AI-based face super resolution for forensic evaluation: a trustworthiness analysis

Rocco Mazzei[1], Massimo Iuliani[2], Marco Fontani[2], Martino Jerian[2], Mauro Barni[1]
1 – University of Siena, 2 – Amped Software

Introduction

Advances in AI suggest that image and video quality can be improved for forensic examinations.
Recent algorithms allow to increase the image resolution thus providing the expectation that original information can be restored even in low quality images captured by CCTV cameras. This would open to various possible applications, including recognizing people via facial super resolution. However, it remains unclear whether these technologies satisfy fundamental forensic requirements and whether processing preserves, rather than altering, the original information. In some recent forensic cases, it was already highlighted that AI-based enhancement led to unreliable and misleading information [1].
We therefore examined a recent AI-based enhancement system, Supir [2], applying it to increase the resolution of faces.
The achieved results show that, in many cases, the tested AI-based super resolution method compromises discriminative features, such as the shape of teeth and ears. Furthermore, the system also occasionally removes, adds, or reinterprets subject-specific elements.
The results also suggest that the main anthropometric features should be treated with care since their reconstruction can be affected by several factors.
This study underscores the need for great caution in developing and using deep-learning based systems for evidentiary purposes, and highlights the importance of training all involved stakeholders so these tools are employed with full awareness of their limitations.

Methodology

We considered 50 high-resolution close-up portraits (more than one megapixel) of male and female real subjects: the first half were taken directly by the authors, and the other half were obtained from Freepik [3]. Both male and female subjects were considered.
The images were then downscaled to a resolution of 64x64 to deliberately remove relevant facial details for the experiment.

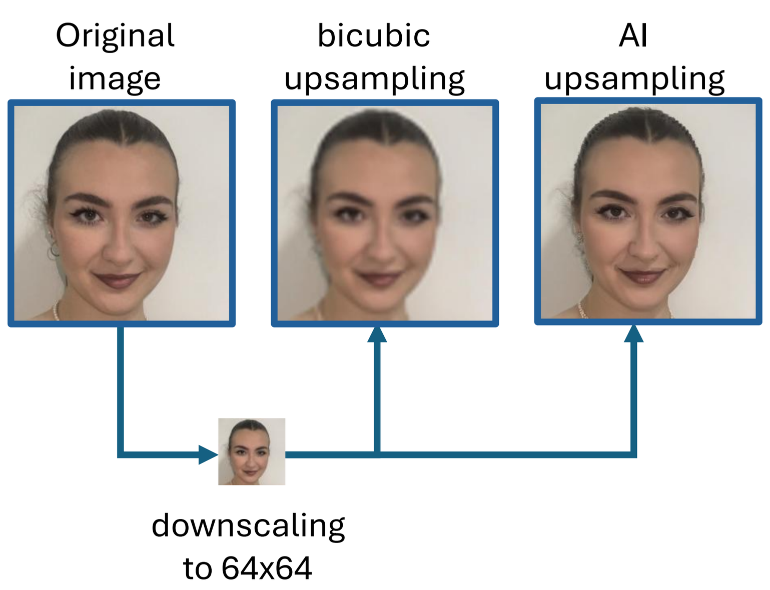

Figure 1 – Images are downscaled and then upscaled with bicubic algorithm and Supir respectively. The reconstructed images are then compared with the original ones.

Subsequently, each downscaled image was subjected to a 16x upscaling using two methods: a classic bicubic upscaling technique (that is considered forensically sound since it cannot negatively affect the result) and the AI-based method Supir [2], designed for image restoration and upscaling. Supir was selected among CNN-based [4], in particular TFMAN [5], GAN-based [6] and diffusion-based methods after a preliminary analysis on super-resolved images from our facial dataset (see Figure 1): Supir appeared to preserve facial identity and visual details most faithfully.
Since Supir intrinsically allows a 4x upscaling, the final 16x upscaling was achieved by applying two consecutive 4x upscaling operations.
For each tested image, we then analyzed the resulting AI-upsampled image to determine if Supir introduces, removes, or changes image features that could affect the subject identification.

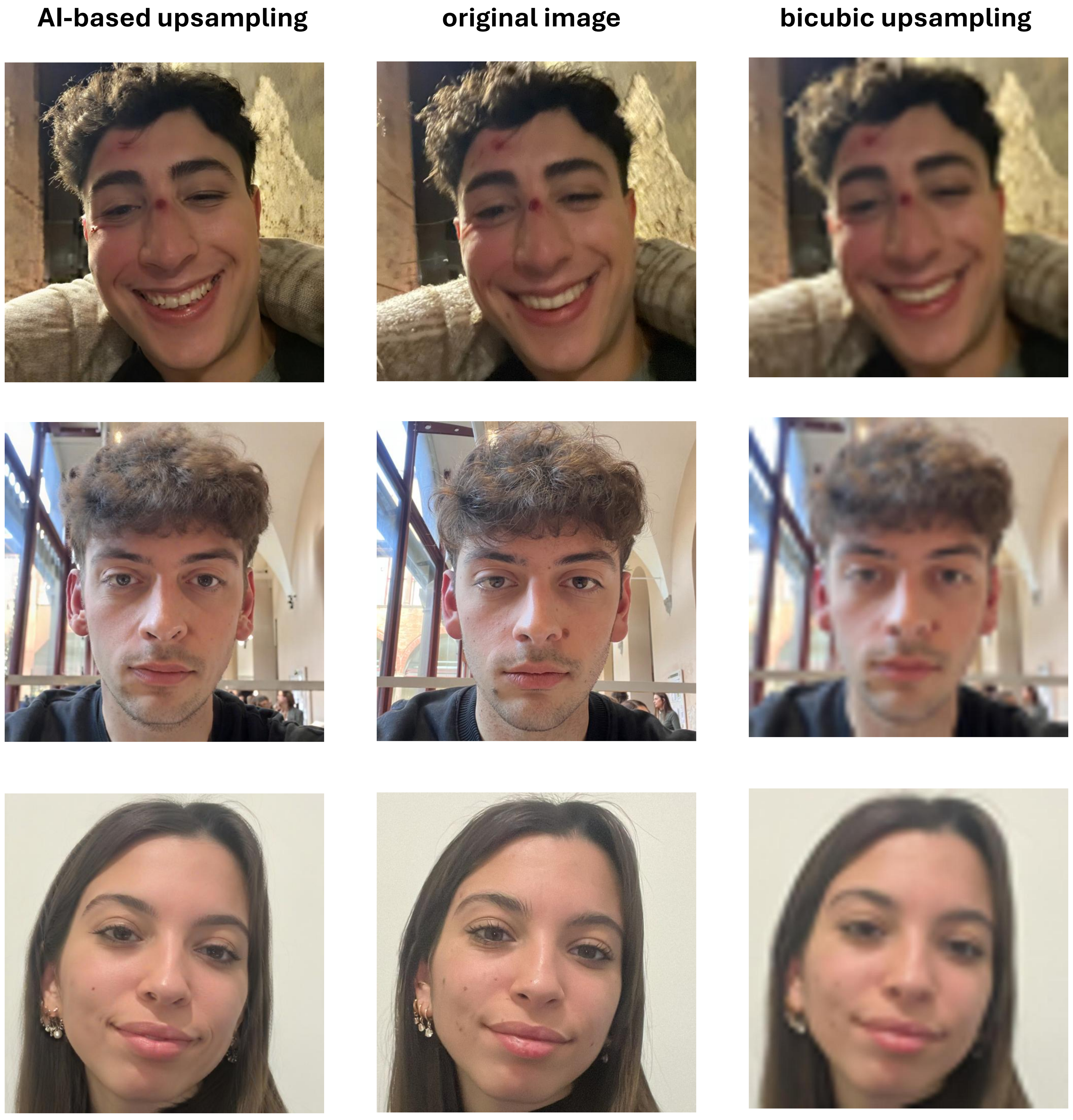

Figure 4 - Three example images are shown: the original image (shown in the central column) is downsampled and then upsampled using Supir and bicubic interpolation (shown in the left and right columns, respectively). It can be observed that the AI-based reconstruction introduces, modifies, or removes relevant discriminative facial features. In the first example, the edges of the teeth are heavily altered; in the second, the shape of the ear is completely changed; in the third, several moles are removed by the AI-based upsampling.

More specifically, we considered:

- Three anthropometric features, according to the FISWG guidelines [7]:
  1. Distance between the center of the eyes (Interpupillary distance);
  2. Distance between the corners of the mouth;
  3. Length of nasal base;
- The number of discriminant features that were visually modified by the AI upsampling, such as the shape of ears, teeth, scars, etc.

Experimental Results

With reference to the three anthropometric features, we computed the difference between the original and the AI-based reconstructed image. We report an example in Figure 2.

Figure 2 – The normalized distance is computed on the original and AI-reconstructed images (top and bottom row respectively). Each measure is normalized with respect to the face width.

To assess the impact of such variations, we need to compare them with the distributions of the intra-subject and inter-subject distances respectively.

We have then collected the inter-distances between subjects captured in controlled environment. Intra-distance variability was not available due to the limited dataset size. We then estimated the intra-variability introduced by user points selection. For this purpose, two different authors measured the three features and the differences between measurements were collected.

The achieved statistics were used to build two ROC curves: the first between the measurement differences (between users) and the inter-distances; the second one between the measurement differences (pre-post AI) and the inter-distances (see Figure 3).

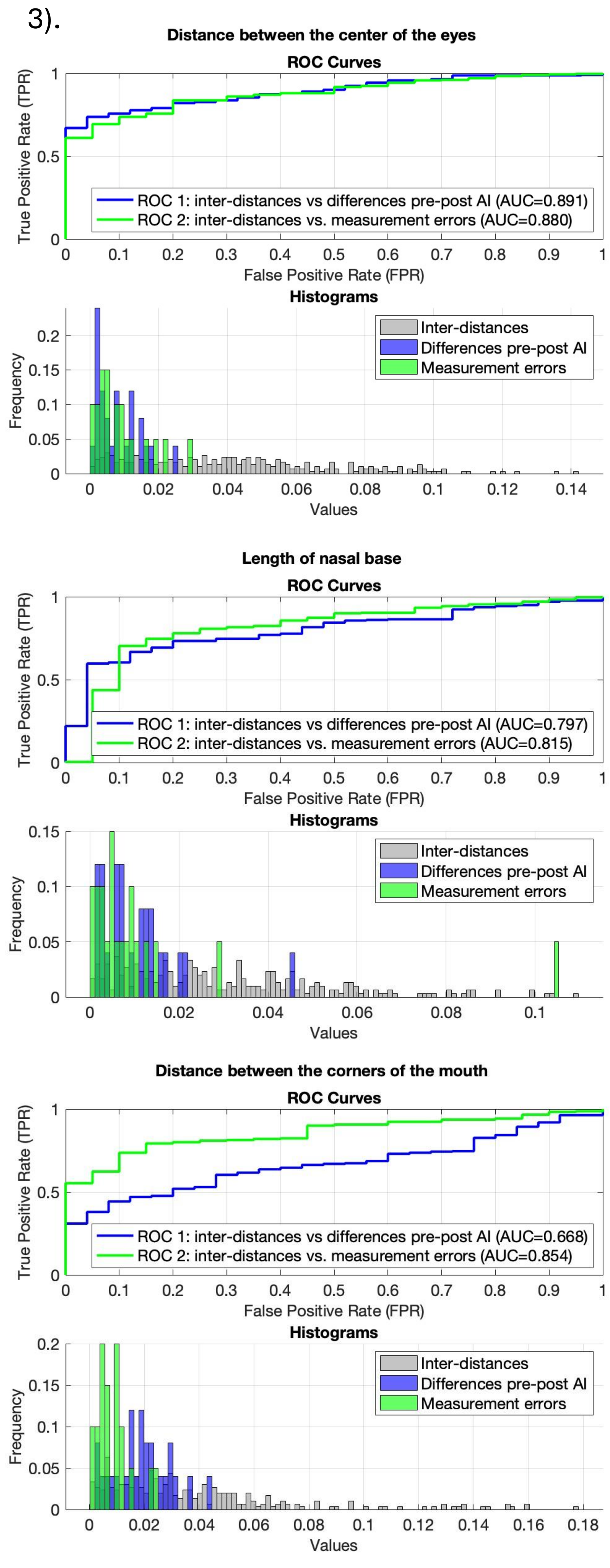

Figure 3 – ROC curves for each facial anthropometric feature comparing statistics from original and super-resolved images.

The similarity between the ROC curves for eye distance and nasal length suggests that AI-based super-resolution does not substantially affect these anthropometric measurements. Conversely, mouth width appears to be negatively impacted by AI super-resolution, as the AUC decreases after AI upsampling.
As a second analysis, we counted the number of discriminative features that were visually modified by AI-based upsampling. Figure 4 shows three examples of real faces (central column) alongside their AI-based and bicubic reconstructions (left and right columns, respectively). In all three cases, the Supir upsampling substantially alters key facial features. In the first image, the edges of the teeth are reinterpreted and visibly modified; in the second, the shape and structure of the ear are completely altered; in the third, several moles are removed by the AI-based upsampling.
This behavior was observed in 27 out of 50 images, regardless of their origin (manually captured or internet). The most frequently affected elements include moles, freckles, ear and teeth deformities, and variations in earrings.

Conclusion

A preliminary analysis was conducted to evaluate whether AI-based super-resolution can be reliably exploited for forensic assessments. We focused on facial super-resolution, examining the effect of a recent AI-based method on 50 close-up facial portraits.

The results indicate that the tested technology often fails to satisfy fundamental forensic requirements, such as preserving—rather than altering—the original information. In more than half of the images, the reconstructed faces exhibited changes in discriminative characteristics, including the shape of the teeth and ears. In addition, the tested system occasionally removes, adds, or reinterprets subject-specific details.

We also assessed the stability of three anthropometric measurements. In two cases, stability was maintained; however, a larger and more statistically representative dataset is needed to draw robust conclusions.

Overall, this study underscores the need for considerable caution when developing and using deep learning–based systems for evidentiary purposes. These tools can produce misleading outputs that may nonetheless be perceived as “high-quality” evidence. It is therefore essential to train all involved practitioners to increase their awareness of the limitations of such technologies.

References

[1] https://www.nbcnews.com/news/us-news/washington-state-judge-blocks-use-ai-enhanced-video-evidence-rcna141932
[2] https://supir.xpixel.group/
[3] https://it.freepik.com
[4] Dong, Chao, et al. "Image super-resolution using deep convolutional networks." *IEEE transactions on pattern analysis and machine intelligence* 38.2 (2015): 295-307.
[5] Chen, Qizhou, and Qing Shao. "Single image super-resolution based on trainable feature matching attention network." Pattern Recognition 149 (2024): 110289,
[6] Goodfellow, Ian, et al. "Generative adversarial networks." Communications of the ACM 63.11 (2020): 139-144.
[7] https://fiswg.org/FISWG_Morph_Analysis_Feature_List_v2.0_20180911.pdf