Improving Facial Landmark Detection via a Super-Resolution Inception Network

Martin Knoche, Daniel Merget, Gerhard Rigoll

Institute for Human-Machine Communication
Technical University of Munich, Germany

Abstract. Modern convolutional neural networks for facial landmark detection have become increasingly robust against occlusions, lighting conditions and pose variations. With the predictions being close to pixel-accurate in some cases, intuitively, the input resolution should be as high as possible. We verify this intuition by thoroughly analyzing the impact of low image resolution on landmark prediction performance. Indeed, performance degradations are already measurable for faces smaller than 50×50 px. In order to mitigate those degradations, a new super-resolution inception network architecture is developed which outperforms recent super-resolution methods on various data sets. By enhancing low resolution images with our model, we are able to improve upon the state of the art in facial landmark detection.

1 Introduction

In the last couple of years, convolutional neural networks (CNNs) have proven to be very powerful in many applications surrounding image processing, computer vision and pattern recognition. While CNNs already outperform humans in the field of image recognition [7, 19] and face recognition [25], they still struggle with facial landmark detection [6]. This is not too surprising given that humans are very experienced and well-trained in detecting and locating human faces. Facial landmark detection is difficult because of the large variety of parameters that need to be considered, for example, shape, pose, gender, age, race, lighting, (self-)occlusion, and many more.

Many facial landmark data sets have acknowledged this variety with "in-the-wild" images [3, 9, 15]. The train and test images are typically of fairly high resolution. This is sensible since the facial landmarks have to be labeled accurately, which requires some minimum resolution. Real-world data, however, may be captured under far worse conditions. For example, surveillance systems often operate at high compression rates, recording people from relatively far away. In that sense, reality can be "even wilder" than common "in-the-wild" data sets. Wang et al. [28] distinguish two principle approaches of dealing with low resolution: Direct and indirect.

Direct Methods try to find appropriate feature representations in the low resolution space. For example a *direct* approach is to create low resolution images from the high resolution data sets via post-processing. This requires additional effort during training and is not common practice. State-of-the-art landmark detection methods are therefore optimized for medium to high resolution images.

Indirect Methods try to restore high resolution information, for example, via statistical models or super resolution. The key advantage of *indirect* approaches is that they operate independently of the landmark detection algorithm, simplifying training and preserving maximum flexibility. For example, there is much more available training data for super resolution than for landmark detection.

In this work, we pursue an indirect approach to facial landmark detection on low resolution images. Our main contributions are the following:

- We analyze the impact of low resolution images on CNN landmark detection.
- We present a new super-resolution inception (SRINC) architecture with slight improvements over the state of the art in super resolution.
- We demonstrate that CNN landmark detection performance can be improved by applying super resolution to low quality images.
- We show that additional performance can be gained by training the super-resolution network in the same domain as the landmark detection algorithm (i.e., faces).

2 Related Work

Facial Landmark Detection on Low Resolution Images has not been addressed in many publications. While there is a wide range of research for face recognition in the context of low image resolution, facial landmark detection has not gained nearly as much attention. The few works that exist mostly focus on direct methods.

Biswas et al. [2] provide an in-depth analysis of pose regression in low resolution images, using five different landmarks. They transform "the poor quality probe images and the high-quality gallery images in such a manner that the distances between them approximate the distances had the probe images been captured in the same condition". In other words, they use a direct approach in the taxonomy of Wang et al. [28].

An elaborate analysis considering the impact of low resolution in the context of facial landmark detection is provided by Seshadri [22]. In the taxonomy of Wang et al. [28], Seshadri also uses a direct approach by adopting the training set resolution to the test set, but also investigates some cross-resolution effects. In contrast to our approach, Seshadri does not consider convolutional neural networks. Indirect approaches are not considered, either.

Super Resolution Since our proposed super-resolution method is based on CNNs, we focus on related work in the field of deep learning. A thorough overview of classical image super resolution methods is given in [4] and [30].

Dong et al. were the first to implement image super resolution using a convolutional neural network (SRCNN) [5] and were able to beat the fairly recent methods A+ [26], super-resolution forests [21], and transformed self-exemplars [10] concerning PSNR and structural similarity (SSIM) [29]. In the following, we will discuss a number of improved network architectures that have since been proposed [12–14, 27].

Tuzel et al. [27] train a global-local upsampling network specifically suited for super resolution of human faces. They demonstrate superior performance compared to SRCNN and other face-specific methods on face data sets. However, they neither provide their model, nor any comparisons for non-facial images, which makes it overall difficult to compare against their results.

Kim et al. propose a very deep super-resolution (VDSR) architecture [13]. In contrast to SRCNN, VDSR is trained on the residual between the ground truth and the interpolated bicubic result. This residual learning strategy is similar to the recently proposed ResNet approach by He et al. [8] but only affects the very last layer, i.e., there is a shortcut connection from input to output but none in between layers. Since VDSR is still shallow compared to ResNet, the single shortcut is already sufficient to boost the network performance considerably. In another work, Kim et al. use the same residual learning target as in [13], but instead of a linear deep architecture they use a recurrent neural network [14].

Another interesting method was proposed by Johnson et al. [12]. Instead of optimizing their network towards PSNR, they introduce a perceptual loss that accounts for the way the human eye perceives image quality. Although their PSNR is worse than simple bicubic interpolation, the results are optically very appealing and realistic. Apart from the fact that the perceptual quality is hard to compare objectively, using a perceptual loss results in more drastically altered images, especially on small scales. This would likely confuse a landmark detection algorithm relying heavily on small scale features. Therefore, we do not consider perceptual loss for our further analysis.

3 Landmark Detection on Low Resolution Images

In this section we investigate the impact of low resolution images on two state-of-the art landmark detection algorithms: TCDCN [35] and CFSS [36]. We use the iBUG [20] (135 images) and HELEN [16] (330 images) data sets for the evaluation. By empirical analysis we found that there are no significant differences for bounding boxes larger than 100 × 100 px. The images are thus sampled down such that the bounding boxes (provided by [36]) are 100 px in width. Images with bounding boxes narrower than 100 px are discarded. These new test sets will be referred to as iBUG-Norm (90 images) and HELEN-Norm (330 images).

The landmark detection performance is measured using the detection error as used in [23]:

$$\varepsilon = \frac{1}{N d_0} \sum_{n=1}^{N} \sqrt{(x_n - x'_n)^2 + (y_n - y'_n)^2}\,, \tag{1}$$

where (x_n, y_n) and (x'_n, y'_n) represent the ground truth and predicted coordinates for the $N = 68$ landmarks, respectively. The error is normalized in relation to the ocular distance d_0.

For the analysis, the images are scaled down and up again by different factors (×2, ×3, and ×4) via bicubic interpolation. Figure 1 depicts the detection error distributions for TCDCN and CFSS on both test sets. The detection error increases considerably at factors ×3 and ×4, but even for factor ×2 there is a slight degradation. Note that CFSS and TCDCN use an internal resizing to 60×60 px and 250×250 px, respectively, which explains why the effect of ×2 scaling is less significant for CFSS. The results indicate that there is a general margin for improving landmark detection on low resolution images: A super resolution algorithm preceding CFSS could theoretically reduce the error for iBug-Norm at ×4 scaling from 12.0% to under 9.7%, which is a relative improvement of 19.5%.

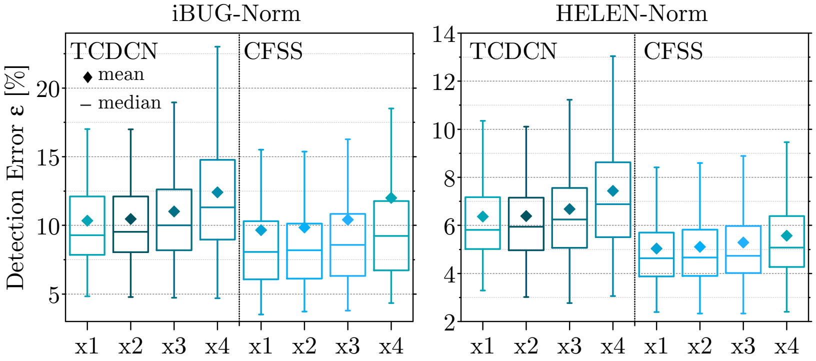

Fig. 1. Landmark detection error (Eq. 1) for TCDCN [35] and CFSS [36] on iBUG-Norm and HELEN-Norm test sets at original scale (×1) and scaling factors ×2 through ×4. The detection error clearly increases for larger scalings, i.e., for lower image quality.

4 Proposed Super-Resolution Network

In order to improve the landmark detection on low resolution images we implement a novel super resolution network, which is described in more detail in the following sections.

4.1 Network Architecture

We follow the general idea of a deep convolution neural network as described by Kim et al. (VDSR) [13] and combine it with inception modules inspired by Szegedy et al. [24] as illustrated in Figure 2. As proposed in [13], the network is trained on the residual by adding the input to the output before calculating the loss. This helps the network to converge much faster. The model was implemented using Microsoft's Cognitive Toolkit (CNTK) [33]. We provide the network configurations and trained models at www.mmk.ei.tum.de/srinc/.

The fact that the network is fully convolutional without any fully-connected layers allows for arbitrary input dimensions and therefore arbitrary input scales. Since the patterns on different scales differ, each scale requires a separate set of filters. The core idea behind using inception modules is to allow the network to combine and select among several filter scales in each layer that account for different object scales.

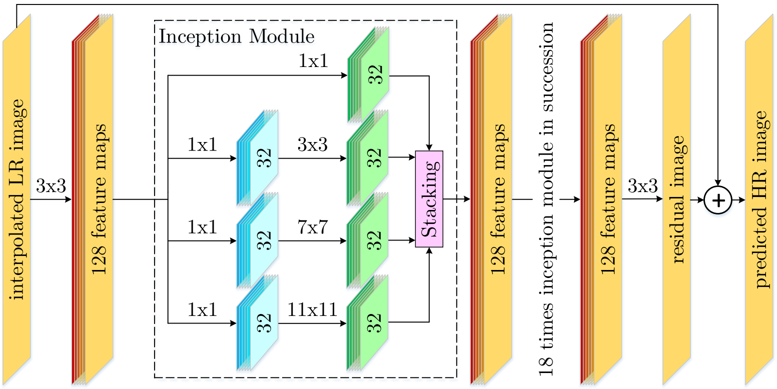

Fig. 2. Proposed super resolution inception (SRINC) architecture: The zero-padded input image is convolved with 128 3 × 3 filters in the first layer. After 19 successive inception modules, a last convolutional layer shrinks the feature maps down to the number of output channels, i.e, in our case one (grayscale) channel. Rectified linear units are used between layers except after the final convolution and addition.

Another benefit lies in the fact that the receptive field of the network is increased. The maximum receptive field of the network is 195 × 195 px via the path of consecutive 11 × 11 convolutions. Nevertheless, we use 51 × 51 px patches for training to limit memory consumption. The patch size obviously does not exploit the full capacity of the network, but we found that it is sufficient for scaling factors up to ×4. This can be explained by two facts: 1) the output of the network depends mostly on local features and 2) the outer parts of the receptive field are supported by far less paths through the network and therefore contain mostly noise. Further tests (not shown) revealed that the patch size has to be increased for scaling factors larger than ×10 because the local regions affecting the pixels are enlarged.

4.2 Training

Our SRINC model is trained on Set291 which is a composition of 91 natural images by Yang et al. [32] and another 200 natural images from the Berkeley Segmentation Dataset [18].

As in Section 3, the training images are sampled down and up, creating a multi-scale training set. Additionally, the data is rotated and flipped, following the same protocol as SRCNN, VDSR and DRCN [5, 13, 14]. Finally the data set is broken down into approximately 1.2 million 51×51 px patches at a stride of 26 px. As described in [13], deeper networks are more likely to fail to converge. For this reason, adjustable gradient clipping is used for training in order to prevent exploding gradients. The clipping threshold per sample is set to 0.01 at a mini-batch size of 32. All weights are initialized according to He et al. [7]. The learning rate is set to 0.0596 and divided by a factor of 3 every 20 epochs. Training 60 epochs on a GTX1080 takes roughly 10 days.

4.3 Results

We benchmark our SRINC model against the state of the art on four widely used test sets for super resolution: Set5 [1], Set14 [34], BSD100 [31] and Urban100 [11]. Table 1 provides a summary of the quantitative evaluation. With only few exceptions, our SRINC model outperforms recent approaches consistently in both PSNR and structural similarity (SSIM) [29]:

$$\mathrm{PSNR} = 10\log_{10}\left(N_{\mathrm{px}} \cdot I_{max}^2 \left(\sum_x \sum_y \Big(I(x,y) - I'(x,y)\Big)^2\right)^{-1}\right) \tag{2}$$

$$\mathrm{SSIM} = \frac{(2\mu_x\mu_y + C)(2\sigma_{xy} + 9C)}{(\mu_x^2 + \mu_y^2 + C)(\sigma_x^2 + \sigma_y^2 + 9C)} \quad \text{with} \quad C = \left(\frac{I_{max}}{100}\right)^2 \tag{3}$$

I_{max} describes the maximum intensity value (i.e., 255 for 8 bits); I and I' are the ground truth and predicted images, respectively; μ_* and σ_* are the means and (co-)variances, respectively.

In order to get a deeper understanding of these results, we conduct a more fine-grained analysis, comparing against VDSR and DRCN as the closest competitors. Instead of the mean PSNR, we take a look at the error distribution. Therefore, we define the cumulative PSNR, considering only errors up to $\delta_{\mathrm{px}} \in [1, I_{max}]$ pixels:

$$\mathrm{PSNR}_{\varSigma} = 10\sum_{n=1}^{\delta_{\mathrm{px}}} \log_{10}\left(N_{\mathrm{px}} \cdot I_{max}^2 \left(\sum_x \sum_y \Big(I(x,y) - I'(x,y)\Big)^2\right)^{-1}\right) \tag{4}$$

$$\forall\, x, y \mid I(x,y) - I'(x,y) = n$$

N_{px} denotes the number of pixel in the image. The differences are emphasized by putting the cumulative PSNR in relation to our SRINC method:

$$\varDelta\mathrm{PSNR}_{\varSigma,<\mathrm{method}>} = \mathrm{PSNR}_{\varSigma,<\mathrm{method}>} - \mathrm{PSNR}_{\varSigma,\mathrm{SRINC}} \tag{5}$$

Table 1. Average PSNR/SSIM on Set5 [1], Set14 [34], BSD100 [31], and Urban100 [11] test sets for scaling factors ×2, ×3 and ×4. The best performance is highlighted in bold.

Method		Bicubic		SRCNN [5]		VDSR [13]		DRCN [14]		SRINC (ours)	
Training Set				Set291		Set291		Set91		Set291	
Data Set	Scale	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Set5	x2	33.66	0.930	36.66	0.954	37.53	**0.959**	**37.63**	**0.959**	37.58	**0.959**
	x3	30.39	0.868	32.75	0.909	33.66	0.921	33.82	**0.923**	**33.92**	**0.923**
	x4	28.42	0.810	30.48	0.863	31.35	0.884	31.53	0.885	**31.55**	**0.886**
Set14	x2	30.24	0.869	32.42	0.906	33.03	**0.912**	33.04	**0.912**	**33.07**	**0.912**
	x3	27.55	0.774	29.28	0.821	29.77	0.831	29.76	0.831	**29.87**	**0.834**
	x4	26.00	0.703	27.49	0.750	28.01	0.767	28.02	0.767	**28.09**	**0.770**
BSD100	x2	29.56	0.843	31.36	0.888	31.90	0.896	31.85	0.894	**31.97**	**0.897**
	x3	27.21	0.738	28.41	0.786	28.82	0.798	28.80	0.796	**28.88**	**0.800**
	x4	25.96	0.668	26.90	0.710	27.29	0.725	27.23	0.723	**27.34**	**0.728**
Urban100	x2	26.88	0.840	29.50	0.895	30.76	0.914	30.75	0.913	**30.89**	**0.915**
	x3	24.46	0.735	26.24	0.799	27.14	0.828	27.15	0.828	**27.29**	**0.832**
	x4	23.14	0.658	24.52	0.722	25.18	0.752	25.14	0.751	**25.31**	**0.758**

The fine-grained results are illustrated in Figure 3. Reflected by the slope from upper left to bottom right, the most evident observation is that both VDSR and DRCN perform better than SRINC for small errors, but worse overall. This behavior persists on all tested data sets and scaling factors. It can be concluded that VDSR and DRCN are more susceptible to generating outlier pixels. In our intuition, outlier robustness is very important for tasks such as facial landmark detection, because those tasks require reliable data down to the pixel level. This should be kept in mind when analyzing the results presented in Section 5.1.

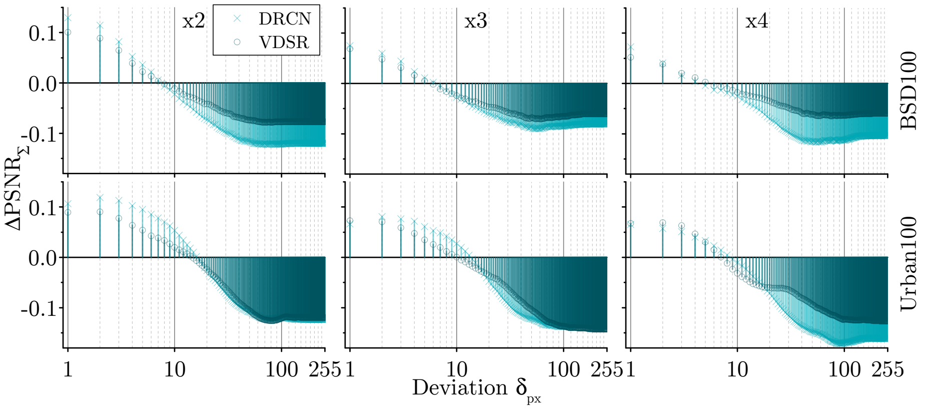

Fig. 3. $\Delta\mathrm{PSNR}_{\Sigma}$ for VDSR [13] and DRCN [14] compared to our approach (SRINC) on test sets BSD100 [31] and Urban100 [11]. Since errors are less likely for large deviations, the abscissa is scaled logarithmically. Both VDSR and DRCN perform clearly better for small deviations, but produce more outliers which dominate the overall PSNR.

5 Super Resolution for Facial Landmark Detection

Putting the theoretical insights from Section 3 into practice, the actual impact of super resolution for facial landmark detection remains to be investigated. For optimal results, it is crucial to choose the training set according to the purpose of the model avoiding domain shift. Hence, we train our SRINC model on a different training set, CelebA [17], containing facial images rather than natural images. Only the first 5k images are used. These are cropped and decomposed into 51×51 px patches at a stride of 26 px. Patches with a bicubic PSNR less than 35.12 for ×3 scaling are discarded for being too blurry, for example, because they contain background. This results in approximately 227k patches total. We refer to this newly trained model as SRINC-F. Except for an additional dropout rate of 10% (all 3×3, 7×7, and 11×11 convolutions), the parameterizations for training SRINC and SRINC-F are identical, see Section 4.2.

5.1 Results

Following the same protocol as in Section 3, we compare VDSR [13] and DRCN [14] with our SRINC and SRINC-F models on the iBUG-Norm and HELEN-Norm data sets, using TCDCN [35] and CFSS [36] for facial landmark detection. In order to highlight the differences between the methods in a more readable fashion, Figure 4 shows the error reduction ($\varepsilon_{\mathrm{SR}} - \varepsilon_{\mathrm{LR}}$) rather than the absolute error ($\varepsilon_{\mathrm{SR}}$, cf. Eq. 1).

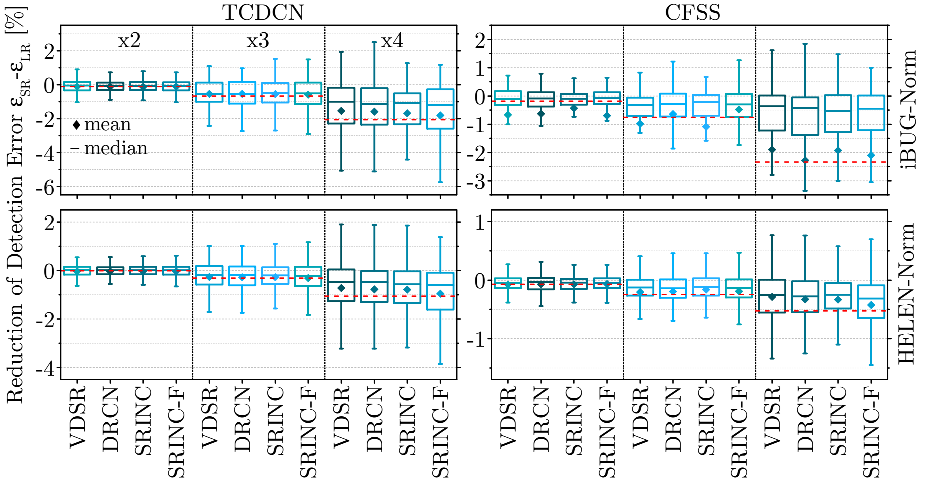

Fig. 4. The landmark detection error reduction (cf. Eq. 1) after applying different super-resolution methods to low resolution images from iBUG-Norm and HELEN-Norm, cf. Section 3. As a reference, the red line indicates the theoretical performance limit, i.e., the mean ground-truth resolution performance according to Figure 1. Best viewed in the digital version.

The results clearly indicate that both TCDCN and CFSS profit from super resolution when the original image resolution is low. The variance is relatively high and no method is strictly dominating. For the iBUG-Norm data set, this can be attributed to the relatively small test set (90 images). The small test set also explains why the super-resolution methods are sometimes able to help perform better than the ground-truth resolution (red line).

The margin for improving landmark detection using super resolution is exploited pretty well by our SRINC-F model. Linking the results to the findings from Section 3, the landmark detection error can be reduced by up to 17.5% (ground-truth resolution 19.5%) relative to the bicubic error, with an average improvement of 13.2% (ground-truth resolution 15.5%) for ×4 scaling. Even for ×2 and ×3 scaling the super-resolved images perform close to the ground-truth resolution with respect to the landmark detection error.

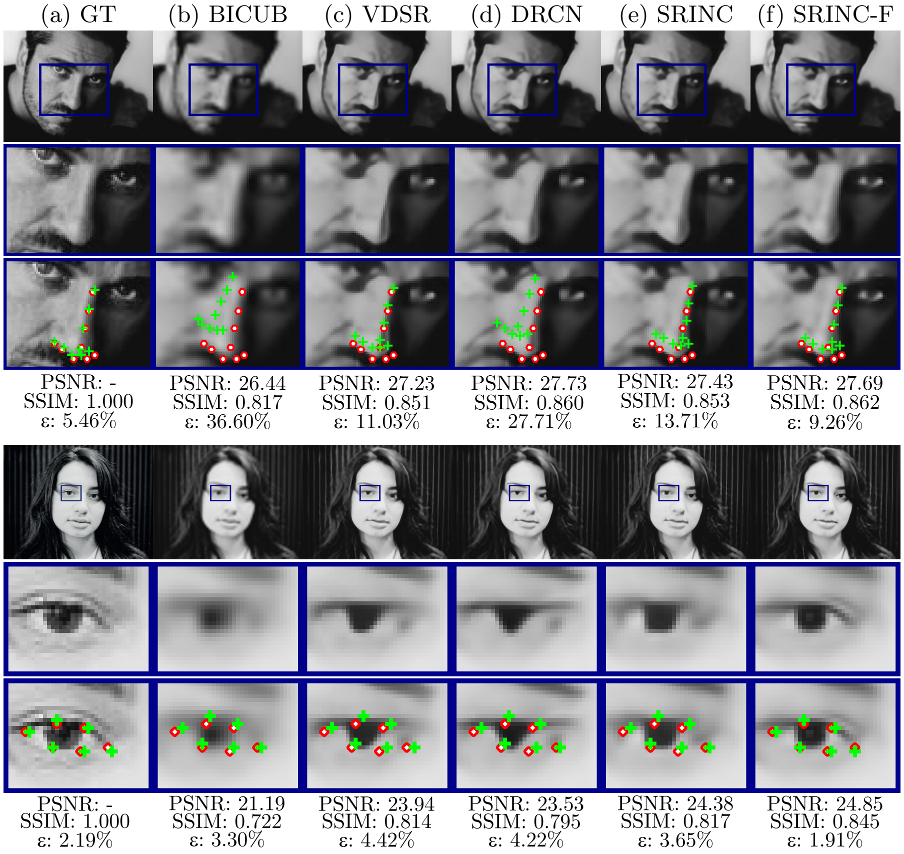
(a) GT (b) BICUB (c) VDSR (d) DRCN (e) SRINC (f) SRINC-F

Fig. 5. Landmark detection examples for CFSS [36] (top) and TCDCN [35] (bottom) on super resolved LR images (×4). Third row also shows ground truth (red dots) and predicted landmarks (green crosses). The provided PSNR and SSIM figures refer to the zoomed patches only. Best viewed in the digital version.

Compared with the other approaches, our SRINC-F model is the most consistent and performs overall best with a clear advantage over SRINC, although the training set is significantly smaller. This underlines that selecting the training data best suited for the problem is of key importance.

Complementary to the quantitative results in Figure 4, Figure 5 depicts two sample images, visualizing the qualitative nature of the different super-resolution methods. Not only do the images look more realistic, but they also explain why landmark detection is positively influenced by SRINC-F. For example, SRINC-F reconstructs the top image with a clearer and more realistic nose contour, which ultimately leads to better landmark predictions.

Landmark detection algorithms are essentially based on pattern matching and are easily confused when the patterns reconstructed by super resolution differ from the expectation. This is most evident in the bottom example of Figure 5. Despite the higher PSNR and SSIM, the standard super resolution approaches are outperformed by a simple bicubic interpolation. This is a hint that PSNR and SSIM alone are no ideal metrics for evaluating the reliability of landmark detection. This correlates well with the findings of Johnson et al. [12] and could be addressed by future research.

While none of the other methods is able to reconstruct the pupil and eye lid correctly, our SRINC-F model predicts a nicely shaped eye. This is mostly explained by the different training sets. Our SRINC-F model blends well with the landmark detection algorithms because it is able to recognize that eyes must be a composite of recurring patterns, for example, a circular pupil.

6 Conclusion

While there is a wide range of research for face recognition considering low image resolution, facial landmark detection has not been thoroughly addressed, yet. Tackling this problem, we first showed that low image resolution degrades facial landmark detection performance, especially for faces smaller than 50 × 50 px, leaving a margin for improvement of up to 19.5%. A new super-resolution inception (SRINC) convolutional neural network architecture was thus presented, beating state-of-the-art super-resolution methods in both PSNR and SSIM. By practical experiments, it was verified that super-resolution indeed helps to improve landmark detection considerably.

Subsequently, in order to achieve the best result possible, the SRINC network was trained on faces rather than natural images. This enables the network to identify recurring patterns such as eyes more accurately and thus enhance the landmark prediction performance even further. Applying our super-resolution network before landmark detection, we are able to improve the average landmark prediction error by up to 17.5% (∅13.2%) which is very close to the ground-truth resolution performance with 19.5% (∅15.5%).

References

1. Bevilacqua, M., Roumy, A., Guillemot, C., Alberi-Morel, M.L.: Low-complexity single-image super-resolution based on nonnegative neighbor embedding (2012)
2. Biswas, S., Aggarwal, G., Flynn, P.J., Bowyer, K.W.: Pose-robust recognition of low-resolution face images. IEEE transactions on pattern analysis and machine intelligence 35(12), 3037–3049 (2013)
3. Burgos-Artizzu, X.P., Perona, P., Dollár, P.: Robust face landmark estimation under occlusion. In: Proc. IEEE Int'l Conf. Computer Vision. pp. 1513–1520 (2013)
4. Chaudhuri, S.: Super-resolution imaging, vol. 632. Springer Science & Business Media (2001)
5. Dong, C., Loy, C.C., He, K., Tang, X.: Image super-resolution using deep convolutional networks. IEEE transactions on pattern analysis and machine intelligence 38(2), 295–307 (2016)
6. Fan, H., Zhou, E.: Approaching human level facial landmark localization by deep learning. Image and Vision Computing 47, 27–35 (2016)
7. He, K., Zhang, X., Ren, S., Sun, J.: Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In: Proc. IEEE international Conf. computer vision. pp. 1026–1034 (2015)
8. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition. pp. 770–778 (2016)
9. Huang, G.B., Ramesh, M., Berg, T., Learned-Miller, E.: Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Tech. rep., Technical Report 07-49, University of Massachusetts, Amherst (2007)
10. Huang, J.B., Singh, A., Ahuja, N.: Single image super-resolution from transformed self-exemplars. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition. pp. 5197–5206 (2015)
11. Huang, J.B., Singh, A., Ahuja, N.: Single image super-resolution from transformed self-exemplars. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition. pp. 5197–5206 (2015)
12. Johnson, J., Alahi, A., Fei-Fei, L.: Perceptual losses for real-time style transfer and super-resolution. In: EU Conf. Computer Vision. pp. 694–711. Springer (2016)
13. Kim, J., Kwon Lee, J., Mu Lee, K.: Accurate image super-resolution using very deep convolutional networks. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition. pp. 1646–1654 (2016)
14. Kim, J., Kwon Lee, J., Mu Lee, K.: Deeply-recursive convolutional network for image super-resolution. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition. pp. 1637–1645 (2016)
15. Köstinger, M., Wohlhart, P., Roth, P.M., Bischof, H.: Annotated facial landmarks in the wild: A large-scale, real-world database for facial landmark localization. In: IEEE Int'l Conf. Computer Vision Workshops (ICCV Workshops). pp. 2144–2151 (2011)
16. Le, V., Brandt, J., Lin, Z., Bourdev, L., Huang, T.S.: Interactive facial feature localization. In: EU Conf. Computer Vision. pp. 679–692. Springer (2012)
17. Liu, Z., Luo, P., Wang, X., Tang, X.: Deep learning face attributes in the wild. In: Proc. Int'l Conf. Computer Vision (ICCV) (2015)
18. Martin, D., Fowlkes, C., Tal, D., Malik, J.: A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: Proc. 8th Int'l Conf. Computer Vision. vol. 2, pp. 416–423 (2001)

19. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., et al.: Imagenet large scale visual recognition challenge. Int'l Journal of Computer Vision 115(3), 211–252 (2015)
20. Sagonas, C., Tzimiropoulos, G., Zafeiriou, S., Pantic, M.: 300 faces in-the-wild challenge: The first facial landmark localization challenge. In: Proc. IEEE Int'l Conf. Computer Vision Workshops. pp. 397–403 (2013)
21. Schulter, S., Leistner, C., Bischof, H.: Fast and accurate image upscaling with super-resolution forests. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition. pp. 3791–3799 (2015)
22. Seshadri, K.T.: Robust Facial Landmark Localization Under Simultaneous Real-World Degradations (2015)
23. Sun, Y., Wang, X., Tang, X.: Deep convolutional network cascade for facial point detection. In: Proc. IEEE Conf. computer vision and pattern recognition. pp. 3476–3483 (2013)
24. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition. pp. 1–9 (2015)
25. Taigman, Y., Yang, M., Ranzato, M., Wolf, L.: Deepface: Closing the gap to human-level performance in face verification. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition. pp. 1701–1708 (2014)
26. Timofte, R., De Smet, V., Van Gool, L.: A+: Adjusted anchored neighborhood regression for fast super-resolution. In: Asian Conf. Computer Vision. pp. 111–126. Springer (2014)
27. Tuzel, O., Taguchi, Y., Hershey, J.R.: Global-local face upsampling network. arXiv preprint arXiv:1603.07235 (2016)
28. Wang, Z., Miao, Z., Wu, Q.J., Wan, Y., Tang, Z.: Low-resolution face recognition: a review. The Visual Computer 30(4), 359–386 (2014)
29. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE transactions on image processing 13(4), 600–612 (2004)
30. Yang, C.Y., Ma, C., Yang, M.H.: Single-image super-resolution: A benchmark. In: European Conf. Computer Vision. pp. 372–386. Springer (2014)
31. Yang, C.Y., Yang, M.H.: Fast direct super-resolution by simple functions. In: Proc. IEEE Int'l Conf. Computer Vision. pp. 561–568 (2013)
32. Yang, J., Wright, J., Huang, T.S., Ma, Y.: Image super-resolution via sparse representation. IEEE transactions on image processing 19(11), 2861–2873 (2010)
33. Yu, D., Eversole, A., Seltzer, M., Yao, K., Kuchaiev, O., Zhang, Y., Seide, F., Huang, Z., Guenter, B., Wang, H., Droppo, J., Zweig, G., Rossbach, C., Gao, J., Stolcke, A., Currey, J., Slaney, M., Chen, G., Agarwal, A., Basoglu, C., Padmilac, M., Kamenev, A., Ivanov, V., Cypher, S., Parthasarathi, H., Mitra, B., Peng, B., Huang, X.: An introduction to computational networks and the computational network toolkit. Tech. rep. (2014)
34. Zeyde, R., Elad, M., Protter, M.: On single image scale-up using sparse-representations. In: Int'l Conf. curves and surfaces. pp. 711–730. Springer (2010)
35. Zhang, Z., Luo, P., Loy, C.C., Tang, X.: Learning deep representation for face alignment with auxiliary attributes. IEEE transactions on pattern analysis and machine intelligence 38(5), 918–930 (2016)
36. Zhu, S., Li, C., Loy, C.C., Tang, X.: Face alignment by coarse-to-fine shape searching. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition. pp. 4998–5006 (2015)