



# Does Deep Learning-Based Super-Resolution Help Humans With Face Recognition?

Erik Velan<sup>1</sup>, Marco Fontani<sup>1\*</sup>, Sergio Carrato<sup>2</sup> and Martino Jerian<sup>1</sup>

<sup>1</sup>Amped Software, Trieste, Italy, <sup>2</sup>Department of Engineering and Architecture, University of Trieste, Trieste, Italy

The last decade witnessed a renaissance of machine learning for image processing. Super-resolution (SR) is one of the areas where deep learning techniques have achieved impressive results, with a specific focus on the SR of facial images. Examining and comparing facial images is one of the critical activities in forensic video analysis; a compelling question is thus whether recent SR techniques could help face recognition (FR) made by a human operator, especially in the challenging scenario where very low resolution images are available, which is typical of surveillance recordings. This paper addresses such a question through a simple yet insightful experiment: we used two state-of-the-art deep learning-based SR algorithms to enhance some very low-resolution faces of 30 worldwide celebrities. We then asked a heterogeneous group of more than 130 individuals to recognize them and compared the recognition accuracy against the one achieved by presenting a simple bicubic-interpolated version of the same faces. Results are somehow surprising: despite an undisputed general superiority of SR-enhanced images in terms of visual appearance, SR techniques brought no considerable advantage in overall recognition accuracy.

**Keywords:** super-resolution, deep learning, face identification, face recognition, face enhancement, AI

## OPEN ACCESS

### Edited by:

Cecilia Pasquini,  
University of Trento, Italy

### Reviewed by:

Duc Tien Dang Nguyen,  
University of Bergen, Norway  
Federica Lago,  
University of Trento, Italy

### \*Correspondence:

Marco Fontani  
marco.fontani@  
ampedsoftware.com

### Specialty section:

This article was submitted to  
Image Processing,  
a section of the journal  
Frontiers in Signal Processing

**Received:** 14 January 2022

**Accepted:** 01 April 2022

**Published:** 20 April 2022

### Citation:

Velan E, Fontani M, Carrato S and  
Jerian M (2022) Does Deep Learning-  
Based Super-Resolution Help Humans  
With Face Recognition?  
Front. Sig. Proc. 2:854737.  
doi: 10.3389/frsip.2022.854737

## 1 INTRODUCTION

Since the beginning of the '90s, the impact of digital images on security and forensics has increased steadily. The rapid deployment of pervasive surveillance systems has made it possible to record billions of hours of video every day and make them accessible to law enforcement agencies (LEAs) when needed. One of the most interesting bindings between digital imaging and forensics is the task of identifying someone by looking at a picture of his or her face. This task is known as Face Recognition (FR) (Zhao et al., 2003), while the term Face Comparison typically refers to the case where a comparison has to be made between two given facial pictures; in the scope of this paper, we will use Face Recognition and Face Identification as synonyms. Notwithstanding the continuous improvements both from the hardware (camera sensors, optics, components) and software (video codecs) points of view, it is still very common that surveillance recordings are of limited quality. Indeed, one very common situation is that a law enforcement officer has to carry out FR but is not able to do so because the resolution is limited.

To combat the low-resolution problem, researchers proposed image super-resolution (SR) to increase the details of a low resolution (LR) image by using statistical approaches, optimization techniques, or machine learning. When multiple temporally close images of the same subject are available, it is possible to combine them, trading time resolution for spatial resolution (Nelson et al., 2012); however, it is very common that SR has to be carried out from a single picture (Li

et al., 2020). The question on whether SR techniques could help with FR naturally followed. Fookes et al. (2012) investigated the impact of image resolution and, most interestingly, single image super-resolution on FR performance. Their comprehensive study concluded that even simple bilinear interpolation could significantly improve the performance of automated FR algorithms, mainly because it facilitated the face alignment phase. They also showed that reconstruction-based face super-resolution algorithms improved FR performance by a moderate amount. Interestingly, however, they also found that the only “training-based” system they considered, based on face hallucination<sup>1</sup> (Baker and Kanade, 2002), always under-performed bilinear interpolation in terms of helping FR.

After some years of a plateau in the performance of face SR techniques, the astounding achievements brought by deep neural networks have opened new perspectives (Wang et al., 2020). In particular, several deep learning-based techniques emerged that are specifically tailored to the SR of facial images, as reported in **Section 2.1** of this paper.

However, the forensic community is still questioning the suitability of deep learning techniques for facial SR. There is little understanding of *how* a deep neural network creates the final image starting from the LR input and from “previously seen” training data, which of course are external to the specific investigation (Narwaria, 2022). Moreover, to the best of our knowledge, no study has been carried out yet to evaluate whether deep learning SR can help *humans* in the face recognition task. Indeed, all works mentioned have always used automated algorithms for the FR phase.

Ueda and Koyama (2010) evaluated the influence of makeup on FR carried out by human recognizers. The results, somehow surprisingly, showed that applying light makeup increased the recognition rate slightly, both for male and female models, compared to the no-makeup case. However, when heavy makeup was applied, performances dropped sensibly. Authors hypothesized that moderate makeup could enhance the distinctive properties of the faces, thus favoring human recognition; on the other hand, heavy makeup could hide and standardize the peculiar properties of faces thus making FR harder. This latter hypothesis appears to be confirmed by a recent study by Lago *et al.*, where authors investigated the ability of human observers in telling apart real facial images from faces created by generative adversarial networks (GANs). Results showed that people tend to classify more likely as “real” the GAN-generated images than the actual real images (Lago et al., 2021), when last generation GANs were used. This outcome may be explained according to Ueda’s hypothesis that GANs tend to create “average faces” without those distinctive and perhaps peculiar elements found in natural images (strange hear shape, different eye size, etc.). That could make an actual face look suspicious.

<sup>1</sup>In Baker and Kanade’s work (Baker and Kanade, 2002), dated 2002, the term “hallucination” did not refer to the use of deep neural networks, which were still impractical for image super-resolution at the time.

In this work, we aim at investigating whether modern, artificial intelligence (AI)-based face SR techniques can help a human observer in the task of FR, compared to more straightforward upscaling techniques such as bicubic interpolation. In particular, we focus on the case where only very low resolution images are available to start with, which is unfortunately a common use case in forensic applications. To this end, we devised a simple yet insightful experiment, detailed in **Section 2**: we created LR versions of the faces of 30 worldwide celebrities and upscaled those images using the standard bicubic interpolation plus two state-of-the-art, deep-learning-based SR methods. Then, we asked a heterogeneous set of 135 individuals to identify each face, showing some users the bicubic-interpolated version and the AI version to others. We then made a critical analysis of the results, reported in **Section 3**, which surprisingly revealed that despite images upscaled with AI “look” generally much better than interpolated ones, on average there is no real advantage in terms of FR performance, and in some cases interpolation performed better than AI methods, somehow confirming what was observed by Fookes *et al.* 20 years ago (Fookes et al., 2012).

## 2 MATERIALS AND METHODS

Our study consisted of the following steps: 1) analyzing the state-of-the-art of super-resolution techniques, with a special focus on face super-resolution; 2) selecting the most appropriate SR techniques to be used from the state-of-the-art; 3) choosing the subjects to be recognized and obtaining a suitable facial image for each of them; 4) creating the interpolated and SR images starting from LR pictures obtained from subjects’ faces; 5) creating surveys that presented interpolated and SR images, submitting such surveys to recognizers (that is, people that were asked to identify the subjects) and collecting their answers; 6) cleaning and analyzing data. The first five phases are detailed in the following, while the analysis of obtained results will be the object of **Section 3**.

### 2.1 Brief Overview of the State-of-the-Art of Super-resolution Applied to Face Recognition

Super-resolution of an LR image can be seen as an ill-posed problem in which a unique solution does not exist since there is a one-to-many relationship between the LR and corresponding SR images (Yang et al., 2019). The evaluation of an SR image can either be subjective or objective. The objective evaluation tends to be faster and more consistent; the subjective evaluation is more user-friendly, since it is the result of human perception. Two of the most frequently used objective evaluation metrics are PSNR (Peak Signal-to-Noise Ratio) and SSIM (Structural Similarity Index (Wang et al., 2004)). PSNR is often considered the standard for objective evaluation metrics (Hore and Ziou, 2010) and is strictly correlated with the mean square error (MSE) of the HR image with respect to the LR one; it is a metric based on the distance between intensity levels of pairs of pixels, and poorly correlates with human perception (Ong

et al., 2005). On the other hand, SSIM better correlates with human perception, since it takes into account the similarity between elements that are considered to be relevant to image perception, such as contrast and structure (Hore and Ziou, 2010).

In the early years of face SR, there have been attempts to solve the problem using well-known techniques which improve the resolution of the inputs, such as Gaussian pyramids (Lai et al., 2018) and principal components analysis (PCA) (Chakrabarti et al., 2007); these techniques had some degree of success but are far off from the recent years' ones.

The next step was to utilize local and global features and structures of the image (Guo et al., 2009); however, computational power has significantly increased since then, and these methods can hardly compete with more recent ones, quality wise.

The SR process has often been based on the minimization of a suitable error, or loss, function. One of the most popular loss functions used in face SR is the  $\mathcal{L}_1$  loss (see, e.g. (He and Cheng, 2022)), which is a pixel-wise mean absolute error between two pictures. Another pixel-wise loss is  $\mathcal{L}_2$ , used e.g. in (Dong et al., 2014), which is similar but uses the mean square error instead. Pixel-wise losses are often seen as non-optimal when dealing with face SR since their results tend to miss high frequency components in the super-resolved images (Zhang et al., 2012).

Perceptual loss in combination with an "adversarial loss" tends to give results that are subjectively pleasing (Wang et al., 2018). The concept of adversarial loss was firstly proposed in Generative Adversarial Networks (GANs) (Goodfellow et al., 2014), and describes the ability of a Generator (G) to generate an SR image which fools the discriminator (D) into assuming that it was a true HR picture and not an SR one. Ideally, a generator should trick the discriminator exactly half of the times, although in practice this rarely happens and the system can easily lead to model collapse during training (Thanh-Tung and Tran, 2020). A one-for-all loss function does not exist, but it should be noticed that a combination of multiple functions may better suit each case (Lucas et al., 2019; Jiang et al., 2021a; Tej et al., 2020).

Several more recent models utilize the fact that human faces have some degree of consistency and a fixed structure, so that some main features can be exploited in order to create a good HR version of a face image from an LR one (Yu et al., 2018; Jiang et al., 2021b). Also, gender, age and other special attributes of the face, such as glasses, can be used to accomplish that, even though, since the SR is an ill-posed problem and the mapping from input and output of the problem is one-to-many, the output can have features that did not belong to the original HR version of the input. To exploit this prior knowledge about human faces, a face SR method has been proposed (Yu et al., 2018) which is based on a Convolutional Neural Network (CNN) which is able to exploit the local and global structure of the image of a face. The first CNNs for face SR, however, did not utilize this face attributes (Hu et al., 2010) but were more focused more on the global and local properties of the image: a suitable combination of the two is able to capture the low-frequency features and the high-frequency ones, respectively.

More recently, researchers started to develop general GANs, trying to optimize the process of the game between

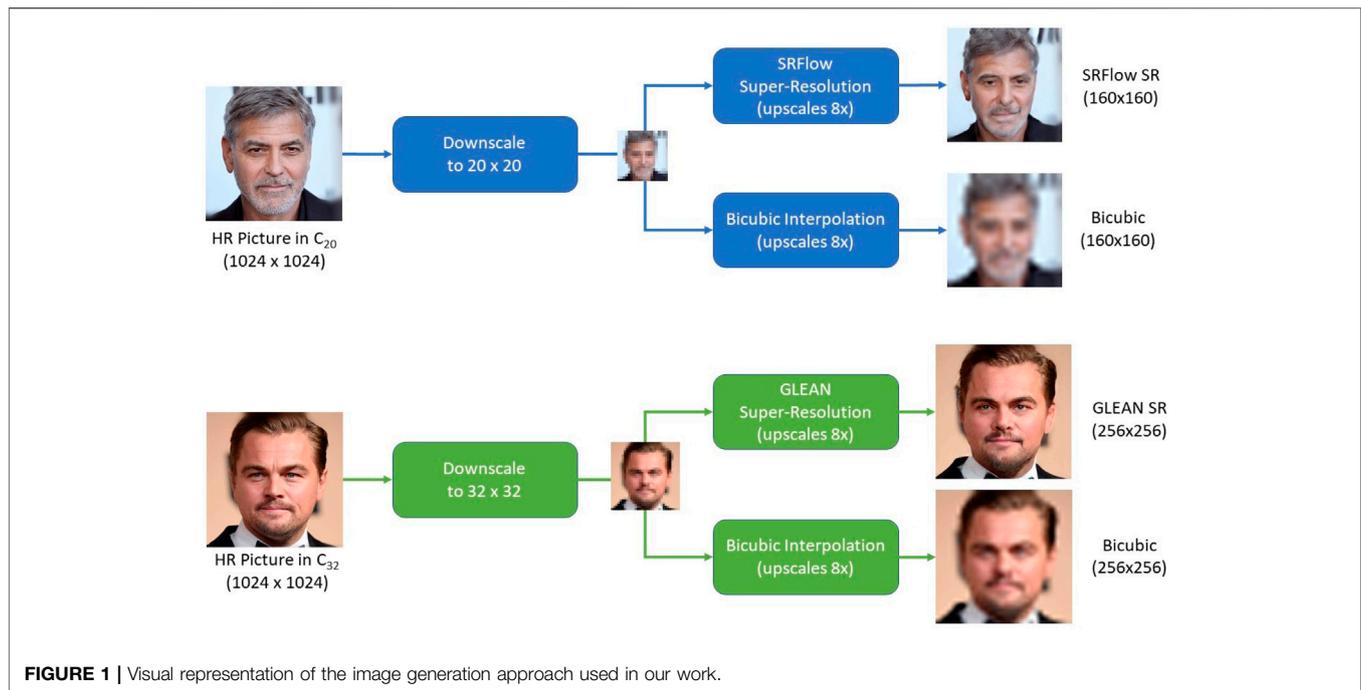
D and G. In (Ko and Dai, 2021) the discriminator was developed with edge maps, while a PCA-based decomposition of the input image was proposed in (Dou et al., 2020), in order to reduce the computational complexity. A shared property of these models is that the high frequency content in SR images is typically poor. One of the solutions proposed to add high frequencies details in SR face images is to rely on a pre-trained GAN as a prior for the model (Karras et al., 2019), which serves as a dictionary of face features to improve the subjective quality of the pictures. One of the problems with such models is that they can easily be too aggressive during the SR process, in the sense that the SR images may happen to have some features that actually do not belong to the original image. In some practical cases, indeed, this can lead to extreme results, where e.g. the SR face appears to have changed gender or largely modified age with respect to the original HR one (Menon et al., 2020).

Two recent super-resolution methods suitable for faces based on a pre-trained GAN are SRFlow (Lugmayr et al., 2020) and GLEAN (Chan et al., 2021). SRFlow uses a deep prior distribution of images, in our particular case of faces, to learn the distribution of high resolution (HR) faces corresponding to low resolution (LR) faces. During training, a negative log-likelihood loss is computed from HR-LR pairs. This also mitigates the notorious instability issues when dealing with GANs. The peculiarity of this architecture is that it learns to super-resolve, during inference, a Gaussian distribution in pair with an LR image from the previous conditional distribution of HR images. On the other hand, GLEAN uses a simple encoder-bank-decoder architecture to super-resolve LR images. Such bank can be seen as a dictionary of face features that gives greater details to the SR image. Those features are provided by a pre-trained model, which can produce portraits of fake human faces.

## 2.2 Selection of the Super-resolution Techniques for our Study

Among the various existing methods, we opted to use GLEAN (Chan et al., 2021) and SRFlow (Lugmayr et al., 2020). Our choice was motivated by several facts. Firstly, authors of these two methods showed that their solutions outperform several other recent techniques, including some of those mentioned in **Section 2.1**, e.g. those proposed by Menon et al. (2020) and Lim et al. (2017). Moreover, GLEAN and SRFlow are designed to work with low-resolution input images ( $32 \times 32$  and  $20 \times 20$ , respectively), while other techniques require a higher starting resolution, thus falling outside the case study considered in this work. Finally, a publicly available implementation of both techniques is available<sup>2</sup>, which helps ensure the reproducibility of the results presented in this work.

<sup>2</sup>SRFlow implementation is available here <https://github.com/andreas128/SRFlow>, while GLEAN implementation is available here <https://github.com/neonbjb/DL-Art-School>. Both resources were verified to be accessible as of 24 March 2022.



**FIGURE 1** | Visual representation of the image generation approach used in our work.

## 2.3 Selection of the Subjects to Be Identified and Gathering of Images

In the considered face recognition task, the observer is presented with a face and is asked to identify the subject without any additional reference image or database for comparison. Therefore, it was essential to select subjects whose face was known to the observer with high probability. We thus opted for worldwide known celebrities, including actors, politicians, and musicians. Since the two employed SR algorithms required different resolutions of the LR input images ( $20 \times 20$  for SRFlow,  $32 \times 32$  for GLEAN), we split the celebrities into two sets:  $C_{20}$ , which contains faces of Angela Merkel, Angelina Jolie, Boris Johnson, Brad Pitt, Bruce Willis, Christian Bale, Denzel Washington, Eminem, Emmanuel Macron, George Clooney, Hillary Clinton, Jack Nicholson, Joe Biden, Robert De Niro; and  $C_{32}$ , which contains faces of Donald Trump, John Travolta, Johnny Depp, Lady Diana, Leonardo Di Caprio, Barack Obama, Michelle Obama, Morgan Freeman, Quentin Tarantino, Samuel Leroy Jackson, Sylvester Stallone, Tom Cruise, Tom Hanks, Tommy Lee Jones, Vladimir Putin, Will Smith.

We searched each of the subjects on Google Images, and downloaded a picture of their face such that: 1) the face was clearly visible in a frontal pose; 2) the resolution was high or very high (minimum 55 pixels of eye-to-eye distance); 3) no significant or anomalous makeup was present. We remark that we only used images that were already publicly available on the web. A list of the source web location for each celebrity's picture is provided in the **Supplementary Material**.

## 2.4 Creation of Interpolated and SR Images

We adopted the classical approach for generating SR images, which is visually depicted in **Figure 1**: we started from a HR

version of celebrities' faces and downscaled them to create the LR version. The LR faces of celebrities in  $C_{20}$  were obtained downscaling faces to  $20 \times 20$  pixels, while faces of celebrities in  $C_{32}$  were downscaled to  $32 \times 32$  pixels. We finally upscaled each LR image by a factor of 8 using:

- bicubic interpolation (BIC from now on);
- the SRFlow algorithm for images in  $C_{20}$ , resulting in pictures of  $160 \times 160$  pixel resolution;
- the GLEAN algorithm for images in  $C_{32}$ , resulting in pictures of  $256 \times 256$  pixel resolution.

We remark that the GLEAN and SRFlow algorithms are both based on neural networks, so the size of the input image is fixed and cannot be changed without redesigning and retraining the network. The same holds for the output resolution. Therefore, we tested the two algorithms in the application scenario they were devised for: SRFlow is designed to upscale  $20 \times 20$  pixels images by a factor of 8, and GLEAN is designed to upscale  $32 \times 32$  pixels images by a factor of 8. The HR, LR, interpolated, and SR versions of some images in  $C_{32}$  are shown in **Figure 2**, which, as expected, tells of an indisputable advantage of using SR over bicubic interpolation in terms of perceptual quality.

In order to evaluate the effectiveness of SR algorithms from an objective point of view, we compared the interpolated and super-resolution images against the HR images. To do so, we used the Peak Signal-to-Noise Ratio (PSNR), the Structural Similarity Index (SSIM) (Wang et al., 2004), already mentioned in **Section 2.1**, and its extended version CW-SSIM (Sampat et al., 2009) that adds robustness to translations and rotations. Given the limited resolution of HR images, in computing the CW-SSIM, we used three decomposition levels and 16 orientations. All



**FIGURE 2** | The top-left picture shows some HR faces of celebrities in  $C_{32}$ , the top-right picture shows the generated LR version ( $32 \times 32$ ), the bottom-left picture shows the super-resolution version obtained with the GLEAN algorithm, and the bottom-right picture shows the interpolated version. The web address of pages containing celebrities faces is provided in the **Supplementary Material**. Sources include *biography.com*, *Wikipedia*, *mubi.com*, *imbd.com*, *theguardian.com*, *nientepopcorn.it* *ilsussidiario.net*, *notizie.it*.

images were converted to grayscale before comparison, as required by the similarity metrics.

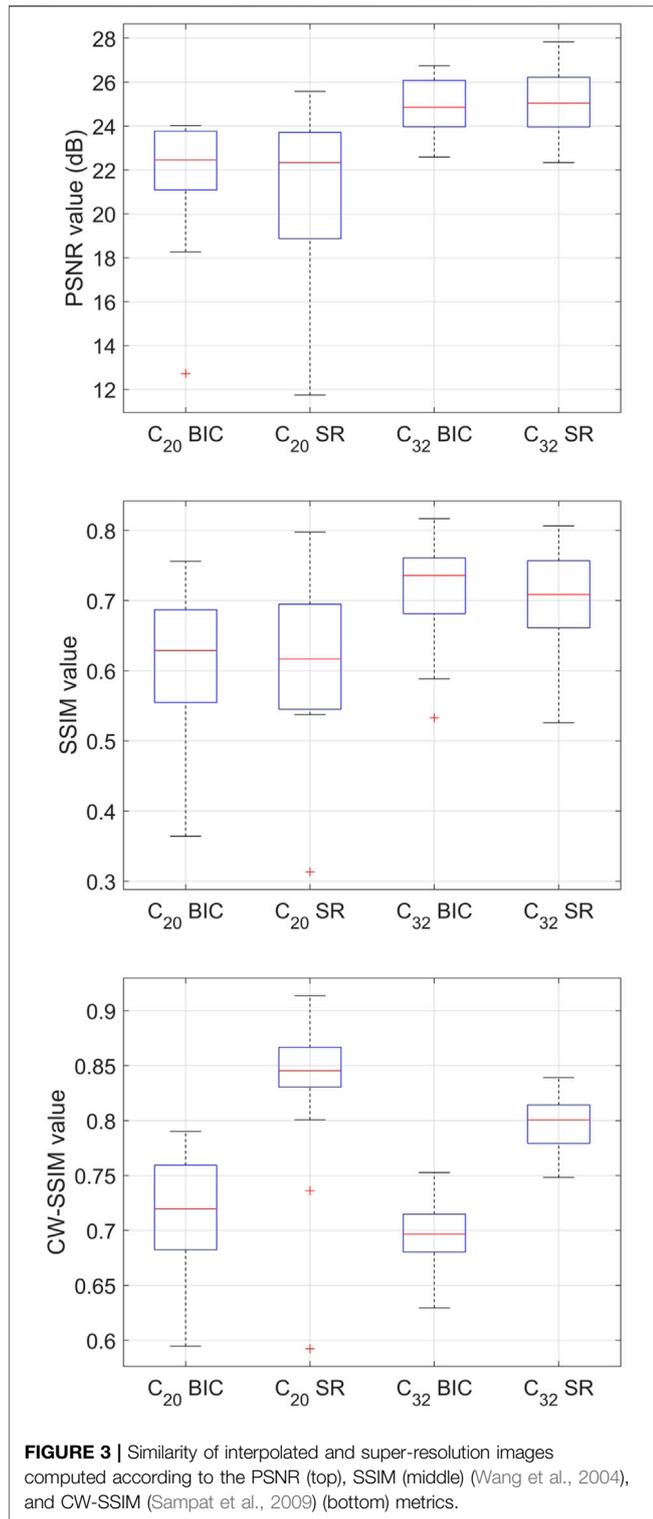
Results are shown in **Figure 3**. As we can see, the advantage of using SR techniques over bicubic interpolation is limited or even detrimental as long as PSNR and SSIM are considered. This fact is not surprising considering that, even though interpolated images are much more blurred than SR images, their structure remains consistent (shape of the face, the position of the eyes and mouth, etc.), which is the most important part for SSIM. However, when the CW-SSIM is considered, the advantage of using SR becomes much more evident.

## 2.5 Surveys Setup and Submission

Taking the superior visual quality of SR images for granted, we then wanted to investigate whether recent SR techniques can help face recognition made by a human operator. We used Google Forms to show images to the recognizers and gather their feedback. We verified using three popular browsers (Google Chrome v. 99.0.4844.74, Microsoft Edge v. 99.0.1150.46, and

Apple's Safari v. 9.1.1) that Google Form displays images at their native resolution if they fit in the page, which is an easily satisfied condition for our pictures given their limited pixel size. Recognizers could watch each face without any time limit, meaning that the form did not force them to switch to a different image after a given amount of time. Participants were allowed to write the name of the identified subject in plain text. In case they could not recognize a subject, they were asked to leave the answer blank. We did not suggest using any image search system as this would have distorted the experiment, which focuses on the subjects' ability to recognize a face without any technological aid besides image interpolation or SR.

In order to avoid the same recognizer seeing multiple versions of the same face, we prepared two distinct forms, defined as  $F_A$  and  $F_B$  in the following. Both forms contained the 30 celebrities' faces in the same order. In  $F_A$ , faces in odd positions were obtained with SR, and faces in even positions were interpolated, while in  $F_B$  the opposite was done. In such a way, each recognizer saw the same amount of interpolated and



SR faces, and no recognizer was asked to identify twice the same celebrity in order to avoid self-bias. In both forms, celebrities in  $C_{20}$  were shown first, and those in  $C_{32}$  followed. Surveys were sent to a heterogeneous group of volunteers, encompassing people from different countries, ages and professions, spanning

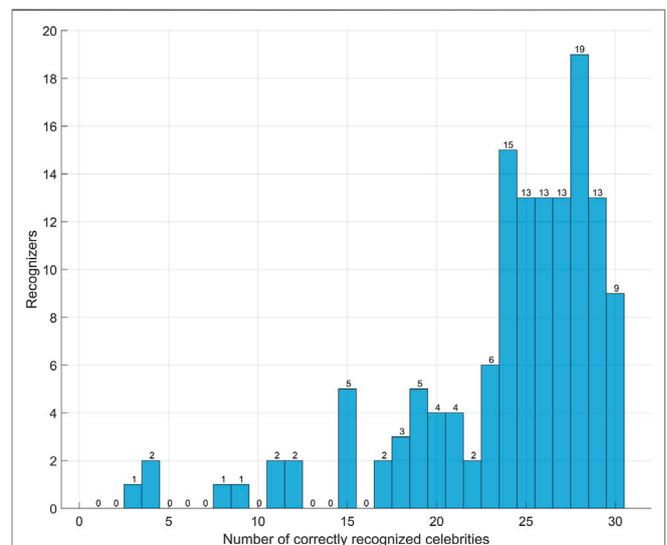
university-level students, law enforcement officers, and the general public. Although we did not collect personal data about the participants, their age range spanned the [20–65] interval, since the group included university-level students, law enforcement operators of mixed ages, general workers, but not retired subjects. As to the participant’s gender, there may be a slight unbalance towards males for the [35–65] age range. Indeed, several participants in such range were from the law enforcement industry, where the gender balance leans towards males (Duffin, 2020), and workers of technological companies or institutions. We received 102 responses for  $F_A$  and 33 responses for  $F_B$ , for a grand total of 135 submissions. The unbalancing in the number of submissions is due to the fact that some groups of participants were more reactive than others, and we did not have an effective way of preemptively balancing the submissions.

### 3 RESULTS

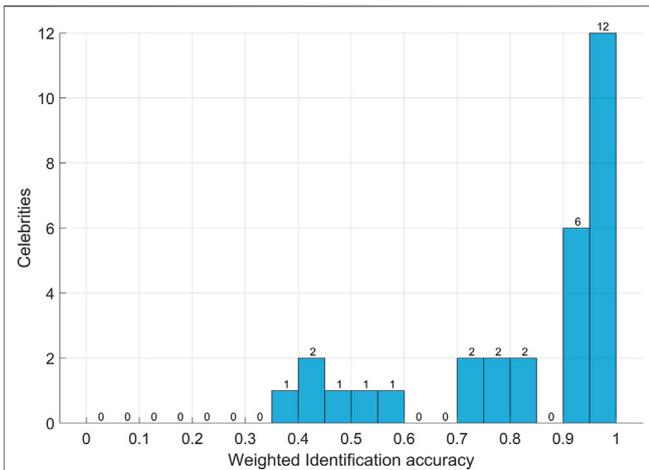
The analysis of collected surveys started with converting textual answers to binary values based on whether the celebrity had been correctly recognized. One of the authors carried out such conversion manually, allowing for a reasonable degree of flexibility in the interpretation. For example, when evaluating answers for Samuel Leroy Jackson, answers such as “Sam Jackson,” “S. L. Jakson”, and other typos or simplifications were all considered valid identifications.

#### 3.1 Preliminary Analysis and Pruning of Anomalies

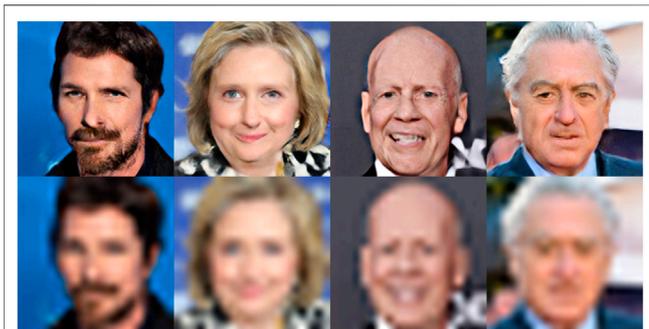
As a preliminary step, we evaluated the overall correct recognition rate for all recognizers, without any marginalization on the employed upsampling algorithm;



**FIGURE 4 |** Histogram of the number of correctly recognized celebrities achieved by all recognizers.



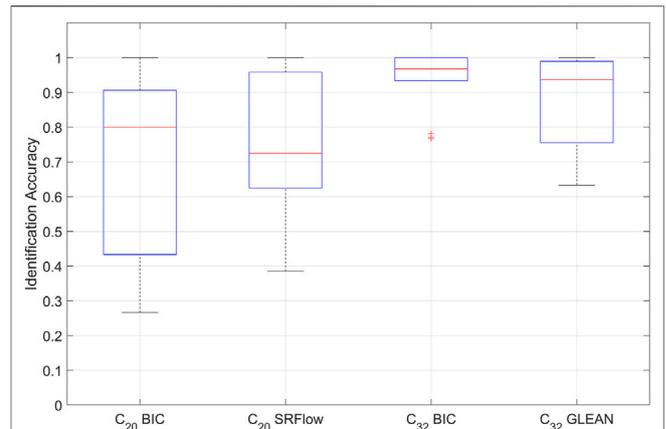
**FIGURE 5** | Histogram of the accuracy score achieved for all celebrities, obtained by weighted averaging the accuracies in  $F_A$  and  $F_B$ .



**FIGURE 6** | SR (top) and interpolated (bottom) images of the celebrities that were recognized by less than half of the participants.

Figure 4 reports the resulting values. It is rather evident that most recognizers were able to identify more than 50% of celebrities, with only nine recognizers performing worse than that. A closer inspection of their submitted surveys revealed that their answers were hilarious or left blank in most cases. We, therefore, decided to exclude these nine recognizers, six of which had compiled  $F_A$ . The total of available surveys thus decreased to 126.

We then analyzed the overall recognition accuracy marginalized over the celebrities. Since the same celebrity’s face was represented differently in the two surveys (once interpolated, once super-resolved), and considering that surveys did not have the same amount of feedback, we made a weighted average of the recognition accuracy achieved in each form and reported it in Figure 5. Once again, the vast majority of celebrities (26 over 30, 86.6%) were correctly identified by more than 50% of recognizers. In contrast, four celebrities proved to be harder to recognize in general, namely Bruce Willis (46.0%), Christian Bale (41.2%), Robert De Niro (39.7%), and Hillary Clinton (44.4%); Figure 6 shows the interpolated and SR versions of such celebrities. Not surprisingly, all the four celebrities belonged to the more challenging  $C_{20}$  set, which means that



**FIGURE 7** | Boxplot of the face recognition accuracies obtained for various celebrities separated according to the employed upsampling method and starting resolution. The red line indicates the median accuracy, the solid blue box includes 25-th to 75-th percentiles, the black dashed intervals include all elements that are not considered outliers, red crosses indicate outliers. The population size is 14 for the two boxes on the left and 16 for the two boxes on the right.

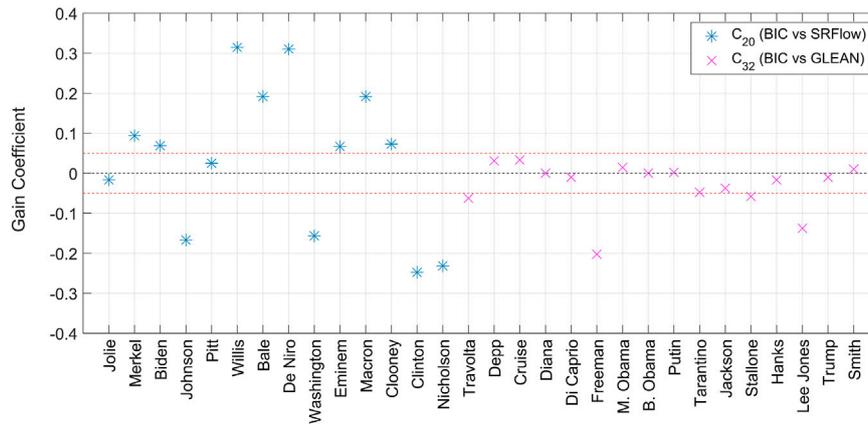
their face was shown at an upsampled resolution of  $160, \times, 160$  pixels. It is also to be noted that, as Figure 2 shows, our surveys contained relatively up-to-date pictures of celebrities, while a “less aged” and more idealized version could have been more familiar for recognizers. However, aging is one of the classical nuisance factors when carrying out face recognition experiments, so we do not consider this a limitation of our study.

Summarizing, after the preliminary analysis, we excluded nine recognizers because they provided hilarious or non-reliable answers, and we identified, but not excluded, four celebrities that were recognized by less than half of the participants.

### 3.2 Impact of Upsampling Techniques on FR Performance

In order to compare the performance of recognizers when presented interpolated versus SR images, in Figure 7 we provide a boxplot of the accuracies achieved for each celebrity, separating the four different kinds of images according to their starting resolution and the upsampling algorithm. The plot shows several insightful results. First of all, both for celebrities in  $C_{20}$  and in  $C_{32}$ , the median accuracy is higher when bicubic interpolation is used. It is important to note that the previous consideration holds independently of the starting resolution, even though the identification is obviously easier for celebrities in  $C_{32}$  compared to those in  $C_{20}$ . It is also apparent that the blue boxes (which cover 25-th to 75-th percentiles) are narrower for images in  $C_{32}$ , indicating a lower variance in the identification accuracies of various celebrities.

After looking at the aggregated results, we find it interesting to analyze the advantage of SR over interpolation on a celebrity-wise basis. In order to do so, we define a performance indicator called *gain coefficient*, indicated with  $\rho$  in the following, which for a given celebrity is obtained by subtracting the accuracy achieved



**FIGURE 8** | Gain coefficient when using SRFlow (blue markers) and GLEAN (pink markers) against bicubic interpolation. The black dashed line indicates zero, while the red dashed lines mark the range  $[-0.5, 0.5]$ .



**FIGURE 9** | Comparison of Bruce Willis' and Hillary Clinton's faces. The original HR image is shown on the left (downscaled to  $160 \times 160$ ), the bicubic interpolated version of the LR image is shown at the center, while the SRFlow super resolution version is shown on the right. For Bruce Willis, the measured gain coefficient was  $+0.32$ , while for Hillary Clinton it was  $-0.25$ .



**FIGURE 10** | Comparison of Morgan Freeman's and Tommy Lee Jones' faces. The original HR image is shown on the left (downscaled to  $256 \times 256$ ), the bicubic interpolated version of the LR image is shown at the center, while the GLEAN super resolution version is shown on the right. The measured gain coefficient was  $-0.20$  and  $-0.14$ , respectively.

when recognizing super-resolved images to the accuracy obtained when recognizing interpolated images. Therefore, celebrities with a positive  $\rho$  were more easily recognized when SR was used, those with a negative  $\rho$  indicate that interpolation was better, and those with a nearly null  $\rho$  indicate that SR and interpolation performed substantially the same in terms of favoring FR. The gain coefficient obtained for all celebrities is reported in **Figure 8**. We inspected the mean and standard deviation of  $\rho$ , weighting the accuracies obtained for each celebrity by the number of received submissions for  $F_A$  (96) and  $F_B$  (30). The obtained average value for the gain coefficient is  $\rho_\mu = -0.0093$  and its standard deviation is  $\rho_\sigma = 0.4586$ . While the average is very close to zero, the standard deviation is remarkable, as visually apparent in **Figure 8**.

Several comments are in order. We can see that, for celebrities in  $C_{20}$ , there is significant variability in the gain coefficient, which

ranges from  $-0.25$  for Hillary Clinton to  $+0.32$  for Bruce Willis. This fact suggests that the employed SRFlow algorithm can either help or hinder FR by a significant amount, depending on the input image. **Figure 9** compares the bicubic and SR versions of the mentioned celebrities: we can indeed observe a considerable amount of artifacts on Clinton's face, including a kind of "rejuvenation" effect, which probably hindered detection; in the case of Willis's face, instead, the SR image is of good visual fidelity if we leave out the mouth region.

When celebrities in  $C_{32}$  are considered, the gain coefficient is mostly centered around 0 and rarely overcomes 0.05 in magnitude. However, for Freeman's and Lee Jones' faces, the coefficient is as low as  $-0.20$  and  $-0.14$ , respectively, indicating that bicubic interpolation helped FR better than the GLEAN algorithm. **Figure 10** compares the bicubic and SR versions of the

mentioned celebrities. Freeman's face is corrupted by artifacts that likely acted against FR. This could be because, as reported by Biswas et al. (2011), humans tend to use peculiar traits such as scars, moles, and similar elements when doing recognition, and the relevant artifacts introduced by SR may interfere with this cognitive process. Instead, the negative gain coefficient obtained for Lee Jones' face is more surprising at first sight since the visual quality of the reconstructed image seems good. However, at a closer side-by-side inspection, allowed by **Figure 10**, it becomes apparent that the SR face has considerable differences in the shape of the eyes and mouth, which could undoubtedly hinder the recognition. This latter example supports the hypothesis that the "normalization" carried out by learning-based SR methods could be detrimental for face recognition, where peculiarities and details are essential. This observation is indeed reminiscent of what was observed by Ueda and Koyama (2010) in their study, where intense makeup acted as a normalization factor and hindered FR performance. This issue is less prominent in other fields, such as object recognition, where peculiarities are of lesser importance, or license plate reading, where the target object is standard by definition. Indeed, researchers demonstrated a clear advantage of deep neural networks in the latter task (Lorch et al., 2019; Kaiser et al., 2021; Rossi et al., 2021).

## 4 CONCLUSION

Aiding face recognition is one of the essential image processing applications in the forensic domain. Given the exciting results achieved by machine learning applied to single image super-resolution, the question of whether such techniques could help the face identification task is undoubtedly relevant. This paper addressed such a question by designing an experiment that compared simple bicubic interpolation to two state-of-the-art, deep learning-based super-resolution algorithms. We asked more than 130 volunteers to recognize 30 worldwide celebrities' upscaled faces and collected their answers. After a preliminary data cleaning process, we could eventually observe that the two tested super-resolution algorithms did not bring a significant

advantage over bicubic interpolation in terms of FR accuracy despite producing pictures with higher visual quality. Noticeably, in some cases, super-resolved images were recognized much better than interpolated images and vice-versa. However, on average, the difference in recognition performance was minimal.

As opposed to existing studies in state-of-the-art, we remark that our work did not use any automated algorithm for face recognition, which humans instead carried out. Our study results may lead to arguing that what was not done by the artificial neural network has been compensated for by the observer's visual system (which, all in all, is based on a biological neural network). We believe this intuition deserves further investigation with a more extensive and articulated study, aiming at understanding to what extent improvements to the visual quality of an image are helpful to a human observer when carrying out face recognition and other advanced image analysis tasks.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

## AUTHOR CONTRIBUTIONS

EV gathered pictures, prepared the LR, interpolated, and SR images, created surveys, and collected the results. MF and SC conducted the statistical analysis of data and manuscript preparation. MJ promoted surveys, reviewed the experiments and the manuscript, and coordinated the work.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/frsip.2022.854737/full#supplementary-material>

## REFERENCES

- Baker, S., and Kanade, T. (2002). Limits on Super-resolution and How to Break Them. *IEEE Trans. Pattern Anal. Machine Intell.* 24, 1167–1183. doi:10.1109/tpami.2002.1033210
- Biswas, S., Bowyer, K. W., and Flynn, P. J. (2011). "A Study of Face Recognition of Identical Twins by Humans," in 2011 IEEE International Workshop on Information Forensics and Security, 1–6. doi:10.1109/WIFS.2011.6123126
- Chakrabarti, A., Rajagopalan, A. N., and Chellappa, R. (2007). Super-resolution of Face Images Using Kernel PCA-Based Prior. *IEEE Trans. Multimedia* 9, 888–892. doi:10.1109/TMM.2007.893346
- Chan, K. C., Wang, X., Xu, X., Gu, J., and Loy, C. C. (2021). "Glean: Generative Latent Bank for Large-Factor Image Super-resolution," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 14245–14254. doi:10.1109/cvpr46437.2021.01402
- Dong, C., Loy, C. C., He, K., and Tang, X. (2014). "Learning a Deep Convolutional Network for Image Super-resolution," in European conference on computer vision (Berlin, Germany: Springer), 184–199. doi:10.1007/978-3-319-10593-2\_13
- Dou, H., Chen, C., Hu, X., Xuan, Z., Hu, Z., and Peng, S. (2020). "PCA-SRGAN: Incremental Orthogonal Projection Discrimination for Face Super-resolution," in Proceedings of the 28th ACM International Conference on Multimedia, 1891–1899. doi:10.1145/3394171.3413590
- Duffin, E. (2020). *Gender Distribution of Full-Time Us Law Enforcement Employees 2019* (Washington, D.C., USA: Federal Bureau of Investigation). Retrieved February 10, 2021.
- Fookes, C., Lin, F., Chandran, V., and Sridharan, S. (2012). Evaluation of Image Resolution and Super-resolution on Face Recognition Performance. *J. Vis. Commun. Image Representation* 23, 75–93. doi:10.1016/j.jvcir.2011.06.004
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al. (2014). Generative Adversarial Nets. *Adv. Neural Inf. Process. Syst.* 27.
- Guo, K., Yang, X., Zhang, R., and Yu, S. (2009). "Learning Super Resolution with Global and Local Constraints," in 2009 IEEE International Conference on Multimedia and Expo (New York, NY, USA: IEEE), 590–593. doi:10.1109/icme.2009.5202565

- He, X., and Cheng, J. (2022). Revisiting L1 Loss in Super-resolution: A Probabilistic View and beyond. *arXiv preprint arXiv:2201.10084*.
- Hore, A., and Ziou, D. (2010). "Image Quality Metrics: PSNR vs. SSIM," in 2010 20th international conference on pattern recognition (Istanbul, Turkey: IEEE), 2366–2369. doi:10.1109/icpr.2010.579
- Hu, Y., Lam, K. M., Qiu, G., and Shen, T. (2010). From Local Pixel Structure to Global Image Super-resolution: A New Face Hallucination Framework. *IEEE Trans. Image Process.* 20, 433–445. doi:10.1109/TIP.2010.2063437
- Jiang, J., Wang, C., Liu, X., and Ma, J. (2021a). Deep Learning-Based Face Super-resolution: A Survey. *ACM Comput. Surv. (Csur)* 55, 1–36.
- Jiang, J., Wang, C., Liu, X., and Ma, J. (2021b). Deep Learning-Based Face Super-resolution: A Survey. *arXiv preprint arXiv:2101.03749*.
- Kaiser, P., Schirmacher, F., Lorch, B., and Riess, C. (2021). "Learning to Decipher License Plates in Severely Degraded Images," in International Conference on Pattern Recognition (Berlin, Germany: Springer), 544–559. doi:10.1007/978-3-030-68780-9\_43
- Karras, T., Laine, S., and Aila, T. (2019). "A Style-Based Generator Architecture for Generative Adversarial Networks," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 4401–4410. doi:10.1109/cvpr.2019.00453
- Ko, S., and Dai, B.-R. (2021). "Multi-laplacian GAN with Edge Enhancement for Face Super Resolution," in 2020 25th International Conference on Pattern Recognition (ICPR) (Milan, Italy: IEEE), 3505–3512. doi:10.1109/icpr48806.2021.9412950
- Lago, F., Pasquini, C., Böhme, R., Dumont, H., Goffaux, V., and Boato, G. (2021). More Real Than Real: A Study on Human Visual Perception of Synthetic Faces [applications Corner]. *IEEE Signal. Process. Mag.* 39, 109–116.
- Lai, W. S., Huang, J. B., Ahuja, N., and Yang, M. H. (2018). Fast and Accurate Image Super-resolution with Deep Laplacian Pyramid Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* 41, 2599–2613. doi:10.1109/TPAMI.2018.2865304
- Li, K., Yang, S., Dong, R., Wang, X., and Huang, J. (2020). Survey of Single Image Super-resolution Reconstruction. *IET image process* 14, 2273–2290. doi:10.1049/iet-ipr.2019.1438
- [Dataset] Lim, B., Son, S., Kim, H., Nah, S., and Lee, K. M. (2017). "Enhanced Deep Residual Networks for Single Image Super-resolution," in 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). doi:10.1109/cvprw.2017.151
- Lorch, B., Agarwal, S., and Farid, H. (2019). Forensic Reconstruction of Severely Degraded License Plates. *Electronic Imaging* 2019 (5), 529–531.
- Lucas, A., Lopez-Tapia, S., Molina, R., and Katsaggelos, A. K. (2019). Generative Adversarial Networks and Perceptual Losses for Video Super-resolution. *IEEE Trans. Image Process.* 28, 3312–3327. doi:10.1109/tip.2019.2895768
- [Dataset] Lugmayr, A., Danelljan, M., Gool, L. V., and Timofte, R. (2020). *SRFlow: Learning the Super-resolution Space with Normalizing Flow*.
- Menon, S., Damian, A., Hu, S., Ravi, N., and Rudin, C. (2020). "Pulse: Self-Supervised Photo Upsampling via Latent Space Exploration of Generative Models," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2437–2445. doi:10.1109/cvpr42600.2020.00251
- Narwaria, M. (2022). Does Explainable Machine Learning Uncover the Black Box in Vision Applications? *Image Vis. Comput.* 118, 104353. doi:10.1016/j.imavis.2021.104353
- Nelson, K., Bhatti, A., and Nahavandi, S. (2012). "Performance Evaluation of Multi-Frame Super-resolution Algorithms," in 2012 International Conference on Digital Image Computing Techniques and Applications (DICTA) (Fremantle, WA, Australia: IEEE), 1–8. doi:10.1109/dicta.2012.6411669
- Ong, E., Yang, X., Lin, W., Lu, Z., and Yao, S. (2005). "Perceptual Quality Metric for Compressed Videos," in Proceedings IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005 (Philadelphia, PA, USA: IEEE), ii–581. Vol. 2.
- Rossi, G., Fontani, M., and Milani, S. (2021). "Neural Network for Denoising and reading Degraded License Plates," in International Conference on Pattern Recognition (Berlin, Germany: Springer), 484–499. doi:10.1007/978-3-030-68780-9\_39
- Sampat, M. P., Zhou Wang, Z., Gupta, S., Bovik, A. C., and Markey, M. K. (2009). Complex Wavelet Structural Similarity: A New Image Similarity index. *IEEE Trans. Image Process.* 18, 2385–2401. doi:10.1109/TIP.2009.2025923
- Tej, A. R., Halder, S. S., Shandeelya, A. P., and Pankajakshan, V. (2020). "Enhancing Perceptual Loss with Adversarial Feature Matching for Super-resolution," in 2020 International Joint Conference on Neural Networks (IJCNN) (Piscataway, NJ, USA: IEEE), 1–8. doi:10.1109/ijcnn48605.2020.9207102
- Thanh-Tung, H., and Tran, T. (2020). "Catastrophic Forgetting and Mode Collapse in GANs," in 2020 International Joint Conference on Neural Networks (IJCNN) (Piscataway, NJ, USA: IEEE), 1–10. doi:10.1109/ijcnn48605.2020.9207181
- Ueda, S., and Koyama, T. (2010). Influence of Make-Up on Facial Recognition. *Perception* 39, 260–264. doi:10.1068/p6634
- Wang, C., Xu, C., Wang, C., and Tao, D. (2018). Perceptual Adversarial Networks for Image-To-Image Transformation. *IEEE Trans. Image Process.* 27, 4066–4079. doi:10.1109/tip.2018.2836316
- Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P. (2004). Image Quality Assessment: from Error Visibility to Structural Similarity. *IEEE Trans. Image Process.* 13, 600–612. doi:10.1109/TIP.2003.819861
- Wang, Z., Chen, J., and Hoi, S. C. (2020). Deep Learning for Image Super-Resolution: A Survey. *IEEE Trans. Pattern Anal. Mach. Intell.* 43 (10), 3365–3387.
- Yang, W., Zhang, X., Tian, Y., Wang, W., Xue, J.-H., and Liao, Q. (2019). Deep Learning for Single Image Super-resolution: A Brief Review. *IEEE Trans. Multimedia* 21, 3106–3121. doi:10.1109/tmm.2019.2919431
- Yu, X., Fernando, B., Ghanem, B., Porikli, F., and Hartley, R. (2018). "Face Super-resolution Guided by Facial Component Heatmaps," in Proceedings of the European conference on computer vision (ECCV), 217–233. doi:10.1007/978-3-030-01240-3\_14
- Zhang, L., Zhang, L., Mou, X., and Zhang, D. (2012). "A Comprehensive Evaluation of Full Reference Image Quality Assessment Algorithms," in 2012 19th IEEE International Conference on Image Processing, 1477–1480. doi:10.1109/ICIP.2012.6467150
- Zhao, W., Chellappa, R., Phillips, P. J., and Rosenfeld, A. (2003). Face Recognition. *ACM Comput. Surv.* 35, 399–458. doi:10.1145/954339.954342

**Conflict of Interest:** Authors EV, MF, and MJ were employed by the company Amped Software.

The remaining author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Velan, Fontani, Carrato and Jerian. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.