

Low-Resolution Face Recognition

Zhiyi Cheng¹, Xiatian Zhu², and Shaogang Gong¹

¹ School of Electronic Engineering and Computer Science,
Queen Mary University of London, London, UK

{z.cheng,s.gong}@qmul.ac.uk

² Vision Semantics Ltd., London, UK

eddy@visionsemantics.com

Abstract. Whilst recent face-recognition (FR) techniques have made significant progress on recognising constrained high-resolution web images, the same cannot be said on natively unconstrained low-resolution images at large scales. In this work, we examine systematically this under-studied FR problem, and introduce a novel Complement Super-Resolution and Identity (CSRI) joint deep learning method with a unified end-to-end network architecture. We further construct a new large-scale dataset *TinyFace* of native unconstrained low-resolution face images from selected public datasets, because none benchmark of this nature exists in the literature. With extensive experiments we show there is a significant gap between the reported FR performances on popular benchmarks and the results on *TinyFace*, and the advantages of the proposed CSRI over a variety of state-of-the-art FR and super-resolution deep models on solving this largely ignored FR scenario. The *TinyFace* dataset is released publicly at: <https://qmul-tinyface.github.io/>.

Keywords: Face Recognition · Low-Resolution · Super-Resolution.

1 Introduction

Face recognition (FR) models have made significant progress on constrained good-quality images, with reported 99.63% accuracy (1:1 verification) on the LFW benchmark [20] and 99.087% rank-1 rate (1:N identification with 1,000,000 distractors in the gallery) on the MegaFace challenge [22]. Surprisingly, in this work we show systematically that FR remains a significant challenge on *natively unconstrained low-resolution (LR)* images – *not artificially* down-sampled from high-resolution (HR) images, as typically captured in surveillance videos [9,47] and unconstrained (unposed) snapshots from a wide field of view at distance [46,44]. In particular, when tested against native low-resolution face images from a newly constructed tiny face dataset, we reveal that the performances of current state-of-the-art deep learning FR models degrade significantly. This is because the LR facial imagery lack sufficient visual information for current deep models to learn expressive feature representations, as compared to HR, good quality photo images under constrained (posed) viewing conditions (Fig. 1).

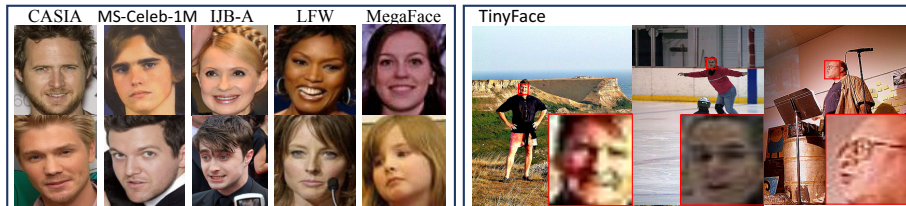


Fig. 1. Examples of (Left) **constrained high-resolution** web face images from five popular benchmarking FR datasets, and (Right) **native unconstrained low-resolution** web face images captured in typical natural scenes.

In general, unconstrained low-resolution FR (LRFR) is severely under-studied versus many FR models tested on popular benchmarks of HR images, mostly captured either under constrained viewing conditions or from “posed” photo-shoots including passport photo verification for airport immigration control and identity check in e-banking. Another obstacle for enabling more studies on LRFR is the lack of large scale *native LR* face image data both for model training and testing, rather than artificially down-sampled synthetic data from HR images. To collect sufficient data for deep learning, it requires to process a large amount of public domain (e.g. from the web) video and image data generated from a wide range of sources such as social-media, e.g. the MegaFace dataset [22,28]. So far, this has only been available for HR and good quality (constrained) web face images, e.g. widely distributed celebrity images [20,30,27].

In this work, we investigate the largely neglected and practically significant LRFR problem. We make three contributions:

1. We propose a novel Super-Resolution and Identity joint learning approach to face recognition in native LR images, with a unified deep network architecture. Unlike most existing FR methods assuming constrained HR facial images in model training and test, the proposed approach is specially designed to improve the model generalisation for LRFR tasks by enhancing the compatibility of face enhancement and recognition. Compared to directly applying super-resolution algorithms to improve image details without jointly optimising for face discrimination, our method has been shown to be effective in reducing the negative effect of noisy fidelity for the LRFR task (Table 5).
2. We introduce a Complement Super-Resolution learning mechanism to overcome the inherent challenge of native LRFR concerning with the absence of HR facial images coupled with native LR faces, typically required for optimising image super-resolution models. This is realised by transferring the super-resolving knowledge from good-quality HR web images to the natively LR facial data subject to the face identity label constraints of native LR faces in every mini-batch training. Taken together with joint learning, we formulate a *Complement Super-Resolution and Identity joint learning (CSRI)* method.

3. We further create a large scale face recognition benchmark, named *TinyFace*, to facilitate the investigation of natively LRFR at large scales (large gallery population sizes) in deep learning. The TinyFace dataset consists of 5,139 labelled facial identities given by 169,403 native LR face images (average 20×16 pixels) designed for 1:N recognition test. All the LR faces in TinyFace are collected from public web data across a large variety of imaging scenarios, captured under uncontrolled viewing conditions in pose, illumination, occlusion and background. Beyond artificially down-sampling HR facial images for LRFR performance test as in previous works, to our best knowledge, this is the first systematic study focusing specially on face recognition of native LR web images.

In the experiments, we benchmark the performance of four state-of-the-art deep learning FR models [30,26,36,42] and three super-resolution methods [11,37,23] on the TinyFace dataset. We observe that the existing deep learning FR models suffer from significant performance degradation when evaluated on the TinyFace challenge. The results also show the superiority of the proposed CSRI model over the state-of-the-art methods on the LRFR tasks.

2 Related Work

Face Recognition. FR has achieved significant progress from hand-crafted feature based methods [1,4,8] to deep learning models [22,24,26,30,42]. One main driving force behind recent advances is the availability of large sized FR benchmarks and datasets. Earlier FR benchmarks are small, consisting of a limited number of identities and images [4,13,14,31,33,35]. Since 2007, the Labeled Faces in the Wild (LFW) [20] has shifted the FR community towards recognising more unconstrained celebrity faces at larger scales. Since then, a number of large FR training datasets and test evaluation benchmarks have been introduced, such as VGGFace [30], CASIA [45], CelebA [27], MS-Celeb-1M [16], MegaFace [22], and MegaFace2 [28]. Benefiting from large scale training data and deep learning techniques, the best FR model has achieved 99.087% on the current largest 1:N face identification evaluation (with 1,000,000 distractors) MegaFace³.

Despite a great stride in FR on the HR web images, little attention has been paid to native LR face images. We found that state-of-the-art deep FR models trained on HR constrained face images do not generalise well to natively unconstrained LR face images (Table 3), but only generalise much better to synthetic LR data (Table 4). In this study, a newly created TinyFace benchmark provides for the first time a large scale native LRFR test for validating current deep learning FR models. TinyFace images were captured from real-world web social-media data. This complements the QMUL-SurvFace benchmark that is characterised by poor quality surveillance facial imagery captured from real-life security cameras deployed at open public spaces [9].

³ <http://megaface.cs.washington.edu/results/facescrub.html>

Low-Resolution Face Recognition. Existing LRFR methods can be summarised into two approaches: (1) Image super-resolution [12,15,18,41,48], and (2) resolution-invariant learning [2,6,10,17,25,34]. The first approach exploits two model optimisation criteria in model formulation: Pixel-level visual fidelity and face identity discrimination [15,41,48,40]. The second approach instead aims to learn resolution-invariant features [2,10,25] or learning a cross-resolution structure transformation [17,34,32,43]. All the existing LRFR methods share a number of limitations: (a) Considering only small gallery search pools (small scale) and/or artificially down-sampled LR face images; (b) Mostly relying on either hand-crafted features or without end-to-end model optimisation in deep learning; (c) Assuming the availability of labelled LR/HR image pairs for model training, which is unavailable in practice with native LR face imagery.

In terms of LRFR deployment, two typical settings exist. One is LR-to-HR which matches LR probe faces against HR gallery images such as passport photos [6,34,7,32]. The other is LR-to-LR where both probe and gallery are LR facial images [40,15,12,48,41]. Generally, LR-to-LR is a less stringent deployment scenario. This is because, real-world imagery data often contain a very large number of “joe public” without HR gallery images enrolled in the FR system. Besides, the two settings share the same challenge of how to synthesise discriminative facial appearance features missing in the original LR input data – one of the key challenges involved in solving the LRFR problem. The introduced TinyFace benchmark adopts the more general LR-to-LR setting.

Image Super-Recognition. Besides, image super-resolution (SR) deep learning techniques [23,11,37] have been significantly developed which may be beneficial for LRFR. At large, FR and SR studies advance independently. We discovered through our experiments that contemporary SR deep learning models bring about very marginal FR performance benefit on native LR unconstrained images, even after trained on large HR web face imagery. To address this problem, we design a novel deep neural network CSRI to improve the FR performance on unconstrained native LR face images.

3 Complement-Super-Resolution and Identity Joint Learning

For native LRFR, we need to extract identity discriminative feature representations from LR unconstrained images. To that end, we propose a deep neural network architecture for **Complement-Super-Resolution and Identity** (CSRI) joint learning. This approach is based on two considerations: (1) Joint learning of Super-Resolution (SR) and FR for maximising their compatibility and complementary advantages; (2) Complement-Super-Resolution learning for maximising the model discrimination on native LR face data at the absence of native HR counterparts in further SR-FR joint learning.

One major challenge in native LRFR is that we have no coupled HR images which are required for optimising the SR component. To address this problem, we consider knowledge transfer from auxiliary HR face data on which LR/HR pairs can be constructed by down-sampling.

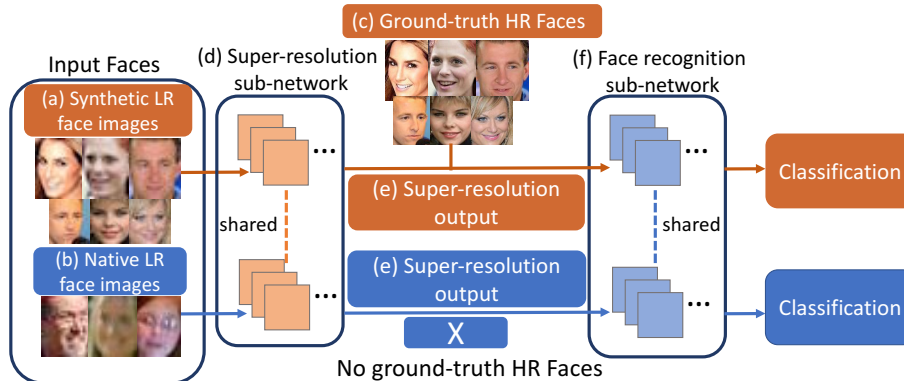


Fig. 2. Overview of the proposed Complement-Super-Resolution and Identity (CSRI) joint learning architecture. The CSRI contains two branches: (Orange): Synthetic LR SR-FR branch; (Blue): Native LR SR-FR branch. The two branches share parameters.

CSRI Overview. Given the CSRI design above, we consider a multi-branch network architecture (Fig. 2). The CSRI contains two branches:

1. A *synthetic LR SR-FR* branch: For improving the compatibility and complementary advantages of SR and FR components by jointly learning auxiliary face data with artificially down-sampled LR/HR pairs (the top stream in Fig. 2);
2. A *native LR SR-FR* branch: For adapting super-resolving information of auxiliary LR/HR face pairs to the native LR facial imagery domain which lacks the corresponding HR faces by complement SR-FR learning (the bottom stream in Fig. 2).

In this study, we instantiate the CSRI by adopting the VDSR [23] for the SR component and the CentreFace [42] for the FR component. We detail these CSRI components as follows.

(I) Joint Learning of Super-Resolution and Face Recognition. To adapt the image SR ability for LRF, we consider a SR-FR joint learning strategy by integrating the output of SR with the input of FR in the CSRI design so to exploit the intrinsic end-to-end deep learning advantage. To train this SR-FR joint network, we use both auxiliary training data with artificially down-sampled LR/HR face pairs $\{(\mathbf{I}^{\text{alr}}, \mathbf{I}^{\text{ahr}})\}$ and face identity labels $\{y\}$ (e.g. CelebA [27]). Formally, a SR model represents a non-linear mapping function between LR and HR face images. For SR component optimisation, we utilise the pixel-level Mean-Squared Error (MSE) minimisation criterion defined as

$$\mathcal{L}_{\text{sr}} = \|\mathbf{I}^{\text{asr}} - \mathbf{I}^{\text{ahr}}\|_2^2 \quad (1)$$

where \mathbf{I}^{asr} denotes the super-resolved face image of \mathbf{I}^{alr} (Fig. 2(a)), and \mathbf{I}^{ahr} denotes the corresponding HR ground-truth image (Fig. 2(c)).

Using the MSE loss intrinsically favours the Peak Signal-to-Noise Ratio (PSNR) measurement, rather than the desired LRFR performance. We address this limitation by concurrently imposing the FR criterion in optimising SR. Formally, we quantify the performance of the FR component by the softmax Cross-Entropy loss function defined as:

$$\mathcal{L}_{\text{fr}}^{\text{syn}} = -\log(p_y) \quad (2)$$

where y is the face identity, and p_y the prediction probability on class y by the FR component. The SR-FR joint learning objective is then formulated as:

$$\mathcal{L}_{\text{sr-fr}} = \mathcal{L}_{\text{fr}}^{\text{syn}} + \lambda_{\text{sr}}\mathcal{L}_{\text{sr}} \quad (3)$$

where λ_{sr} is a weighting parameter for the SR loss quantity. We set $\lambda_{\text{sr}} = 0.003$ by cross-validation in our experiments. In doing so, the FR criterion enforces the SR learning to be identity discriminative simultaneously.

(II) Complement-Super-Resolution Learning. Given the SR-FR joint learning as above, the CSRI model learns to optimise the FR performance on the synthetic (artificially down-sampled) auxiliary LR face data. This model is likely to be sub-optimal for native LRFR due to the inherent visual appearance distribution discrepancy between synthetic and native LR face images (Fig. 6).

To overcome this limitation, we further constrain the SR-FR joint learning towards the native LR data by imposing the native LR face discrimination constraint into the SR component optimisation. Specifically, we jointly optimise the SR and FR components using both auxiliary (with LR/HR pairwise images) and native (with only LR images) training data for adapting the SR component learning towards native LR data. That is, we concurrently optimise the synthetic and native LR branches with the parameters shared in both SR and FR components. To enforce the discrimination of labelled native LR faces, we use the same Cross-Entropy loss formulation.

Overall Loss Function. After combining three complement SR-FR learning loss quantities, we obtain the final CSRI model objective as:

$$\mathcal{L}_{\text{csri}} = (\mathcal{L}_{\text{fr}}^{\text{syn}} + \mathcal{L}_{\text{fr}}^{\text{nat}}) + \lambda_{\text{sr}}\mathcal{L}_{\text{sr}} \quad (4)$$

where $\mathcal{L}_{\text{fr}}^{\text{nat}}$ and $\mathcal{L}_{\text{fr}}^{\text{syn}}$ measure the identity discrimination constraints on the native and synthetic LR training data, respectively. With such a joint multi-task (FR and SR) formulation, the SR optimisation is specifically guided to be more discriminative for the native LR facial imagery data.

Model Training and Deployment. The CSRI can be trained by the standard Stochastic Gradient Descent algorithm in an end-to-end manner. As the auxiliary and native LR data sets are highly imbalanced in size, we further propose to train the CSRI in two steps for improving the model convergence stability: (1) We first pre-train the *synthetic LR SR-FR* branch on a large auxiliary face data (CelebA [27]). (2) We then train the whole CSRI network on both auxiliary and native LR data.

In deployment, we utilise the *native LR SR-FR* branch to extract the feature vectors for face image matching with the Euclidean distance metric.



Fig. 3. Example TinyFace images auto-detected in unconstrained images.

4 TinyFace: Low-Resolution Face Recognition Benchmark

4.1 Dataset Construction

Low-Resolution Criterion. To create a native LR face dataset, we need an explicit LR criterion. As there is no existing standard in the literature, in this study we define LR faces as those $\leq 32 \times 32$ pixels by following the tiny object criterion [38]. Existing FR datasets are all $> 100 \times 100$ pixels (Table 1).

Face Image Collection. The TinyFace dataset contains two parts, face images with *labelled* and *unlabelled* identities. The *labelled* TinyFace images were collected from the publicly available PIPA [46] and MegaFace2 [28] datasets, both of which provide unconstrained social-media web face images with large variety in facial expression/pose and imaging conditions. For the TinyFace to be realistic for LRFR test, we applied the state-of-the-art HR-ResNet101 model [19] for automatic face detection, rather than human cropping. Given the detection results, we removed those faces with spatial extent larger than 32×32 to ensure that all selected faces are of LR.

Face Image Filtering. To make a valid benchmark, it is necessary to remove the false face detections. We verified exhaustively every detection, which took approx. 280 person-hours, i.e. one labeller needs to manually verify detected tiny face images 8 hours/day consistently for a total of 35 days. Utilising multiple labellers introduces additional tasks of extra consistency checking across all the verified data by different labellers. After manual verification, all the remaining PIPA face images were then *labelled* using the identity classes available in the original data. As a result, we assembled 15,975 LR face images *with* 5,139 distinct identity labels, and 153,428 LR faces *without* identity labels. In total, we obtained 169,403 images of labelled and unlabelled faces. Fig. 3 shows some examples randomly selected from TinyFace.

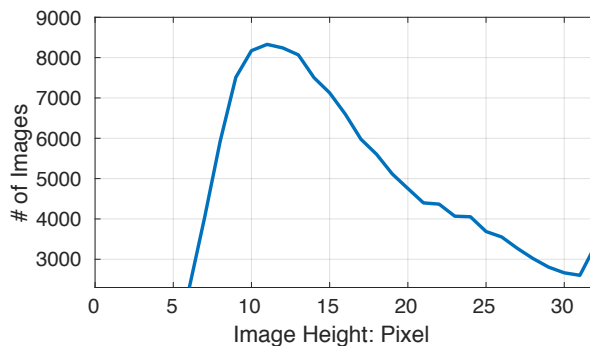


Fig. 4. Distribution of face image height in TinyFace.

Face Image Statistics. Table 1 summarises the face image statistics of TinyFace in comparison to 9 existing FR benchmarks. Fig. 4 shows the distribution of TinyFace height resolution, ranging from 6 to 32 pixels with the average at 20. In comparison, existing benchmarks contain face images of ≥ 100 in average height, a $\geq 5\times$ higher resolution.

Benchmark	Mean Height	# Identity	# Image
LFW [20]	119	5,749	13,233
VGGFace [30]	138	2,622	2.6M
MegaFace [22]	352	530	1M
CASIA [45]	153	10,575	494,414
IJB-A [24]	307	500	5,712
CelebA [27]	212	10,177	202,599
UMDFaces [3]	>100	8,277	367,888
MS-Celeb-1M [16]	>100	99,892	8,456,240
MegaFace2 [28]	252	672,057	4,753,320
TinyFace (Ours)	20	5,139	169,403

Table 1. Statistics of popular FR benchmarks.

Data	All	Training Set	Test Set		
			Probe	Gallery Match	Gallery Distractor
# Identity	5,139	2,570	2,569	2,569	Unknown
# Image	169,403	7,804	3,728	4,443	153,428

Table 2. Data partition and statistics of TinyFace.

4.2 Evaluation Protocol

Data Split. To establish an evaluation protocol on the TinyFace dataset, it is necessary to first define the training and test data partition. Given that both

training and test data require labels with the former for model training and the latter for performance evaluation, we divided the 5,139 known identities into two halves: one (2,570) for training, the other (2,569) for test. All the unlabelled distractor face images are also used as test data (Table 2).

Face Recognition Task. In order to compare model performances on the MegaFace benchmark [28], we adopt the same face identification (1:N matching) protocol as the FR task for the TinyFace. Specifically, the task is to match a given probe face against a gallery set of enrolled face imagery with the best result being that the gallery image of a true-match is ranked at top-1 of the ranking list. For this protocol, we construct a probe and a gallery set from the test data as follows: (1) For each test face class of multiple identity labelled images, we randomly assigned half of the face images to the probe set, and the remaining to the gallery set. (2) We placed all the unlabelled distractor images (with unknown identity) into the gallery set for enlarging the search space therefore presenting a more challenging task, similar to MegaFace [28]. The image and identity statistics of the probe and gallery sets are summarised in Table 2.

Performance Metrics. For FR performance evaluation, we adopt three metrics: the *Cumulative Matching Characteristic* (CMC) curve [24], the *Precision-Recall* (PR) curve [39], and mean Average Precision (mAP). Whilst CMC measures the proportion of test probes with the true match at rank k or better, PR quantifies a trade-off between precision and recall per probe with the aim to find all true matches in the gallery [21]. To summarise the overall performance, we adopt the *mean Average Precision* (mAP), i.e. the mean value of average precision of all per-probe PR curves.

4.3 Training vs Testing Data Size Comparison

To our knowledge, TinyFace is the largest native LR web face recognition benchmark (Table 1). It is a challenging test due to very LR face images ($5\times$ less than other benchmarks) with large variations in illumination, facial pose/expression, and background clutters. These factors represent more realistic real-world low-resolution face images for model robustness and effectiveness test.

In terms of training data size, TinyFace is smaller than some existing HR FR model *training* datasets, notably the MegaFace2 of 672,057 IDs. It is much more difficult to collect *natively* LR face images with label information. Unlike celebrities, there are much less facial images of known identity labels from the general public available for model training.

In terms of testing data size, on the other hand, the face identification *test* evaluation offered by the current largest benchmark MegaFace [22] contains *only 530 test face IDs* (from FaceScrub [29]) and 1 million gallery images, whilst TinyFace benchmark consists of 2,569 test IDs and 154,471 gallery images. Moreover, in comparison to LFW benchmark there are 5,749 face IDs in the LFW designed originally for 1:1 verification test [20], however a much smaller gallery set of 596 face IDs of LFW were adopted for 1:N matching test (open-set) with 10,090 probe images of which 596 true-matches (1-shot per ID) and 9,494 distractors

[5]. Overall, TinyFace for 1:N test data has 3~4× more test IDs than MegaFace and LFW, and 15× more distractors than LFW 1:N test data.

5 Experiments

In this section, we presented experimental analysis on TinyFace, the *only* large scale native LRFR benchmark, by three sets of evaluations: **(1)** Evaluation of generic FR methods *without* considering the LR challenge. We adopted the state-of-the-art deep learning FR methods (Sec. 5.1); **(2)** Evaluation of LRFR methods. For this test, we applied super-resolution deep learning techniques in addition to the deep learning FR models (Sec. 5.2); **(3)** Component analysis of the proposed CSRI method (Sec. 5.3).

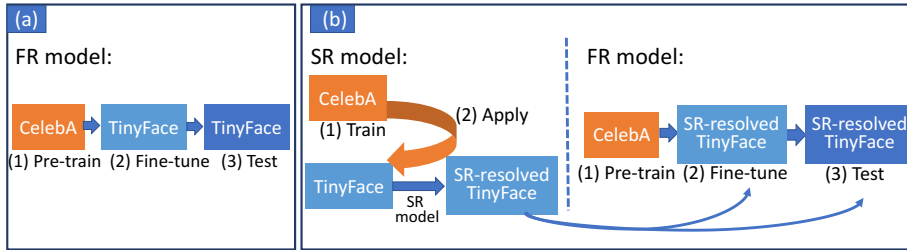


Fig. 5. Overview of training (a) generic FR models and (b) low-resolution FR models (Independent training of Super-Resolution (SR) and FR models).

5.1 Evaluation of Generic Face Resolution Methods

Metric (%)	Rank-1	Rank-20	Rank-50	mAP
DeepID2 [36]	17.4	25.2	28.3	12.1
SphereFace [26]	22.3	35.5	40.5	16.2
VggFace [30]	30.4	40.4	42.7	23.1
CentreFace [42]	32.1	44.5	48.4	24.6

Table 3. *Generic* FR evaluation on TinyFace (Native LR face images).

In this test, we evaluated four representative deep FR models including DeepID2 [36], VggFace [30], CentreFace [42] and SphereFace [26]. For model optimisation, we first trained a given FR model on the CelebA face data [27] before fine-tuning on the TinyFace training set⁴ (see Fig. 5(a)). We adopted the parameter settings suggested by the original authors.

Results. Table 3 shows that the FR performance by any model is significantly inferior on TinyFace than on existing high-resolution FR benchmarks. For example, the best performer CentreFace yields Rank-1 32.1% on TinyFace *versus*

⁴ The SphereFace method fails to converge in fine-tuning on TinyFace even with careful parameter selection. We hence deployed the CelebA-trained SphereFace model.

65.2% on MegaFace [22], i.e. more than half performance drop. This suggests that the FR problem is more challenging on natively unconstrained LR images.

FR Model	Dataset	Rank-1	Rank-20	Rank-50	mAP
VggFace [30]	TinyFace	30.4	40.4	42.7	23.1
	SynLR-MF2	34.8	46.8	49.4	26.0
CentreFace [42]	TinyFace	32.1	44.5	48.4	24.6
	SynLR-MF2	39.2	63.4	70.2	31.4

Table 4. Native (TinyFace) vs. synthetic (SynLR-MF2) LR face recognition.

Native vs Synthetic LR Face Images. For more in-depth understanding on *native* LRFR, we further compared with the FR performance on *synthetic* LR face images. For this purpose, we created a synthetic LR face dataset, which we call *SynLR-MF2*, using 169,403 HR MegaFace2 images [28]. Following the data distribution of TinyFace (Table 2), we randomly selected 15,975 images from 5,139 IDs as the labelled test images and further randomly selected 153,428 images from the remaining IDs as the unlabelled distractors. We down-sampled all selected MegaFace2 images to the average size (20×16) of TinyFace images. To enable a like-for-like comparison, we made a random data partition on SynLR-MF2 same as TinyFace (see Table 2).

Table 4 shows that FR on synthetic LR face images is a less challenging task than that of native LR images, with a Rank-20 model performance advantage of 6.4% (46.8-40.4) by VggFace and 18.9% (63.4-44.5) by CentreFace. This difference is also visually indicated in the comparison of native and synthetic LR face images in a variety of illumination/pose and imaging quality (Fig. 6). This demonstrates the importance of TinyFace as a native LRFR benchmark for testing more realistic real-world FR model performances.

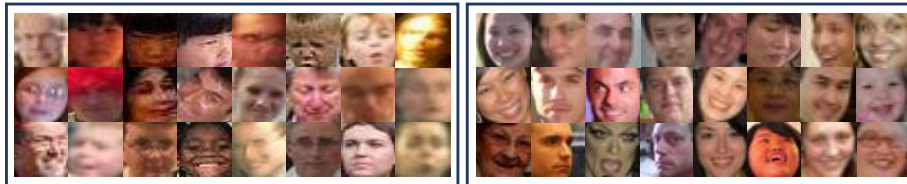


Fig. 6. Comparison of (left) native LR face images from TinyFace and (right) synthetic LR face image from SynLR-MF2.

5.2 Evaluation of Low-Resolution Face Resolution Methods

In this evaluation, we explored the potential of contemporary super-resolution (SR) methods in addressing the LRFR challenge. To compare with the proposed CSRI model, we selected three representative deep learning generic-image SR models (SRCNN [11], VDSR [23] and DRRN [37]), and one LRFR deep model

FR	Method		Rank-1	Rank-20	Rank-50	mAP
CentreFace	No		32.1	44.5	48.4	24.6
	SR	SRCNN [11]	28.8	38.6	42.3	21.7
		VDSR [23]	26.0	34.5	37.7	19.1
		DRRN [37]	29.4	39.4	43.0	22.2
VggFace	No		30.4	40.4	42.7	23.1
	SR	SRCNN [11]	29.6	39.2	41.4	22.7
		VDSR [23]	28.8	38.3	40.3	22.1
		DRRN [37]	29.4	39.8	41.9	22.4
RPCN [41]		18.6	25.3	27.4	12.9	
CSRI (Ours)		44.8	60.4	65.1	36.2	

Table 5. Native *Low-Resolution* FR evaluation on TinyFace.

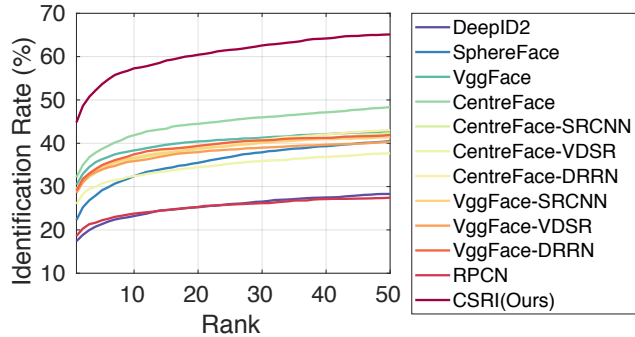


Fig. 7. Performance comparison of different methods in CMC curves on the TinyFace dataset.

RPCN [41] (also using SR). We trained these SR models on the CelebA images [27] (202,599 LR/HR face pairs from 10,177 identities) with the authors suggested parameter settings for maximising their performance in the FR task (see Fig. 5(b)). We adopted the CentreFace and VggFace (top-2 FR models, see Table 3) for performing FR model training and test on super-resolved faces generated by any SR model. Since the RPCN integrates SR with FR in design, we used both CelebA and TinyFace data to train the RPCN for a fair comparison.

Results. Table 5 Fig. 7 show that: (1) All SR methods *degrade* the performance of a deep learning FR model. The possible explanation is that the artifacts and noise introduced in super-resolution are likely to hurt the FR model generalisation (see Fig. 8). This suggests that applying SR as a separate process in a simplistic approach to enhancing LRFR not only does not offer any benefit, but also is more likely a hindrance. (2) The RPCN yields the worst performance although it was specially designed for LR face recognition. The possible reason is two-fold: (a) This method exploits the SR as model pre-training by design, which leads to insufficient FR supervision in the ID label guided model fine-tuning. (b) Adopting a weaker base network with 3 conv layers. These results suggest that existing methods are ineffective for face recognition on natively low-resolution

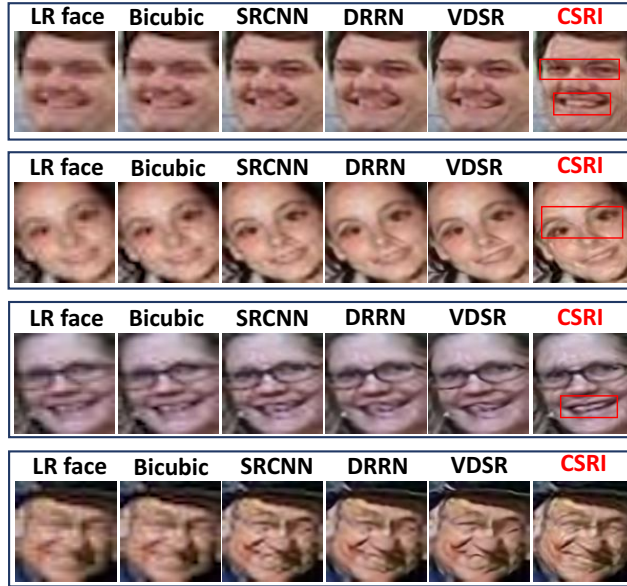


Fig. 8. Examples of super-resolved faces.

images and when the test gallery population size becomes rather large. **(3)** The CSRI outperforms significantly all the competitors, e.g. the Rank-1 recognition performance gain by CSRI over CentreFace is significant at 12.7% (44.8-32.1). This shows the advantage of the CSRI model design in enabling FR on natively LR face images over existing generic FR models.

5.3 Component Analysis of CSRI

To better understand the CSRI’s performance advantage, we evaluated the individual model components on the TinyFace benchmark by incrementally introducing individual components of the CSRI model.

SR-FR joint learning was examined in comparison to SR-FR independent learning (same as in Sec. 5.2). For fair comparison, we used the VDSR [23] and CentreFace [42] which are adopted the components of CSRI. For SR-FR joint learning, we first trained the CSRI *synthetic LR SR-FR* branch on the CelebA data, followed by fine-tuning the FR part on TinyFace training data. Table 6 shows that SR-FR joint learning has a Rank-1 advantage of 10.1% (36.1-26.0) and 4.0% (36.1-32.1) over SR-FR independent learning and FR only (i.e. CentreFace in Table 3), respectively. This suggests the clear benefit of SR-FR joint learning due to the enhanced compatibility of SR and FR components obtained by end-to-end concurrent optimisation.

Complement SR learning was evaluated by comparing the full CSRI with the above SR-FR joint learning. Table 7 shows a Rank-1 boost of 8.7% (44.8-36.1), another significant benefit from the complement SR learning.

SR-FR	Rank-1	Rank-20	Rank-50	mAP
Independent Learning	26.0	34.5	37.7	19.1
Joint Learning	36.1	49.8	54.5	28.2

Table 6. Joint vs. independent learning of super-resolution and face recognition.

CSR	Rank-1	Rank-20	Rank-50	mAP
\times	36.1	49.8	54.5	28.2
\checkmark	44.8	60.4	65.1	36.2

Table 7. Effect of complement super-resolution (CSR) learning.

6 Conclusions

In this work, we presented for the first time a large scale *native* low-resolution face recognition (LRFR) study. This is realised by joint learning of Complement Super-Resolution and face Identity (CSRI) in an end-to-end trainable neural network architecture. By design, the proposed method differs significantly from most existing FR methods that assume high-resolution good quality facial imagery in both model training and testing, whereas ignoring the more challenging tasks in typical unconstrained low-resolution web imagery data. Furthermore, to enable a proper study of LRFR, we introduce a large LRFR benchmark TinyFace. Compared to previous FR datasets that focus on high-resolution face images, TinyFace is uniquely characterised by *natively low-resolution and unconstrained* face images, both for model training and testing. Our experiments show that TinyFace imposes a more challenging test to current deep learning face recognition models. For example, the CentreFace model yields 32.1% Rank-1 on TinyFace *versus* 65.2% on MegaFace, i.e. a 50+% performance degradation. Additionally, we demonstrate that synthetic (artificially down-sampled) LRFR is a relatively easier task than the native counterpart. We further show the performance advantage of the proposed CSRI approach to native LRFR. Extensive comparative evaluations show the superiority of CSRI over a range of state-of-the-art face recognition and super-resolution deep learning methods when tested on the newly introduced TinyFace benchmark. Our more detailed CSRI component analysis provides further insights on the CSRI model design.

Acknowledgement

This work was partially supported by the Royal Society Newton Advanced Fellowship Programme (NA150459), Innovate UK Industrial Challenge Project on Developing and Commercialising Intelligent Video Analytics Solutions for Public Safety (98111-571149), Vision Semantics Ltd, and SeeQuestor Ltd.

References

1. Ahonen, T., Hadid, A., Pietikainen, M.: Face description with local binary patterns: Application to face recognition. *IEEE TPAMI* (2006)
2. Ahonen, T., Rahtu, E., Ojansivu, V., Heikkila, J.: Recognition of blurred faces using local phase quantization. In: *ICPR* (2008)
3. Bansal, A., Nanduri, A., Castillo, C., Ranjan, R., Chellappa, R.: Umdfaces: An annotated face dataset for training deep networks. *arXiv* (2016)
4. Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J.: Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE TPAMI* (1997)
5. Best-Rowden, L., Han, H., Otto, C., Klare, B.F., Jain, A.K.: Unconstrained face recognition: Identifying a person of interest from a media collection. *IEEE Transactions on Information Forensics and Security* (2014)
6. Biswas, S., Bowyer, K.W., Flynn, P.J.: Multidimensional scaling for matching low-resolution facial images. In: *BTAS* (2010)
7. Biswas, S., Bowyer, K.W., Flynn, P.J.: Multidimensional scaling for matching low-resolution face images. *IEEE TPAMI* (2012)
8. Chen, D., Cao, X., Wen, F., Sun, J.: Blessing of dimensionality: High-dimensional feature and its efficient compression for face verification. In: *CVPR* (2013)
9. Cheng, Z., Zhu, X., Gong, S.: Surveillance face recognition challenge. *arXiv preprint arXiv:1804.09691* (2018)
10. Choi, J.Y., Ro, Y.M., Plataniotis, K.N.: Color face recognition for degraded face images. *IEEE TSMC(Part B)* (2009)
11. Dong, C., Loy, C.C., He, K., Tang, X.: Learning a deep convolutional network for image super-resolution. In: *ECCV* (2014)
12. Fookes, C., Lin, F., Chandran, V., Sridharan, S.: Evaluation of image resolution and super-resolution on face recognition performance. *Journal of Visual Communication and Image Representation* (2012)
13. Georghiades, A.S., Belhumeur, P.N., Kriegman, D.J.: From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE TPAMI* (2001)
14. Gross, R., Matthews, I., Cohn, J., Kanade, T., Baker, S.: Multi-pie. *IVC* (2010)
15. Gunturk, B.K., Batur, A.U., Altunbasak, Y., Hayes, M.H., Mersereau, R.M.: Eigenface-domain super-resolution for face recognition. *IEEE TIP* **12**(5) (2003)
16. Guo, Y., Zhang, L., Hu, Y., He, X., Gao, J.: Ms-celeb-1m: A dataset and benchmark for large-scale face recognition. In: *ECCV* (2016)
17. He, X., Cai, D., Yan, S., Zhang, H.J.: Neighborhood preserving embedding. In: *ICCV* (2005)
18. Hennings-Yeomans, P.H., Baker, S., Kumar, B.V.: Simultaneous super-resolution and feature extraction for recognition of low-resolution faces. In: *CVPR* (2008)
19. Hu, P., Ramanan, D.: Finding tiny faces. *arXiv* (2016)
20. 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., University of Massachusetts (2007)
21. Jegou, H., Douze, M., Schmid, C.: Product quantization for nearest neighbor search. *IEEE TPAMI* **33**(1), 117–128 (2011)
22. Kemelmacher-Shlizerman, I., Seitz, S.M., Miller, D., Brossard, E.: The megaface benchmark: 1 million faces for recognition at scale. In: *CVPR* (2016)
23. Kim, J., Kwon Lee, J., Mu Lee, K.: Accurate image super-resolution using very deep convolutional networks. In: *CVPR* (2016)

24. Klare, B.F., Klein, B., Taborsky, E., Blanton, A., Cheney, J., Allen, K., Grother, P., Mah, A., Jain, A.K.: Pushing the frontiers of unconstrained face detection and recognition: Iarpa janus benchmark a. In: CVPR (2015)
25. Lei, Z., Ahonen, T., Pietikäinen, M., Li, S.Z.: Local frequency descriptor for low-resolution face recognition. In: FG (2011)
26. Liu, W., Wen, Y., Yu, Z., Li, M., Raj, B., Song, L.: Spheraface: Deep hypersphere embedding for face recognition. arXiv (2017)
27. Liu, Z., Luo, P., Wang, X., Tang, X.: Deep learning face attributes in the wild. In: ICCV (2015)
28. Nech, A., Kemelmacher-Shlizerman, I.: Level playing field for million scale face recognition. arXiv preprint arXiv:1705.00393 (2017)
29. Ng, H.W., Winkler, S.: A data-driven approach to cleaning large face datasets. In: ICIP (2014)
30. Parkhi, O.M., Vedaldi, A., Zisserman, A.: Deep face recognition. In: BMVC (2015)
31. Phillips, P.J., Scruggs, W.T., O'Toole, A.J., Flynn, P.J., Bowyer, K.W., Schott, C.L., Sharpe, M.: Fvt 2006 and ice 2006 large-scale experimental results. IEEE TPAMI **32**(5), 831–846 (2010)
32. Ren, C.X., Dai, D.Q., Yan, H.: Coupled kernel embedding for low-resolution face image recognition. IEEE TIP **21**(8), 3770–3783 (2012)
33. Samaria, F.S., Harter, A.C.: Parameterisation of a stochastic model for human face identification. In: IEEE Workshop on Applications of Computer Vision (1994)
34. Shekhar, S., Patel, V.M., Chellappa, R.: Synthesis-based recognition of low resolution faces. In: IJCB (2011)
35. Sim, T., Baker, S., Bsat, M.: The cmu pose, illumination, and expression (pie) database. In: FG. pp. 53–58 (2002)
36. Sun, Y., Chen, Y., Wang, X., Tang, X.: Deep learning face representation by joint identification-verification. In: NIPS (2014)
37. Tai, Y., Yang, J., Liu, X.: Image super-resolution via deep recursive residual network. In: CVPR (2017)
38. Torralba, A., Fergus, R., Freeman, W.T.: 80 million tiny images: A large data set for nonparametric object and scene recognition. IEEE TPAMI **30**(11) (2008)
39. Wang, D., Otto, C., Jain, A.K.: Face search at scale. IEEE TPAMI (2016)
40. Wang, X., Tang, X.: Face hallucination and recognition. In: International Conference on Audio-and Video-Based Biometric Person Authentication (2003)
41. Wang, Z., Chang, S., Yang, Y., Liu, D., Huang, T.S.: Studying very low resolution recognition using deep networks. In: CVPR (2016)
42. Wen, Y., Zhang, K., Li, Z., Qiao, Y.: A discriminative feature learning approach for deep face recognition. In: ECCV. Springer (2016)
43. Wong, Y., Sanderson, C., Mau, S., Lovell, B.C.: Dynamic amelioration of resolution mismatches for local feature based identity inference. In: ICPR (2010)
44. Yang, S., Luo, P., Loy, C.C., Tang, X.: Wider face: A face detection benchmark. In: CVPR (2016)
45. Yi, D., Lei, Z., Liao, S., Li, S.Z.: Learning face representation from scratch. arXiv (2014)
46. Zhang, N., Paluri, M., Taigman, Y., Fergus, R., Bourdev, L.: Beyond frontal faces: Improving person recognition using multiple cues. In: CVPR (2015)
47. Zhu, X.: Semantic Structure Discovery in Surveillance Videos. Ph.D. thesis, Queen Mary University of London (2016)
48. Zou, W.W., Yuen, P.C.: Very low resolution face recognition problem. IEEE TIP **21**(1), 327–340 (2012)