Upscaling Method for Enhancing Low-Resolution Image Classification using HAT Real-SRGAN

Yufis Azhara) and Gilang Nuril Haqimb)

Department of Informatics Engineering, Universitas Muhammadiyah Malang, Malang, Indonesia

b) Corresponding author: [gilangn89@gmail.com](mailto:gilangn89@gmail.com)

a) [yufis@umm.ac.id](mailto:yufis@umm.ac.id)

**Abstract.**  Low-resolution images present a significant challenge in image classification due to their inadequate visual information, which often leads to poor feature extraction and, consequently, reduced accuracy in classification tasks. This study focuses on the application of the image upscaling method using the pretrained Hybrid Attention Transformer Real-Super Resolution Generative Adversarial Network (HAT Real-SRGAN) to enhance the accuracy of low-resolution image classification. The dataset used is Corel-10k, which comprises 10,000 low-resolution images across 100 categories. Images from this dataset were upscaled using the HAT Real-SRGAN model prior to classification with the VGG-16 model. The results of this study demonstrate a significant increase in classification accuracy, from 69.0% with non-upscaled images to 79.4% with upscaled images. This increase indicates that upscaling can effectively enhance the quality of low-resolution images, thereby facilitating better feature extraction by the classification model.

**Keywords:** Super-resolution, Generative Advesarial Network, Convolutional Neural Network, Classification, Deep Learning

# INTRODUCTION

With the rapid advancement of technology, the production and storage of large-scale photographs are becoming more prevalent. The diversity of images, including those found in digital libraries, medical imaging, and other domains, is more intricate. Nevertheless, the manual management of these photographs by people is becoming progressively inefficient and time-consuming [1].

Hence, the advancement of systems like image classification and image retrieval is crucial for properly handling these images with high storage capacity. Image classification is a subfield of image processing and Machine Learning that aims to predict the class or attributes of a given image [2]. The process of image classification involves training a model to recognize patterns in images and then using that model to classify objects in previously unseen images. Deep learning technology has made a substantial contribution to improving the accuracy of image classification tasks [3].

Deep learning is a specific area within the broader science of Machine Learning. It employs deep artificial neural networks to identify and extract intricate characteristics from images [4]. A key component of deep learning is the Convolutional Neural Network (CNN), which works by directly extracting features from a collection of raw images [5]. While this method is recognized for its ability to accurately extract features from images, the issue of low-resolution images continues to be a common obstacle when using deep learning for image classification [6].

Low resolution refers to images with a restricted pixel count, resulting in less ability to extract information [7]. This can lead to the omission of crucial features in the images, hence impacting the accuracy of classification. Li Pei et al. highlights that low-resolution images significantly impact the results of object detection [8]. Specifically, images with lower resolutions, particularly those smaller than 32x32 pixels, introduce substantial accuracy challenges for face recognition systems. The study found that as the resolution decreases, the performance of these systems deteriorates rapidly, making it difficult to accurately detect and recognize faces

Several approaches can be employed to enhance the quality of an image, and one such approach is known as upscaling. Upscaling is a method that enables the enhancement of image resolution without compromising the information present within [9]. This method is categorized into two distinct types: classical and deep learning-based. Classical methods include interpolation algorithms such as bilinear, bicubic, or Lanczos, along with edge-directed methods that prioritize the preservation of contours and lines in images. Deep learning-based methods are algorithms that use deep learning models like Super Resolution-Generative Adversarial Network (SRGAN) [10], Enhanced Super Resolution-Generative Adversarial Network (ESRGAN) [11], Stable Diffusion [12], Super Resolution-Convolutional Neural Network (SRCNN) [9], and other similar models to enhance the resolution of images.

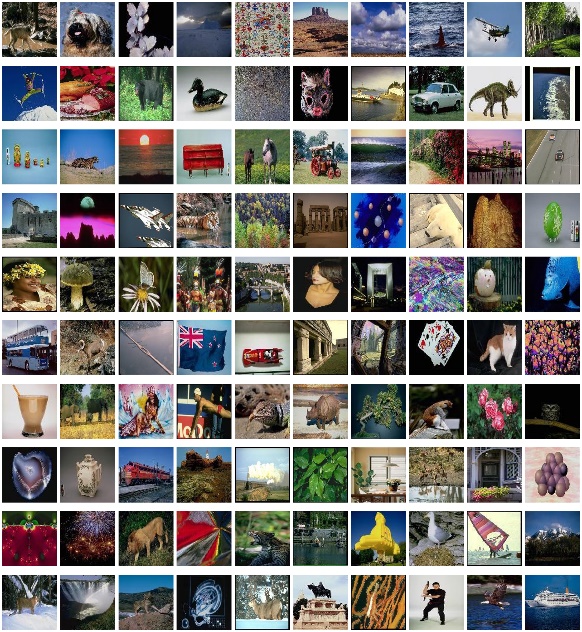
Multiple research works have been carried out on the categorization of low-resolution images. Yichao Xu et al. applied several GAN models such as Conditional GAN (CGAN) and Conditional Wasserstein GAN-Gradient Penalty (CWGAN-GP) to underwater sonar datasets [13]. Juan Wen et al. applied the ESRGAN model to low-resolution images of plant diseases [13]. Evan K. et al. conducted a study where they compared the outcomes of employing Multi-scale Super Resolution and Nearest Neighbor Interpolation for object recognition in satellite data [14] However, none of the previously stated studies have addressed the use of upscaling techniques for classifying low-resolution images, especially on the Corel-10k dataset.

Thus, in order to address the lack of information in prior research, the main focus is put on implementing the upscaling method using the HAT Real-SRGAN model for the purpose of classifying low-resolution images in the Corel-10k dataset. This approach aims to improve the quality of low-resolution images by increasing their resolution, with the goal of advancing the development of more accurate and efficient classification methods.

# METHODS

## DATASETS

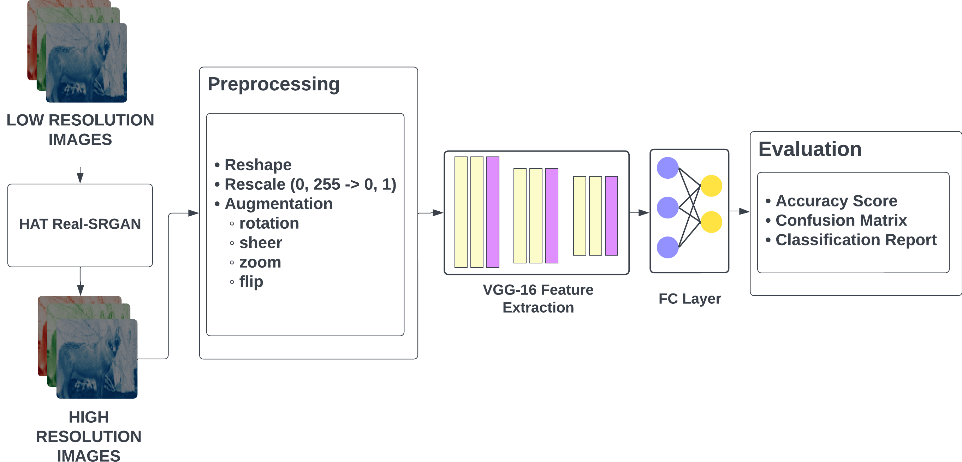
This research utilizes the Corel-10k dataset, which consists of 10,000 images spanning 100 categories, each category containing 100 images with a variety of subjects such as abstract paintings, sea waves, monkeys, deer, cards, drinks, trains, plants, hot air balloons, and more. Each image within this dataset originally has a low resolution of 128 x 192 or 192 x 128, which poses a challenge for effective feature extraction and classification accuracy. Representative images from the Corel-10k dataset are displayed in **FIGURE 1**.



**FIGURE 1.** Sample of Corel-10K Datasets

## PROPOSED METHODS

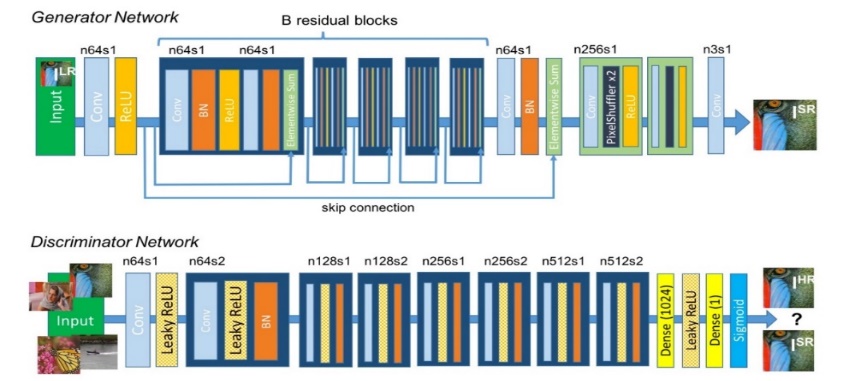
The primary objective of this study is to classify low-resolution images from the Corel-10K dataset, using an upscaling method to convert these images into high-resolution versions for more accurate classification. Following the upscaling process, the images undergo preprocessing and are then trained using the pretrained VGG-16 model. The performance is subsequently evaluated using several metrics. The overall experimental design is depicted in **FIGURE 2**.



**FIGURE 2**. Research workflow

The preprocessing stage occurs after the images have undergone upscaling. Initially, images are resized to 224 x 224 pixels to match the requirements of the model. Subsequently, rescaling transforms the color range of the images from 0-255 to 0-1. Augmentation techniques such as rotation, shearing, zooming (in/out), and flipping are then applied to enhance the size and quality of the training data, thereby improving model performance [15]. Following data augmentation, the dataset is divided into three parts: training, validation, and testing sets. This division is crucial for determining the accuracy of the classification [16].

To improve the resolution and quality of these images, the HAT Real-SRGAN model is employed. This model was introduced by Xiangyu Chen et al in the study "HAT: Hybrid Attention Transformer for Image Restoration." [17]. The model consists of three main parts: shallow feature extraction, deep feature extraction, and image reconstruction. Shallow feature extraction is the process that transforms input from low-dimensional space to high-dimensional space while achieving high-dimensional representation for each pixel. Deep feature extraction consists of several groups of residual hybrid attention (RHAG) and 3 x 3 convolutional layers. After that, a global residual connection is added to combine shallow features with deep features with the aim of building high-resolution images. The architecture of the HAT Real-SRGAN model can be seen in **FIGURE 3**. Each image in the dataset will be upscaled one by one, thus creating new images in the form of high-resolution images. The results of this process will be used as new data to be trained.



**FIGURE 3.** HAT Real-SRGAN Architecture

# RESULTS AND DISCUSSION

In order to compare the performance between low resolution and high resolution datasets, two experiments were conducted. In the first experiment, raw datasets were used without undergoing any upscaling process, while in the second experiment, datasets that had undergone upscaling were utilized. Both experiments employed a VGG 16 [18] classification model configured with specific layers: an input layer of dimensions 224 × 224 × 3, followed by 13 convolutional layers of size 3 × 3 with ReLU activation function, 5 Max Pooling layers, and 3 Dense layers. The Fully Connected (FC) layers consisted of 1 Average Pooling layer, 1 Dense layer with 512 neurons, and an output layer with 100 classes (num\_classes). This setup allowed for a comprehensive evaluation of the classification model's performance under varying dataset resolutions.

In this research, the evaluation metrics used are the classification report and accuracy score, focusing on the performance of the classification task. Due to the small (low-resolution) dataset, metrics such as PSNR or SSIM were not utilized as they are less effective in assessing performance in this context.

|  |  |
| --- | --- |
|  |  |
| (a) | (b) |

**FIGURE 4.** Model Accuracy and loss: (a) before and (b) after upscaling

Figure 3 illustrates the results of training the VGG-16 model for the two types of data described earlier. The model training was conducted for a maximum of 100 epochs with early stopping to halt training if there was no progress (training would stop after 5 epochs). In the first experiment (low resolution), the graph (**FIGURE 4a**) shows that training accuracy improved initially and gradually until reaching about 0.7 in the final epoch, while validation accuracy also increased but started to plateau significantly after the 25th epoch. The loss values in this experiment decreased sharply at the beginning of training and dropped to reach 1.0 for training loss and 1.1 for validation loss.

In the second experiment (high resolution) using the HAT Real-SRGAN model, training lasted only up to the 70th epoch, see **FIGURE 4b**. Training accuracy in this experiment showed significant improvements from the beginning until reaching about 0.9 at the end of the epoch, whereas validation accuracy did not improve and remained consistent around 0.76 to 0.79. The training loss value dropped drastically to below 0.5, and validation loss remained unchanged, consistent around 1.0 to 0.8. However, the accuracy and loss curves in this experiment exhibited slight overfitting on the training data.

**TABLE 1**. F1-Score improvement between low-resolution and high-resolution

|  |  |  |
| --- | --- | --- |
| Category | F1-Score | |
| low-resolutioon | high-resolution |
| 4\_sea\_nature | 0.18 | 0.63 |
| 23\_castle | 0.48 | 0.82 |
| 31\_greek | 0.61 | 0.59 |
| 34\_air\_balloon | 0.50 | 0.76 |
| 2\_deer | 0.55 | 0.82 |
| 18\_view\_art | 0.43 | 0.76 |
| 16\_sailboat | 0.53 | 0.84 |
| 3\_astronaut | 0.48 | 0.82 |
| 63\_tree | 0.38 | 0.62 |
| 64\_arctic | 0.25 | 0.70 |
| 87\_coyote | 0.12 | 0.29 |
| 88\_bighorn\_sheep | 0.26 | 0.61 |
| 99\_wolf | 0.20 | 0.57 |

The results obtained on the low-resolution image dataset included several categories that achieved very good scores, but many categories still showed relatively low scores. However, the results using upscaled data (high resolution) show that several categories with previously low F1-scores experienced significant improvements. For instance, the sea nature category saw a drastic increase in its F1-score from 0.18 to 0.63, the air balloon category improved from 0.5 to 0.76, view art increased from 0.4s3 to 0.76, and tree improved from 0.38 to 0.62. The enhancement of image resolution through upscaling helped the model capture more details and features, thereby improving classification accuracy. This indicates that image resolution plays a crucial role in the performance of image classification models. The improvements in these categories, among others previously mentioned, can be further examined in **TABLE 1**.

Despite the significant improvements observed in many categories after upscaling the data, some categories still show relatively low scores. These include cougar, fish, coyote, and bird. From the prediction results with 10 images per category, the model misclassified cougar as coyote and bighorn sheep once each, indicating a visual similarity between these three categories. This issue also occurred with the coyote category, where the model misclassified it as wolf four times. **FIGURE 5** provides a visual representation of the categories cougar, coyote, and wolf.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Output image |  | Output image |  | Output image |
| (a) |  | (b) |  | (c) |

**FIGURE 5.** Visual of similar images which consist of (a) cougar, (b) coyote, (c) wolf

The accuracy results for the average of all categories, as shown in **TABLE 2**, indicate that the classification model using the test data without upscaling (low resolution) achieved an accuracy of 69.0%. Meanwhile, in the second experiment using the upscaled dataset (high resolution), the accuracy significantly increased to 79.4%. This demonstrates that using upscaling can significantly enhance image classification accuracy, making this method a viable consideration in image processing for classification purposes. Additionally, the HAT Real-SRGAN model [19] used in this study was able to compete with other upscaling methods such as bicubic interpolation and SRGAN [10], as demonstrated in **TABLE 3.**

**TABLE 2**. Accuracy score with and without upscaling

|  |  |
| --- | --- |
| Method | Accuracy Score (%) |
| Raw Image Dataset (low resolution) | 69.0 |
| Upscaled Image (high resolution) using HAT Real-SRGAN | 79.4 |

**TABLE 3**. Accuracy score for each upscaling method

|  |  |
| --- | --- |
| Model Upscaling | Accuraci Score (%) |
| Bicubic Interpolation | 74.4 |
| SRGAN [10] | 71.6 |
| **HAT Real-SRGAN [19]** | **79.4** |

# CONCLUSIONS

In this study, the upscaling method was employed to enhance the accuracy of low-resolution image classification. The method utilized a GAN model named HAT Real-SRGAN developed by Xiangyu Chen et al. in the research paper "HAT: Hybrid Attention Transformer for Image Restoration" [19]. There were two experiments conducted: the first using raw datasets without upscaling, resulting in a classification accuracy of 69.0%. After implementing upscaling using the HAT Real-SRGAN model in the second trial, the accuracy improved significantly to 79.4%. Nevertheless, despite the overall enhancement in accuracy, the model found difficulties in accurately predicting certain categories due to visual similarities. The categories consist of cougar, fish, coyote, and bird.

However, applying upscaling methods proved to be successful in enhancing accuracy in this study. Even traditional upscaling techniques such as bicubic interpolation attained an accuracy of 71.4%, while SRGAN produced a little higher accuracy of 71.6%. This shows the substantial influence that image resolution can exert on the accuracy of image classification.

# Acknowledgments

The authors would like to express their gratitude to the anonymous reviewers for their constructive comments that helped improve the clarity and rigor of this manuscript.

# References

1. Lorente, Ò., I. Riera, and A. Rana, *Image classification with classic and deep learning techniques.* arXiv preprint arXiv:2105.04895, 2021.

2. Ahmed, A., A. Jalal, and K. Kim, *A novel statistical method for scene classification based on multi-object categorization and logistic regression.* Sensors, 2020. **20**(14): p. 3871.

3. Pavithra, L.K., T. Sree Sharmila, and P. Subbulakshmi, *Texture image classification and retrieval using multi-resolution radial gradient binary pattern.* Applied Artificial Intelligence, 2021. **35**(15): p. 2298-2326.

4. Peryanto, A., A. Yudhana, and R. Umar, *Rancang bangun klasifikasi citra dengan teknologi deep learning berbasis metode convolutional neural network.* Format J. Ilm. Tek. Inform, 2020. **8**(2): p. 138.

5. Krishna, S.T. and H.K. Kalluri, *Deep learning and transfer learning approaches for image classification.* International Journal of Recent Technology and Engineering (IJRTE), 2019. **7**(5S4): p. 427-432.

6. Yang, W., et al., *Deep edge guided recurrent residual learning for image super-resolution.* IEEE Transactions on Image Processing, 2017. **26**(12): p. 5895-5907.

7. Zhou, L., et al. *Improving low-resolution image classification by super-resolution with enhancing high-frequency content*. in *25th International Conference on Pattern Recognition (ICPR)*. 2021. IEEE.

8. Li, P., et al., *Face recognition in low quality images: A survey.* arXiv preprint arXiv:1805.11519, 2018.

9. Dong, C., et al., *Image super-resolution using deep convolutional networks.* IEEE transactions on pattern analysis and machine intelligence, 2015. **38**(2): p. 295-307.

10. Ledig, C., et al. *Photo-realistic single image super-resolution using a generative adversarial network*.

11. Wang, X., et al. *Esrgan: Enhanced super-resolution generative adversarial networks*.

12. Rombach, R., et al. *High-resolution image synthesis with latent diffusion models*.

13. Xu, Y., et al., *Underwater sonar image classification using generative adversarial network and convolutional neural network.* IET Image Processing, 2020. **14**(12): p. 2819-2825.

14. Koester, E. and C.S. Sahin, *A comparison of super-resolution and nearest neighbors interpolation applied to object detection on satellite data.* arXiv preprint arXiv:1907.05283, 2019.

15. Volkova, S. *An Overview on Data Augmentation for Machine Learning*. Springer.

16. Kavzoglu, T., *Increasing the accuracy of neural network classification using refined training data.* Environmental Modelling & Software, 2009. **24**(7): p. 850-858.

17. Chen, X., et al., *Hat: Hybrid attention transformer for image restoration.* arXiv preprint arXiv:2309.05239, 2023.

18. Simonyan, K. and A. Zisserman, *Very deep convolutional networks for large-scale image recognition.* arXiv preprint arXiv:1409.1556, 2014.