JOURNAL OF MULTIDISCIPLINARY & TRANSLATIONAL RESEARCH

Journal of Multidisciplinary and Translational Research (JMTR)

journal homepage: https://journals.kln.ac.lk/jmtr/

Micro facial expression recognition using generative adversarial network generated super resolution images

Janotheepan Mariyathas[1]*, Yanusha Mehendran[1], Mohamed Muath[1], Fathima Lasna[1], Sangaralingam Jathukulan[1] and Jayalath Ekanayake[1]

[1]*Department of Computer Science and Informatics, Uva Wellassa University of Sri Lanka, Badulla, Sri Lanka*

Abstract

Recent years have seen a remarkable acceleration in the development of micro facial expression recognition techniques. The recordings made in daily lives engaging with the actual environment provide the prime sources for many studies, yet the quality of these data is sometimes low. As a result, low resolution micro expression images have become a new area of research. Due to the lack of clear understanding to differentiate inter class features and to accurately recognize the micro facial expression among several categories is challenging. The ability to distinguish between micro-facial features is further diminished by the low resolution of such images which doubles the recognition challenge. This work provides a novel method that makes use of the Generative Adversarial Network to apply a super resolution methodology to overcome the low-resolution conflicts on facial micro expression. The overall performance of the model was assessed with the recognition accuracy achieved using a support vector machine. Additionally, the image quality was measured employing different methods to assess reconstruction performance. The proposed approach was tested using low resolution images simulated from the CASME-II, Spontaneous Micro-expression database (SMIC-HS) and SMIC-subHS dataset, demonstrating the utility of the method. The ESRGAN model achieves the best reconstruction performance for micro expressions, with image metrics of 30.887 dB PSNR, 0.000865 MSE, and 0.938 SSIM, considering the scale factor in five defined classes as happiness, surprise, disgust, depression, and others.

Keywords: Micro expression, General Adversarial Network, Super resolution images

Article info

Article history:
Received 27th January 2025
Received in revised form 25th February 2025
Accepted 30th March 2025
Available online 12th June 2025

ISSN (E-Copy): ISSN 3051-5262
ISSN (Hard copy): ISSN 3051-5602
Doi: https://doi.org/10.4038/jmtr.v10i1.78
ORCID iD: https://orcid.org/0009-0007-8403-8725
*Corresponding author:
E-mail address: janotheepan.m@uwu.ac.lk
(Janotheepan Mariyathas)

Introduction

Emotional expression is a basic way to express feelings and basically comes naturally to everyone. Perceived human emotions are defined as verbal and non-verbal expressions. When these emotions are expressed facially, they are referred to as non-verbal expressions and are known as facial expressions. It is extremely challenging for a person to control the appearance of micro facial expressions, because micro expressions will be held for a small period of time. Humans can identify the state of mind and behavior of a person by facial expressions which is important for interpersonal communication.

Micro Expression Recognition (MER) is important in medical imaging, satellite imaging, surveillance, psychology, crowd scenarios, airport security and criminal investigations. There is a wide range of applications, where low-quality photos must be analyzed. In such situations, the quality of images can be compromised due to poor lighting conditions, the use of cheap, low-quality imaging devices, low resolution of downloaded images, and restricted memory capacity. These factors make it extremely difficult for both humans and machines to accurately interpret the available information.

Consequently, a new research direction has emerged to address the problems arising from low resolution (LR) images, with a particular focus on micro expressions (Zhao & Li, et al., 2019). Deep learning-based super resolution (SR) models have been intensively researched in recent years due to the improvement of deep learning techniques, and advent of state-of-the-art performance of many SR benchmarks. Convolutional Neural Networks (CNN), an early deep learning technique, as well as more modern, promising Generative Adversarial Network-based SR approaches, have all been used to solve SR tasks.

Micro expression recognition techniques have advanced to an extraordinary milestone in the past decades. As a result, low resolution micro expression images have become a new area of research. Due to less obvious inter-class discriminative features, it is extremely difficult to determine a specific class of micro expression among multiple classes. The ability to distinguish between little facial features is further reduced by the low resolution of such images. This study focused on developing a novel deep learning-based pipeline for handling low resolution micro expression recognition difficulties and to analyze generative adversarial network (GAN)-based method developed for addressing low resolution micro expression problems. The overall performance was assessed based on recognition accuracy discovered using classifier models. In addition to that, the final model show high accuracy on predicting micro expressions by upscaling the LR images using SR and deep learning techniques. The proposed method was tested using SR images generated from CASME-II dataset, Spontaneous Micro-expression database (SMIC-HS) and SMIC-subHS.

Recent technologies clearly show advanced facial micro expression detection methods employing both shallow and deep learning techniques (Chang, C., & Lin, C., 2011). The low resolution of the images makes it very hard for both machines and people to make use of the information they contain. Therefore, additional research is needed to address the LR problem, especially for micro expressions (Goodfellow et al. 2014; Hung K. et al. 2019). Compared to images with standard resolution, LR images have dispersed and weakly aligned pixels, which produce fewer image details. Image super resolution (ISR) is the process of predicting an HR image by rebuilding an image from an LR input image, and the reconstructed image is known as a super-resolved image. There are currently several cutting-edge deep convolutional neural networks (e.g., CNNs) that applied versions of residual dense networks (RDN), residual dense blocks (RDB), and recursive learning architectures (Ledig C. et al. 2017; Li G. et al. 2019; Oh Y. et al. 2018; Sharma P. et al. 2021), and the SR problem with success. Recovery of crucial facial information is the primary motive of super resolving LR facial images.

The difficulty for face SR algorithms is to reconstruct the face while maintaining attribute consistency with the underlying high-resolution images. Therefore, to enable facial expression analysis, it is essential for face SR algorithms to restore face details in the reconstructed image. There is practically little variation in facial expressions amongst the many kinds of micro expressions. These expressions are extremely delicate and exhibit less pronounced interclass discriminative characteristics. Overall, micro expression recognition (MER) is further compounded by the insufficient availability of information related to LR and the lack of a comprehensive LR micro expression dataset (Goodfellow et al. 2014).

It has been previously suggested to use a face hallucination method on individual frames to reconstruct HR images from LR images to overcome some of the issues with LR micro expressions (Goodfellow et al. 2014). Currently, the datasets for micro expression (ME) only include HR pictures. For instance, the Spontaneous Micro-expression Database (SMIC-HS) (Wang X. et al. 2019) micro expression dataset includes HR images with a resolution of roughly 190 × 230, while LR images typically have a resolution of less than 50 x 50 (Goodfellow et al. 2014). Hence, in their work, Goodfellow et al. (2014), simulated three existing HR micro expression image datasets to obtain the LR micro expression image datasets i.e., CASME-II (Xiaobai Li et al. 2013), Spontaneous Micro-expression database (SMIC-HS) and SMIC-subHS (Wang X. et al. 2019). On these datasets, improved overall classification accuracy was attained through experimental findings. However, this was also accompanied by low accuracy for specific classes.

According to the data, expressions with particularly LR showed a sharp fall in recognition accuracy. Datasets from CASMEII and SMIC-HS produced misclassifications that were of higher magnitude than those from SMIC-subHS. SR on micro expression is a novel idea proposed by Goodfellow et al. (2014), while prior work using macro expression has been used in face SR expression analysis (Goodfellow et al. 2014). By expanding on this idea, deep learning methodology has been introduced into the LR micro expression recognition framework (Yan W. et al. 2014).

Generative Adversarial Network (GAN) technique and its variation (Zhang, Y et al. 2018) was suggested to assess how well it performs in addressing the problem of low-resolution targeted micro expression (Yan W. et al. 2014). This work represents the first attempt to apply GANs explicitly to low resolution micro expression problems. Low resolution micro expression images were obtained by simulating data from the SMIC-HS dataset (Yan W. et al. 2014).

This study assessed the suggested strategy using low resolution ME pictures created by imitating facial expression recognition (open source) data. LR, SR techniques can be used on both videos and pictures to produce the matching super-resolved versions of each. SR images can also be generated from LR video. When conducting the reconstruction process only the LR images were used (Yan W. et al. 2014). To identify the micro facial features, LBP-TOP and LPQ-TOP along with a support vector machine (SVM) (Zhao, G., & Li, X., 2019) was used to classify the entire data into the defined emotion classes such as positive, negative and surprise (Yan W. et al. 2014).

Macro expressions typically last longer than micro expressions, which usually last between 0.04 to 0.2 of a second (Sharma P. et al. 2021). The process involved in MER is considerably more challenging than macro expression. Nevertheless, recent technologies clearly show advanced facial micro expression detection methods employing both shallow and deep learning techniques (Chang, C., & Lin, C., 2011). In a low-resolution image it is a challenging task to identify micro expressions. Three types of class labels of facial expressions have been predicted from LR images using SMIC-HS dataset (Yan W. et al. 2014). Because these class labels are insufficient to identify the emotions of an individual, , the study considered increasing the class labels of facial expressions . SMIC-HS dataset has 164 samples from 16 participants. Since the number of samples and the number of participants in this dataset were not sufficient, , other datasets were also included in the study. CASME II dataset which contains 8,042 samples from 26 participants was used to predict the micro facial expression. . Five types of class labels of facial expressions were predicted from LR images using facial expression recognition (open source) dataset. The study utilized deep learning methods to accurately detect facial expression recognition.

Methodology

Data collection and preprocessing were conducted as follows. CASME II, SMIC-HS, and SMIC-subHS datasets were considered for the study. Among the datasets 70%, 20%, 10% were allocated for training, testing, and validation purposes, respectively. The degradation module was used to create the low-resolution dataset. To Deep learning algorithms were implemented in Python programming language due to its responsive and adaptive environments for easy automation and analyzing tasks. Python supported many libraries and frameworks in its environment for providing efficient performance. OpenCv, matplotlib, Keras, TensorFlow, Numpy, and Scipy libraries were employed.

Image processing was simply referred to as OpenCv. It is a large open-source library for computer vision, machine learning, and image processing. The main purpose of image processing was to gain useful information or to enhance the original image by applying some operations. This library was used to perform image degradation (down sample, gaussian blur and median blur). TensorFlow was a Python-friendly, open-source library for numerical computation that streamlines machine learning and neural network development. NumPy was a Python library designed for working with arrays.

Google Colab offered the easiest way to perform Python and deep learning on a single machine. It also worked with the open-source libraries and packages for convenient implementation.

Image Degradation

The study employed the dataset, a well-known micro expression dataset, to evaluate the suggested approach. As was previously mentioned, this dataset contains HR images shown in

Figure 1. In this study, first, HR images were used (×128) . Hence, the SR technique could not be applied directly. Then, the LR micro expression image dataset was created using the existing HR micro expression image datasets because these approaches are appropriate for LR images.

Down sampling was used again (×32) on the received image dataset to subject the HR images to further degradation by Gaussian blurring on the current HR micro expression images of the dataset to create degraded images. Then, median blur was randomly applied on the sub folders of the dataset.

This process of creating lower quality images from its HR images is known as image degradation. Such a degradation model cause the images to become blurrier and lose quality like the image shown in the Figure 2. The effects and the degree of loss of details of the degraded images can be seen in the resultant images. Both HR and LR images have different quality and resolution levels. The newly created datasets, containing degraded images simulated from existing high-resolution datasets, were now suitable for use with SR algorithms.

Figure 1. High resolution image

Figure 2. Low-resolution image output

Image Reconstruction

The aim of this process was to upscale LR images using SR techniques. To reach the expected outcome, the study considered the ESRGAN model. These high-resolution methods in the research work simulated LR pictures from the CASME-II dataset. A generator, a discriminator, and a loss function were the ideal components of a GAN-based SR technique. The discriminator determined whether the generated image shown in Figure 3, was sufficiently realistic while the generator function estimates an HR image for the associated LR image.

The loss function also significantly aided in the optimization of the entire GAN framework. Similar to other GAN architectures, the Super Resolution GAN (SRGAN) consists of two parts: the Generator and the Discriminator. The Generator produced data based on probability distribution, while the Discriminator determined whether the data comes from the input dataset or the Generator. The Generator then optimized the data generated to better fool the Discriminator.

Figure 3. Super resolution image output

Micro facial feature extraction and feature classification

In this study, CNN Architecture was adopted to extract micro facial features illustrated in Figure 4. The datasets were separated into two sets as training dataset and testing dataset. Individual instances of data in the training set contain various attributes and class labels. Using this information along with the attributes of the test data, CNN constructed a model that can predict the class labels for each test data instance. In the experiments conducted in this work, classification is achieved using a classifier. The standard architecture of a convolutional neural network consisted of an input layer, a few convolutional layers, a few pooling layers, a few fully connected layers, and an output layer. In this research, three (03) convolution layers and two (02) hidden layers were used for micro facial extraction. Pooling layers and fully connected layers are used for feature classification.

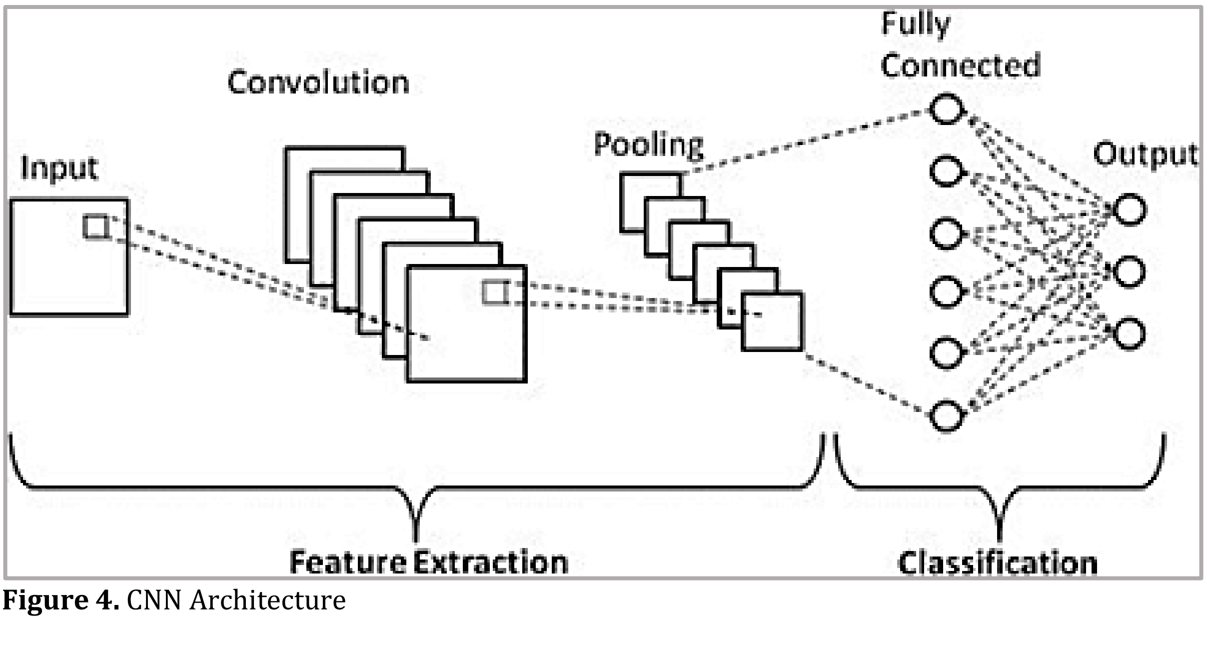

Figure 4. CNN Architecture

Results and Discussion

As previously mentioned, experiments were conducted using datasets, which contain 256 micro expression samples classified as repression (27), happiness (31), surprise (25), disgust (63), and other (99) class labels. These HR images have roughly 280 x 340 facial resolution. By using the degradation method, the study simulated the LR dataset to produce images that are appropriate for the SR algorithm. This section provides the outcomes on the implementations and the parameters considered for several SR models and feature extraction techniques in this research work.

F-irst the study preprocessed dataset using class labels. After that, HR images were degraded to LR images using down sample, gaussian blur, and median blur methods. Targeting LR images to recognize Gaussian blur method shows high performance to mute the noise in the image. Median blur showeds efficient performance on removing noises in the image without damaging the image edges. Theimage size (32×32 pixel) was then decreased and guassian blur ×3 kernel size was applied. The best low resolution images were obtained through this process. The process is shown in Figure 5. Finally GAN model was created using ESRGAN algorithm. The resolution of the SR images reconstructed using the proposed approach were proportional to the input resolution used. Calculated PSNR, SSIM and MSE values were better than the previously reported methods.

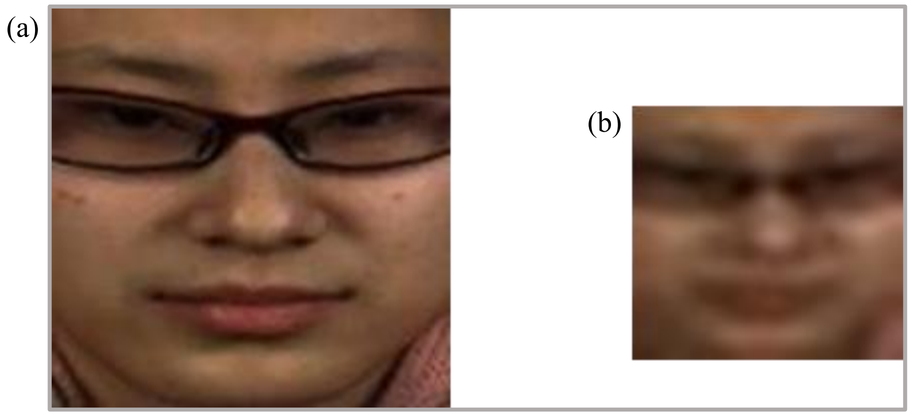

Figure 5. Instance of (a) HR image at 128x128 and (b) LR image at 32x32 simulated from CASME-II dataset by applying image degradation.

The resolution of the original high-resolution (HR) images was standardized to a size of 128 x 128 to obtain LR images. Specifically, these HR images were down sampled by a factor of 4 to obtain LR images of size 32x32, referred to as LR32 henceforth. Instances of these LR32 images were then upscaled using different SR algorithms, with a scaling factor set to four (x4), as described above, to obtain 128 x 128 super-resolved images, referred to as SR32 henceforth. Table 1 describes the performance of the SR algorithm evaluated using three quality metrics: peak signal-to-noise ratio (PSNR), structural similarity index measure (SSIM), and mean squared error (MSE). Essentially, a low MSE value, a high PSNR value (in decibels, dB), and an SSIM value closer to 1 indicate better quality reconstructed images. The PSNR and SSIM values were calculated by comparing the HR images with the reconstructed images. The average PSNR, SSIM and MSE values obtained for resultant image sequences at various resolutions are presented in Table 1 for various SR models.

Table 1: PSNR, SSIM and MSE Value for 32x32 SR image performance in SRGAN method

Method	SRGAN
Resolution	SR32
PSNR Value	30.887
SSIM Value	0.938
MSE Value	0.000865

The quality of the super-resolved images produced by the proposed approach is directly related to the input resolution. This is supported by numerical evidence, as the PSNR and SSIM values at SR32 are higher by 1.387 and 0.088, respectively, compared to previous research using the ESRGAN method. The ESRGAN model achieves the best reconstruction performance for micro expressions, with image metrics of 30.89 dB PSNR, 0.000865 MSE, and 0.938 SSIM, considering the scale factor. Previous studies reporteded 29.5 dB PSNR, 0.00085 MSE and 0.8502SSIM values (Li, G., Shi, J., Peng, J., & Zhao, G., 2019).

Conclusion

This study, provides a thorough analysis of a GAN-based method for identifying micro facial expressions in LR images. The study integrates deep learning-based reconstruction method into the MER framework with the goal of producing a high-quality up-scaled image. The technique was effective in micro facial picture details, which implicitly improved overall identification accuracy under all circumstances. The ultimate goal of this thorough study was to present a brand-new deep learning-based pipeline for resolving low resolution MER problems as well as to analyze GAN- based techniques specifically for low resolution ME problems. The obtained results unmistakably suggest that GAN and its variants can be further utilized for optimizing this particular challenge focusing on micro expressions. Despite the fact that the research work is still in its early stages, the outcomes are encouraging. This study shows a significant performance on ESRGAN model with 30.887 dB PSNR, 0.000865 MSE, and 0.938 SSIM values. Due to the lack of a dataset with LR micro expressions, experimentation with them was difficult. Therefore, HR images were used in the study with degradation technique to make the dataset effective. It must be noted that the dataset imbalance that exists for the majority of the accessible ME datasets is not addressed by the current study; hence, resolving this limitation can undoubtedly be another objective for future efforts to achieve more robust performance.

Acknowledgement

Sincere gratitude is extended to all who supported this research through valuable guidance, constructive feedback, and insightful perspectives. M. Janitheepan, A.M. Muath, K.F. Lasna, S. Jathukulan, M. Yanusha, and Prof. E.M.U.W.J.B. Ekanayake made significant contributions to its success, with each member handling specific aspects of the study. Appreciation is also extended to Mrs. S.D.H.S. Wickramarathne, Head of the Department of Computer Science and Informatics, Uva Wellassa University of Sri Lanka, for her continued support.

Sources of Funding

Secondary data was utilized for the completion of the research. All other aspects of the project were carried out using self-funded resources.

Conflict of Interest

The authors declare no conflicts of interest.

References

Chang, C., & Lin, C. (2011). LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST), 2*(3), 1–27. https://doi.org/10.1145/1961189.1961199

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. , "s. (2014). Generative adversarial network. *Communications of the ACM, 63*, 139-144. https://doi.org/10.1145/3422622

Hung, K., Wang, K., & Jiang, J. (2019). Image interpolation using convolutional neural networks with deep recursive residual learning. *Multimedia Tools and Applications, 78*, 22813–22831. https://doi.org/10.1007/s11042-019-7633-1

Ledig, C., Theis, L., Huszar, F., Caballero, J., Cunningham, A., Acosta, A. (2017). Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. *CVPR*, 105-114. https://doi.org/10.1109/CVPR.2017.19

Li, G., Shi, J., Peng, J., & Zhao, G. (2019). Micro-expression Recognition Under Low-resolution Cases. *14th International Joint Conference on Computer Vision Imaging and Computer Graphics Theory and ICPRAM.* https://doi.org/10.5220/0007373604270434

Oh, Y., See, J., Le Ngo, A. C., Phan, N. W., & Baskaran, V. M. (2018). A survey of Automatic Facial Micro-Expression Analysis. *Databases, Methods, and Challenges Frontiers Media*, *9.* SA. Retrieved from https://doi.org/10.3389/fpsyg.2018.01128

Sharma, P., Coleman, S., Yogarajah, P., Taggart, L., & Samarasinghe, P. (2021). Magnifying Spontaneous Facial Micro Expressions for Improved Recognition. *ICPR.* https://doi.org/10.1109/ICPR48806.2021.9412585

Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C. (2019). Enhanced Super-Resolution Generative Adversarial Networks. *Computer Vision - ECCV, 11133*, 63-79. Retrieved from https://doi.org/10.1007/978-3-030-11021-5_5

Xiaobai Li, Pfister, T., Xiaohua Huang, Guoying Zhao, & Pietikainen, M. (2013). Inducement, collection and baseline. *A Spontaneous Micro-expression Database,*, 1-6. https://doi.org/10.1109/FG.2013.6553717

Yan, W., Li, X., Wang, S., Zhao, G., Liu, Y., Chen, Y. (2014). An Improved Spontaneous Micro-Expression Database and the Baseline Evaluation. *CASME II*. Retrieved from https://doi.org/10.1371/journal.pone.0086041

Zhang, Y., Tian, Y., Kong, Y., Zhong, B., & Fu, Y. (2018). Residual Dense Network for Image SuperResolution. *CVPR*, 2472-2481. https://doi.org/10.1109/CVPR.2018.00262

Zhao, G., & Li, X. (2019). Automatic Micro-Expression Analysis. *Open Challenges Frontiers Media*, *10.* SA. Retrieved from https://doi.org/10.3389/fpsyg.2019.01833