.

[8] J. Shermeyer and A. Van Etten, “The effects of super-resolution on object detection performance in satellite imagery,” in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Jun. 2019, vol. 2019-June, pp. 1432–1441, doi: 10.1109/CVPRW.2019.00184.

[9] A. Deshpande, V. V. Estrela, and P. Patavardhan, “The DCT-CNN-ResNet50 architecture to classify brain tumors with super-resolution, convolutional neural network, and the ResNet50,” *Neuroscience Informatics*, vol. 1, no. 4, p. 100013, Dec. 2021, doi: 10.1016/j.neuri.2021.100013.

[10] C.-Y. Yang, C. Ma, and M.-H. Yang, “Single-image super-resolution: a benchmark,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 8692 LNCS, no. PART 4, Springer International Publishing, 2014, pp. 372–386.

[11] W. Li, D. He, Y. Liu, F. Wang, and F. Huang, “Super-resolution reconstruction, recognition, and evaluation of laser confocal images of hyperaccumulator Solanum nigrum endocytosis vesicles based on deep learning: comparative study of SRGAN and SRResNet,” *Frontiers in Plant Science*, vol. 14, Mar. 2023, doi: 10.3389/fpls.2023.1146485.

[12] C. Ledig *et al.*, “Photo-realistic single image super-resolution using a generative adversarial network,” in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jul. 2017, vol. 2017-Janua, pp. 105–114, doi: 10.1109/CVPR.2017.19.

[13] Y. Lu and Y. Wang, “Research on super-resolution image processing based on deep learning,” in *2023 6th International Conference on Computer Network, Electronic and Automation (ICCNEA)*, Sep. 2023, pp. 207–210, doi: 10.1109/ICCNEA60107.2023.00052.

[14] W. Zhang, Y. Liu, C. Dong, and Y. Qiao, “RankSRGAN: super resolution generative adversarial networks with learning to rank,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 10, pp. 7149–7166, Oct. 2022, doi: 10.1109/tpami.2021.3096327.

[15] L. Dong and K. Inoue, “Super-resolution reconstruction based on two-stage residual neural network,” *Machine Learning with Applications*, vol. 6, p. 100162, Dec. 2021, doi: 10.1016/j.mlwa.2021.100162.

[16] W. Shi *et al.*, “Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2016, vol. 2016-Decem, pp. 1874–1883, doi: 10.1109/CVPR.2016.207.

[17] F. Cao and H. Liu, “Single image super-resolution via multi-scale residual channel attention network,” *Neurocomputing*, vol. 358, pp. 424–436, Sep. 2019, doi: 10.1016/j.neucom.2019.05.066.

[18] J. Du, Q. Liu, and K. Chen, “License plate super-resolution reconstruction based on improved ESPCN network,” in *2019 IEEE 3rd Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC)*, Oct. 2019, pp. 122–125, doi: 10.1109/IMCEC46724.2019.8984050.

[19] Z. Chu *et al.*, “A generalizable sample resolution augmentation method for mechanical fault diagnosis based on ESPCN,” *Journal of Sensors*, vol. 2021, no. 1, Jan. 2021, doi: 10.1155/2021/7496007.

[20] Y. Wang, F. Niu, X. Zhang, J. Xiao, and C. Xu, “Super resolution reconstruction of digital core image based on transfer learning,” *Energy Reports*, vol. 8, pp. 87–94, Nov. 2022, doi: 10.1016/j.egyr.2022.10.069.

[21] Y. Zhang, W. Yu, and D. Zhu, “Terrain feature-aware deep learning network for digital elevation model superresolution,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 189, pp. 143–162, Jul. 2022, doi: 10.1016/j.isprsjprs.2022.04.028.

[22] S. Ullah and S. H. Song, “SRResNet performance enhancement using patch inputs and partial convolution-based padding,” *Computers, Materials and Continua*, vol. 74, no. 2, pp. 2999–3014, 2023, doi: 10.32604/cmc.2023.032326.

[23] D. Zhou, R. Duan, L. Zhao, and X. Chai, “Single image super-resolution reconstruction based on multi-scale feature mapping adversarial network,” *Signal Processing*, vol. 166, p. 107251, Jan. 2020, doi: 10.1016/j.sigpro.2019.107251.

[24] D. Kesrarat, K. Thakulsukanant, and V. Patanavijit, “A novel elementary spatial expanding scheme form on SISR method with modifying Geman&McClure function,” *Telkomnika (Telecommunication Computing Electronics and Control)*, vol. 17, no. 5, pp. 2554–2560, Oct. 2019, doi: 10.12928/TELKOMNIKA.v17i5.12799.

[25] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: from error visibility to structural similarity,” *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, Apr. 2004, doi: 10.1109/TIP.2003.819861.

[26] S. Mascarenhas and M. Agarwal, “A comparison between VGG16, VGG19 and ResNet50 architecture frameworks for image classification,” in *2021 International Conference on Disruptive Technologies for Multi-Disciplinary Research and Applications (CENTCON)*, Nov. 2021, pp. 96–99, doi: 10.1109/CENTCON52345.2021.9687944.

[27] Y. Hamid, S. Wani, A. B. Soomro, A. A. Alwan, and Y. Gulzar, “Smart seed classification system based on MobileNetV2 architecture,” in *2022 2nd International Conference on Computing and Information Technology (ICCIT)*, Jan. 2022, pp. 217–222, doi: 10.1109/ICCIT52419.2022.9711662.

BIOGRAPHIES OF AUTHORS

Mohammad Faisal Riftiarrasyid is a graduate's student at Bina Nusantara University Doctoral Program, Indonesia. He majored in doctor of computer science with experience in computer vision and robotics research when pursuing his bachelor's degree and enhance his experience in internet of things when pursuing his master's degree. After finishing his bachelor's degree in 2018 and master's degree in 2024, he decided to pursue higher education to achieve a career in the educational field. His research areas are internet of things and image processing. His field of study is inspired by his interest in technological advances in robotics. He envisions to achieve an expertise in drone development and drone path automation. He can be contacted at email: mohammad.riftiarrasyid@binus.ac.id.

Rico Halim is a master's student at Bina Nusantara University, Indonesia. He majored in computer science with experience in computer visions research when pursuing his bachelor's degree. After finishing his bachelor's degree in 2018, he decided to pursue higher education to achieve a career in the educational field. His research areas are deep learning, and computer vision. Inspired by his interest in images, he aims to pursue a career in education, driven by his dedication to advancing these fields. He can be contacted at email: rico.halim@binus.ac.id.

Andien Dwi Novika is a lecturer in the computer science program at Bina Nusantara University, having completed her master's degree in computer science there in June 2024. She previously earned her bachelor's degree in computer science from Diponegoro University, bringing a solid foundation in academic knowledge and practical experience. Her professional background includes roles as an integration staff at NicePay, a full stack developer intern at Maven Digital Asia, and a back-end developer intern at Gagas Cipta Persada. She can be contacted at email: andien.novika@binus.ac.id.

Amalia Zahra is a lecturer at the master of information technology, Bina Nusantara University, Indonesia. She received her bachelor's degree in computer science from the Faculty of Computer Science, University of Indonesia (UI) in 2008. She does not have a master's degree. Her Ph.D. was obtained from the School of Computer Science and Informatics, University College Dublin (UCD), Ireland in 2014. Her research interests cover various fields in speech technology, such as speech recognition, spoken language identification, speaker verification, speech emotion recognition, and so on. Additionally, she also has interest in natural language processing (NLP), computational linguistics, machine learning, and artificial intelligence. She can be contacted at email: amalia.zahra@binus.edu.