Advances in Computational Intelligence (2022) 2:24
https://doi.org/10.1007/s43674-022-00035-x

ORIGINAL ARTICLE

Comparative analysis of super-resolution reconstructed images for micro-expression recognition

Pratikshya Sharma[1] · Sonya Coleman[1] · Pratheepan Yogarajah[1] · Laurence Taggart[2] · Pradeepa Samarasinghe[3]

Received: 8 December 2021 / Revised: 18 March 2022 / Accepted: 7 April 2022 / Published online: 14 May 2022

Abstract

It is an established fact that the genuineness of facial micro-expression is an effective means for estimating concealed emotions (Li et al. in Micro-expression recognition under low-resolution cases. SciTePress, Science and Technology Publications, Setúbal, 2019). Conventionally, analysis of these expressions has been performed using high resolution images which are ideal cases. However, in a real-world scenario, capturing expressions with high resolution images may not always be possible particularly using low-cost surveillance cameras. Faces captured using such cameras are often very tiny and of poor resolution. Due to the loss of discriminative features these images may not be of much use particularly for identifying certain minute facial details. To make these images useful, enhancing the textural information becomes essential and super-resolution algorithms can be ideal to achieve this. In this work, we utilize algorithms based on deep learning and generative adversarial network for transforming low-resolution micro-expression images into super-resolution images and examine their fitness particularly for micro-expression recognition. The proposed approach is tested on simulated dataset obtained from two popular spontaneous micro-expression datasets namely CASME II and SMIC-VIS; the experimental results demonstrate that the method achieved favourable results with the best recognition performance recorded as 61.63%. The significance of this work is: first, it thoroughly investigates reconstruction performance of several deep learning super-resolution algorithms on simulated low-quality micro-expression images; second, it provides a comprehensive analysis of the results obtained employing these reconstructed images to determine their contribution in addressing image quality issues specifically for micro-expression recognition.

Keywords Micro-expression · Image super-resolution · Low resolution · Deep learning · Generative adversarial network · Micro-expression reconstruction

1 Introduction

Several forms of expressions exist through which humans generally interact and convey their emotions. Among verbal and non-verbal expressions, evaluating non-verbal expressions is significantly more challenging. Research on non-verbal expressions involving face and facial entities is an ongoing process for several decades. According to (Rinn 1984), facial expressions are a constant negotiation between two neurological pathways, pyramidal and extrapyramidal tract, which are sourced from two different section of brain. Facial movements that are voluntary in nature are caused by pyramidal tract whereas involuntary ones are a result of extrapyramidal tract (Rinn 1984). In a high-stake situation when a person tries to control expression, both these tracts get activated which creates a neural conflict leading to a quick leakage of expressions, called micro-expressions

✉ Pratikshya Sharma
sharma-p3@ulster.ac.uk

Sonya Coleman
sa.coleman@ulster.ac.uk

Pratheepan Yogarajah
p.yogarajah@ulster.ac.uk

Laurence Taggart
l.taggart@ulster.ac.uk

Pradeepa Samarasinghe
pradeepa.s@sliit.lk

[1] School of Computing, Engineering and Intelligent Systems, Ulster University, Magee, Northern Ireland, UK

[2] School of Nursing, Ulster University, Magee, Northern Ireland, UK

[3] Department of Information Technology, Sri Lanka Institute of Information Technology, Malabe, Sri Lanka

(Rinn 1984). Such facial expressions are short-lived (1/25 to 1/2 s) and not easily picked by naked eyes (Liong et al. 2018). Due to extremely low muscle movement intensity, the visibility of lines, folds, and wrinkles on the face during muscle contraction phase become obscure but are effective in reflecting one's emotional state and intention. Established as an emotion indicator, examining such expressions in humans can play a crucial role for addressing various emotional concerns. Some common application areas include mental health, psychology, depression level estimation and deception detection (Ekman and Friesen 1969; Ekman 2009). For instance, analysing micro-expressions of students in a classroom can be beneficial for instructors to realize the effectiveness of their teaching methods (Pei and Shan 2019). Similarly, examining micro-expression of individuals in an airport can help identify any security issues (Li et al. 2018). Its application can also be extended to the medical field where examining micro-expressions in patients can be used for identifying pain or mental health issues (Lucey et al. 2011). Evidently, we notice that these expressions play a vital role in our day-to-day life, thereby attracting researchers from diverse fields including computer vision, fuelling development of efficient automated micro-expression recognition (MER) system.

The contemporary solutions that exist for solving MER problems have considered images taken from datasets that are produced in ideal conditions with good lighting, no interference of illumination variations, full frontal view with no obstructions and good resolutions ($\sim$ above 150 × 150) (Oh et al. 2018; Takalkar et al. 2018). However, facial images captured in real scenarios using commonly available surveillance cameras may fail to produce good resolutions images due to external factors like ill pose, meagre lighting conditions, non-uniform illumination, etc. Moreover, face captured using such cameras often appear alongside several other objects hence are likely to take up only a limited space in the entire image. Subsequently, such faces will be small sized and of low resolution, i.e., 50 × 50 or below (Li et al. 2019). A significant surge in the use of surveillance cameras, especially for monitoring public domains, has created a new challenge for recognizing micro-expression collected under shallow lighting conditions. For these cameras, more emphasis is laid on capturing reliable recordings for longer period which is generally achieved by making a significant compromise on image/video resolution, thereby raising the need for algorithms that can deal with such resolution concerns (Yue et al. 2016). Super resolution (SR) is one such medium capable of addressing resolution challenges that are often engrained in images acquired using ordinary imaging devices. Reconstructed images obtained using SR algorithms are expected to have improved pixel density subsequently offering more image details. Achieving good resolution using superior hardware is not always cost effective and, therefore, employing image processing algorithm seem more feasible (Yue et al. 2016). The absence of discriminative facial details in micro-expression (ME) along with faint muscle movement intensity prevailing for extremely short span is already a challenge for recognition techniques. Therefore, resolution of such ME images can be a pivotal factor during recognition process. Extracting informative attributes from such low-resolution (LR) micro-expression images becomes effortful due to further loss in the availability of salient information which may have unfavourable influence on the performance of overall MER systems. In general, exploring LR micro-expression can be extremely beneficial particularly for crowd scenarios and poorly illuminated areas.

In computer vision the need for good resolution images is vital for algorithms to achieve reliable and superior performance. Therefore, several experiments transforming images from LR into SR using various deep learning (DL) and generative adversarial network (GAN) algorithms have already been conducted with promising outcomes for macro expressions (Li and Deng 2020). However, at present there exists only one work that addresses LR facial micro-expression recognition (Li et al. 2019). In their work, low-resolution ME was super resolved using a face hallucination method by blending patch-based and pixel-based regularization but does not explore deep learning methods. Building on this concept in our previous work (Sharma et al. 2022) we introduced several DL and GAN-based SR approach to deal with LR images containing micro-expression and utilize them for recognition process. Chosen SR techniques were tested on LR images simulated from SMIC-HS database (Li et al. 2013).

Extending this concept, we further test these SR techniques on LR images simulated from Chinese academy of sciences micro-expression, CASMEII (Yan et al. 2014) and spontaneous micro-expression database, SMIC-VIS (Li et al. 2013) database in this work. Additionally, here we also employ bicubic interpolation SR technique for a fair and comprehensive comparison of these SR methods and investigate their usefulness for micro-expression. Therefore, the key focus of this research is to test various SR techniques that commit performance boost and review their performance for LR micro-expression images. Keeping the previous pipeline intact, features are extracted using two spatiotemporal methods, namely local binary pattern on three orthogonal planes (LBP-TOP) (Zhao and Pietikainen 2007) and local phase quantization on three orthogonal planes (LPQ-TOP) (Ojansivu and Heikkilä 2008; Päivärinta et al. 2011). Corresponding sets of features obtained using these methods are then individually fed to the support vector machine (SVM) (Chang and Lin 2011) in various sets of experiment.

The main contributions of this work are: (1) extend the use of deep learning and GAN super-resolution models and test them on simulated LR micro-expression images; (2) present exhaustive performance analysis of various SR algorithms for micro-expression image reconstruction; (3) present a comparative analysis of performance of the overall approach on three simulated ME databases.

The remaining content of this paper is structured as follows. In Sect. 2 we present a brief review on advancements of methods for micro-expression analysis. Section 3 provides a detailed description of our proposed pipeline. Experiments conducted and results obtained using the proposed approach on all databases employed in our work are discussed in Sect. 4, followed by concluding remarks presented in Sect. 5.

2 Related works

We have surveyed some of the techniques that have evolved for MER in recent years and study (Oh et al. 2018; Takalkar et al. 2018) shows two notable pre-processing methods that lead to significant boost in recognition accuracy for micro-expressions are video magnification and interpolation (Wang et al. 2017; Li et al. 2018; Peng et al. 2019; Sharma et al. 2021). In one of the early works using magnification process, researchers have achieved recognition accuracy as high as 75.3% for micro-expressions (Wang et al. 2017). Likewise, temporal interpolation model (TIM) is very effective in achieving unified frame length by up-sampling videos containing few frames (Li et al. 2018; Peng et al. 2019). For micro-feature extraction the most widely used technique is LBP-TOP and was employed for baseline evaluation of MER on CASME II dataset (Yan et al. 2014). Another histogram-based method LPQ-TOP has also been explored for micro-expression by (Zong et al. 2019; Sharma et al. 2019, 2021). Apart from these, two gradient-based methods popular for solving MER problems include histogram of oriented gradient on three orthogonal planes (HOG-TOP) and histogram of image gradient orientation on three orthogonal planes (HIGO-TOP). Both these methods have been extensively examined by (Li et al. 2018) along with video magnification. The HIGO method augmented with magnification was able to achieve a remarkable recognition rate of 78.14% on CASME II dataset. Optical flow-based approach named bi-weighted oriented optical flow (Bi-WOOF) was proposed in (Liong et al. 2018) to extract optical features of micro-expressions from a single apex frame, which is believed to possess the most discriminative features in comparison to other frames. The method was tested on the CASME II and SMIC datasets and produced results comparable with other methods with accuracy of 61% and 62%, respectively. A region of interest-based main direction mean optical flow (MDMO) method was proposed in (Liu et al. 2016) for MER. The method was immune from the influence of translation, rotation, and illumination variance. Experimental results demonstrated that this method performed better than baseline LBP-TOP.

Moving beyond these techniques, attempts have also been made to apply deep learning-based approaches into MER framework. One of the earliest works that attempted to utilize the deep learning (DL) concept for MER was by (Patel et al. 2016). To apply transfer learning, a convolutional neural network (CNN) model was trained on ImageNet facial expression dataset, then transferred the appropriate features for further processing. CNN was explored further in (Takalkar and Xu 2017) along with data augmentation to generate exhaustive data for training purposes. Authors from (Gan et al. 2019) also utilized CNN with optical flow method which employed an apex frame and an onset frame to derive optical flow features which were then sent to a CNN model. To deal with the low volume of data in ME databases, (Liong et al. 2020) employed GAN to generate fake micro-expression images. The results obtained implied that GAN technique employed in their framework was able to realize optical flow changes on both vertical and horizontal directions. The method was able to obtain recognition performance comparable with existing state-of-the-art techniques. These works clearly suggest utilization of DL approaches for micro-expression-based experiments, with plenty of scope for further exploration.

Studies indicate micro-expression analysis have achieved immense success in recent years (Oh et al. 2018; Takalkar et al. 2018). However, they have attempted to recognize micro-expressions from datasets containing images with good resolution, approximately above 100 × 100. Therefore, they fail to take a real-world scenario into account where, poor lighting conditions can severely affect image resolutions and quality. The very first work that studied the effect of resolution for micro-expression was by Merghani et al. (2018). The original images taken from CASMEII database were downscaled to 75%, 50% and 25%. Performance of three feature extraction techniques were tested on these downscaled images. At the lowest downscale level 3DHOG technique performed the best whereas at high resolution (HR), the LBP-TOP method seemed to perform much better. At half the resolution histogram of optical flow orientation (HOOF) method gave the best performance in comparison to other two techniques. The work successfully realized the effects of resolution for micro-expression but did not consider the influence of image quality for such expressions. To have more relevance with real-life applications, (Li et al. 2019) proposed using deteriorated ME images that was both blurred and down sampled. Three levels for LR were considered, i.e., 16 × 16, 32 × 32 and 64 × 64. These LR micro-expression images were then super-resolved using patch-based and pixel-based face hallucination techniques and was the first work to perform MER using deteriorated image quality. The LR micro-expression recognition was performed for three spontaneous micro-expression databases SMIC (HS and subHS) and CASME II. Fast LBP-TOP was used for extracting the features and SVM was employed for classifying them. The results indicated that employing significantly LR images at 16 × 16 level makes it extremely difficult to achieve decent recognition results. Their approach worked comparatively better on SMIC-subHS with less misclassification reported than other two databases. Employing SMIC-HS database images at 16 × 16, a drastic improvement on the recognition results was reported particularly for positive label. In contrast, substantially higher misclassification results were reported for CASMEII database. Another observation made for this database was that most of its data were misclassified into "others" category. When recognition accuracy obtained for SR images were compared with their corresponding LR images a significant improvement was noticed for all the three databases at all chosen resolution. By analysing the overall reconstruction performance, observed through structural similarity index (SSIM), it was clear that the method produced best reconstruction results for SMIC-HS database followed by CASMEII and SMIC-subHS at 64 × 64 level. Same trend was observed for other two levels also. However, observing the peak signal to noise ratio (PSNR) suggested that reconstruction performance on SMIC-subHS was better than on CASMEII database at both 64 × 64 and 32 × 32 level. Though reconstruction values obtained for SMIC-subHS database was slightly less compared to other databases, yet it successfully produced best recognition results recorded at 74.65% which is much higher than that obtained for SMIC-HS and CASMEII database at 52.44% and 48.18%, respectively. Lower volume of data samples, slightly balanced along with fewer class categories in SMIC-subHS might have worked in its favour thereby producing better recognition results in comparison.

3 The overall proposed pipeline

To build a system capable of recognizing micro-expressions captured under low or poor lighting conditions we have embedded DL and GAN-based image SR module along with other components as depicted in Fig. 1. The entire pipeline can be divided into various blocks namely, image degradation, micro-expression reconstruction, micro-facial feature extraction and feature classification. An overview of each of these modules are presented in this section.

3.1 Image degradation

The principal factor that determines the quality of any image is spatial resolution and is represented by the total pixel count per unit area in a given image. Our pipeline is designed to address quality issues in micro-expression images but at present there is an absence of LR micro-expression database. Therefore, by applying degradation on existing databases, i.e., CASMEII and SMIC-VIS we simulate the required LR databases. To achieve this, we introduce noise and reduce image size by applying downscaling and Gaussian blurring on the HR images contained in these databases. The degradation applied is expressed as (Li et al. 2019):

$$X = DBY + x. \tag{1}$$

Here, Y and X represent HR input image and its corresponding LR image. Blurring is denoted by B, while D denotes down sampling and x is noise from external factors. By applying this model, the images obtained in the simulated database are of reduced quality. This reduction in image quality affects the discriminative attributes of micro-expression due to loss of image details. Figure 2a, c, e represent HR images taken from CASME II, SMIC-HS and SMIC-VIS databases, respectively. Corresponding LR images obtained by applying degradation on these HR images are depicted in Fig. 2b, d, f, respectively. Subtle expressions illustrated in Fig. 2a, c, e are more obvious compared with the expressions on the degraded images presented in Fig. 2b, d, f. Evidently, we can notice loss of image details in the low-quality images generated by the degradation model. This degradation is applied to all the image sequences for CASME II, and SMIC-VIS to generate a new set of databases with LR micro-expression images. These low-quality images are then suitable for use with SR algorithms.

3.2 Micro-expression image reconstruction

Deep learning and GAN-based approaches have been selected specifically due to their proven efficiency in delivering high end results for super-resolution tasks (Zhang et al. 2018; Wang et al. 2019; Rakotonirina and Rasoanaivo 2020). Taking inspiration from these works, we explored five different DL or GAN-based SR algorithms in our previous work (Sharma et al. 2022) to recover micro-expression image details lost due to poor image quality and reduced size. These methods have already demonstrated their effectiveness for SMIC-HS database, here we are extending its application and test for two other micro-expression databases. The three GAN-based SR approach are: (i) enhanced super-resolution generative adversarial network (ESRGAN) (Wang et al. 2019), (ii) further improving enhanced super-resolution generative adversarial network (nESRGAN+) (Rakotonirina et al. 2020), (iii) artefact cancelling generative adversarial network model (noise-cancel) (Ledig et al., 2017; Zhang et al. 2018). Further, two DL-based peak signal to noise ratio (PSNR) methods (Zhang et al. 2018) known as, psnr-large,

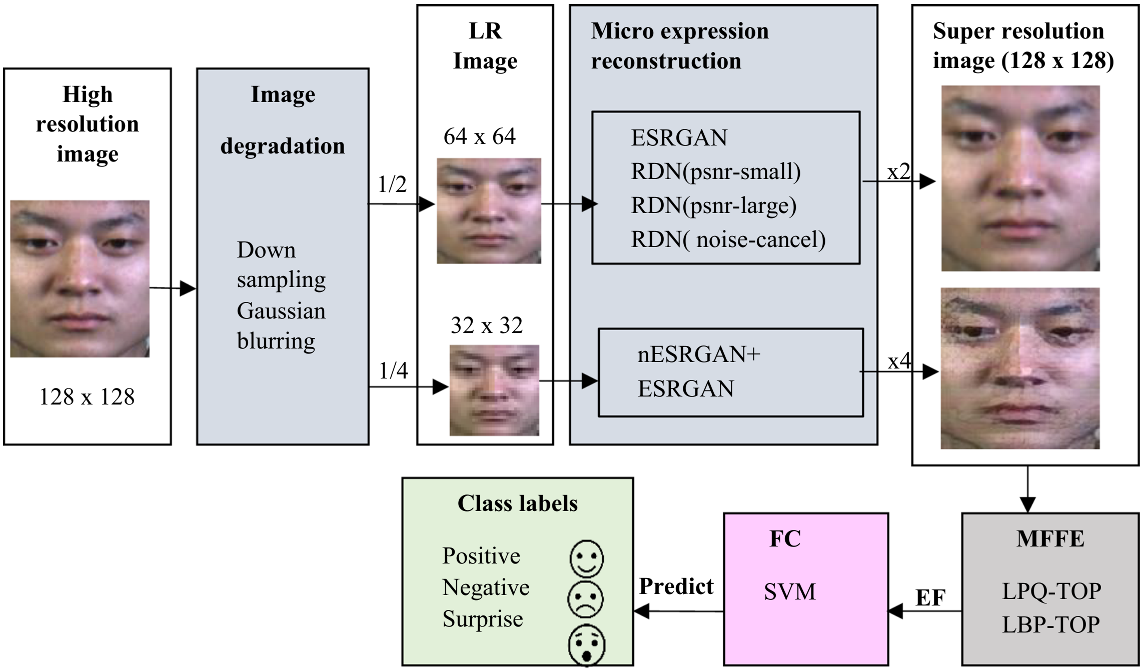


Fig. 1 Pipeline to reconstruct micro-expression images from low-quality data and its recognition process. *EF* extracted features, *ESRGAN* enhanced super-resolution generative adversarial network, *FC* feature classification, *LBP-TOP* local binary pattern on three orthogonal planes, *LPQ-TOP* local phase quantization on three orthogonal planes, *LR* low resolution, *MFFE* micro-facial feature extraction, *nESRGAN+* further improving enhanced super-resolution generative adversarial network, *RDN* residual dense network, *SVM* support vector machine

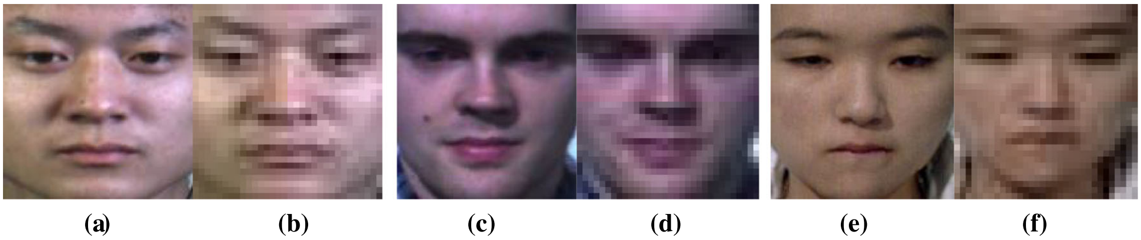


Fig. 2 Before and after applying degradation on CASME II (**a, b**), SMIC-HS (**c, d**) (Sharma et al. 2022) and SMIC-VIS (**e, f**) databases

and psnr-small are also used. In addition to these five techniques, we have also performed super-resolution experiments using bicubic interpolation (Russell 1995). The objective of applying these methods is to recover essential facial attributes when given a low-quality ME image input. We compare the quality of super-resolved images generated by these methods and provide useful inferences.

3.2.1 Super-resolution models based on peak signal to noise ratio

A generic architecture for generating super-resolution images is given in Fig. 3 with its components. To generate upscaled images using PSNR approach we have employed psnr-small and psnr-large models, differing in terms of training, similar to our previous work (Sharma et al. 2022). To train the network for each of these models, random batches of image patches are selected. Extending the previous work, we use small images patches with lower PSNR values for psnr-small model whereas larger image patches with bigger PSNR values were utilized for psnr-large model during training. Both these models use residual dense network (RDN) (Zhang et al. 2018) with architecture as depicted in Fig. 4. Four components of this architecture include shallow feature extraction, residual dense blocks, dense feature fusion followed by up-sampling network. The network utilizes Residual Dense Block (RDB) and exploits a contiguous memory (CM) mechanism as depicted in Fig. 5. Here, every layer of present RDB receives the state of its immediate predecessor RDB and uses hierarchical features to extract both local as well as global features (Zhang et al. 2018). Several RDB is used to build the RDN architecture as depicted in

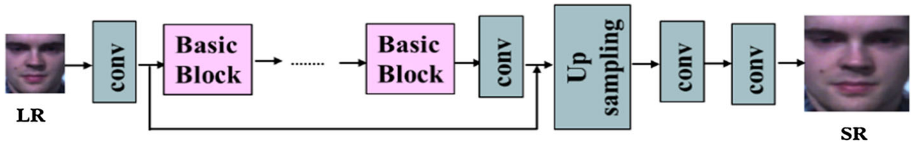


Fig. 3 Basic architecture of super-resolution network (Wang et al. 2019)

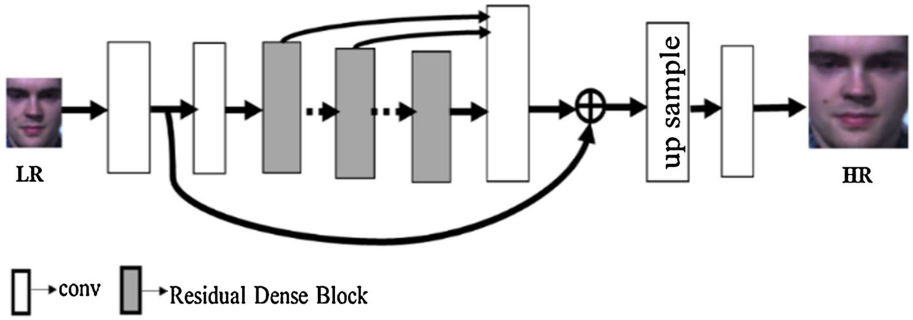


Fig. 4 Illustration of residual dense network (RDN) (Zhang et al. 2018)

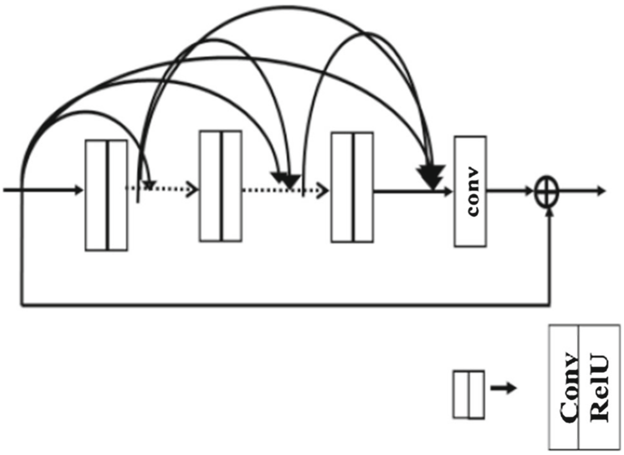


Fig. 5 Illustration of residual dense block (RDB) (Zhang et al. 2018)

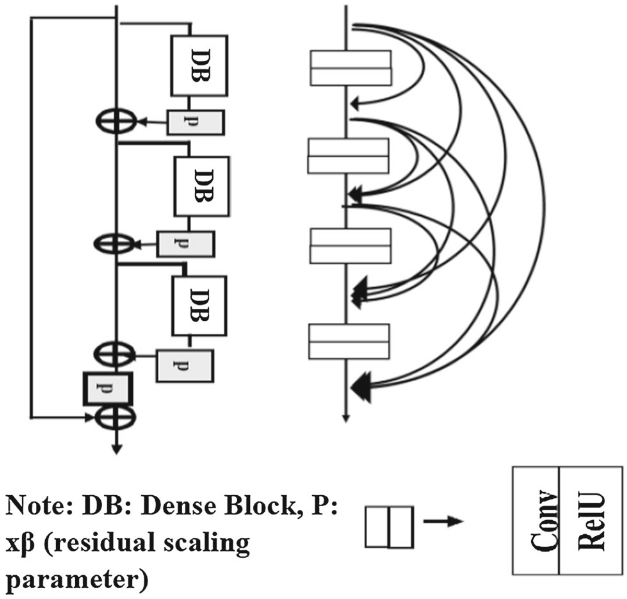


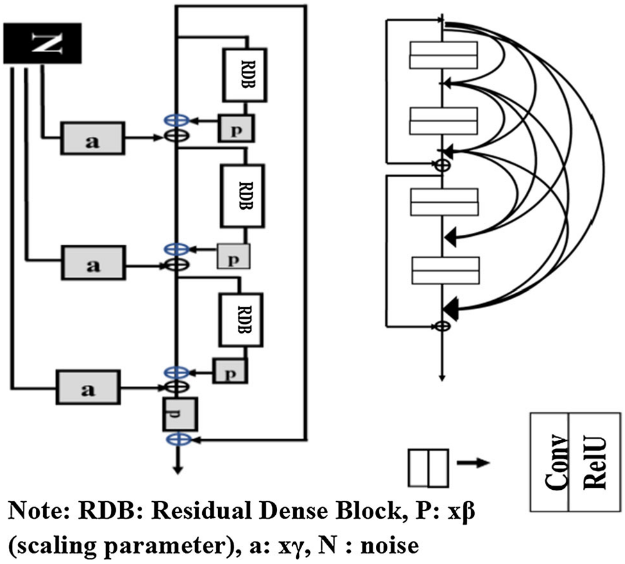
Fig. 6 Residual in residual dense block (RRDB) (Wang et al. 2019)

Fig. 4. Each RDB consists of several convolutional layers and every convolutional layer that exists within the RDB consists of feature maps. Feature maps also exist for every output of RDB as well as convolutions that reside out of RDB. Further details of architecture for both RDN and RDB can be found in (Zhang et al. 2018).

3.2.2 Generative adversarial network for image super-resolution

The basic architecture of general adversarial network (GAN) comprises of the generator, discriminator, and loss function which was initially developed by (Goodfellow et al. 2014). Broadening its application, it has been experimented for image super-resolution too (Wang et al. 2019). The basic architecture of GAN-based SR is depicted in Fig. 3 which consists of feed forward CNN generator network. The noise-cancel, ESRGAN and nESRGAN + models utilize this basic SR architecture with appropriate modifications. For instance, in noise-cancel the basic block is built using RDB (Fig. 5), whereas ESRGAN uses residual-in-residual dense block (RRDB) (Fig. 6) and nESRGAN + uses residual-in-residual dense residual block (RRDRB) (Fig. 7) with residual learning.

The noise-cancel model utilizes GAN architecture where the basic blocks use the same RDB design as given in Fig. 5, within the RDN architecture illustrated in Fig. 4. Therefore, the RDN and RDB architecture is same as that employed in the PSNR model, but both are trained differently. For training this noise-cancel model, distinct sets of training are undertaken employing multiple sets of data. The approach utilizes VGG feature loss as well as adversarial loss and attempts to eliminate the noise.



Note: RDB: Residual Dense Block, P: xβ (scaling parameter), a: xγ, N : noise

Fig. 7 nESREGAN + architecture employed for super-resolution (Rakotonirina et al. 2020)

The next two SR method employed in our work based on GAN is ESRGAN (Wang et al. 2019) and nESRGAN + (Rakotonirina and Rasoanaivo 2020). ESRGAN utilizes residual-in-residual dense network (RRDN) architecture built using residual-in-residual blocks (RRDB) as illustrated in Fig. 6. Here, several RRDBs are employed, where each of these RRDBs are built using several RDB. Further, each RDB consists of several convolutional layers and every convolutional layer that exists within the RDB consists of feature maps same as that discussed in Sect. 3.2.1 and utilizes residual scaling denoted by β. The nESRGAN + uses further denser network by employing dense blocks with RRDRB arrangement, i.e., an extra layer of residual learning is augmented compared to ESRGAN architecture. These residuals are added at an interval of every two layers. An overview of this further improved design is illustrated in Fig. 7. Moreover, Gaussian noise is also injected after each residual in this architecture. Details regarding architecture of ESRGAN can be found in (Wang et al. 2019) and for nESRGAN + in (Rakotonirina and Rasoanaivo 2020).

3.2.3 Bicubic interpolation

The bicubic interpolation method is employed in this work to generate super-resolution images. By applying a third order polynomial function this method ensures that within four corner points the required surface can be fitted. It utilizes the value of intensity at these four points in addition to the derivatives along three directions, i.e., diagonal, vertical and horizontal. The interpolated area is represented using Eq. (2) (Russell 1995):

$$f_i(x,\ y) = \sum_{j=0}^{3}\sum_{j=0}^{3} a_{ij}x^i y^j. \tag{2}$$

Here, $f_i(x,\ y)$ denotes the interpolated area for the point $(x,\ y)$ and a_{ij} denotes the coefficients. Sixteen coefficients are computed in total among them four are computed from the intensity values at four corners. From the diagonal derivates four other coefficients are computed. Lastly, from the horizontal and vertical directions utilizing their spatial derivative information eight coefficients are computed. These sixteen coefficients are also commonly known as neighbours.

3.3 Feature extraction and classification

Throughout the experiments performed in this work we have used two different feature extraction techniques namely LBP-TOP and LPQ-TOP at different instances. Successful use of both these extraction techniques for micro-expression has already been demonstrated by various research (Yan et al. 2014; Li et al. 2018; Sharma et al. 2019, 2021; Zong et al. 2019); hence we choose to employ these two techniques in our experiments. The LBP-TOP method extract features from image sequences along three planes that are orthogonal in nature commonly referred to as XY, XT and YT. It utilizes information from a set of neighbouring pixels to compute the desired binary pattern for each of these planes, then generates a histogram to represent the features. By including XT and YT planes it helps in representing temporal variations of its subjects. On the other hand, LPQ-TOP uses short term Fourier transform (STFT) to describe the texture arrangements in an image (Ojansivu et al. 2008; Päivärinta et al. 2011). Like previous extraction technique, this method also takes the three orthogonal planes into consideration and stacks them into a histogram. By incorporating the time domain in both the methods the dynamic variations within the image sequences are also captured, thus effectively picking up minute changes wherever present. Both these methods use histogram to describe the features extracted from the image sequences which are ultimately reduced to a single feature vector for each instance. These feature vectors are further given to classification technique to train and test their model, SVM is chosen in this work to achieve this. To keep the extraction and classification procedure simple we have chosen these respective methods. Using this classification technique, for SMIC databases the samples are to be classified as positive, negative and surprise while for CASMEII they are to be classified as happy, surprise, disgust, repression and others.

4 Experiments, results and analysis

The details of experiments performed along with parameters used is presented in this section. Results obtained by employing the proposed pipeline on CASMEII and SMIC-VIS is presented along with the results for SMIC-HS from our earlier work (Sharma et al. 2022). A comprehensive performance analysis of the reconstruction algorithms and overall recognition pipeline on all three databases is discussed here.

4.1 Database preparation

An overview of three databases used in our work is presented in Table 1. CASMEII database contains greater number of samples as well as class distribution. However, data distribution is not even, for instance expression classified with "other" label is 99 whereas for "surprise" it is only 25. By carefully observing the table we can see that both SMIC-HS and SMIC-VIS database have more uniform distribution of data than CASMEII, but SMIC-VIS contains comparatively lesser data. From the specifications presented in this table we can clearly see that facial resolution for all three databases vary. Therefore, to maintain uniformity across all datasets we set all HR images to 128 × 128 following the work by (Li et al. 2019). We also simulate new sets of databases consisting of LR images using image degradation model described in Sect. 3.1. These HR 128 × 128 image sequences for all three databases will be referred to as HR128 in our work. A down sampling factor of two and four is applied on these HR128 images to obtain down scaled sets of image sequences at 64 × 64 and 32 × 32 to be referred as LR64 and LR32. Instance of LR images obtained at these two resolutions for three databases is presented in Fig. 8. Super-resolution method is then applied to these LR images to recover facial details and obtain improved image sequences. To maintain uniform final resolution and have fair comparisons, all resultant images are set to a standard size of 128 × 128. SR resolution images obtained from LR64 will be referred to as SR64, whereas those obtained from LR32 will be referred to as SR32 in our work.

4.2 Super-resolution experiments

As mentioned earlier, we have experimented using five different SR models to reconstruct super-resolved images. By employing RDN architecture three sets of experiments were performed following (Cardinale and Tran 2018), where two approaches are based on PSNR and third is GAN based. In first approach, the RDN network was trained on large image patches with large PSNR value, to be addressed as psnr-large. Each residual block in the architecture consisted of six convolution layers. The network was built using a total of 20 residual blocks, each with 64 output filters and another 64 output filters inside the RDB. This model was trained to generate super-scaled images by applying scale factor two. Likewise, in second PSNR-based approach, the RDN network was built by training the models using smaller image patches having smaller PSNR values, to be addressed as psnr-small. The model was built using three convolution layers with ten residual blocks, the number of filters used, and scale factor was same as that of psnr-large approach. For both these models, image sequences of 64 × 64 were fed as input and the model returned reconstructed image upscaled by scale factor two to a final size of 128 × 128.

In the third approach, GAN with RDN architecture consisted of six convolution layer, twenty residual blocks, 64 RDB output filters, another 64 convolution output filters inside the RDB and scale factor of two. The model was built by training the network with both VGG feature loss as well as adversarial loss and is referred to as noise-cancel. In this model different datasets were employed to perform training at different sessions. ESRGAN is the fourth approach tested in our work built using ten RRDB, with three RDB in each of these RRDB. Further each of these RDB is built using four convolution layers and inside each RDB there are 32 convolution output filters. Additionally, the architecture is fitted with 32 output filters for every RDB. With learning rate at 0.004,100 decay frequency and decay factor at 0.5 the training parameters were set. The network was optimized using Adam optimizer and leaky version of rectified linear unit was used as activation function. The model built was capable of upscaling images and supported scale factors two and four. The implementation of four models mentioned above was adapted from (Cardinale and Tran 2018).

To train the fifth model, i.e., nESRGAN + , the loss function set at 0.005, decay factor at 0.01, learning rate set to 1×10^{-4} was considered. The Adam optimizer with parameters β_1 and β_2 set to 0.9 and 0.999, respectively, was also used. The model built was trained to upscale images by a scale factor set to 4. These parameter settings for this model have been adapted from Rakotonirina and Rasoanaivo (2020).

To summarize, two sets of LR64 databases, each simulated from CASME II, and SMIC-VIS were used to perform the SR experiments at scale factor two. For each simulated instance of LR64 databases, all four SR algorithms (i.e., psnr-small, psnr-large, noise-cancel and ESRGAN) were employed in four different sets of experiments to obtain four corresponding sets of super-resolved images.

Similarly, another two sets of LR32 databases simulated from CASMEII and SMIC-VIS were used to perform the SR experiments at scale factor 4. For this case ESRGAN and nESRGAN + models were employed to obtain corresponding SR images.

Images generated by each of these methods for both databases are presented in Figs. 9 and 10 along with results obtained on SMIC-HS database in our previous work

Table 1 Spontaneous micro-expression dataset used

Dataset	Subjects	Data count	Classes	Class label and distribution	Facial resolution	Speed (fps)
CASMEII	26	246	5	Happy—32, Surprise—25, Disgust—63, Repression—27, Others—99	280 × 340	200
SMIC-VIS	8	71	3	Positive—28, Negative—23, Surprise—20	130 × 160	25
SMIC-HS	16	164	3	Positive—51, Negative—70, Surprise—43	190 × 300	100

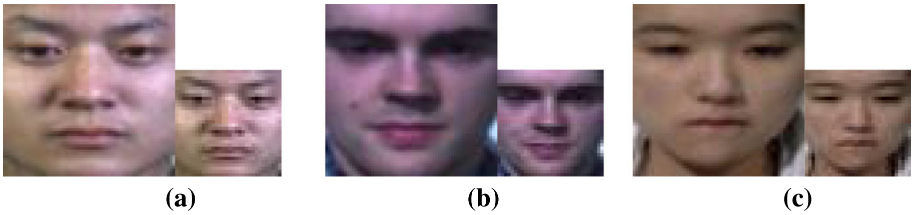


Fig. 8 Low resolution image at 64 × 64 (left) and 32 × 32 (right) obtained using image degradation model on HR128 image sequence of **a** CASMEII, **b** SMIC-HS (Sharma et al. 2022), and **c** SMIC-VIS

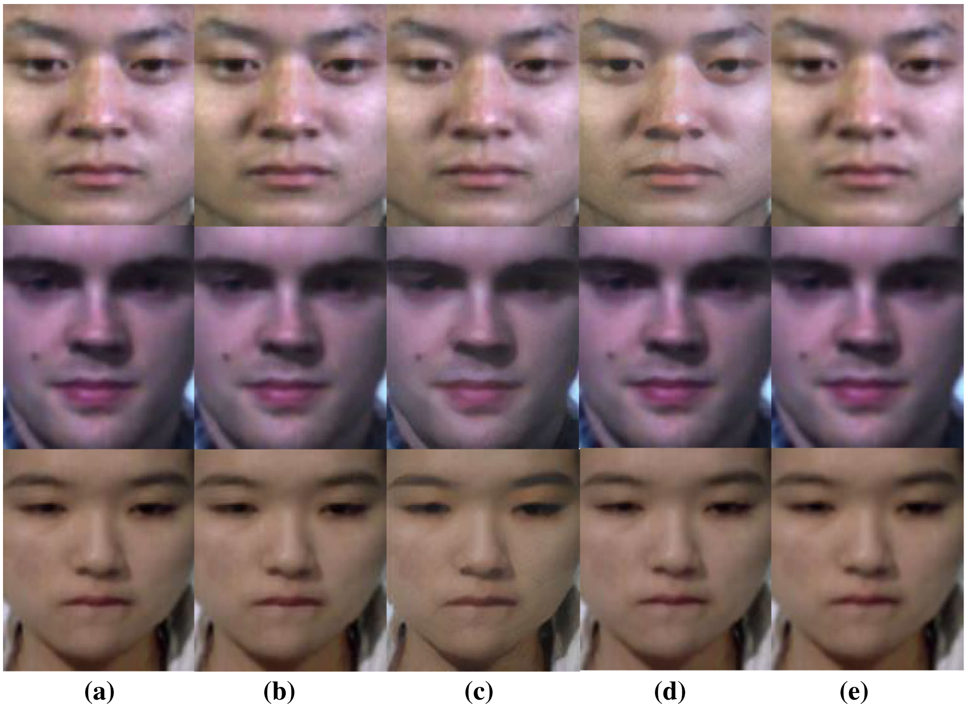


Fig. 9 Images reconstructed using super-resolution algorithms **a** psnr-small, **b** psnr-large, **c** noise-cancel, **d** ESRGAN and **e** bicubic interpolation with scale factor two for CAMSEII (top horizontal layer), SMIC-HS (middle horizontal layer) (Sharma et al. 2022), SMIC-VIS (bottom horizontal layer)

(Sharma et al. 2022). From these we can observe that images generated by SR algorithm with scale factor set to two is visually much clearer than those obtained at scale factor four. The visual perception while assessing SR and HR image quality through human eyes may not always seem consistent hence, image quality needs to be assessed using quality metrics for these reconstructed images.

4.3 Image quality assessment

Two widely used methods to assess SR image quality are structural similarity index measure (SSIM) and peak signal to noise ratio (PSNR) (Horé and Ziou 2010; Wang et al. 2019), so we choose these methods to assess quality of the super-resolved images obtained in the experiment. Values obtained for PSNR (measured in decibels, dB) and SSIM reflect the quality and rate of distortion of the reconstructed images compared with their corresponding HR128 images. In simple terms they estimate structural correlation between original and input image. SSIM is based on those structures that are typically visible in an image. The maximum value for SSIM is one which means closer the SSIM values are to one better is the reconstructed image quality. Higher the PSNR value better is the quality of reconstructed images. PSNR can be estimated by comparing the reconstructed image with an

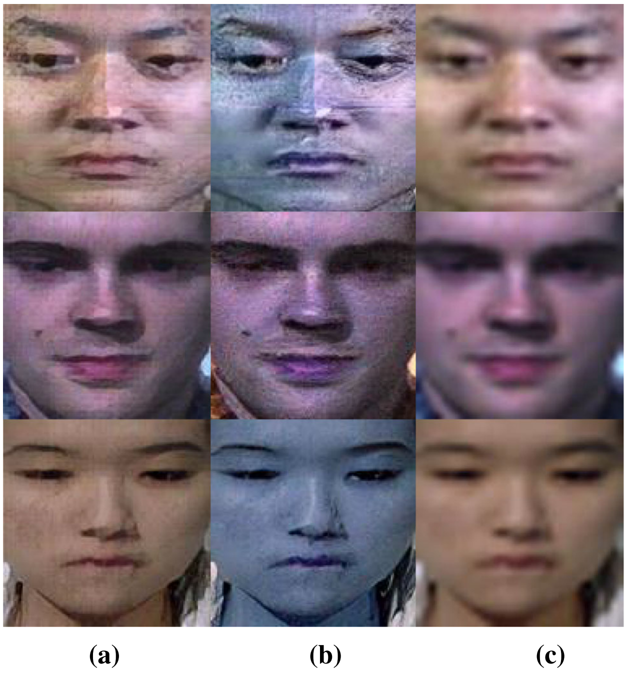


Fig. 10 Images reconstructed using super-resolution algorithms ESRGAN (**a**), nESRGAN + (**b**) and bicubic interpolation (**c**); with scale factor set to 4 for CASMEII (top horizontal layer), SMIC-HS (middle horizontal layer) (Sharma et al. 2022) and SMIC-VIS (bottom horizontal layer)

ideal image as follows (Horé and Ziou, 2010).

$$\mathrm{PSNR} = 10\log10\left(\frac{\max^2}{\mathrm{MSE}}\right) \tag{3}$$

$$\mathrm{MSE} = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}\left(I(i, j) - I'(i, j)\right)^2 \tag{4}$$

In Eq. (3), max refers to maximum possible pixel intensity for a given input image and MSE refers to mean squared error. In Eq. (4) the number of rows is given by m, the number of columns is given by n,I is the HR original image and I' is the degraded image, i is the row index and j is the column index. SSIM is estimated using the following (Horé and Ziou 2010).

$$SSIM(f, g) = l(f, g)c(f, g)s(f, g) \tag{5}$$

Here $l(f, g)$ estimates mean luminance closeness between two images f and g and is known as the luminance comparison function. Similarly, $c(f, g)$ estimates contrast closeness between two images and is known as the contrast comparison function. The correlation coefficient is estimated using the structure comparison function $s(f, g)$ between two images, i.e., f and g. For SSIM, positive values can range between 0 and 1, where 0 means no correlation between two images and value one means high correlation between two images. These two measures, i.e., PSNR and SSIM computed for various SR algorithms for all three databases will be discussed in next section.

4.4 Reconstruction result analysis

Here, we discuss the performance of all five SR algorithms based on the numerical values obtained by computing PSNR and SSIM metrics. Every set of super-resolved image sequence obtained were then compared with their corresponding HR images to obtain PSNR and SSIM values by utilizing Eqs. (3), (4) and (5). The average PSNR and SSIM values computed for all reconstructed SR instances at SR64 and SR32 for all three databases are listed in Tables 2 and 3.

Observing these image metrics, we can say that psnr-small model was able to generate higher quality super-resolved images across all databases at 64 × 64 image resolution. Specifically, the best reconstruction performance was obtained for SMIC-HS database with this model achieving PSNR value at 37.41 dB and SSIM value at 0.9827. On CASMEII database the performance of psnr-large model was very close to psnr-small model with PSNR/SSIM values behind by a very small value of 0.11 dB/0.001, respectively. For SMIC-HS database the images produced by psnr-small and ESRGAN model were structurally very close with a difference of 0.0001 but, psnr-small method produced 1.62 dB higher PSNR value than ESRGAN. However, observing PSNR value alone we can see that reconstruction performance of psnr-small and psnr-large on SMIC-HS database is almost equal with a difference of only 0.83 dB. Similar observation regarding PSNR metric can be made between psnr-large and ESRGAN model where later is lacking by a nominal value, i.e., 0.79 dB. Examining PSNR/SSIM values, a competitive performance between psnr-small, psnr-large and ESRGAN models can be observed on SMIC-VIS database, where psnr-small model is ahead by 0.55 dB/0.0029 and 1.31/0.0034, respectively. When observing SSIM value alone for this database, structural performance of psnr-large and ESRGAN is almost same with later lacking by a value as small as 0.0005.

At 64 × 64 for bicubic and other reconstruction methods, inspecting PSNR values we notice that its performance is exactly same as that of psnr-small method on CASMEII. All other reconstruction methods have produced image with lesser quality than bicubic method on this database. On SMIC-HS and SMIC-VIS database the reconstruction performance of both psnr-small and psnr-large are superior to bicubic method. Though reconstruction performance of ESRGAN is below bicubic but is still closer in comparison to noise-cancel method. Examining SSIM values in Table 3, we notice that reconstruction performance of all the methods is inferior to bicubic method on CASMEII.

Table 2 PSNR (dB)

SR method	CASME II	SMIC-HS (Sharma et al 2022)	SMIC-VIS	Resolution
psnr-small	**34.7**	**37.41**	**36.67**	64 × 64
psnr-large	34.59	**36.58**	**36.12**	64 × 64
Noise-cancel	31.06	30.38	31.15	64 × 64
ESRGAN	33.71	35.79	35.36	64 × 64
Bicubic	34.7	36.45	35.57	64 × 64
ESRGAN	27.65	29.5	28.95	32 × 32
nESRGAN +	14.83	23.08	15.73	32 × 32
Bicubic	30.83	32.3	31.64	32 × 32

Bold indicate better reconstruction values obtained in comparison to bicubic method

Table 3 SSIM

Resolution	SR method	CASME II	SMIC-HS (Sharma et al 2022)	SMIC-VIS
64 × 64	psnr-small	0.954	**0.9827**	**0.9701**
64 × 64	psnr-large	0.953	**0.9789**	**0.9672**
64 × 64	Noise-cancel	0.9261	0.9412	0.925
64 × 64	ESRGAN	0.9503	**0.9826**	**0.9667**
64 × 64	Bicubic	0.9555	0.9771	0.9616
32 × 32	ESRGAN	0.7811	0.8502	0.8189
32 × 32	nESRGAN +	0.6559	0.7601	0.7527
32 × 32	Bicubic	0.9032	0.9365	0.9111

Bold indicate better reconstruction values obtained in comparison to bicubic method

Best value obtained on this database is 0.954, which is still 0.001 below bicubic method. On SMIC-HS and SMIC-VIS database the reconstruction performance of three methods, i.e., psnr-small, psnr-large and ESRGAN are superior to bicubic method, thus these methods seem to perform well on both these databases. Clearly performance of the reconstruction methods on CASMEII is not satisfactory in comparison to other two databases.

Moving onto values obtained for 32 × 32 images we can clearly see that ESRGAN is able to outperform nESRGAN + model across all databases. The best reconstruction performance given by ESRGAN at this resolution is for SMIC-HS with 29.5 dB/0.8502 metric values. The reconstruction performance of nESRGAN + is far behind with its best performance metrics at 23.08 dB/0.7601 for the same database.

In overall, we can say that images reconstructed using psnr-small, psnr-large and ESRGAN model was almost similar where psnr-small model was ahead by a narrow margin. Also, all these three models produced comparatively superior results compared to noise-cancel model for 64 × 64 images. Similarly, at 32 × 32 ESRGAN produced far better results than nESRGAN + model but overall performance was still lower than those obtained at higher resolution of 64 × 64. While comparing results obtained using ESRGAN at both 64 × 64 and 32 × 32 levels, we can observe that the model is able to produce comparatively better result by employing higher resolution images. For instance, highest performance given by ESRGAN is 35.79 dB/0.9826 on SMIC-HS database for 64 × 64 images, however, using the same model a dip in performance is noticed when lower resolution images (i.e., 32 × 32) are employed with metrics value 29.5 dB/0.8502. This clearly strengthens the common belief that the resolution employed at input directly affects the reconstruction performance of SR algorithms and same can be observed for micro-expression images as well. Comparing reconstruction performance using PSNR/SSIM values of all these methods with bicubic we can notice that at this level their performance is less superior than bicubic technique. Thus, the bicubic method seems to perform far too well on all the databases at this level.

4.5 Recognition result analysis before super-resolution

Analysis made in this section is based on the recognition performance obtained before introducing SR algorithms into the pipeline as depicted in Table 4. On the standard HR128 images of SMIC-VIS database employing LPQ-TOP approach, 8% higher recognition accuracy was achieved

Table 4 Accuracy obtained with various instances of super-resolution algorithms

Resolution	SR method (our approach)	Accuracy %							
		CASME II			SMIC-HS			SMIC-VIS	
		LBP-TOP (ours)	LPQ-TOP (ours)	Fast LBP-TOP (Li et al. 2019)*	LBP-TOP (ours) (Sharma et al 2022)	LPQ-TOP (ours) (Sharma et al, 2022)	Fast LBP-TOP (Li et al. 2019)*	LBP-TOP (ours)	LPQ-TOP (ours)
HR128	–	**48.16**	47.17	48.18	50.06	**52.43**	50.00	53.26	**61.26**
SR64	psnr-small	**47.74**	**46.37**	48.18	**51.45**	**52.43**	52.44	**59.62**	**61.63**
	psnr-large	47.34	**46.01**		**50.67**	52.00		**56.67**	**61.40**
	Noise-cancel	46.50	43.54		49.39	51.82		**55.57**	60.33
	ESRGAN	**47.93**	**45.96**		**51.43**	**52.43**		**56.60**	61.01
	Bicubic	47.74	45.56		49.87	52.43		55.23	61.40
SR32	ESRGAN	43.05	41.93	44.53	**49.82**	50.60	51.83	51.69	59.15
	nESRGAN +	40.04	34.67		49.24	50.00		49.40	56.73
	Bicubic	44.35	42.75		49.35	51.02		52.11	60.03
LR64	–	43.05	43.14	44.94	49.2	49.39	50.00	49.40	**57.26**
LR32	–	43.00	41.05	44.13	44.25	**48.17**	46.95	**45.54**	41.18

Bold indicates best values obtained in our work

*Super-resolution method used is patch-based and pixel-based regularization which is different from our deep learning-based approach

compared to LBP-TOP method. Likewise, for LR64 images the LPQ-TOP was higher by 7.86% whereas for LR32 the LBP-TOP seemed to work well by 4.36%. Observing results for CASMEII database we notice that for HR128 images by employing LBP-TOP method recognition performance was slightly better by 0.99% than LPQ-TOP. Meanwhile for LR64 images recognition performance obtained using LBP-TOP was leading by a small margin of 0.09% whereas for LR32 the lead was by 1.95%. For SMIC-HS database with HR128 images phase method was better by 2.37% than the binary method. Similar trend was seen with LR64 at 0.19% and 3.92% higher than binary method.

Comparing the recognition performance for all the three databases we notice that by employing LPQ-TOP on SMIC-VIS and SMIC-HS database produced the best recognition accuracy for HR128 and LR64; but for CASMEII database employing LBP-TOP method produced better recognition performance. This recognition performance comparison at various resolution levels employing both feature extraction techniques on all three databases is graphically demonstrated in Fig. 11. With a gradual decrease of resolution, dip in recognition performance can be clearly noticed across all databases. Thus, results obtained from this section shows that resolution and quality of image can influence the overall recognition process for micro-expression.

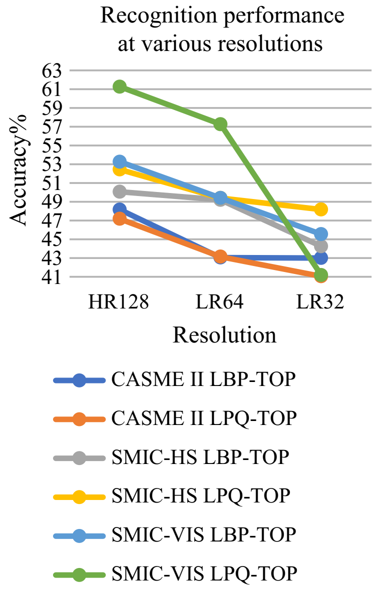


Fig. 11 Recognition performance analysis on three databases at different resolutions before introducing super-resolution

4.6 Recognition result analysis employing super-resolution

In this section, we discuss the overall recognition performance of the proposed pipeline after introducing various

SR algorithms on all three databases, recorded in Table 4. The discussion presented is in reference to the two databases employed in this work along with their performance comparison with SMIC-HS database.

4.6.1 Performance analysis on SMIC-VIS

Utilizing super-resolved images, the best recognition performance recorded for the SMIC-VIS database is 61.63%. This is an increase of 4.37% using the psnr-small method reconstructed images with LPQ-TOP method at scale factor 2 compared to its corresponding LR64. The next best recognition was obtained employing psnr-large with the same extraction method resulting in an increase of 4.14% compared with its corresponding LR64. This was followed by ESRGAN with an increase of 3.75%. The lowest performance was given by noise-cancel method with an increase of 3.07%.

Employing the LBP-TOP method, the best recognition performance was obtained at 59.62%. This is a boost of 10.22% obtained by employing psnr-small at scale factor 2 compared to its corresponding LR64. For the same extraction method when combined with the psnr-large method, the reconstructed images produced a recognition boost of 7.27% followed by the ESRGAN method with a boost of 7.2%. With a boost of 6.17%, noise-cancel produced the lowest improvement overall. Performance boost is obtained for all cases here but is still lower compared with that obtained employing LPQ-TOP method.

Reconstructing images with a scale factor of 4 with the ESRGAN method obtained a boost of 6.15% and 17.97% with LBP-TOP and LPQ-TOP, respectively, whereas with nESRGAN + the accuracy was increased by 3.86% and 15.55%, respectively.

Therefore, boost in recognition performance obtained after employing SR algorithms at both scale factors is undeniable for this database. This analysis of recognition performance using the SMIC-VIS database employing various SR and extraction methods is illustrated in Fig. 12.

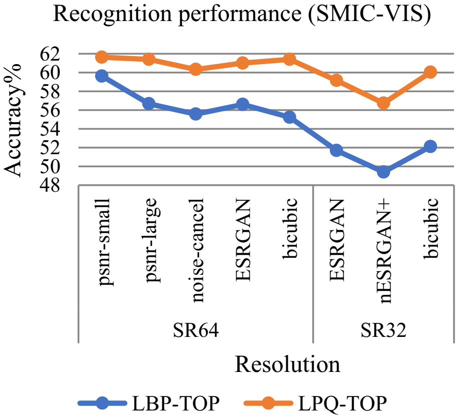


Fig. 12 Recognition performance analysis on SMIC-VIS database after introducing super-resolution

4.6.2 Performance analysis on CASMEII

Using the CASMEII database, the best recognition performance was obtained after introducing the SR model was 47.93% using the ESRGAN super-resolution algorithm along with the LBP-TOP extraction method for a scale factor of 2. This reflects an obvious boost in recognition performance of 4.88% compared with its corresponding LR64 after employing SR. The next best performance at the same scale factor was given when images reconstructed using psnr-small were used with the LBP-TOP method, a boost of 4.69%. This was followed by images reconstructed by psnr-large, with 4.29% boost in recognition performance. Once again, the lowest performance at this scale factor was obtained using the noise-cancel approach with a boost of 3.45%.

Employing the LPQ-TOP method for images reconstructed using psnr-small produced the best recognition performance of 46.37%, which is an increase of 3.23% over its corresponding LR64 and yet it is still 1.37% lower than the performance obtained using the LBP-TOP method. The boost in recognition obtained employing this extraction method with psnr-large is 2.83%, noise-cancel is 0.4% and ESRGAN is 2.82%. Although performance boost is achieved in all these cases compared with its corresponding LR64 images, the accuracies are lower than those obtained employing LBP-TOP approach. Therefore, for these cases recognition obtained by using LBP-TOP seemed to perform better than the phase method.

At scale factor four, images reconstructed with ESRGAN method employed with LBP-TOP obtained performance boost of 0.05% whereas with LPQ-TOP it was 0.88%. Though a higher boost is achieved using LPQ-TOP method, yet the overall recognition performance is still better with LBP-TOP method for this case.

The lowest performance was obtained for nESRGAN + reconstructed images with recognition values recorded below its corresponding LR32 images for both the extraction methods on this database. This analysis of performance on CASMEII database employing various SR and extraction methods is illustrated in Fig. 13.

4.6.3 Performance comparison across all databases

Observing the values presented in Table 4, we notice that super-resolution images at a scale factor of 2, when employed with LPQ-TOP, produced much better recognition performance than the binary method using SMIC-VIS and slightly better for SMIC-HS database. For instance, examining the best performance on all databases we see that

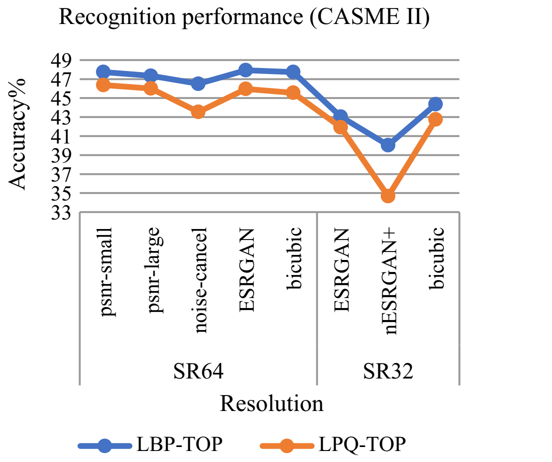


Fig. 13 Recognition performance analysis on CASME II database after introducing super-resolution

recognition is higher by 2.01% using SMIC-VIS and 0.98% using SMIC-HS employing LPQ-TOP, compared with the binary method. However, using the CASMEII database, the LBP-TOP method seems to perform better than the phase quantisation approach. For methods utilizing the RDN architecture, employing images reconstructed by the psnr-small method seems to consistently give the best recognition performance across all databases. The performance of ESRGAN employed images was also on par with images constructed using psnr-small with the SMIC-HS database. However, using the SMIC-VIS database, the psnr-large model constructed images performed slightly better than the ESRGAN model constructed images. The lowest performance was consistently obtained by employing images reconstructed by the noise-cancel approach across all three databases. Therefore, at this scale factor all three SR approaches, i.e., psnr-small, ESRGAN and psnr-large, seem to be very competitive and performed consistently better than noise-cancel method.

Considering a higher scale factor of 4, images reconstructed using ESRGAN seem to consistently outperform the improvised variant across all databases. Once again both methods performed better on SMIC-VIS and SMIC-HS. The lowest performance was obtained using the CASMEII database for both SR methods.

To have a fair comparison among results obtained using various SR methods in our work we compare them with bicubic interpolation results. For SMIC-VIS database, SR images reconstructed at scale factor 2 by all SR methods using LBP-TOP seems to work fairly with recognition accuracies higher than those obtained using the bicubic method. Employing LPQ-TOP on the SMIC-VIS database, the psnr-small and psnr-large methods performed better than bicubic, whereas the noise-cancel and ESRGAN performances were lower than the bicubic method. At scale factor four recognition performance using images reconstructed by the bicubic method was better which is consistent with the image quality metrics obtained for this instance.

For the SMIC-HS database, both the psnr-small and ESRGAN methods were able to produce results on par with the bicubic method at scale factor two when combined with the LPQ-TOP method. Likewise, when combined with the LBP-TOP method the images reconstructed using all SR methods performed better than the bicubic reconstructed images with the exception of noise-cancel which performed lower than bicubic. At scale factor four recognition performances of both SR methods were lower than bicubic method when combined with LPQ-TOP.

Using the CASMEII database at scale factor two, psnr-small and ESRGAN, when used with the LBP-TOP method produced recognition performances better or equal to that of the bicubic method, however, using the psnr-large and noise-cancel approaches, recognition was marginally lower than the bicubic method. Most of the SR methods when combined with the LPQ-TOP approach for this database performed equal or better than the bicubic method whereas with the LBP-TOP method only psnr-small and ESRGAN seemed to perform better than bicubic. At scale factor four the bicubic method was slightly better than ESRGAN-based approach but much better than nESRGAN + based approach.

To summarize, both psnr-small and ESRGAN methods when combined with LBP-TOP work best on CASMEII at a scale factor of 2, whereas at a scale factor of 4 ESRGAN worked best. Using the SMIC-HS database psnr-small and ESRGAN, when combined with LPQ-TOP, worked the best. Using the SMIC-VIS database psnr-small and psnr-large combined with LPQ-TOP worked will. Therefore, most of the SR methods were able to produce results better than the bicubic method at a scale factor of 2 across all databases, but at a scale factor of 4 this was only the case when using the SMIC-HS database and ESRGAN was combined with LBP-TOP. The overall comparison of recognition accuracy employing different super-resolution and feature extraction techniques on three micro-expression database is illustrated in Fig. 14.

5 Conclusions

In this paper, we give an exhaustive performance analysis of various deep learning-based SR for micro-expression recognition. By introducing such SR algorithms into the pipeline, MER have achieved favourable results. It compares the reconstruction performance of various SR models for ME images and also gives a comprehensive analysis on the basis of recognition performance obtained with various combinations of SR and feature extraction methods on three databases. Results illustrate recognition improvement obtained by utilizing these SR images in almost all

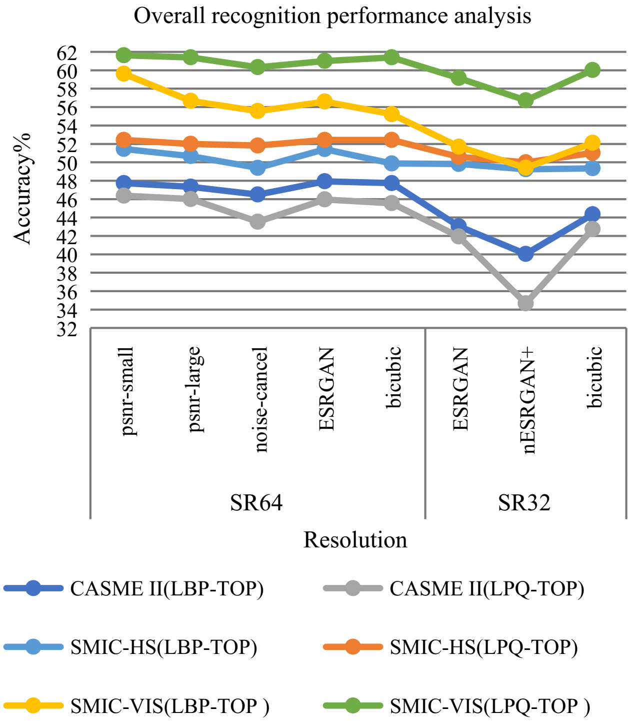


Fig. 14 Comparison of recognition accuracy employing various instances of super-resolution and feature extraction techniques on three micro-expression databases

the combinations of SR and extraction methods when compared with their corresponding LR images. Clearly, all the SR models employed have been able to successfully reconstruct the facial details, though the image quality obtained is varying. The aim of this study was to thoroughly examine if such SR algorithms can be a good fit for low-resolution ME images. Evidently, results obtained demonstrate that the proposed pipeline works comparatively well on SMIC-HS and SMIC-VIS than on CASMEII database. Examining the overall performance, these positive results are a good indicator to ascertain the effectives of the basic model and GAN-based SR technique for boosting facial micro-expression image details. Certainly, the classification accuracy was influenced by the size and quality of image reconstructed across all databases, and same is reflected in the results obtained. Higher volume of data and uneven class distribution present in CASMEII database may have been a disadvantage, thus producing inferior recognition performance. On the other hand, more uniform data distribution and lesser volume of data has worked in favour of both versions of SMIC database. Though acquiring adequate number of ME data remains a challenge even today yet we have managed to simulate required data to perform these sets of experiments. Two limitations of this work are, first we have tested only two resolutions for LR images, however, in future resolutions lower than these can also be assessed. Second, data imbalance has not been addressed as such the work can be substantiated by incorporating a suitable approach with a greater number of datasets in the future. The results achieved are promising and can be extended further by evaluating more SR algorithms with additional scale factors. This study serves as a channel providing researchers with an insight into the application of popular deep learning-based SR algorithms for poor quality ME images. It can also be used as a general guideline to widen the usage of suitable deep learning-based SR techniques for such specific applications. Acquiring good facial resolution with low-cost surveillance cameras may not always be realistic in day-to-day life especially when faces to be captured are distant from the camera, this directly affects the quality of facial details obtained. Therefore, to overcome resolution issues that may exist in ME obtained in unfavourable settings, utilizing such SR-based recognition approach seems a feasible option with abundant room for growth.

Funding This study was funded by Ulster University.

Data availability All the data and material are available upon request.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

Cardinale F, Tran D (2018) Image super resolution. https://github.com/idealo/image-super-resolution

Chang C, Lin C (2011) LIBSVM. ACM Trans Intell Syst Technol 2(3):1–27. https://doi.org/10.1145/1961189.1961199

Ekman P, Friesen WV (1969) Nonverbal leakage and clues to deception. Psychiatry 32(1):88–106. https://doi.org/10.1080/00332747.1969.11023575

Ekman P (2009) Telling lies: clues to deceit in the marketplace, politics, and marriage, rev edn. WW Norton & Company

Gan YS, Liong S, Yau W, Huang Y, Tan L (2019) OFF-ApexNet on micro-expression recognition system. Signal Process Image Commun 74:129–139. https://doi.org/10.1016/j.image.2019.02.005

Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y (2014) Generative adversarial nets. In: Ghahramani Z, Welling M, Cortes C, Lawrence ND,. Weinberger

KQ (eds) Advances in neural information processing systems, vol 27. Curran Associates, Inc., Boston, pp 2672–2680

Horé A, Ziou D (2010) Image quality metrics: PSNR vs. SSIM. In: Paper presented at the—2010 20th international conference on pattern recognition, pp 2366–2369. https://doi.org/10.1109/ICPR.2010.579

Ledig C, Theis L, Huszar F, Caballero J, Cunningham A, Acosta A et al (2017) Photo-realistic single image super-resolution using a generative adversarial network. In: Paper presented, pp 105–114. https://doi.org/10.1109/CVPR.2017.19

Li S, Deng W (2020) Deep facial expression recognition: a survey. IEEE Trans Affect Comput. https://doi.org/10.1109/TAFFC.2020.2981446

Li G, Shi J, Peng J, Zhao G (2019) Micro-expression recognition under low-resolution cases. SciTePress-Science and Technology Publications. https://doi.org/10.5220/0007373604270434

Li X, Pfister T, Huang X, Zhao G, Pietikäinen M (2013) A spontaneous micro-expression database: inducement, collection and baseline. In: Paper presented at the 2013 10th IEEE international conference and workshops on automatic face and gesture recognition (FG), pp 1–6. https://doi.org/10.1109/FG.2013.6553717

Li X, Hong X, Moilanen A, Huang X, Pfister T, Zhao G, Pietikäinen M (2018) Towards reading hidden emotions: a comparative study of spontaneous micro-expression spotting and recognition methods. In: IEEE transactions on affective computing, vol 9(4), pp 563–577. https://doi.org/10.1109/TAFFC.2017.2667642

Liong S-T, See J, Wong KS, Phan RCW (2018) Less is more: micro-expression recognition from video using apex frame. Signal Process Image Commun 62:82–92. https://doi.org/10.1016/j.image.2017.11.006

Liong ST, Gan YS, Zheng D, Li S, Xu H, Zhang H et al (2020) Evaluation of the spatio-temporal features and GAN for micro-expression recognition system. J Signal Process Syst 92(7):705–725. https://doi.org/10.1007/s11265-020-01523-4

Liu Y, Zhang J, Yan W, Wang S, Zhao G, Fu X (2016) A main directional mean optical flow feature for spontaneous micro-expression recognition. IEEE Trans Affect Comput 7:299–310

Lucey P, Cohn JF, Prkachin KM, Solomon PE, Matthews I (2011) Painful data: the UNBC-McMaster shoulder pain expression archive database. In: IEEE international conference on automatic face & gesture recognition (FG), pp 57–64. https://doi.org/10.1109/FG.2011.5771462

Merghani W, Davison AK, Yap MH (2018) A review on facial micro-expressions analysis: datasets, features and metrics. http://arxiv.org/abs/1805.02397

Oh YH, See J, Le Ngo AC, Phan RC, Baskaran VM (2018) A survey of automatic facial micro-expression analysis: databases, methods, and challenges. Front Psychol 9:1128. https://doi.org/10.3389/fpsyg.2018.01128

Ojansivu V, Heikkilä J (2008) Blur insensitive texture classification using local phase quantization. In: Elmoataz A, Lezoray O, Nouboud F, Mammass D (eds) Image and signal processing. ICISP 2008. Lecture notes in computer science, vol 5099. Springer, Berlin. https://doi.org/10.1007/978-3-540-69905-7_27

Päivärinta J, Rahtu E, Heikkilä J (2011) Volume local phase quantization for blur-insensitive dynamic texture classification. In Proceedings of the 17th Scandinavian conference on image analysis (SCIA'11). Springer, Berlin, pp 360–369

Patel D, Hong X, Zhao G (2016) Selective deep features for micro-expression recognition. In: Paper presented at the—2016 23rd international conference on pattern recognition (ICPR), pp 2258–2263. https://doi.org/10.1109/ICPR.2016.7899972

Pei JY, Shan P (2019) A micro-expression recognition algorithm for students in classroom learning based on convolutional neural network. In: Traitement du signal, vol 36(6), pp 557–563. https://doi.org/10.18280/ts.360611

Peng W, Hong X, Xu Y, Zhao G (2019) A boost in revealing subtle facial expressions: a consolidated Eulerian framework. In:14th IEEE international conference on automatic face & gesture recognition, Lille, pp 1–5. https://doi.org/10.1109/FG.2019.8756541

Rakotonirina NC, Rasoanaivo A (2020) ESRGAN+ : further improving enhanced super-resolution generative adversarial network. In: Paper presented at the—ICASSP 2020—2020 IEEE international conference on acoustics, speech and signal processing (ICASSP), pp 3637–3641. https://doi.org/10.1109/ICASSP40776.2020.9054071

Rinn WE (1984) The neuropsychology of facial expression: a review of the neurological and psychological mechanisms for producing facial expressions. Psychol Bull 95(1):52–77. https://doi.org/10.1037/0033-2909.95.1.52

Russell WS (1995) Polynomial interpolation schemes for internal derivative distributions on structured grids. Appl Numer Math 17(2):129–171. https://doi.org/10.1016/0168-9274(95)00014-L

Sharma P, Coleman S, Yogarajah P, Taggart L (2019) Dublin micro expression classification accuracy assessment micro expression classification accuracy. Assessment. https://doi.org/10.21427/kbny-0a41

Sharma P, Coleman S, Yogarajah P, Laurence T, Samarasinghe P (2021) Magnifying spontaneous facial micro expressions for improved recognition. In: 25th ICPR 2021

Sharma P, Coleman S, Yogarajah P, Laurence T, Samarasinghe P (2022) Evaluation of generative adversarial network generated super resolution images for micro expression recognition. In: Proceedings of the 11th international conference on pattern recognition applications and methods, ISBN 978-989-758-549-4, ISSN 2184-4313, pp 560–569

Takalkar M, Xu M, Wu Q et al (2018) A survey: facial micro-expression recognition. Multimed Tools Appl 77:19301–19325. https://doi.org/10.1007/s11042-017-5317-2

Takalkar MA, Xu M (2017) Image based facial micro-expression recognition using deep learning on small datasets. In: International conference on digital image computing: techniques and applications (DICTA), pp 1–7. https://doi.org/10.1109/DICTA.2017.8227443

Wang Y, See J, Oh Y, Phan-Rahulamathavan RC et al (2017) Effective recognition of facial micro-expressions with video motion magnification. Multimed Tools Appl 76(20):21665–21690. https://doi.org/10.1007/s11042-016-4079-6

Wang X, Yu K, Wu S, Gu J, Liu Y, Dong C et al (2019) ESRGAN: enhanced super-resolution generative adversarial networks. In: Proceedings of the European conference on computer vision (ECCV) workshops, 2018, pp 63–79

Yan WJ, Li X, Wang SJ, Zhao G, Liu YJ, Chen YH, Fu X (2014) CASME II: an improved spontaneous micro-expression database and the baseline evaluation. PLoS ONE 9(1):e86041. https://doi.org/10.1371/journal.pone.0086041

Yue L, Shen H, Li J, Yuan Q, Zhang H, Zhang L (2016) Image super-resolution: the techniques, applications, and future. Signal Process 128:389–408. https://doi.org/10.1016/j.sigpro.2016.05.002

Zhang Y, Tian Y, Kong Y, Zhong B, Fu YR (2018) Residual dense network for image super-resolution. In: IEEE/CVF conference on computer vision and pattern recognition, pp 2472–2481. http://arxiv.org/abs/1802.08797

Zhao G, Pietikainen M (2007) Dynamic texture recognition using local binary patterns with an application to facial expressions. IEEE Trans Pattern Anal Mach Intell. https://doi.org/10.1109/TPAMI.2007.1110

Zong Y, Zheng W, Hong X, Tang C, Cui Z, Zhao G (2019) Cross-database micro-expression recognition: a benchmark. In: Proceedings of the 2019 on international conference on multimedia. https://doi.org/10.1145/3323873.3326590