Susceptibility to Image Resolution in Face Recognition and Training Strategies to Enhance Robustness

Martin Knoche
Technische Universität München, Arcisstrasse 21 80333 München, Deutschland

Stefan Hörmann
Technische Universität München, Arcisstrasse 21 80333 München, Deutschland

Gerhard Rigoll
Technische Universität München, Arcisstrasse 21 80333 München, Deutschland

Abstract

Many face recognition approaches expect the input images to have similar image resolution. However, in real-world applications, the image resolution varies due to different image capture mechanisms or sources, affecting the performance of face recognition systems. This work first analyzes the image resolution susceptibility of modern face recognition. Face verification on the very popular LFW dataset drops from 99.23% accuracy to almost 55% when image dimensions of both images are reduced to arguable very poor resolution. With cross-resolution image pairs (one HR and one LR image), face verification accuracy is even worse. This characteristic is investigated more in-depth by analyzing the feature distances utilized for face verification. To increase the robustness, we propose two training strategies applied to a state-of-the-art face recognition model: 1) Training with 50% low resolution images within each batch and 2) using the cosine distance loss between high and low resolution features in a siamese network structure. Both methods significantly boost face verification accuracy for matching training and testing image resolutions. Training a network with different resolutions simultaneously instead of adding only one specific low resolution showed improvements across all resolutions and made a single model applicable to unknown resolutions. However, models trained for one particular low resolution perform better when using the exact resolution for testing. We improve the face verification accuracy from 96.86% to 97.72% on the popular LFW database with uniformly distributed image dimensions between 112 × 112 px and 5 × 5 px. Our approaches improve face verification accuracy even more from 77.56% to 87.17% for distributions focusing on lower images resolutions. Lastly, we propose specific image dimension sets focusing on high, mid, and low resolution for five well-known datasets to benchmark face verification accuracy in cross-resolution scenarios.

2012 ACM Subject Classification Computing methodologies → Neural networks
Keywords and Phrases recognition, resolution, cross, face, identification
Digital Object Identifier 10.4230/LITES.8.1.1
Supplementary Material *Dataset (Evaluation Protocols)*: `https://github.com/martlgap/btm-stm`
Funding This work was partially supported by “Bayerische Staatsministerium für Wirtschaft, Energie und Technologie” within the framework of a funding program of “Informations- und Kommunikationstechnik” for the project “Grundrechtskonforme Gesichtserkennung im öffentlichen Raum” (e-freedom).
Received 2020-12-15 **Accepted** 2022-01-25 **Published** 2022-11-16
Editor Samarjit Chakraborty and Qing Rao
Special Issue Special Issue on Embedded Systems for Computer Vision

1 Introduction

Over the last few years, face recognition has gained progressively more attraction. Szegedy et al. [24] introduced one of the first deep-learning-based approaches in 2014 and applied a 9-layer convolutional neural network. Since then, deep-learning-based approaches have evolved more and more due to the growing availability of powerful GPUs and novel large datasets, e.g., Microsoft’s

Leibniz Transactions on Embedded Systems, Vol. 8, Issue 1, Article No. 1, pp. 01:1–01:20
LITES Leibniz Transactions on Embedded Systems
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

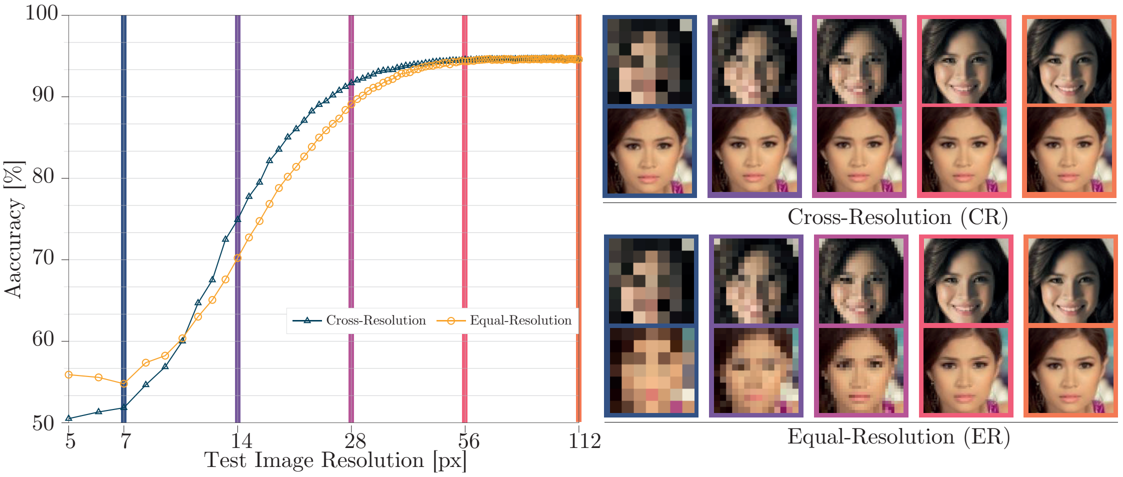

Figure 1 Average face verification accuracy across five popular datasets (LFW [9], AgeDB [17], CFP-FP [22], CALFW [36] and CPLFW [35]) for cross-resolution and equal-resolution (left). One example image pair for both scenarios in five image resolutions (right).

Celeb Dataset (MS1M) [6] with up to 87k identities. These networks are trained to map a facial image, typically after head pose normalization, into a feature space, in which intra-class features distances are minimized, and inter-class feature distances are maximized.

In Figure 1, we show that image accuracy drops for lower image resolution. Hence, we argue that the learned features depend on the training image resolution. Popular approaches learn a projection into a distinct feature space with datasets containing mainly high resolution (HR) images. However, in real-world applications, the image quality is often inferior. Besides external factors like illumination or the subject's distance to the camera, sensor characteristics or image compression affect the image quality. For example, surveillance cameras capture faces at very low resolutions, in contrast to very high-quality mug-shots-like passport images. Another example is social media, which tries to recognize HR faces in the foreground and tiny low resolution (LR) faces in the background. In this work we focus on the most important characteristic of image quality - the image resolution.

LR face recognition [13, 2, 1] addresses the verification and identification of faces on images with the same coarse resolution. However, in real-world scenarios, the image resolution is arbitrary and unknown. Cross-Resolution (CR) face recognition addresses this problem of comparing images with varying resolutions, but has yet found minor attraction by the research community.

In this work, we first investigate the verification performance of a state-of-the-art face recognition network [3] on different image resolutions. We differentiate between CR and LR verification scenarios in our analysis. Figure 1 demonstrates that the performance is significantly worse in CR and LR scenarios across several datasets. At resolutions below $10 \times 10\,\text{px}$, the accuracy is slightly above 50%, which is only barely above guessing. Therefore, we assume a possibility for improvements, especially for very low image resolutions.

A major drawback of the works [5, 18] in CR face recognition is their focus on one specific image resolution, which assumes the image resolution to be known. Moreover, one needs several models to face a wide range of image resolutions, which are likely to occur in real-world applications. Zeng et al. [32] use a mix of two/four different image resolutions during training.

Our work distinguishes between two-resolution (i.e., training a network specifically with images in high resolution and one particular low resolution) and multi-resolution (i.e., feeding the model with HR and multiple LR images) training.

In summary, our main contributions are:

- We analyze the susceptibility for different image resolution on face verification in-depth.
- We propose two intuitive, straightforward approaches and show performance improvements on several datasets for CR face verification, especially at very low image resolutions.
- Lastly, we propose and publish three evaluation protocols to measure face recognition robustness against CR images. That is, to the best of our knowledge, the first benchmark for CR.

2 Related Work

2.1 Generic Face Recognition

In recent years, face recognition research has focused on loss functions applied mainly on ResNet [7] architectures. In [14], the authors propose an angular softmax loss with a multiplicative angular margin and in [27] an additive cosine margin. Deng et al. [3] applied an additive angular margin loss function, which can effectively extend the discriminating power of features. Recently, Kim et al. [12] presented with GroupFace a novel architecture that utilizes multiple group-aware representations to improve the quality of the features. Wang et al. [28] proposed a hierarchical pyramid diverse attention network. Schroff et al. [21] introduced the triplet loss to maximize the distances between an anchor image and its genuine sample (same identity) while minimizing the distance between an anchor image and its imposter sample (different identity).

2.2 Image Resolutions

To the best of our knowledge, no large training dataset provides different resolution versions of the same facial image. Furthermore, large datasets are often crawled from the web, and thus they lack very LR images on which the identity is unrecognizable. However, such a dataset is crucial to train a network, which is robust against varying image resolutions. The generation of LR images from HR images is an essential component in super-resolution. According to Zhou and Süsstrunk [37], a mapping from LR to HR images is often learned by synthetically downsampled HR images to retrieve target-oriented training data. They further state that the frequently used bicubic interpolation [10] significantly differs from real-world camera-blur and is not optimal. Nevertheless, simple bicubic downsampling is a cheap, reproducible, and effective way to lower the image resolution.

2.3 Cross-Resolution Face Recognition

According to [23], existing CR approaches can be grouped into two areas: 1) Transformation-based methods [34, 4, 19, 8] aim to transform images/features from low resolution to high resolution or vice versa to project them in a common space. 2) Non-transformation-based approaches [16, 32] intend to extract scale-invariant features directly into a common feature space. Wang et al. [29] show an exhaustive review of those methods for addressing CR face recognition, and Figure 2 gives a brief functional overview of those two methods.

Transformation-based Methods

Lu et al. [15] presented a deep coupled ResNet model containing one trunk network and a two-branch network. The trunk network extracts features, whereas the two-branch network transforms HR and the corresponding LR features into a space in which their difference can be minimized.

Zangeneh et al. [31] proposed a two-branch deep convolutional neural network. While the LR branch consists of a super-resolution network combined with a feature extraction network, the HR branch is only a feature extraction network. Both branches are trained in three different

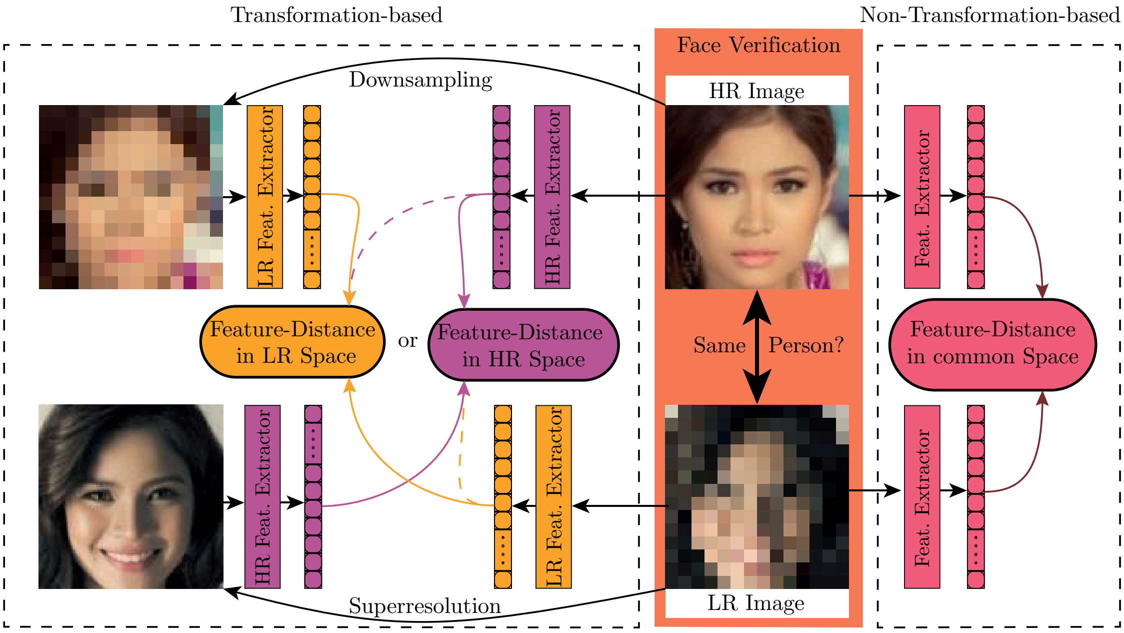

Figure 2 Transformation-based approaches (left) transform either learned image features (dashed path) or images into a shared space (solid path). Non-Transformation-based (right) methods aim to project scale-invariant image features directly.

training phases. In the benchmark, images are assigned to a particular branch depending on their resolution. A similar approach was used in [11]. They trained a U-Net with a combination of reconstruction and identity preserving loss to super-resolve multi-scale LR images. For feature extraction, they utilized a pretrained Inception-ResNet.

The authors of [25] proposed a coupled GAN network structure, which comprises two sub-nets, one for high resolution and one for low resolution. The correlation between the sub-net-generated features is maximized. Moreover, they considered facial attributes by implicitly matching facial details for both resolutions.

Non-Transformation-based Methods

In [32], Zeng et al. presented a resolution-invariant deep network and trained it directly with unified LR and HR images. However, they used only resolutions in the range of 24 to 60 pixels for LR images.

Massoli et al. [16] proposed a student-teacher network approach. They showed that their approach can be more effective rather than preprocessing images with super-resolution techniques.

The authors of [18] report that their deep CNN architecture can address the problem of CR face recognition. They present a two-branch network architecture, which is trained in a pair-wise manner with multiple classification and contrastive loss functions.

In [5], Ge et al. focused on low computational costs in LR face recognition. Therefore, they introduced a new learning approach via selective knowledge distillation. A two-stream technique (large teacher model and a light-weight student model) is employed to transfer selected knowledge from the teacher model to the student model.

3 Experimental Setup

3.1 Baseline Network

As our baseline network, we choose a network structure comprising a modified ResNet-50 [7] as proposed in ArcFace [3], pretrained on ImageNet [20], and an ArcFace layer for classification.

The backbone network (ResNet-50) consists of a set of stacked residual blocks, which are repeated four times and contains in total 50 convolutional layers. It squeezes the input image from 112 × 112 × 3 px down to 4 × 4 × 2048 px utilizing multiple convolutions. After flattening the output from the backbone network, dropout is added. A bottleneck layer (512-dimensional fully connected layer), which represents the extracted features is added following [30, 14, 27]. Finally, a fully connected layer with 87 k (number of identities in our training set) neurons is added. We then apply Additive Angular Margin Loss to the network following [3].

For training, we select the cleaned Microsoft MS1M [6] dataset containing about 5.8M images from about 87k identities. We perform random brightness and saturation variations, left-right flipping, and random cropping of images as data augmentation. All training parameters are set according to [3] except for a smaller batch-size of 128 due to hardware limitations. The learning rate is set to 0.01 and is decreased by a factor of 10 after epoch 9 and epoch 13. In total, we train for 16 epochs with momentum SGD optimizer. The dropout rate and weight decay are set to 0.5 and $5 \cdot 10^{-4}$, respectively.

3.2 Testing Datasets

We select five popular dataset (cf. Table 1) for evaluating face verification performance.

Table 1 Statistics for five popular test datasets.

	LFW [9]	AgeDB [17]	CFP-FP [22]	CALFW [36]	CPLFW [35]
Identities	5749	568	500	5749	5749
Images	13233	16488	7000	13233	13233
Pairs	6000	6000	7000	6000	6000

We use the aligned face, which is cropped to 112 × 112 × 3 px afterwards for all testing datasets mentioned in Table 1. In this paper, we exclusively deal with images having equal width and height. For the sake of simplicity, we denote the image resolution by naming only the first dimension, i.e., a resolution of 112 px defines a 112 × 112 × 3 px image.

3.3 Reduction of Image Resolution

The baseline network requires HR input images $\mathbf{I}_{\mathrm{HR}}$ of the dimension 112 px. We simulate a resolution reduction by performing the following two steps: 1) Downsample $\mathrm{F}_{\mathrm{down},r}(\cdot)$ images to an image dimension r in pixels followed by 2) upsampling $\mathrm{F}_{\mathrm{up},r}(\cdot)$ the images back to the original image dimension and denote the resulting LR images as $\mathbf{I}_{\mathrm{LR}}$. The complete process can be formulated as follows:

$$\mathbf{I}_{\mathrm{LR}} = \mathrm{F}_{\mathrm{up},112}(\mathrm{F}_{\mathrm{down},r}(\mathbf{I}_{\mathrm{HR}})) \tag{1}$$

For both sampling processes, bicubic interpolation [10] is applied. To reduce unwanted artifacts, typically introduced by the downsampling process, standard anti-aliasing techniques are employed. In subsection 4.1, we further investigate these effects.

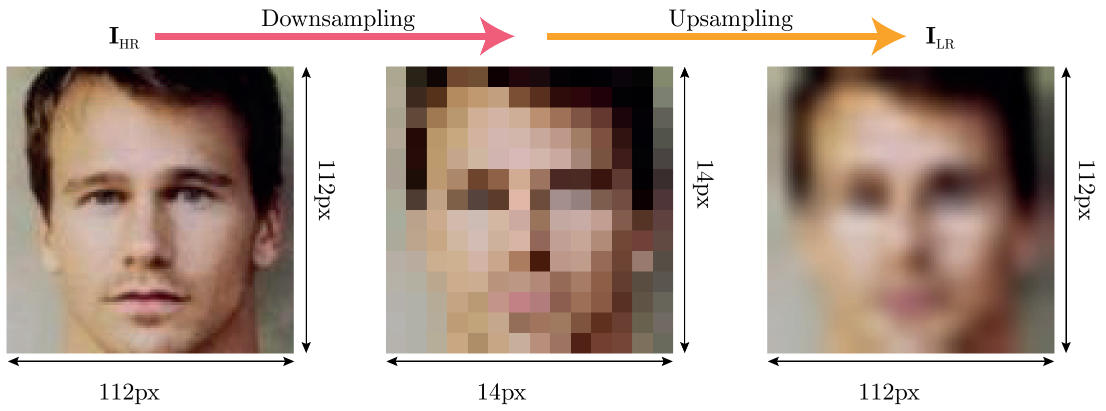

Figure 3 Bicubic down- and up-sampling process to reduce the image resolution but keeping the image dimension.

Figure 3 illustrates the synthetic image resolution reduction. The left image $\mathbf{I}_{\text{HR}}$ is a sample taken from the MS1M dataset with a resolution $r = 112$. In the center, the downsampled image $F_{\text{down},14}(\mathbf{I}_{\text{HR}})$ with image dimension $r = 14$ is depicted. Finally, the upsampled image $\mathbf{I}_{LR}$ is shown on the right and has qualitatively considered an image resolution of $r = 14$ but technically the same image dimension as the $\mathbf{I}_{\text{HR}}$ image. It is evident that all the high-frequency information is removed by this synthetically resolution reduction. Simultaneously, the image dimension is equal to the original image, which is the required image dimension for our networks.

3.4 Accuracy in Face Verification

We report accuracy in all experiments, which denotes the face recognition rate in terms of face verification. To calculate the accuracy value for a given dataset, we first take the cosine distances d between features of every image pair $(\mathbf{I}_1, \mathbf{I}_2)$ extracted from a model $\mathrm{M}(\cdot)$ according to N image pairs defined in the specific evaluation protocol for each dataset. respectively:

$$d = 1 - \frac{\mathrm{M}(\mathbf{I}_1) \cdot \mathrm{M}(\mathbf{I}_2)}{\|\mathrm{M}(\mathbf{I}_1)\|^2 \, \|\mathrm{M}(\mathbf{I}_2)\|^2} \tag{2}$$

Then, we use 10-fold cross-validation to find optimal thresholds that can separate feature distances of genuine from imposter pairs. The number of correctly identified genuine and imposter samples from a total number of samples N are then named as true positives TP and true negatives TN, respectively. We then calculate an accuracy score Acc as follows:

$$Acc = \frac{TP + TN}{N} \tag{3}$$

For all experiments in the CR scenario we generate two evaluation protocols by flipping the pairwise matching resolution from

$$\Big(\mathrm{M}(\mathrm{F}_{\text{up},112}(\mathrm{F}_{\text{down},r}(\mathbf{I}_1))), \mathrm{M}(\mathbf{I}_2)\Big)$$

to

$$\Big(\mathrm{M}(\mathbf{I}_1), \mathrm{M}(\mathrm{F}_{\text{up},112}(\mathrm{F}_{\text{down},r}(\mathbf{I}_2))\Big)$$

We then calculate the accuracy score for both test datasets and then compute the mean.

4 Analysis of Image Resolution Susceptibility

In this section, we first investigate the effect by reducing the resolution across five test datasets. Then, we examine the performance of the baseline network under LR conditions in CR and ER scenarios. Afterward, we take a closer look at the extracted features, especially at the cosine distance between the image pairs, which is used to classify them as genuine or imposter.

4.1 Resolution Reduction on several Datasets

To better understand what happens when performing resolution reduction synthetically, we analyze the effect of downsampling on several testing datasets. Hence, we calculate a mean image across the whole dataset and highlight the mean pixel difference between LR and HR images. The mean images in Figure 4 are computed as follows:

$$\mathbf{I}_{\mathrm{HR}}^{\mathrm{mean}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{I}_{\mathrm{HR},i} \tag{4}$$

where N denotes the number of elements of the dataset.

Below each mean HR image, we denote the mean absolute pixel differences D_r between synthetically reduced images $I_{LR,r}$, and original I_{HR} images across each dataset. We retrieve those images for four resolutions $r \in \{7, 14, 28, 56\}$ according to:

$$\mathbf{D}_{r}^{\mathrm{mean}} = \frac{1}{N}\sum_{i=1}^{N}\left(\left|\mathrm{F}_{\mathrm{up},112}(\mathrm{F}_{\mathrm{down},r}(\mathbf{I}_{\mathrm{HR},i})) - \mathbf{I}_{\mathrm{HR},i}\right|\right) \tag{5}$$

As expected, the resolution reduction process in all datasets is heavily affected by eye, nose, and mouth regions. High detail information in those regions is lost. This reasonably results in worse face verification performance as we show later in the next section. The maximum derivation of a single LR image pixel concerning its corresponding pixel in the HR image is about 50%. In all pixel-difference images grid-style artifacts occur, which in our opinion result from the anti-aliasing method of the bicubic interpolation algorithm.

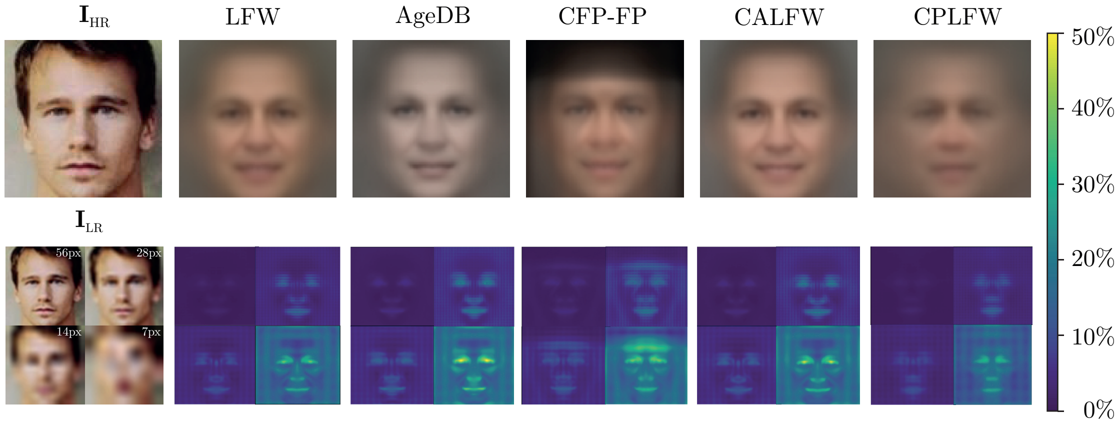

Figure 4 The left column shows a high resolution sample image $\mathbf{I}_{\mathrm{HR}}$ from MS1M and its corresponding reduced-resolution images $\mathrm{F}_{\mathrm{down},r}(\mathrm{F}_{\mathrm{up},112}(I_{\mathrm{HR}}))$ for four resolutions $r \in \{7, 14, 28, 56\}$. In the first row are then the mean images $\mathbf{I}_{\mathrm{HR}}^{\mathrm{mean}}$ for several datasets shown. Below are the pixel difference images $\mathbf{D}_{r}^{\mathrm{mean}}$ for specific resolutions.

The mean dataset HR images are quite different across all datasets. Pose variations in CPLFW dataset result in blurrier areas of the image. In contrast, the CALFW and LFW dataset images seem to be very accurately aligned and show almost a clear and detailed average face. Interestingly, the background in the CFP-FP dataset is very dark compared to other datasets. This results from cropped faces which are padded in black, especially in the top image region. Also, the pose variation is visible in the average face. Some ghosting effects are also present in that image.

The mean absolute pixel difference images show the same pattern across all datasets. With decreasing resolution the difference is more visible, especially in the high frequency regions (eyes, nose, and mouth).

4.2 Face Verification Accuracy

As depicted in Figure 5, the performance on all datasets drops for lower resolutions as expected. The accuracy on the LFW dataset is best for high resolution but drops heavily for lower resolutions. The worst performance on high resolutions can be seen on the CPLFW dataset. A reason for this behavior can be the large pose variations in this test set, which are not occurring in the training set and therefore unknown to the network.

Interestingly, we see different decreasing characteristics between the CR and ER scenarios. To a particular resolution, all datasets show worse performance in the ER scenario than in the CR scenario. This performance gap is reasonable since more pixel information is present in a CR pair than in an ER image pair. Against intuition, this trend reverses for very low resolutions except for the AgeDB dataset. We explain this behavior with a more significant domain shift for the network necessary within the CR image pairs than within the ER image pairs. Our network is familiar with HR images, and down to a specific resolution, it can interpret lower quality faces quite well. Whereas beneath a threshold, both LR images are unfamiliar to the network, and thus, features represent different ID characteristics compared to HR features. However, in the AgeDB dataset is a significant age gap within the pairs, which implicates that on LR images, for example, large hair-style variations or the effect of gray-scale vs. color images, might confuse the network for positive image pairs.

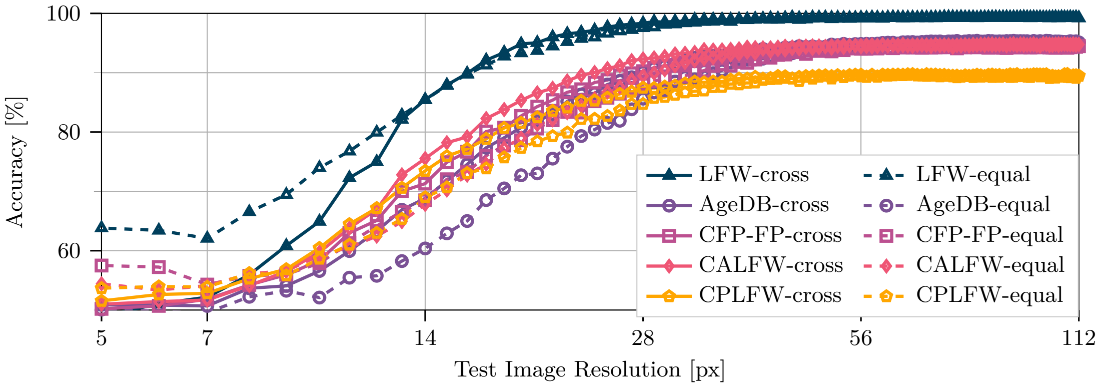

Figure 5 Face verification accuracy across several datasets for different image resolutions in cross-resolution (high resolution vs. low resolution image) and equal-resolution (low resolution vs. low resolution image) scenario.

4.3 Feature Distances

Since the accuracy depends on a distance threshold, which classifies sample pairs as genuine or imposter, the distance between both features vectors is crucial for the verification accuracy. Hence, we look at the average feature distance for all genuine and imposter image pairs of the LFW dataset (cf. Figure 6. We divide the diagram roughly into three sections: 1) $r > 60$ px, 2) > 20 px $< r < 60$ px, and 3) $r < 20$ px. In the first section, feature distances between genuine and imposter image pairs seem to be independent of the image resolution. The average distance is about 0.3 within genuine pairs and 1.0 within imposter pairs, which means that the high dimensional feature vectors are almost orthogonal. The second section reveals, that in both CR and ER scenario the distance of genuine image pairs tends to increase, whereas the distance for imposter image pairs is only slightly decreasing. A small reduction of image resolution causes repelling features. However, reducing the image resolution more (section 3), all LR image features are projected closely together (far away from HR features), which results in small distances for ER and high distances in the CR scenario. Considering that almost all pairs are then categorized as genuine (in the CR scenario) or imposter (in the ER scenario), the face verification performance is merely guessing.

For the CR scenario, we conduct that our network is not able to extract accurate features for the very LR images. Hence, this results in a large distance between features because the HR image features are still very distinctive. However, in the ER scenario, both images are unfamiliar to the network, which results in resembling extracted features.

Figure 7 captures the cosine feature distance distributions for the LFW dataset. The center violin plots represent the feature distance distribution for HR image pairs. Distances for genuine and imposter image pairs are clearly distinguishable. The genuine distances are mainly in a range between 0.1 and 0.6, whereas imposter distances are mostly in the field of 0.6 and 1.4. Both classes can be separated effectively with a threshold of about 0.6, and thus, the accuracy for HR face verification is best (cf. Figure 5). On the left side, distributions for the CR scenario are shown, whereas on the right side, ER feature distributions are plotted. For images resolutions of 56 px and 28 px the distributions in both scenarios is similar to the HR distribution. The fact that the peak feature distance for genuine image pairs even exceeds the maximum distance for imposter pairs in the CR scenario at very low resolution 5 px, leads to the conclusion that image resolution

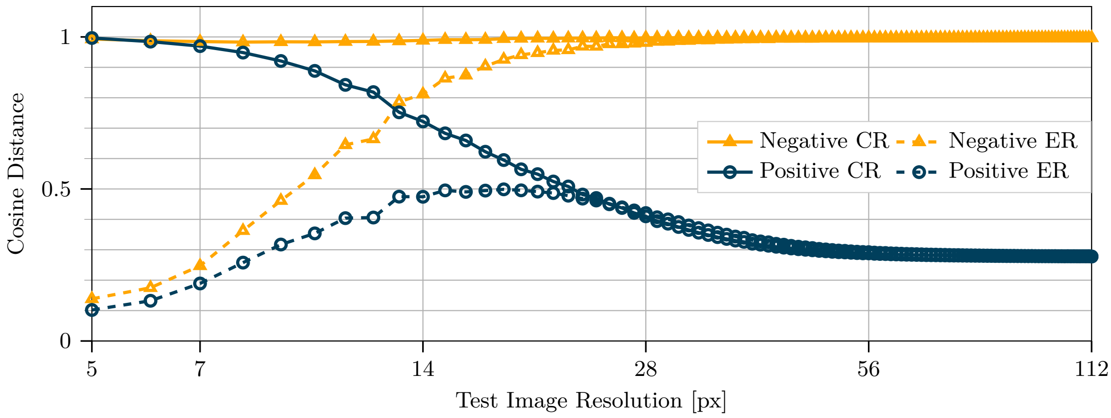

Figure 6 Average cosine feature distances between image pairs for genuine (○) and imposter (△) pairs in the LFW dataset. Dashed lines shows distances in the equal-resolution scenario, while solid lines represent distances in the CR scenario.

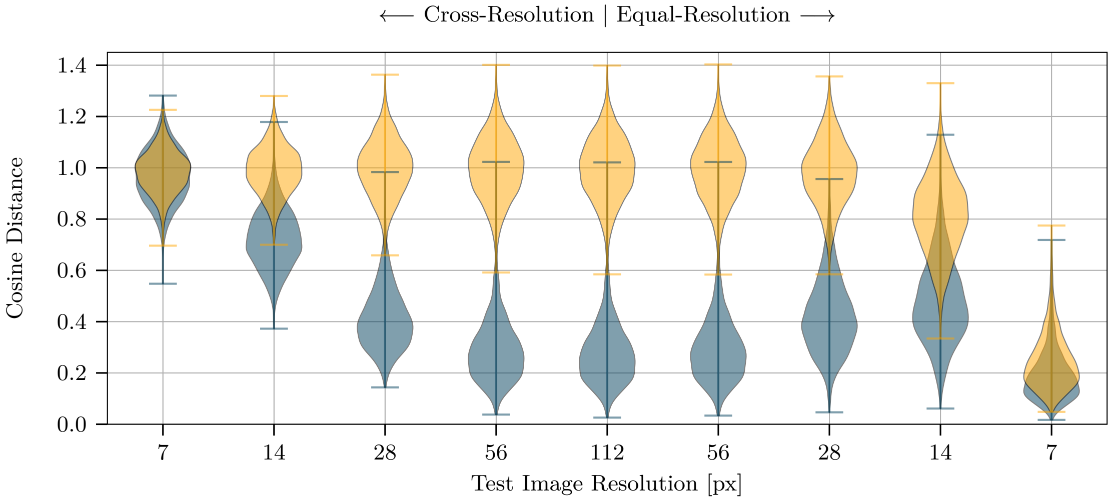

Figure 7 Cosine feature distance distributions for genuine (blue) and imposter (yellow) cross-resolution (left) and equal-resolution (right) pairs in the LFW dataset. Five different resolutions are shown for our baseline model.

has a more significant impact on the distance than the identity itself. The gap between CR and ER accuracy for very low resolutions is therefore reasonable. Although the small distances for both kinds of image pairs in the ER case, more genuine feature distance still have a smaller value. This behavior explains a higher accuracy for very low resolutions in the ER scenario compared to CR scenario. Further experiments with CFP-FP, AgeDB, CALFW, and CPLFW datasets underline this trend.

5 Training Methods

To improve the separability between feature distances of genuine and imposter image pairs, and hence, the accuracy, we pursue two intuitive non-transformation-based methods (cf. Figure 8): 1) CR batch training and 2) CR siamese network training.

In all training sessions, we use the MS1M dataset and train in total for 16 epochs. All training parameters are set according to our baseline (cf. section 3) for a fair comparison.

5.1 Cross-Resolution Batch Training

Motivated by [32, 16], we first propose a straightforward batch CR training approach to tackle the susceptibility to image resolutions. Instead of using HR images only, we randomly select half of the images per batch and synthetically reduce their resolution (cf. subsection 3.3).We train several specific networks specializing each on a particular resolution. For the sake of simplicity, we refer to these models according to the following rule: $BT\text{-}r$ where r denotes the specific LR value during training. We apply resolutions $r \in \{x \in \mathbb{N} \,|\, 5 \leq x \leq 22\} \cup \{28, 56\}$ in our experiments. Since we only take half of the images per batch for resolution reduction, the network is still exposed to HR images and thus learns to extract features from HR and LR images at the same gradient update step.

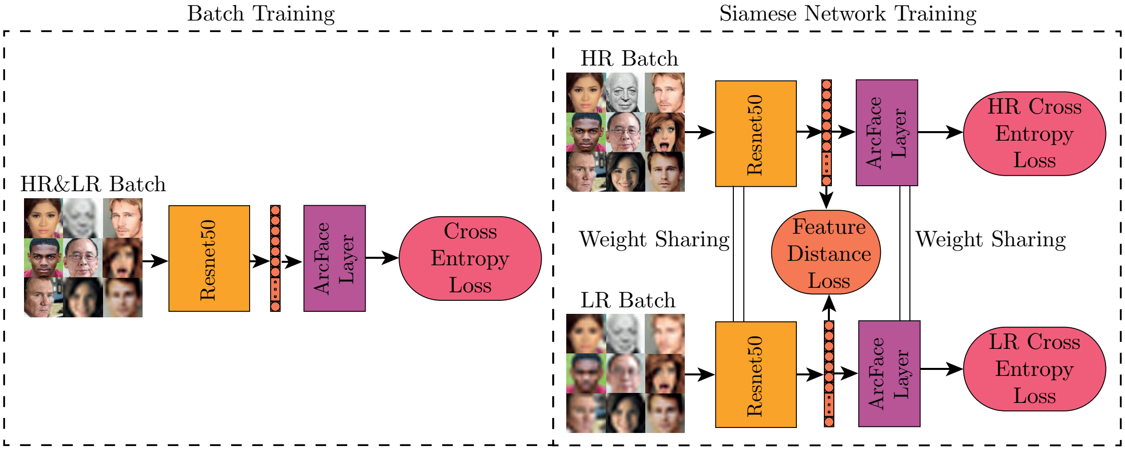

Figure 8 Overview of our proposed methods. The left part shows the cross-resolution batch training method, whereas the right part shows the siamese network cross-resolution training approach.

5.2 Siamese Network Cross-Resolution Training

Inspired by Tang et al. [26], we implement a siamese network structure (cf. Figure 8 CR training. Each branch of the network consists of our baseline architecture and trains the network for a specific image resolution. With weight-sharing across all branches, we keep the same number of parameters and ensure similar inference during test-time compared to *BT-r*. We construct two branches for training with exactly two resolutions (the high resolution and one low resolution). Our objective is to closely project the corresponding features from all branches for a specific image with an arbitrary resolution. We add a new loss function to the network to enforce this, which penalizes a high cosine distance between both features. We employed the cosine distance metric to match the evaluation protocol. Applying a HR image to the ArcFace network (HR branch), f_{HR} then denotes the corresponding output feature vector, and images from the particular LR branches are named f_{LR} accordingly. The cosine feature distance loss L_{dist} is then calculated as:

$$\mathcal{L}_{\text{dist}} = 1 - \frac{f_{\text{HR}} \cdot f_{\text{LR}}}{\|f_{\text{HR}}\|^2 \, \|f_{\text{LR}}\|^2} \tag{6}$$

For both branches, we calculate the cross-entropy loss $\mathcal{L}_{\text{CE,HR}}$ and $\mathcal{L}_{\text{CE,LR}}$, respectively. We weigh all three losses approximately equally and multiply the feature distance loss by a factor of 25. Finally, we conclude the total loss function $\mathcal{L}$ for the siamese training approach as follows:

$$\mathcal{L} = \mathcal{L}_{\text{CE,HR}} + \mathcal{L}_{\text{CE,LR}} + 25 \cdot \mathcal{L}_{\text{dist}} \tag{7}$$

Due to the siamese network architecture, both images, in high resolution and low resolution, need to be propagated through each branch. Thus, the training time is about double in the two resolution training scenario. In our experiments, we select the following resolutions $r \in \{5, 6, 7, 8, 12, 14, 20, 28, 56\}$ to train specific resolution models. In the following, we refer to this training technique as *ST-r*.

6 Experimental Results

In this section, we present and discuss the results of our proposed approaches. Firstly, we focus on the two-resolution scenario, i.e., high resolution (112 px) and one specific low resolution. Secondly, focus on simultaneously training with multiple image resolutions, i.e., high resolution (112 px) and multiple low resolutions (7 px, 14 px, 28 px, and 56 px) in one training. We analyze the accuracy on five popular datasets and compare the distances of the resulting features for all methods. Moreover, we introduce a new evaluation protocol to measure the performance of a model for multiple resolutions in the test dataset. We conclude this section with a comparison of all methods proposed in this paper, especially concerning the differences in accuracy and training time.

6.1 Two-Resolution Training Scenario

As previously mentioned in section 5, we now analyze the CR batch training approach BT-r and the siamese network CR training approach ST-r concerning the face verification accuracy on five popular datasets. This two-resolution training scenario trains each model with exactly two specified resolutions and compares the results to the baseline network concerning accuracy and feature distances.

Face Verification Accuracy

As introduced in subsection 3.4, accuracy is a standard metric to measure the performance of a face verification model. Figure 9 depicts the average face verification accuracy across five common datasets of the BT-r and ST-r model compared to our baseline model. Note that BT-@ and ST-@ data points represent different models trained explicitly for the test resolution. Both approaches outperform the baseline model for low image resolutions. For very low resolutions, i.e., 5 px to 8 px, the performance can be increased from ≈ 50% up to 70%. Above $r \approx 40$ px, no significant difference exists between all approaches, which affirms our expectations since the LR images are visually hardly distinguishable from the original images, and the absolute pixel difference is minimal (cf. subsection 3.3).

Generally, the performance improvement is increasing with decreasing resolutions. The BT-r method performs slightly better than the ST-r method, from which we conclude that the siamese approach might concentrate too much on projecting the features of the same image in different resolutions into the same space than on classifying the correct identity regardless of the resolution. For applications with a known fixed resolution, a $BT-@$ are the better choices.

Moreover, we compare our results on the very popular LFW dataset with two other approaches (cf. Table 2): First, the selected knowledge distillation technique proposed by Ge et al. [5], and second, the attribute-guided coupled GAN approach introduced by Talreja et al. [25]. Our systems

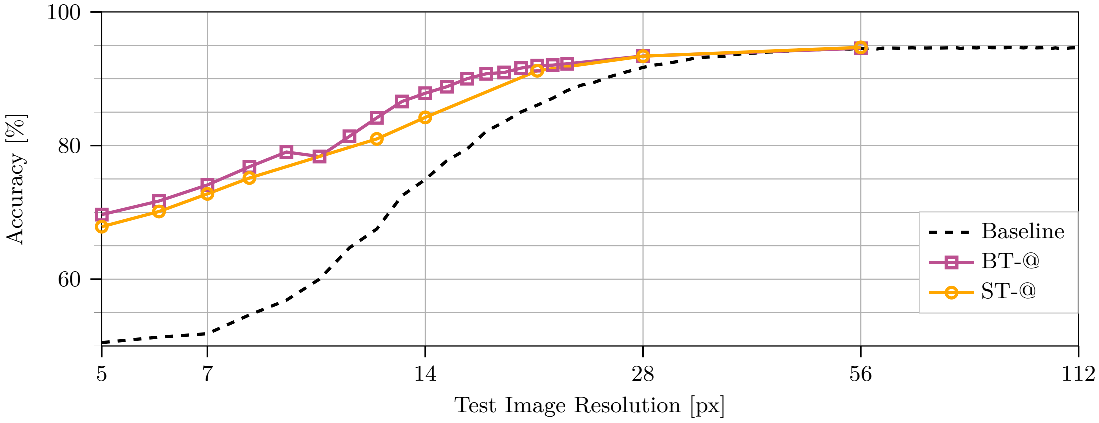

Figure 9 Evaluation of average face verification accuracy across five popular datasets for different resolution with several models.

Table 2 Face verification accuracy on the LFW dataset. The best performance of each image resolution is marked bold.

Image Resolution	Model	Accuracy
64 px	*BT-64* (ours)	**99.38%**
	ST-64 (ours)	99.35%
	S-64-sc [5]	92.83%
	Talreja et al. [25]	94.92%
32 px	*BT-32* (ours)	**99.08%**
	ST-32 (ours)	98.32%
	S-32-sc [5]	89.72%
	Talreja et al. [25]	91.08%
16 px	*BT-16* (ours)	**98.17%**
	ST-16 (ours)	97.8%
	S-16-sc [5]	85.87%

clearly outperform both competitors. However, the comparison to Ge et al.'s approach is not fair. Their baseline model (teacher model) only reaches an accuracy of 97.15%, which is not comparable to our baseline and state-of-the-art. On the other hand, the model's number of parameters also differs. They only trained both models for three different resolutions, showing only a few snapshots and not the whole performance curve. The lowest resolution, (16 px) is relatively high compared to our analysis, so we cannot fully exploit our strengths here.

Feature Distances

Similar to subsection 4.3, we pick five different resolutions and take a closer look at the features themselves. To be more precise, we plot the distance distributions for genuine and imposter image pairs from the LFW dataset.

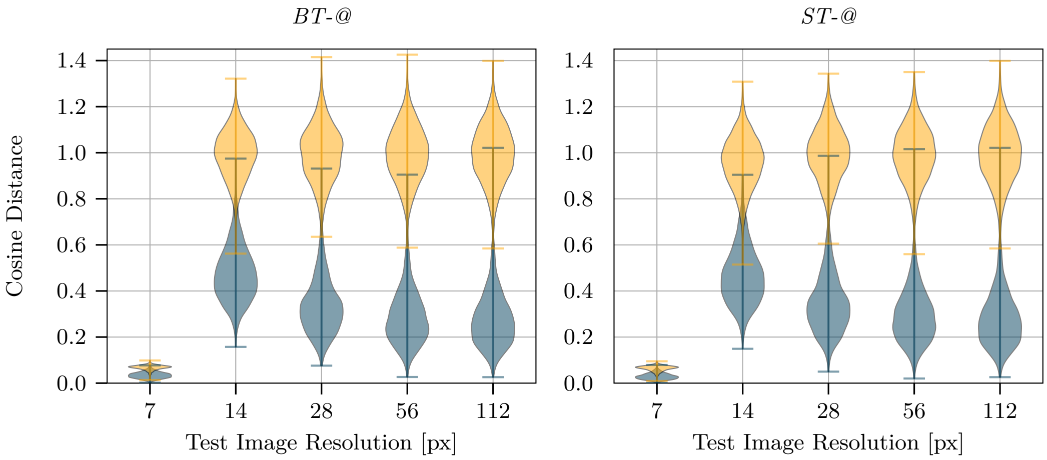

Figure 10 Cosine feature distance distributions for genuine (blue) and imposter (yellow) cross-resolution pairs in the LFW dataset. For both models, @ denotes that the training resolution matches the test resolution.

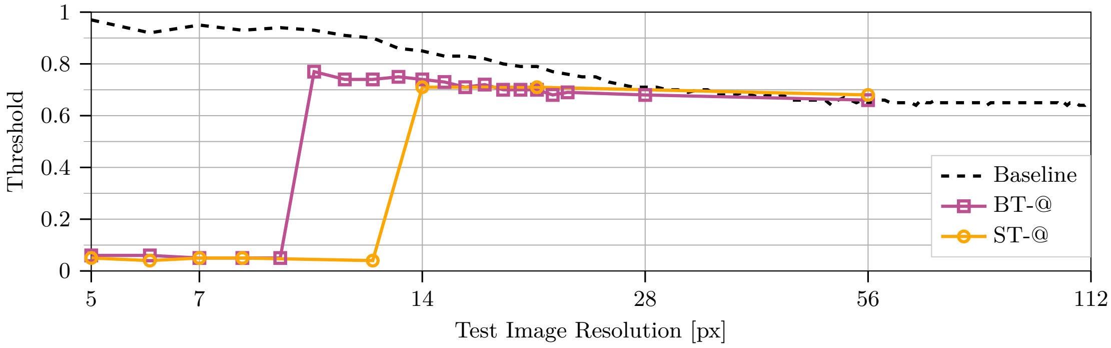

Figure 11 Best thresholds selected for calculating the accuracy on the LFW dataset using different models. In the model description an @ denotes, that the training resolution matches the test resolution.

Figure 10 shows that distances of genuine and imposter pairs are much better separable for low resolutions 14 px, 28 px and 56 px than the baseline results (cf. Figure 7). The main difference compared to the baseline is the shift of genuine and imposter distances to a range of almost 0 and 0.1 in the very low resolution scenario (7 px). This behavior is remarkable and shows that both networks learn to project features from very different resolutions into the same space. Although all distances are small, imposter distances are still greater than genuine, and the distributions are separable, consistent with the accuracy improvement discussed in the previous subsection (cf. Figure 9). Furthermore, there is no significant difference between both proposed methods. This is in line with with the last section's accuracy values.

To understand and determine the exact resolution where the feature distances drop so much, we calculate the optimal threshold and analyze the corresponding accuracy values (cf. subsection 3.4). Figure 11 depicts the threshold values for the baseline, *BT-r*, and *ST-r* models on all tested image resolution in the CR scenario. Thresholds for our baseline model are increasing for lower resolutions. This trend is consistent with our results in subsection 4.3, where genuine and imposter feature distances increase for lower resolutions. Our two-resolution training networks show a significant drop at $r = 9$ px for *BT-r* and $r \approx 12$ px for *ST-r*. From these points on, both models behave differently in the training sessions and project features for significant resolution differences more closely.

6.2 Multi-Resolution Training Scenario

We propose multiple-resolution training for both approaches to simulate a more applicable model, which is capable of handling arbitrary resolutions. We train the *BT-r* model with more than two resolutions simultaneously by randomly picking a different resolution in $\{7, 14, 28, 56, 112\}$ to generate a LR image. Each batch contains HR images and multiple LR images with different resolutions. We find that those five resolutions equally represent the range of image resolutions. This range reflects, for example, equivalent distances from subjects to the camera in real life. The probability of each resolution is set to be equal. We name this approach *BT-M* in the following.

In the *ST-r* approach, we apply two different methods for training with multiple resolutions simultaneously. First, for the LR branch, we randomly pick a resolution from the numbers $\{7, 14, 28, 56\}$ and feed the LR branch with LR images of different resolutions within each batch. The HR branch always takes 112 px images. This training with in total five different resolutions simultaneously doubles the training time. In the following, this approach will be referred to as

ST-M1. The second method *ST-M2* extends the siamese network to five branches, each branch representing a particular defined resolution (7 px, 14 px, 28 px, 56 px, and 112 px). Consequently, four feature distance losses are calculated each between the HR and the corresponding LR branch. Moreover, we also calculate the cross-entropy loss for each branch. All feature distance losses are weighted each with a factor of 25, to be in the same order of magnitude as the cross-entropy losses. The training time for this experiment is about five times longer than the baseline because it is scaling with the number of defined resolutions for training.

Face Verification Accuracy

Figure 12 presents the face verification accuracy for *BT-M*, *ST-M1*, and *ST-M2* model across arbitrary image resolutions. All three approaches perform significantly better than the baseline model in resolutions below 13 px and worse above a resolution of 28 px. Note that there is a significant peak at a resolution of 14 px, especially for *BT-M*. One reason for this could be that this specific resolution was used during training, and hence, this effect is also visible at resolutions 7 px, 28 px, and 56 px.

Another finding is that the siamese network CR training outperforms the CR batch training for low resolutions ($r < 16$ px) and vice-versa for mid and high resolutions ($r > 16$ px). For $r = 7$ px, the *ST-M1* model achieves an accuracy score of $\approx 75\%$, which is almost 25% above the baseline performance and even higher than *ST-7*. At the same time, that approach loses about 8% performance at high resolutions $r = 56$ px. For a more scale-comprehensive performance score, we will introduce three new evaluation protocols in subsection 6.3.

Figure 13 investigates the performance at and close to two selected resolutions, 7 px and 14 px. On the left side, we can see that *BT-7* and *ST-7* optimized the performance strictly for the 7 px resolution, and hence they perform worse in the neighboring regions. *BT-@* and *ST-@*, which represent specific resolution trained models, perform best at each scale, and this is reasonable due to the training with that particular image resolution. The performance loss for all multiple-resolution trained approaches (*BT-M*, *ST-M1*, and *ST-M2*) is compensated by the benefit of having a single model for arbitrary resolutions. The right part of Figure 13 shows an excerpt of resolutions from 10 px to 18 px. Here, the wave effect of *BT-14* and *ST-14* is also slightly visible, meaning that those two models perform relatively best on exactly 14 px resolution.

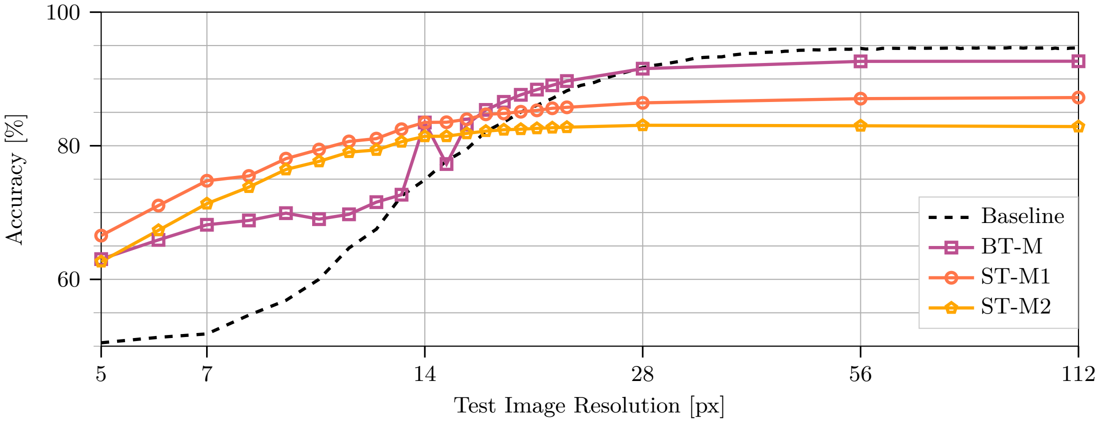

Figure 12 Average face verification accuracy across five popular datasets for different image resolutions with several models. Except for the Baseline all models were trained using multiple image resolutions.

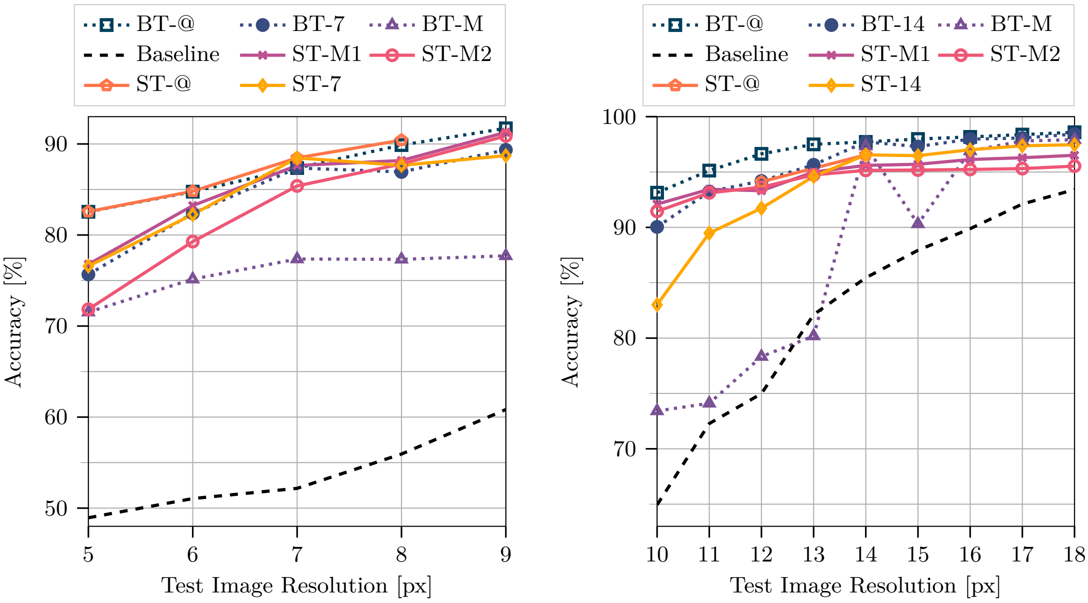

Figure 13 Accuracy on the LFW dataset for several models trained with different image resolutions. In the model description an @ denotes, that the training resolution matches the test resolution.

Feature Distances

Interestingly, in terms of feature distance distributions (cf. left part of Figure 14), the multi-resolution batch training is not behaving similarly to the two resolution batch training. Specifically for *BT-r*, at very low resolutions ($r = 7\,\text{px}$), the feature distance distributions for genuine and imposter pairs are even larger than for all other resolutions. This characteristic fits to the *BT-r* accuracy at that scale (cf. Figure 13). In contrast to the two-resolution case, both siamese training approaches (*ST-M1* and *ST-M2*) project features for all resolutions closer together. We conduct this from very low distances across all scales (center and right parts in Figure 14). The maximum feature distance for both approaches is about 0.1.

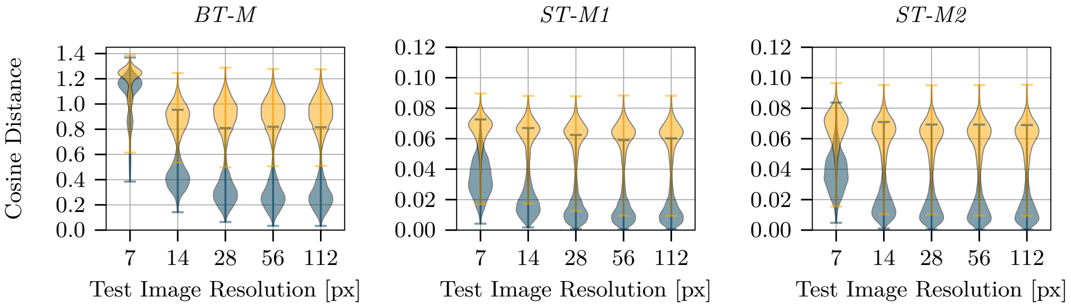

Figure 14 Cosine feature distance distributions for genuine (blue) and imposter (yellow) cross-resolution pairs in the LFW dataset at different test-set image resolutions. For training, multiple image resolutions were used.

6.3 Evaluation Protocols for Multiple Resolutions

Evaluation protocols for common public datasets are not taking into account the image resolution. In the previous sections, we only considered a specific image resolution to calculate the face verification accuracy. With single networks (cf. subsection 6.2) capable of handling arbitrary image resolutions at once, there is a need for a more meaningful evaluation considering multiple resolutions. Therefore, we propose specific evaluation protocols for all five datasets, with a focus on four different resolution ranges:

- Low resolutions: $r \in \{x \in \mathbb{N} \,|\, 5 \leq x \leq 10\}$
- Mid resolutions: $r \in \{x \in \mathbb{N} \,|\, 11 \leq x \leq 40\}$
- High resolutions: $r \in \{x \in \mathbb{N} \,|\, 41 \leq x \leq 112\}$
- All resolutions: $r \in \{x \in \mathbb{N} \,|\, 5 \leq x \leq 112\}$

The evaluation protocols define the resolution for each image in each pair for the corresponding dataset, and we keep the probability for each resolution in the generation process equal. All protocols are available at: `https://github.com/martlgap/btm-stm`.

6.4 Comparison of the proposed Methods

To conclude this chapter, we provide a comparison between all introduced methods. First, we analyze the verification performance on HR images for all proposed methods and compare them to the baseline approach. Figure 15 shows the accuracy of the LFW dataset for each epoch. We select models *BT-7*, *BT-56*, *ST-7*, and *ST-56* to represent both, shallow and relatively high resolutions. After the first epoch, our baseline model achieves about 98% accuracy, followed by almost peak accuracy already after the second epoch. During epochs 3 and 16, no significant changes in accuracy are visible. The *BT-56* starts with equal accuracy after the first epoch and then takes another two epochs to reach almost peak accuracy. The *ST-56* gets only after epoch 4 approximately peak performance. This model needs significantly more samples than both previously mentioned models to achieve similar accuracy. One reason could be the additional feature distance loss, which forces the network to minimize feature distances and learn a reasonable classification.

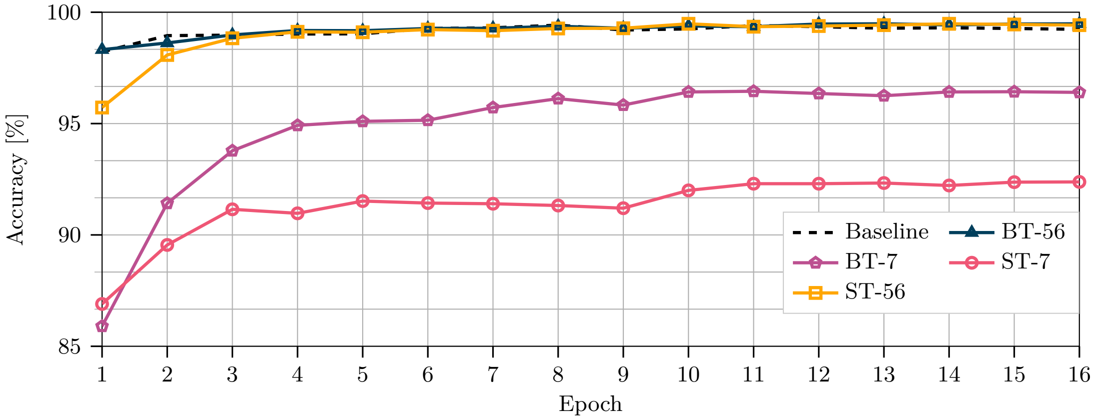

Figure 15 Evaluation of face verification accuracy on the original LFW dataset for high resolution images.

Table 3 Comparison of training time per epoch and accuracy on LFW dataset for different test image resolution protocols. Bold numbers denote the best performance across all methods.

Model	Training Time per Epoch [h]	Accuracy [%] for Test Resolution					
		112 px	all_res	high_res	mid_res	low_res	5 px
Baseline	2	99.23	96.86	99.20	95.89	77.57	54.65
BT-M	2	**99.30**	**97.72**	**99.33**	**97.78**	87.17	71.53
ST-M1	4	97.40	96.76	97.35	96.98	**91.50**	**76.78**
ST-M2	20	95.62	95.07	95.62	95.51	88.72	71.84

The peak performance for both methods *BT-7* and *ST-7* are significantly lower compared to the other approaches. This decrease could evolve from too little information in the shallow LR images, which might probably be just too little resolution to be able to learn a proper feature extraction. Moreover, models converge slower and need at least 10 epochs to reach the overall maximum accuracy region.

Second, Table 3 compares all presented methods to their training time per epoch and performance in the multi-resolution scenarios and depicts accuracy values on the LFW dataset. We conduct that compared to the two resolution techniques, both *ST-M1* and *ST-M2* models clearly outperform the baseline and the *BT-M* for low resolutions. Focusing on higher resolutions, we conclude that *BT-M* is the best performing method. Even for the original image resolution of 11 px, the *BT-M* model performs better than the baseline. We think this is reasonable because using lower resolutions additionally during training can be seen as extra data augmentation and hence, can improve the performance. One also has to compromise that for an absolute performance improvement of about 14% in the low_res protocol, the performance for high_res drops about 2%. In the second siamese training approach, *ST-M2* is performing worse in all categories than *ST-M1*. Therefore, we conclude that the much greater effort for training is not worth it. It seems to be less important to force a network to learn close features for the same image in different resolutions, within each batch, than across several batches.

Lastly, the number of parameters, and hence the inference time, is equal for all models, thus making the comparison fair and reasonable.

7 Conclusions and Future Work

This work analyzes the impact of different image resolutions on face verification performance utilizing a state-of-the-art approach. The distances between extracted features are investigated in detail. Our findings are that facial features extracted from established face recognition networks are not scale-invariant, and hence the performance decreases substantially for lower image resolutions.

To obtain the best performance, the resolution of the testing images must be the same as in the corresponding training dataset for the network. To overcome this problem, we propose two intuitive methods to learn scale-invariant features directly: 1) Training our network with batches containing an equal amount of LR and HR images. Experiments across five conventional test datasets show improvements up to 24.80% for for very low image resolutions of 5 px. 2) Training a siamese network structure, which additionally minimizes feature distances between LR and HR versions of the same image besides the cross-entropy loss. Evaluations across five conventional test datasets indicate an improvement in performance up to 31.77%.

Furthermore, we train our proposed models with several resolutions at once. Hence, a single model can be applied to arbitrary image scales, making it more applicable. We also report a considerable improvement of 17.96% with our best model *ST-M1* for CR verification performance, especially for low resolutions. Compared to the simple batch training method, the siamese network CR training performs better for low resolutions and worse for mid and high resolutions. For applications with a known fixed resolution, the latter method is the better choice.

Lastly, we introduce and release three different evaluation protocols for five popular datasets, defining multiple resolutions for CR scenarios.

Our work showed that a loss on feature distances helps to mitigate the resolution susceptibility in face verification. Therefore, in the future, we want to employ a specifically designed triplet loss variant, which minimizes intra-class and maximizes inter-class feature distances. We also want to extend the downsample process by using arbitrary blur kernels described in [33] and applying them in our work.

References

1 Omid Abdollahi Aghdam, Behzad Bozorgtabar, Hazim Kemal Ekenel, and Jean-Philippe Thiran. Exploring factors for improving low resolution face recognition. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 2363–2370. IEEE, 2019.

2 Zhiyi Cheng, Xiatian Zhu, and Shaogang Gong. Low-resolution face recognition. *CoRR*, abs/1811.08965, 2018. `arXiv:1811.08965`.

3 Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4690–4699, 2019.

4 Berk Dogan, Shuhang Gu, and Radu Timofte. Exemplar guided face image super-resolution without facial landmarks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 0–0, 2019.

5 Shiming Ge, Shengwei Zhao, Chenyu Li, and Jia Li. Low-resolution face recognition in the wild via selective knowledge distillation. *IEEE Transactions on Image Processing*, 28(4):2051–2062, 2018.

6 Yandong Guo, Lei Zhang, Yuxiao Hu, Xiaodong He, and Jianfeng Gao. Ms-celeb-1m: A dataset and benchmark for large-scale face recognition. In *European conference on computer vision*, pages 87–102. Springer, 2016.

7 Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

8 Chih-Chung Hsu, Chia-Wen Lin, Weng-Tai Su, and Gene Cheung. Sigan: Siamese generative adversarial network for identity-preserving face hallucination. *IEEE Transactions on Image Processing*, 28(12):6225–6236, 2019.

9 E. G. Huang, G. B. Learned-Miller. Labeled Faces in the Wild: Updates and New Reporting Procedures. Technical Report UM-CS-2014-003, University of Massachusetts, Amherst, May 2014.

10 Robert Keys. Cubic convolution interpolation for digital image processing. *IEEE transactions on acoustics, speech, and signal processing*, 29(6):1153–1160, 1981.

11 Vahid Reza Khazaie, Nicky Bayat, and Yalda Mohsenzadeh. Ipu-net: Multi scale identity-preserved u-net for low resolution face recognition. *arXiv preprint*, 2020. `arXiv:2010.12249`.

12 Yonghyun Kim, Wonpyo Park, Myung-Cheol Roh, and Jongju Shin. Groupface: Learning latent groups and constructing group-based representations for face recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5621–5630, 2020.

13 Pei Li, Loreto Prieto, Domingo Mery, and Patrick J Flynn. On low-resolution face recognition in the wild: Comparisons and new techniques. *IEEE Transactions on Information Forensics and Security*, 14(8):2000–2012, 2019.

14 Weiyang Liu, Yandong Wen, Zhiding Yu, Ming Li, Bhiksha Raj, and Le Song. Sphereface: Deep hypersphere embedding for face recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 212–220, 2017.

15 Ze Lu, Xudong Jiang, and Alex Kot. Deep coupled resnet for low-resolution face recognition. *IEEE Signal Processing Letters*, 25(4):526–530, 2018.

16 Fabio Valerio Massoli, Giuseppe Amato, and Fabrizio Falchi. Cross-resolution learning for face recognition. *Image and Vision Computing*, page 103927, 2020.

17 Stylianos Moschoglou, Athanasios Papaioannou, Christos Sagonas, Jiankang Deng, Irene Kotsia, and Stefanos Zafeiriou. Agedb: the first manually collected, in-the-wild age database. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 51–59, 2017.

18 Sivaram Prasad Mudunuri, Soubhik Sanyal, and Soma Biswas. Genlr-net: Deep framework for very low resolution face and object recognition with generalization to unseen categories. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 602–60209. IEEE, 2018.

19 Nasrabadi NM et al. Identity-aware deep face hallucination via adversarial face verification. In *IEEE International Conference on Biometrics Theory Applications and Systems*, 2019.

20 Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng

Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115(3):211–252, 2015.

21 Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 815–823, 2015.

22 Soumyadip Sengupta, Jun-Cheng Chen, Carlos Castillo, Vishal M Patel, Rama Chellappa, and David W Jacobs. Frontal to profile face verification in the wild. In *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1–9. IEEE, 2016.

23 Maneet Singh, Shruti Nagpal, Richa Singh, Mayank Vatsa, and Angshul Majumdar. Magnifyme: Aiding cross resolution face recognition via identity aware synthesis. *arXiv preprint*, 2018. `arXiv:1802.08057`.

24 Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.

25 Veeru Talreja, Fariborz Taherkhani, Matthew C Valenti, and Nasser M Nasrabadi. Attribute-guided coupled gan for cross-resolution face recognition. *arXiv preprint*, 2019. `arXiv:1908.01790`.

26 Su Tang, Shan Zhou, Wenxiong Kang, Qiuxia Wu, and Feiqi Deng. Finger vein verification using a siamese cnn. *IET Biometrics*, 8(5):306–315, 2019.

27 Hao Wang, Yitong Wang, Zheng Zhou, Xing Ji, Dihong Gong, Jingchao Zhou, Zhifeng Li, and Wei Liu. Cosface: Large margin cosine loss for deep face recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5265–5274, 2018.

28 Qiangchang Wang, Tianyi Wu, He Zheng, and Guodong Guo. Hierarchical pyramid diverse attention networks for face recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8326–8335, 2020.

29 Zhifei Wang, Zhenjiang Miao, QM Jonathan Wu, Yanli Wan, and Zhen Tang. Low-resolution face recognition: a review. *The Visual Computer*, 30(4):359–386, 2014.

30 Yandong Wen, Kaipeng Zhang, Zhifeng Li, and Yu Qiao. A discriminative feature learning approach for deep face recognition. In *European conference on computer vision*, pages 499–515. Springer, 2016.

31 Erfan Zangeneh, Mohammad Rahmati, and Yalda Mohsenzadeh. Low resolution face recognition using a two-branch deep convolutional neural network architecture. *Expert Systems with Applications*, 139:112854, 2020.

32 Dan Zeng, Hu Chen, and Qijun Zhao. Towards resolution invariant face recognition in uncontrolled scenarios. In *2016 International Conference on Biometrics (ICB)*, pages 1–8. IEEE, 2016.

33 Kai Zhang, Wangmeng Zuo, and Lei Zhang. Deep plug-and-play super-resolution for arbitrary blur kernels. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1671–1681, 2019.

34 Kaipeng Zhang, Zhanpeng Zhang, Chia-Wen Cheng, Winston H Hsu, Yu Qiao, Wei Liu, and Tong Zhang. Super-identity convolutional neural network for face hallucination. In *Proceedings of the European conference on computer vision (ECCV)*, pages 183–198, 2018.

35 Tianyue Zheng and Weihong Deng. Cross-pose lfw: A database for studying cross-pose face recognition in unconstrained environments. *Beijing University of Posts and Telecommunications, Tech. Rep*, 5, 2018.

36 Tianyue Zheng, Weihong Deng, and Jiani Hu. Cross-age lfw: A database for studying cross-age face recognition in unconstrained environments. *arXiv preprint*, 2017. `arXiv:1708.08197`.

37 Ruofan Zhou and Sabine Susstrunk. Kernel modeling super-resolution on real low-resolution images. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2433–2443, 2019.