Assessing the usefulness of super-resolution algorithms for cross-resolution face recognition

Anonymous FG 2018 submission
Paper ID 104

***Abstract*— With recent advancements in deep learning and convolutional neural networks (CNNs), face recognition has seen significant performance improvements over the last few years. However, low-resolution images still remain challenging, with CNNs performing relatively poorly compared to humans. One possibility to improve performance in these settings often advocated in the literature is the use of super-resolution (SR). In this paper, we explore the usefulness of SR algorithms for cross-resolution face recognition in experiments on the Labeled Faces in the Wild (LFW) and SCface datasets using four recent deep CNN models. We conduct experiments with synthetically down-sampled images as well as real-life low-resolution imagery captured by surveillance cameras. Our experiments show that image super-resolution can improve face recognition performance considerably on very low-resolution images (of size 24×24 or 32×32 pixels), when images are artificially down-sampled, but has a lesser (or sometimes even a detrimental) effect with real-life images leaving significant room for further research in this area.**

I. Introduction

Automated face recognition systems have recently been shown to match and in some cases even surpass human recognition performance given facial images of appropriate quality [1], [2], [3]. In cross-resolution settings, however, where low-resolution (LR) images need to be matched against high-resolution (HR) galleries, contemporary systems still lag behind human capabilities. This is a crucial consideration for surveillance and security applications, where reliable recognition from low-resolution imagery is key [4].

Existing work on automated cross-resolution face recognition can be grouped into techniques that exploit: *i) joint multi-resolution feature learning* and try to learn computational models that extract identical features from images at different sizes/resolutions [5], [6], [7], *ii) low-frequency feature extraction* and strive to build models that focus on blur-invariant low-frequency features which are not affected by low image resolution [8], [9], and *iii) super-resolution (SR) algorithms*, which try to modify the low-resolution probe images to reduce the dissimilarity with the high-resolution gallery/target data [10], [11], [12]. The latter group of techniques is becoming increasingly popular mainly due to the successes of recent supper-resolution algorithms.

Contemporary SR approaches exhibit considerable ability to reconstruct high-resolution details from low-resolution images [13], [14], [15], [16], and, therefore, offer a straightforward way of boosting the performance of cross-resolution face recognition. As illustrated in Fig. 1, a desirable characteristic of using SR at the pre-processing level is that it can be used independently from existing recognition models.

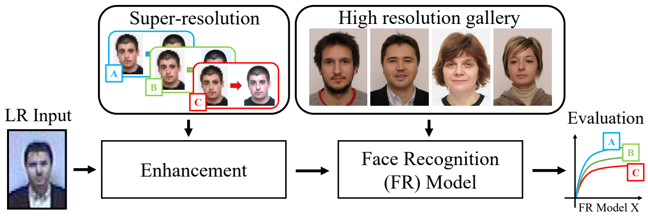


Fig. 1. We evaluate the use usefulness of super-resolution for cross-resolution face recognition using the experimental setup shown above. With this setup super-resolution algorithms reconstruct a HR image from the LR probe. The HR reconstructions are then used with a pre-trained face recognition model to assess performance. While such an setup is common in the literature (e.g., [17], [18], [19]), it's usefulness beyond synthetically down-sampled images is still under-explored.

Thus, SR models can be deployed with existing commercial products and existing models and can aid in improving performance in difficult settings where the input data is of (prohibitively) low-resolution.

The benchmarks typically used in SR research involve reconstructing synthetically-degraded images via the optimization of metrics such as peak signal-to-noise ratio (PSNR) or the structural similarity measure (SSIM, [20]). Such an approach is inherently focused on the perceptual quality of the super-resolved images and less so on the semantics of the reconstructions. Even when SR is used to improve the semantic content of the images (e.g., pre-processing for face recognition), most of the existing work focuses only on synthetically-degraded and not real-life low-resolution data.

In this paper we try to address this gap and study the utility of existing image SR algorithms on synthetic and real low-resolution images as the basis for further research into the use of SR for (automated and human) cross-resolution face recognition. To evaluate the general applicability of example- (or learning-) based SR algorithms beyond their training samples, we first generate artificial low-resolution face images from the Labeled Faces in the Wild (LFW) dataset [21]. We process the degraded images using different SR algorithms and perform recognition experiments with four state-of-the-art deep recognition models using the super-resolved images. To assess the performance of SR algorithms on real-life low-resolution images, we use the SCface [22] dataset of surveillance-camera images containing face images captured in real-life surveillance scenarios.

Our research results in the following main contributions:

- A comprehensive experimental assessment of the usefulness of SR algorithms for cross-resolution face recognition with a particular emphasis on recent SR approaches and deep recognition models.

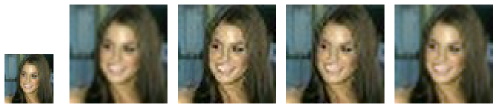


Fig. 2. Visual examples of the upscaling capabilities of the considered super-resolution algorithms. The figure shows an input image of size 32×32 pixels (left most image) and super-resolved images upscaled by a factor of $2\times$ using different SR techniques (the four images on the right).

- A characterization of the generalization properties of SR algorithms via a comparison of their performance on synthetically down-sampled and real-life LR images.
- A discussion of open problems and pointers to research needs in the field of super-resolution.

II. Related Work

Super-resolution algorithms have progressed from the early interpolation-based methods to more recent example-based SR techniques which aim at learning the relationship between low- and high-resolution images from training data [23], [24], [25]. Especially with advancements made in deep-learning, interest in SR has gained additional momentum and numerous algorithms with remarkable reconstruction capabilities have been proposed in the literature [14], [15], [16], [26], [27], [28], [29], [30], [31], [32], [33], [34], [35]. Despite its primary goal as a digital zooming tool aimed at improving the perceived quality of the super-resolved image data, a significant body of work exploit it s a pre-processing step for face recognition [11], [17], [18], [19], [36], [37]. However, most of this work centers around artificially down-sampled images, leaving the question of performance on real-life LR images mostly unexplored.

III. Methodology

A. Experimental setup

To evaluate the usefulness of SR techniques for cross-resolution face recognition we use the experimental setup illustrated in Fig. 1. Here, the low resolution (LR) probe image is first processed and enhanced with a SR algorithm to generate a high-resolution (HR) input that better matches the characteristics of the HR gallery. A pre-trained deep model is then used to compute an image descriptor from the enhanced probe and employed to generate similarity scores (based on the cosine similarity) that are exploited for identity inference.

All super-resolution techniques considered in the evaluation are used with a fixed upscaling factor of $2\times$ regardless of the input image size. To further upscale the input image to the size required by the different deep models, bicubic interpolation is used. This setup allows us to make full use of the super-resolution algorithms, which commonly perform best with small upscaling factors [16], [29], [30], [31], and to super-resolve the smallest image size used in our experiments (i.e., 24×24 pixels) to a size that can be handled by all tested recognition models (as evidenced by our results).

B. Super-resolution algorithms

Using the presented setup, we assess the usefulness of four super-resolution algorithms, i.e., Bicubic interpolation [38], the Naive Bayes Super-resolution Forest (NBSRF, [28]), the Superresolution Convolutional Neural Network (SRCNN, [14]), and the Very Deep SuperResolution Network (VDSR, [15]). For the example-based super-resolution algorithms (i.e., NBSRF, SRCNN, VDSR), which learn image upscaling from training data, we use publicly available pre-trained models from the web. A comparison of the upscaling capabilities of the four SR techniques is presented in Fig. 2 and a brief description is given below.

TABLE I

High-level comparison of the considered deep models

Model	#parameters	input dim.	output dim.	#layers
AlexNet [41]	58,282,752	(3, 224, 224)	4096	7
VGG-Face [2]	117,479,232	(3, 224, 224)	4096	15
GoogLeNet [42]	21,577,728	(3, 299, 299)	2048	37
SqueezeNet [43]	3,753,856	(3, 224, 224)	2048	12

Bicubic interpolation: Interpolation methods are parameter free procedures used to change the sampling rate of signals. Bicubic interpolation [38] upsamples an image by interpolating missing pixel values using Lagrange polynomials, cubic splines, or other similar functions. The interpolation is commonly available in most image editing applications, and has the favorable property (compared to nearest-neighbor and bilinear interpolation methods) of preserving continuous values and gradients given a continuous sampling grid. Bicubic interpolation represents our baseline, as it is commonly used to ensure a correct input size for various recognition models.

NBSRF: This approach uses a tree ensemble to learn hierarchical dictionaries of LR image patches and corresponding (locally linear) regressors needed for HR patch estimation [28]. Unlike previous dictionary learning approaches [13], [39] which use flat (non-hierarchical) patch dictionaries with sparse coding constraints, the hierarchical tree-ensemble provides advantages in both runtime and performance due to the Bayesian tree selection criterion that ensures that a single tree is selected during inference-time.

SRCNN: This algorithm represents one of the first attempts at example-based super-resolution using deep neural networks. Here [14], a relatively simple neural net with three convolutional layers that reflects the LR patch extraction, non-linear mapping and HR patch selection stages from dictionary-learning-based SR approaches is used. Its advantage over prior dictionary-based methods is efficient end-to-end training combined with greater expressivity due to localized constraints and regularizations enabled by the gradient-based training procedure.

VDSR: In [40], the idea of shortcut connections and residual learning was introduced to the field of deep learning. VDSR [15] applies this idea to super-resolution - instead of the super-resolution algorithm having to reconstruct the entire high-resolution image, by adding an interpolated low-resolution image to its output, it effectively only learns the missing high-resolution details, greatly improving expressivity and performance given the same network complexity.

C. Face recognition models

We use four diverse deep-learning-based face recognition models for our evaluation, i.e., the AlexNet [41], GoogLeNet [42], VGG-FACE [2] and SqueezeNet [43] convolutional neural networks (CNNs). The models chosen differ in terms of architecture, parameter-space size, size of the input image they are able to process, the size of the output descriptor and number of network layers. A brief high-level overview of the models is presented in Table I, while a more comprehensive comparison is found in [44]. We train all four models for the face classification task on around 1.8 million images (of appx. 2600 subjects) that we are able to collect from the web based on the URLs provided in the VGG Face dataset [2]. We use data augmentation techniques including random cropping, blurring, translating and rotating to avoid over-fitting and once trained remove the classification layer from each model and exploit the activations of the penultimate layer as the image descriptor of the input images.

D. Datasets, protocols and performance metrics

We use two popular face datasets for our experiments. The first is the Labeled Face in the Wild (LFW [21]) dataset, where we introduce artificial down-sampling to simulate different image resolutions. The second one in the SCface dataset [22], which contains real-life low resolution images recorded by a number of surveillance cameras at different distances of the cameras to the subjects. The datasets and corresponding experimental protocols are presented below.

LFW: The dataset consists of $13,233$ face images of size 256×256 pixels belonging to 5749 subjects. We use the dataset for our super-resolution experiments by down-sampling the images to smaller sizes, thus, simulating lower resolutions. The entire dataset is progressively down-sampled to $24\times24, 32\times32, 48\times48, 64\times64, 96\times96, 128\times128, 160\times160$ and 192×192 pixels using area-rule interpolation, and the down-sampled images are blurred using a Gaussian kernel with width 5 and standard deviation of 0.5. The down-sampled images are then restored to the appropriate input size for the face recognition models using various super-resolution algorithms (using an upscaling factor of $2\times$ followed by bicubic interpolation). We follow the so-called "unrestricted outside data" protocol in our experiments that defines 6000 image pairs for which verification has to be performed. The validation pairs are split into 10 folds, and each fold is partitioned equally between genuine and impostor pairs, where the held-out folds are used to determine the decision threshold for the similarity function. For the experiments, the target (gallery) images are kept unchanged, whereas the probe (test, query) images are down-sampled and then super-resolved for the evaluation. We measure the performance of the experiments with the verification accuracy over 10 folds, as defined by the LFW protocol.

SCface: The dataset [22] contains images of 130 distinct subjects. The available images are split between a gallery set, containing 130 high-resolution frontal mugshots (1 per subject), and a larger probe set of surveillance-camera images. The daylight camera set, which we use for our experiments, consists of images from 5 different security cameras, as shown in Fig. 3. Each subject is recorded by each camera at 3 different distances, resulting in a total of

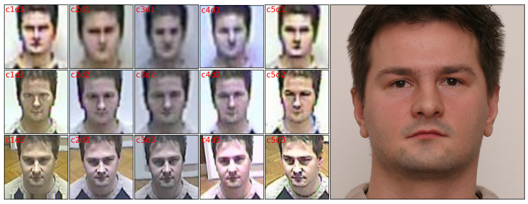


Fig. 3. Sample SCface images. The images on the left show probes captured with different cameras (columns) and at different distances of the subject to the camera (rows). The right image shows a HR gallery image. The labels $cMdN$ denote images recorded with camera M at distance N.

$130 \times 5 \times 3 = 1950$ probe set images. The average size of the facial area used for the experiments recorded at distance $d1$ is 21×21, at distance $d2$ is 36×36, and at distance $d3$ is 62×62 pixels. The evaluation protocol of the SCface dataset defines a series of identification experiments, where the goal is to identify the subjects in the probe set using the HR images in the gallery set. For the evaluation, we pre-process all images of the dataset by cropping the facial area using the provided eye, nose and mouth coordinates. We then upsample the probe images from the database using various super-resolution algorithms at a $2\times$ magnification setting, and perform identification accuracy tests against the gallery as proposed by the authors of the dataset. We report the recognition performance in the form of the identification accuracy (or rank one recognition rate). We report the results separately for each category of probe images, which are grouped in the dataset by camera and distance of subject to camera (see labels at the top of the images in Fig. 3).

IV. Experiments and Results

LFW results: The results of the experiments on the LFW dataset are presented in Fig. 4 in the form of box plots of the verification accuracy (VA) computed over 10 folds. As can be seen, most of the SR methods ensure performance improvements (in terms of median VA) over the baseline bicubic interpolation with the synthetically down-sampled LFW images at the lowest two image resolutions for all tested methods. However, the improvements are increasingly insignificant as the size of the down-sampled image approaches the original image size. Note also that results for image sizes between 48×48 and 192×192 pixels are not included in the figure to keep it uncluttered and because the performance plateaued for image sizes beyond 48×48 pixels. From Fig. 4 we see that for the AlexNet model, NBSRF and VDSR offer the highest improvements in performance at low resolutions. Among the face recognition models we tested, AlexNet is also the least affected by the decrease in image resolution with the verification performance still reaching more than 80% at the smallest image size. This may be explained by the large convolutional filters and strides at the lowest layers of the AlexNet architecture. The GoogleNet results on LFW demonstrate the greatest differences between super-resolution methods, with NBSRF improving the performance by over 10% at the lowest image size and maintaining a significant advantage over the remaining methods even beyond 48×48 pixels. From the SqueezeNet results we see that the performance of this model

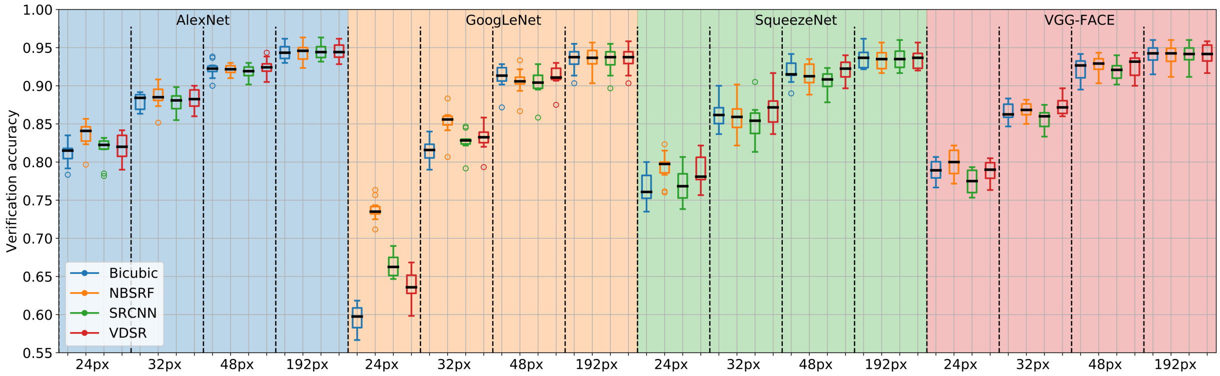
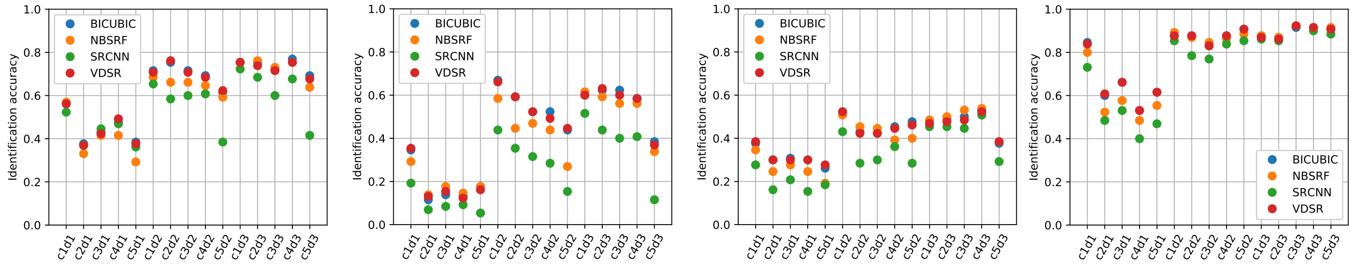


Fig. 4. Results of the experiments on the LFW dataset with synthetically down-sampled images. The box plots show the effect of different super-resolution algorithm on the LFW verification performance at different image sizes, e.g., 24px stands for a size of 24×24 pixels. The results are best viewed in color.



(a) AlexNet (b) GoogLeNet (c) SqueezeNet (d) VGG-FACE

Fig. 5. The results of the experiments on the SCface database. Here, the data points $cMdN$ correspond to the probe image series shot with camera M at distance N. Only VDSR marginally improves on the performance of the bicubic interpolation, SRCNN and NBSRF have mostly a detrimental effect.

is only affected by image SR algorithms up to a relatively small size, since the effect of different SR algorithms is negligible past an image size of 48×48 pixels. The VGG-FACE model shows similar behaviour and is also affected by the SR techniques only to a limited extent.

Overall, the common theme we observe with synthetic down-sampling with all tested face recognition models is that, depending on the face recognition network architecture, we see some performance gains from SR algorithms at very low image sizes (or resolutions), but these fade away quickly with an increase of the probe image size.

SCface results: From the SCface results in the Fig. 5 we observe that SR algorithms also affect face recognition performance with real-life LR images. However, here the results suggest that the best performing SR algorithm, i.e., VDSR, is always very close and on average (across all cameras, distances and models) only slightly better than bicubic interpolation. The remaining two SR algorithms have almost always a detrimental effect on performance.

We observe that different cameras result in large changes in performance even at the same distance, which is expected given the different image quality each of the cameras produces - see Fig. 3 for some visual examples. We also see that the performance gains due to image super-resolution are most noticeable at large distances to the camera (i.e., at low image resolution). On top of that, however, we notice that in better-quality images, for example, at the distance 2 and distance 3 series of images, SR methods degrade performance compared to interpolation in several cases, most prominently in the case of GoogLeNet, as seen in Fig. 5 (b). This is in stark contrast to the experiment with artificially degraded LFW images, where interpolation and example-based SR methods tended to converge to the same performance.

Discussion: Several important findings can be made from the presented results: *i)* There is discrepancy between the results obtained on synthetically down-sampled images and real surveillance data. This suggest that better models for approximating the image capturing process at large distances (and consequently low-resolutions) are needed. Given the success of deep learning in various areas recently, the down-sampling transforms could be learned from data as well [45]. *ii)* Super-resolution algorithms aim only at reconstructing plausible HR data that adequately explains the LR input (perceptual improvements), which may not necessarily aid recognition. To make SR algorithms suitable for recognition models, recognition constraints need to be incorporated into the reconstruction step itself, as for example in [46], [47].

V. Conclusion

We have evaluated the effects of different super-resolution algorithms with several face recognition models in cross-resolution recognition experiments, and found them to improve performance over interpolation on synthetically down-sampled images of very low-resolution (24×24 and 32×32 pixels). On real-life low-resolution images the effect of super-resolution was limited. This makes a strong case for continued research into the interaction between the fields of face recognition and image super-resolution. Our future work will, therefore, be focused on the joint training of super-resolution and face recognition models, as well as face-specific super-resolution algorithms, i.e., face hallucination.

REFERENCES

[1] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "DeepFace: Closing the gap to human-level performance in face verification," in *CVPR*, 2014, pp. 1701–1708.

[2] O. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," in *BMVC*, 2015, vol. 1, p. 6.

[3] J. Phillips and A. O'Toole, "Comparison of human and computer performance across face recognition experiments," *Image and Vision Computing*, vol. 32, no. 1, pp. 74–85, 2014.

[4] J. Neves and H. Proença, "Icb-rw 2016: International challenge on biometric recognition in the wild," in *ICB*. IEEE, 2016, pp. 1–6.

[5] B. Li, H. Chang, S. Shan, and X. Chen, "Low-resolution face recognition via coupled locality preserving mappings," *IEEE Signal Processing Letters*, vol. 17, no. 1, pp. 20–23, 2010.

[6] W. Zou and P. Yuen, "Very low resolution face recognition problem," *IEEE Transactions on Image Processing*, vol. 21, no. 1, pp. 327–340, 2012.

[7] H. Chang, J. Han, C. Zhong, An Snijders, and J. Mao, "Unsupervised transfer learning via multi-scale convolutional sparse coding for biomedical applications," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017.

[8] Z. Lei, T. Ahonen, M. Pietikäinen, and S. Li, "Local frequency descriptor for low-resolution face recognition," in *FG*, 2011, pp. 161–166.

[9] E. Saad and K. Hirakawa, "Defocus blur-invariant scale-space feature extractions," *IEEE Transactions in Image Processing*, vol. 25, no. 7, pp. 3141–3156, 2016.

[10] S. Lee, J. Park, and S. Lee, "Low resolution face recognition based on support vector data description," *Pattern Recognition*, vol. 39, no. 9, pp. 1809–1812, 2006.

[11] B. Gunturk, A. Batur, Y. Altunbasak, M. Hayes, and R. Mersereau, "Eigenface-domain super-resolution for face recognition," *IEEE Transactions in Image Processing*, vol. 12, no. 5, pp. 597–606, 2003.

[12] L. Chen, R. Hu, Z. Han, Q. Li, and Z. Lu, "Face super resolution based on parent patch prior for VLQ scenarios," *Multimedia Tools and Applications Journal*, vol. 76, no. 7, pp. 10231–10254, 2017.

[13] J. Yang, J. Wright, T. Huang, and Y. Ma, "Image super-resolution as sparse representation of raw image patches," in *CVPR*, 2008, pp. 1–8.

[14] C. Dong, C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 38, no. 2, pp. 295–307, 2016.

[15] J. Kim, K. Jung Lee, and K. Mu Lee, "Accurate image super-resolution using very deep convolutional networks," in *CVPR*, 2016, pp. 1646–1654.

[16] C. Ledig, L. Theis, F. Huszár, et al., "Photo-realistic single image super-resolution using a generative adversarial network," *arXiv preprint 1609.04802*, 2016.

[17] E. Bilgazyev, B. Efraty, S. K. Shah, and I. A. Kakadiaris, "Improved face recognition using super-resolution," in *IJCB*. IEEE, 2011, pp. 1–7.

[18] F. Lin, C. Fookes, V. Chandran, and S. Sridharan, "Super-resolved faces for improved face recognition from surveillance video," *ICB*, pp. 1–10, 2007.

[19] J. Zhao, Y. Mao, Q. Fang, Z. Liang, F. Yang, and S. Zhan, "Heterogeneous face recognition based on super resolution reconstruction by adaptive multi-dictionary learning," in *CCBR*. Springer, 2015, pp. 143–150.

[20] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[21] G. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," Tech. Rep., University of Massachusetts, Amherst, 2007.

[22] M. Grgic, K. Delac, and S. Grgic, "SCface–surveillance cameras face database," *Multimedia Tools and Applications Journal*, vol. 51, no. 3, pp. 863–879, 2011.

[23] J. D. Van Ouwerkerk, "Image super-resolution survey," *Image and vision Computing*, vol. 24, no. 10, pp. 1039–1052, 2006.

[24] J. Tian and K. Ma, "A survey on super-resolution imaging," *Signal, Image and Video Processing*, vol. 5, no. 3, pp. 329–342, 2011.

[25] K. Nasrollahi and T. B. Moeslund, "Super-resolution: a comprehensive survey," *Machine Vision and Applications*, vol. 25, no. 6, pp. 1423–1468, 2014.

[26] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei, and S. Savarese, "Social LSTM: Human trajectory prediction in crowded spaces," in *CVPR*, 2016, pp. 961–971.

[27] J. Kim, K. Jung Lee, and K. Mu Lee, "Deeply-recursive convolutional network for image super-resolution," in *CVPR*, June 2016, pp. 1646–1654.

[28] J. Salvador and E. Perez-Pellitero, "Naive bayes super-resolution forest," *ICCV*, vol. 11-18, no. 1, pp. 325–333, 2015.

[29] X. Yu and F. Porikli, "Hallucinating very low-resolution unaligned and noisy face images by transformative discriminative autoencoders," in *CVPR*, July 2017.

[30] Y. Tai, J. Yang, and X. Liu, "Image super-resolution via deep recursive residual network," in *CVPR*, July 2017.

[31] W. Lai, J. Huang, N. Ahuja, and M. Yang, "Deep laplacian pyramid networks for fast and accurate super-resolution," in *CVPR*, July 2017.

[32] M. Sajjadi, B. Schölkopf, and M. Hirsch, "Enhancenet: Single image super-resolution through automated texture synthesis," in *ICCV*, 2017.

[33] X. Xu, D. Sun, J. Pan, Y. Zhang, H. Pfister, and M. Yang, "Learning to super-resolve blurry face and text images," in *ICCV*, 2017.

[34] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," *CoRR*, vol. abs/1707.02921, 2017.

[35] R. Dahl, M. Norouzi, and J. Shlens, "Pixel recursive super resolution," *CoRR*, vol. abs/1702.00783, 2017.

[36] A. Chakrabarti, A.N. Rajagopalan, and R. Chellappa, "Super-resolution of face images using kernel pca-based prior," *IEEE Transactions on Multimedia*, vol. 9, no. 4, pp. 888–892, 2007.

[37] S. P. Mudunuri and S. Biswas, "Low resolution face recognition across variations in pose and illumination," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 38, no. 5, pp. 1034–1040, 2016.

[38] R. Keys, "Cubic convolution interpolation for digital image processing," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 29, no. 6, pp. 1153–1160, Dec 1981.

[39] R. Timofte, V. De Smet, and L. Van Gool, "A+: Adjusted anchored neighborhood regression for fast super-resolution," in *ACCV*, 2014, pp. 111–126.

[40] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *CVPR*, 2016, pp. 770–778.

[41] A. Krizhevsky, I. Sutskever, and G. Hinton, "Imagenet classification with deep convolutional neural networks," in *NIPS*, 2012, pp. 1097–1105.

[42] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *CVPR*, 2016, pp. 2818–2826.

[43] F. Iandola, S. Han, M. Moskewicz, K. Ashraf, W. Dally, and K. Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and 0.5 mb model size," *arXiv 1602.07360*, 2016.

[44] K. Grm, V. Struc, A. Artiges, M. Caron, and H.K. Ekenel, "Strengths and weaknesses of deep learning models for face recognition against image degradations," *IET Biometrics*, 2017.

[45] T. Köhler, M. Bätz, F. Naderi, A. Kaup, A. K. Maier, and C. Riess, "Benchmarking Super-Resolution Algorithms on Real Data," *ArXiv e-prints*, Sept. 2017.

[46] S. Baker and T. Kanade, "Limits on super-resolution and how to break them," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 24, no. 9, pp. 1167 – 1183, 2002.

[47] S. Baker and T. Kanade, "Limits on super-resolution and how to break them," in *CVPR*, 2000.