PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE
SCHOOL OF ENGINEERING

DATASET AND EXPERIMENTAL PROTOCOL FOR FACE RE-IDENTIFICATION WITH LOW RESOLUTION IMAGES

LORETO PRIETO HURTADO

Thesis submitted to the Office of Research and Graduate Studies
in partial fulfillment of the requirements for the degree of
Master of Science in Engineering

Advisor:
DOMINGO MERY

Santiago de Chile, November 2020

PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE
SCHOOL OF ENGINEERING

DATASET AND EXPERIMENTAL PROTOCOL FOR FACE RE-IDENTIFICATION WITH LOW RESOLUTION IMAGES

LORETO PRIETO HURTADO

Members of the Committee:

DOMINGO MERY

KARIM PICHARA

CLAUDIO PÉREZ

SEBASTÍAN VICUÑA

DocuSigned by: Domingo Mery 17BD1484D706424...
DocuSigned by: karim Pichara C2D12839E1A54C9...
DocuSigned by: Claudio Pérez DE7749F954E84A1...
DocuSigned by: Sebastián Vicuña 8795A809EDCD423...

Thesis submitted to the Office of Research and Graduate Studies in partial fulfillment of the requirements for the degree of Master of Science in Engineering

Santiago de Chile, November 2020

Gratefully to my family

ACKNOWLEDGEMENTS

I would like to thank Domingo Mery for everything he taught me and his infinite patience throughout this work and everything it implied.

Also, Pei Li and Patrick Flynn for everything they taught me in Notre Dame.

Finally, my family, especially my husband and my father, for their immense support.

This work was in part by the Seed Grant Program of The Colleges of Engineering at the Pontificia Universidad Catolica de Chile and at the University of Notre Dame, and in part by Fodecyt-Grant 1191131 from Chilean Science Foundation.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

An experimental protocol consist of agreed upon guidelines for a type of experiment. In research, protocols are often neglected and researchers that utilize previous work on their own projects have to allocate a significant amount of time into replicating inadequately described methods. On the other hand, research in low-resolution face re-identification has been increasing in the past few years. Low-resolution face re-identification refers to the problem of identifying if the same person's face appears in two images, one low resolution (LR, e.g. from a surveillance camera) and another high resolution (HR, e.g. from a government issued identification). Previous work in low-resolution face re-identification can be divided into three categories: 1) HR-mapping, 2) methods employing LR and HR robust features, 3) methods learning a unified space representation.

In this work we are proposing a protocol definition for face re-identification HR-mapping experiments and dataset protocol for all methods in low-resolution face re identification. The dataset protocol has been used as a guideline to create a set of training and testing pairs for face re-identification using VGG Face 2. A baseline algorithm has been put together with another 12 experiments to validate the experimental protocol. Obtaining a $d' = 1.236$ and $AUC = 0.81$ for a 14x14 LR pixel size set, $d' = 1.421$ and $AUC = 0.84$ for a 28x28 LR pixel size set and $d' = 1.547$ and $AUC = 0.87$ for a 56x56 LR pixel size set. Our objective is for researchers in the field to use this work as a set of guidelines for building comparable and replicable work of their own.

Keywords: experimental protocol, low-resolution face re-identification, HR-mapping, siamese network, GAN.

RESUMEN

Un protocolo experimental consiste en pautas acordadas para un tipo de experimento. En la investigación, los protocolos a menudo se descuidan y los investigadores que utilizan trabajos anteriores en sus propios proyectos tienen que dedicar una cantidad significativa de tiempo a replicar métodos descritos de manera inadecuada. Por otro lado, la investigación en re-identificación de rostros de baja resolución ha ido aumentando en los últimos años. La re-identificación de rostro de baja resolución se refiere al problema de identificar si el rostro de una misma persona aparece en dos imágenes, una de baja resolución (BR) y otra de alta resolución (AR). El trabajo previo en re-identificación de rostros de baja resolución se puede dividir en tres categorías: 1) mapeo de AR, 2) métodos que emplean características robustas de BR y AR, 3) métodos que aprenden una representación espacial unificada.

En este trabajo, estamos proponiendo la definición de un protocolo para experimentos de mapeo de AR de re-identificación facial y un protocolo de base de datos para todos los métodos de re-identificación facial de baja resolución. El protocolo de la base datos se ha utilizado como una guía para crear un conjunto de pares de entrenamiento y prueba para la re-identificación de rostro utilizando la base de datos VGG Face 2. Además, se ha elaborado un algoritmo de referencia con otros 12 experimentos para validar el protocolo experimental. Se obtuvo un $d' = 1.236$ y $AUC = 0.81$ para un conjunto de tamaño de 14x14 píxeles, $d' = 1.421$ y $AUC = 0.84$ para un conjunto de tamaño de 28x28 píxeles y $d' = 1.547$ y $AUC = 0,87$ para un conjunto de tamaño de 56x56 píxeles. Nuestro objetivo es que los investigadores en el campo utilice este trabajo como un conjunto de pautas para construir su propio trabajo de manera comparable y replicable.

Palabras Claves: protocolo experimental, re-identificación de rostros de baja resolución, mapeo de AR, red siamesa, GAN.

1. INTRODUCTION

Humans fail to recognize someone they know in a regular basis. So it is not surprising algorithms have similar problems. In 2014, the first algorithm to outperform humans in face recognition was created by Chaochao Lu (C. Lu & Tang, 2014). The amount of work done in this research area has caused for it to be even in our common apps, such as Google Photos. Although face recognition with some unconstrained conditions and cooperative subjects can effectively yield good results, face recognition in low resolution image has a long way to go.

Face re-identification refers to the problem of, when given two face photos captured at different times and with different cameras, how to determine if they are of the same person. With the current face recognition methods the task for high-resolution images scenarios can be solve with relative ease (Parkhi, Vedaldi, & Zisserman, 2015; Schroff, Kalenichenko, & Philbin, 2015; Sun, Liang, Wang, & Tang, 2015). Yet within this task there are certain scenarios in which one face is high resolution (HR) but the second face is low resolution (LR). This scenarios have proven quiet challenging (Cheng, Dong, Gong, & Zhu, 2020; Tang, Yang, Wang, Song, & Gao, 2020; Saha & Das, 2019) and have been referred to as Low Resolution Person Re-Identification.

One of the most common cases where LR person re-identification is present, is suspect lookout in security cameras. Both private and public sector have a big interest in this topic, especially when there is not enough security personal. Surveillance cameras constantly capture face images of subjects which tend to be low resolution, as the cameras cover a wide area rather than a specific point, and present changes in pose and light.

In the last decade many research teams have dedicated their efforts to solve this problem. Previous work in LR face re-identification can be divided into three main categories: 1) HR-mapping, 2) methods employing LR and HR robust features, 3) methods learning a unified space representation.

HR-mapping: refers to methods that up-scales the low-resolution (LR) image to the high-resolution (HR) space before any comparison is done. Recent works combine several

novel techniques to achieve this task (Li, Prieto, Mery, & Flynn, 2019; Li, Prieto, Flynn, & Mery, 2017; Cheng et al., 2020). A simplified visual representation of this structure can be seen in figure 1.1.

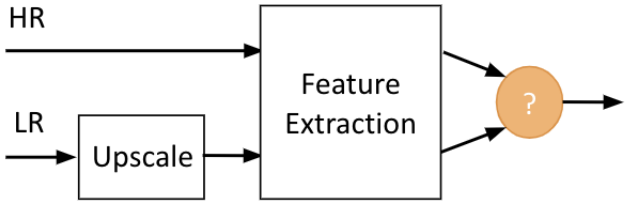


FIGURE 1.1. Simplified visual representation of HR-mapping methods.

LR and HR robust features: Several robust face feature can be obtained for a high resolution face (Deng, Guo, Niannan, & Zafeiriou, 2019; Schroff et al., 2015). Recent works have focused in calculating LR robust face features, these are generally handcrafted and based in texture or color. Some examples are (Tang et al., 2020; Z. Lu, Jiang, & Kot, 2018; Hernández-Durán, Cheplygina, & Plasencia-Calaña, 2015). A simplified visual representation of this methods can be seen in figure 1.2.

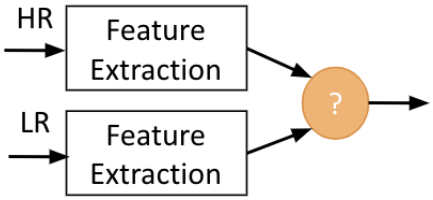


FIGURE 1.2. Simplified visual representation of LR and HR robust features methods.

Unified space representation: methods in this category focused on learning from both the HR and LR images. Afterwards, the images are projected into the learned space and similarity is measure with the projections. Both linear and non-linear spaces can be used. Some examples are (Heinsohn, Villalobos, Prieto, & Mery, 2019; K. Zhao, Xu, & Cheng, 2019; Al-Maadeed, Bourif, Bouridane, & Jiang, 2016; Z. Wang, Yang, & Ben, 2015). A simplified visual representation of these methods can be seen in figure 1.3.

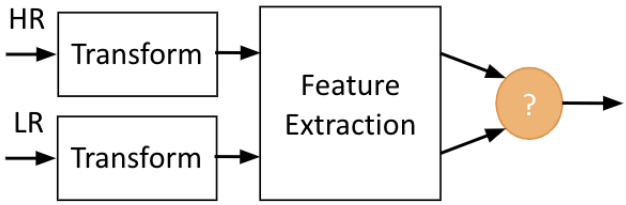


FIGURE 1.3. Simplified visual representation of unified space representation methods.

An experimental protocol consist of agreed upon guidelines for a type of experiment. Protocols are enablers to build the necessary collaborative work. In complex experiments, as low resolution face recognition, collaboration and comparison is mandatory for proving hypothesis. As David L. Selwood said in (Selwood, 2019), protocols are often relegated as they are partly not considered important enough. This negligence then falls on the researchers themselves who tried to utilize previous works for their own projects only to find methods are inadequately described, or worse, don't work at all. Protocols seek to address this shortcoming.

This work presents a protocol definition for both HR-mapping methods and datasets preparation for person re-identification in low resolution settings. Following a Knowledge Discovery on Databases based methodology for the protocols we will first define a realistic dataset with low-resolution images using training, validation and testing sets. Then, we will test several open-source computer vision algorithms and we will design and train new computer vision models for this task. Lastly, we will also define an evaluation protocol to compare these algorithms.

In addition, a baseline algorithm has been created with another 12 experiments to challenge its results. In several of this experiments we have used two novel methods as components, build with GAN and Siamese networks techniques. For each experiment we evaluated their performance as specified by the protocols and compared and discussed their results. In particular, we trained a set of GANs networks to up-scale the images from 14x14, 28x28 and 56x56 pixels to 112x112 pixels and a Siamese network for the face matching.

One of our hypothesis is that the combination of these networks can make a contribution to solving LR face re-identification.

The recount of the effort accomplished is divided into five sections.

- In section 2, a portrayal of the previous works done in HR-mapping
- In section 3, we detail the proposed experimental protocol for HR-mapping, it is our intention that anyone wanting to research in HR-mapping has this section as a guide
- In section 4, the proposed dataset protocol is described, it should be noted that this protocol is relevant for all categories of LR face re-identification
- In section 5, we have constructed a set of experiments to further detail how the proposed protocols should be applied
- In section 6 we will present some concluding remarks, trends and future works

Throughout this work, the following papers have been written:

(i) Li, P., Prieto, L., Mery, D., Flynn, P. (2018). Face recognition in low quality images: a survey. arXiv preprint arXiv:1805.11519. Submitted to ACM Computing Surveys in 2018 (in Second Review).

(ii) Li, P., Prieto, L., Mery, D., Flynn, P.J. (2019). On low-resolution face recognition in the wild: Comparisons and new techniques. IEEE Transactions on Information Forensics and Security, 14(8), 2000–2012.

(iii) Li, P., Prieto, M. L., Flynn, P. J., Mery, D. (2017). Learning face similarity for re-identification from real surveillance video: A deep metric solution. In 2017 IEEE International Joint Conference on Biometrics (IJCB) (pp. 243–252).

(iv) Heinsohn,D., Villalobos,E., Prieto,L., Mery,D. (2019). Face recognition in low-quality images using adaptive sparse representations. Image and Vision Computing, 85, 46–58

2. RELATED WORKS

The accuracy of face recognition algorithms lowers when low resolution is present in at least one of the images. As mention before, in the last decade many research teams have dedicate their efforts to solve this problem. Previous work in LR face re-identification can be divided into three main categories: 1) HR-mapping, 2) methods employing LR and HR robust features, 3) methods learning a unified space representation. The most intuitive approach is up-scaling the LR image so more data can be present (HR-mapping). The complexity of this approach comes when adding data to the low resolution image that actually helps the comparison rather than distracts it. In this section we will recount previous works on HR-mapping, divided into two categories: focus on reconstructing visually HR images in general and up-scaling for face recognition explicitly. For further details see (Li, Prieto, Mery, & Flynn, 2018).

Up-scaling LR images introduces more data on the pixel level which should aid the recognition rate. However, it risks adding artifacts to the image which can affect recognition rate. For recognition purposes, using up-scaling usually return better results than comparing the LR and HR images directly.

For face super-resolution, several models using deep learning methods have surpassed works with handcrafted solutions. Z. Wang (Z. Wang, Liu, Yang, Han, & Huang, 2015) questions whether large-capacity and data-driven methods have outperformed handcrafted work by proposing an end-to-end CNN based sparse modeling method that achieved better performance by employing a cascade of super-resolution convolutional nets (CSCN) trained for small scaling factors.

In other works, J. Jiang (Jiang, Yu, Hu, Tang, & Ma, 2018; Jiang et al., 2017; Jiang, Ma, Chen, Jiang, & Wang, 2016) was able to alleviate the noise effect during up-scaling by proposing a CNN to learn the denoiser prior, which is then plugged into a model-based optimization to jointly benefit the merits of model-based optimization and discriminative

inference, as well as proposing an improved patch-based method which leverages contextual information to develop a more robust and efficient context-patch face hallucination algorithm.

Traditional image up-scaling methods treat image noise on a pixel level without considering the whole image structure. T. Lu (T. Lu, Xiong, Zhang, Wang, & Lu, 2017) proposes a unified framework for representation-based face super-resolution by introducing a locality-constrained low-rank representation scheme.

G. Jia (Jia, Li, Zhuo, & Liu, 2016) proposes a recognition oriented feature hallucination method to map the features of a LR facial image to its High Resolution (HR) version. They extract PCA features from both images and then applied CCA to establish the coherent subspaces between both PCA features. Then employ a recognitioon rate guided prediction model with adaptative Piecewise Kernel Partial Least Squares to map the LR feaure to the HR version. Finally, a weighted combination of the hallucinated PCA features and the Local Binary Pattern Histogram (LBPH) features are adopted for face recognition.

Conventional super-resolution methods rely heavily on alignment of the face and in low-resolution unconstrained environments have a diverse range of poses and expressions. X. Yu (Yu & Porikli, 2017) proposed an end-to-end transformative discriminative neural network, employing the spatial transformation layers to devise super-resolving for unaligned and very small face images. Later, they developed an attribute-embedded upsampling network (Yu, Fernando, Hartley, & Porikli, 2018) utilizing supplementing residual images or feature maps that came from the difference of the HR and LR images.

SP. Mudunuri explored, in (Mudunuri, Sanyal, & Biswas, 2018), how different kinds of constraints at different stages of the architecture systematically affected the recognition. Based on the result, they proposed an inter-intra classification loss for the mid-level features combined with super-resolution loss at the low-level feature in the training procedure.

Z. Cheng (Cheng et al., 2020) recognizes super-resolution as an effective approach for low resolution person re-identification. However, he notes that it is limited due to dramatically more difficult gradients back-propagation during training. He proposes a model

training regularisation method, called Inter-Task Association Critic, which discovers an association between image super-resolution and person re-identification adding it as an extra learning constrain.

In the last decade we have observed tremendous improvements in face re-identification. However, it is clear face re-identification is far from perfect and still an open problem when tackling low-resolution face images. To our knowledge no protocols for low-resolution face re-identification have been published to this day. It is essential for agile research that this protocol is created and adopted so further research can encourage innovation through collaboration.

3. PROPOSED PROTOCOL AND METHODS FOR HR-MAPPING

In this section we will explained the proposed protocol for HR-mapping. Followed by a definition of a baseline and description for 12 experiments to challenge said baseline.

3.1. PROPOSED PROTOCOL

As mentioned before, protocol incentives collaboration, a necessary process for solving complex problems. While researching related work for HR-mapping we observed that these could easily be divided into four key interchangeable components. We have defined that for every experiment in a HR-mapping there is going to be two inputs, HR and LR image and four algorithmic components: 1) up-scaling of the LR image to the same resolution as HR, 2) feature extraction for both images, 3) feature comparison and 4) binary answer to the following question: Are they the same person?. A visual representation of this components can be found in figure 3.1.

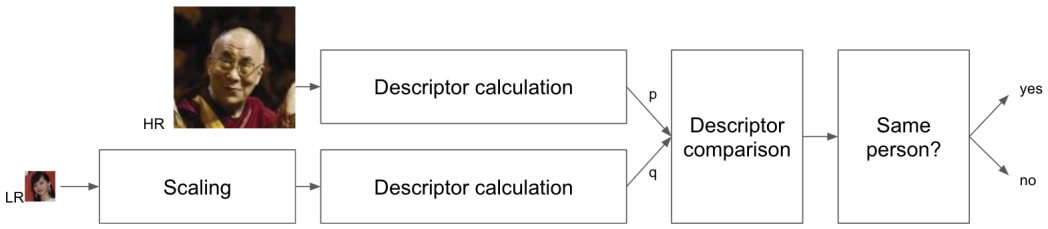


FIGURE 3.1. Visual representation of the proposed HR-Mapping protocol.

It is our intention that by following this protocol, the creation of new experiments can be performed in an easier way and enriched from related works. For example, if a new super-resolution method is created obtaining better results than all previous up-scaling methods, then this super-resolution method can be easily inserted to the state-of-the-art algorithm by replacing the up-scaling component. In the following four subsections we explained the components in further details.

3.1.1. Up-scaling

The main goal of this component is to up-scale the image to the same resolution as the HR. This can be achieved by different methods: learning-based methods such as super-resolution (Sønderby, Caballero, Theis, Shi, & Huszár, 2016; L. Zhao et al., 2019; Ledig et al., 2017; X. Wang et al., 2018; Y. Wang et al., 2018; Kim, Kim, Kwon, & Kim, 2019) or handcrafted algorithms (e.g. bicubic interpolation, Lanczos interpolation over 8x8 neighborhood, re-sampling using pixel area relation, nearest neighbor interpolation).

3.1.2. Feature extraction

Comparing pixel by pixel would be a terrible idea in many cases, especially if any pose, light or expression variation is present. There are abundant methods to extract a feature or embedding from an image. In this component the goal is to use one of this methods to get features or embedding from both HR image and the up-scaled LR image . It is recommended to use a method specific for faces (Schroff et al., 2015; Deng et al., 2019; Parkhi et al., 2015; Viola & Jones, 2004) but it is not a restriction (He, Zhang, Ren, & Sun, 2016; Simonyan & Zisserman, 2014; Szegedy, Vanhoucke, Ioffe, Shlens, & Wojna, 2016).

3.1.3. Feature comparison

Once we have two features, one for each image, we need a method to compare them. Calculating the distance between the two features (vectors) is the best way to get a score (e.g. Euclidean distance as shown in equation 3.1, cosine distance as shown in equation 3.2). With this distance we can setup a scoring system where a smaller distance means a higher probability both images belong to the same person.

$$\text{Euclidean distance}(p, q) = \sqrt{(p_1 - q_1)^2 + ... + (p_n - q_n)^2}, \tag{3.1}$$

where n is the feature size.

$$\text{Cosine distance}(p, q) = (\frac{p}{||p||} \bullet \frac{q}{||q||}) \tag{3.2}$$

3.1.4. Matching

Once we have all the scores we need to determine a threshold to determined when are going to consider a pair of images belongs to the same person. If the images score is lower than the threshold then we consider them of the same person. Otherwise the pair of images belongs to different people. It should be noted, that depending on the scoring systems used this could change so that if the pair is higher than the threshold we consider it a match (e.g. if the score was correlated to the confidence the pair of images belong to the same person). For this a few metrics need to be calculated: genuine/impostor curve with it's d', ROC curve with the area under the curve, the FNMR and FMR. Depending on the specific problem we want to solve we will choose different thresholds. For example, in a surveillance application with very limited resources a false positive is high cost, since the resources needed used to access the suspect can leave other areas unattended. For this cases we want to choose a threshold with a FMR closer to 0.

Genuine and impostor distance distribution can be used to visualize the results, an example can be observed in figure 3.2. For this, we separate the testing results into a set with all the scores belonging to genuine pairs (both faces belong to the same person) and impostors pairs (the faces belong to different people). Then we assume a normal distribution for each set and plot them. The more separate the curves are the best the algorithm. From this data we can calculate the d', as define by equation 3.3, for a less ambiguous comparison.

$$d'(g,i) = \left| \frac{\mu_g - \mu_i}{\sqrt{(\sigma_g^2 + \sigma_i^2)/2}} \right|, \tag{3.3}$$

where g and i are vectors with all the scores for genuine and impostors pairs respectively, μ_n is the mean value for a vector n and σ_n is the standard deviation for a vector n.

In the other hand a Receiver Operating Characteristic (ROC) curve shows how a classification model performs at all thresholds, an example can be observed in figure 3.3. For

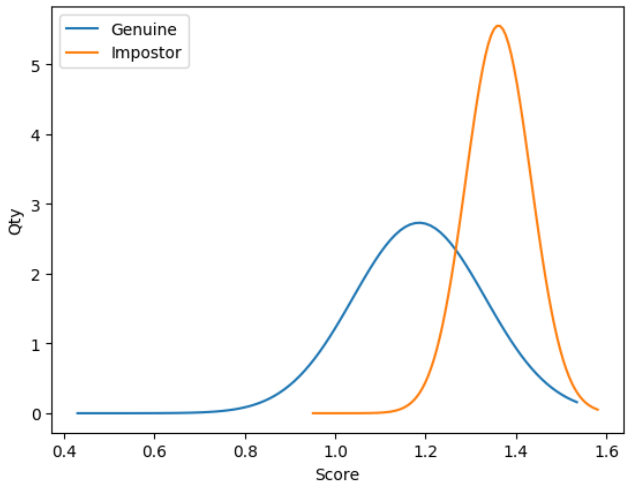


FIGURE 3.2. Example of a genuine/impostor curve.

this, we need to plot the False Match Rate (FMR, the rate at which the experiments miscategorizes a genuine pair as an impostor) at the X-axis and the False Non-Match Rate (FNMR, the rate at which the experiments miscategorizes a impostor pair as an genuine) at the Y-axis. An important metric to obtain from this curve is the area under the curve (AUC) for clear comparison between experiments.

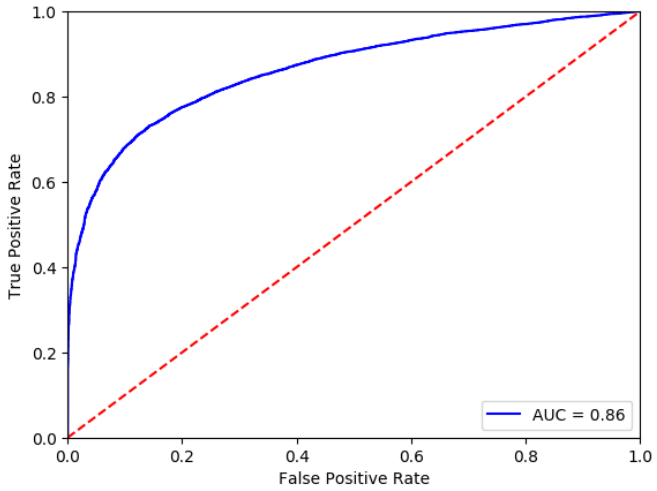


FIGURE 3.3. Example of a ROC curve.

3.2. PROPOSED METHODS

To make use of the protocol we need to choose algorithms for the up-scaling, feature extraction and feature comparison and compute the metrics as explain in 3.1.4. In this section we will describe combinations of several algorithms, two made specially for this work, creating a baseline for HR-mapping and another 12 methods to defy this baseline. Followed by the details of the two algorithms created for this work. In chapter 5 we will detail the results of this methods using the curves and metrics described in 3.1.4.

3.2.1. Methods for protocol evaluation

In this subsection we will explain the baseline and 12 methods to depict the use of the protocol. For each methods and the baseline we have clearly stated the algorithm used for up-scaling, feature extraction and feature comparison as seen in table 3.1.

3.2.1.1. Baseline

A baseline is a minimum point use for comparisons. For this purposes we decided to use bicubic interpolation as an up-scaling method, as it is widely used. For feature extraction we used Arcface (Deng et al., 2019) a state-of-the-art deep network for face recognition. Finally, for feature comparison we used Euclidean distance between the normalized vectors.

3.2.1.2. Bicubic Interpolation + Facenet

As a second choice for state-of-the-art feature extraction we decided to use Facenet (Schroff et al., 2015). For this we up-scaled both images to 160x160 pixels with bicubic interpolation. Finally, for feature comparison we used Euclidean distance between the normalized vectors.

3.2.1.3. Trained GAN + Arcface

For a first experiment we took the baseline and changed the up-scaling method for the Trained GAN from 3.2.2.1. With this we were looking to see if a different up-scaling method affects the results obtained.

TABLE 3.1. Summary of all the experiments and their components.

	Experiment name	Up-scaling method	Feat. extraction	Feat. comparison
1	Baseline	Bicubic interpolation	Arcface	Euclidean dist
2	Bicubic + Facenet	Bicubic interpolation	Facenet	Euclidean dist
3	Trained GAN + Arcface	Trained GANs	Arcface	Euclidean dist
4	Trained GAN + GAN Trained Siamese	Trained GANs	GAN Trained Siamese	Euclidean dist
5	Area + GAN Trained Siamese	re-sampling using pixel area relation	GAN Trained Siamese	Euclidean dist
6	Bicubic + GAN Trained Siamese	Bicubic interpolation	GAN Trained Siamese	Euclidean dist
7	Lanczos + GAN Trained Siamese	Lanczos interpolation over 8x8 neighborhood	GAN Trained Siamese	Euclidean dist
8	Nearest + GAN Trained Siamese	Nearest neighbor interpolation	GAN Trained Siamese	Euclidean dist
9	Area + Area Trained Siamese	re-sampling using pixel area relation	Area Trained Siamese	Euclidean dist
10	Bicubic + Bicubic Trained Siamese	Bicubic interpolation	Bicubic Trained Siamese	Euclidean dist
11	Lanczos + Lanczos Trained Siamese	Lanczos interpolation over 8x8 neighborhood	Lanczos Trained Siamese	Euclidean dist
12	Nearest + Nearest Trained Siamese	Nearest neighbor interpolation	Nearest Trained Siamese	Euclidean dist
13	ESRGAN + GAN Trained Siamese	ESRGAN	GAN Trained Siamese	Euclidean dist

3.2.1.4. Trained GAN + GAN Trained Siamese

In this experiment, we are using the Trained GAN from 3.2.2.1 to upscale all the images in the three lower resolutions. Then used this up-scaled images as test subjects for the GAN Trained Siamese as explained in 3.2.2.2. The objective of this experiment was to see how neuronal networks trained specific for the task would behave. A visual representation of this method for a (28x28, 112x112) images pair can be seen in figure 3.4.

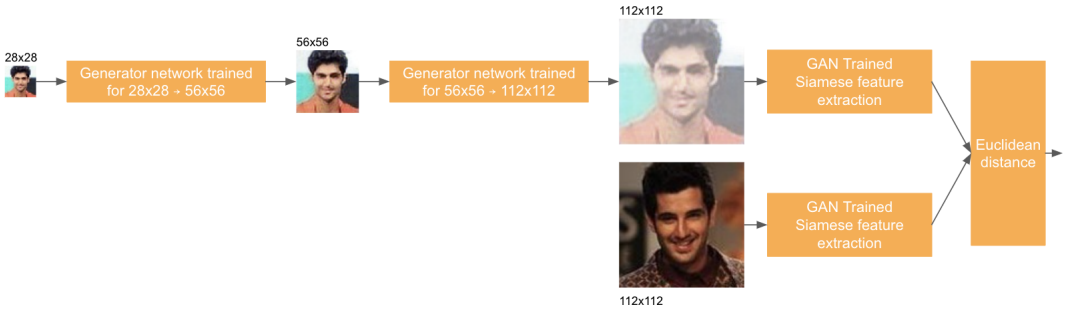


FIGURE 3.4. Visual representation of Trained GAN + GAN Trained Siamese method for a (28x28, 112x112) images pair.

3.2.1.5. Interpolations + GAN Trained Siamese

For this experiment we wanted to challenge the need of a neuronal network for the up-scaling method. In particular, we ran 4 experiments with known interpolations described in the following paragraphs. For each interpolation we obtained the up-scaled images for the three lower resolutions. Then we used this up-scaled images and the GAN Trained Siamese as explained in 3.2.2.2 to get the features.

Area Interpolation: a re-sampling using pixel area relation is calculated as implemented by OpenCV (Bradski, 2000).

Bicubic Interpolation: (Keys, 1981)interpolates on a interval using a polynomial as implemented by OpenCV (Bradski, 2000).

Lanczos Interpolation: Lanczos interpolation over 8x8 neighborhood as implemented by OpenCV (Bradski, 2000).

Nearest Neighbor Interpolation: selects the value of the neighboring point as implemented by OpenCV (Bradski, 2000).

3.2.1.6. Interpolations + Interpolations Trained Siamese

As the experiment in 3.2.1.5 wasn't challenging the need of neuronal networks enough (the Siamese network was trained using the images up-scale with the GANs in 3.2.2.1), we trained the same Siamese architecture using the interpolation up-scaled pairs as input. This training was done separately for each of the interpolations.

3.2.1.7. ESRGAN + Trained Siamese

ESRGAN (X. Wang et al., 2018) is a neuronal network based in SRGAN with improvements in three aspects: 1) adopting a deeper model using Residual-in-Residual Dense Block without batch normalization layers, 2) employing a Relativistic average GAN instead of the vanilla GAN and 3) modifying the perceptual loss by using the features before activation. For this experiment we used the code and trained models provided by the author to upscale our dataset images. Then used the Siamese network as described in 3.2.2.2.

3.2.2. Proposed components

In this section we will explain the two components mentioned in 3.2.1. First is a set of GAN networks used for the up-scaling part of the protocol. Secondly a Siamese network for feature extraction.

We chose a GAN structure for up-scaling because results for GAN-based methods result in a more human appealing image than similar structures without GANs, as seen in (Ledig et al., 2017). GAN solutions apply a generative network to create new information from a seed and utilize a discriminator network to defy the generated image by discriminating between real and generated HR. More details on the implementation can be seen in 3.2.2.1.

For feature extraction we chose a Siamese convolutional network architecture because it uses the same weights while working through a HR and up-scaled-LR pair creating comparable output vectors. In addition to a predefined distance metric, Euclidean distance in our case, the learning goal is to minimize the distance for genuine pairs and minimize it for impostor pairs.

3.2.2.1. Trained GAN

We have created a GAN network to up-scale images x2 based on SRGAN architecture (Ledig et al., 2017). A visual representation of the architecture can be found in figure 3.5. Each Residual Block is composed of a convolutional layer, followed by a parametric ReLu layer, a second convolutional layer and finally a element-wise sum layer. Differently, a Discriminator Block is composed of a convolutional layer and a leaky ReLu.

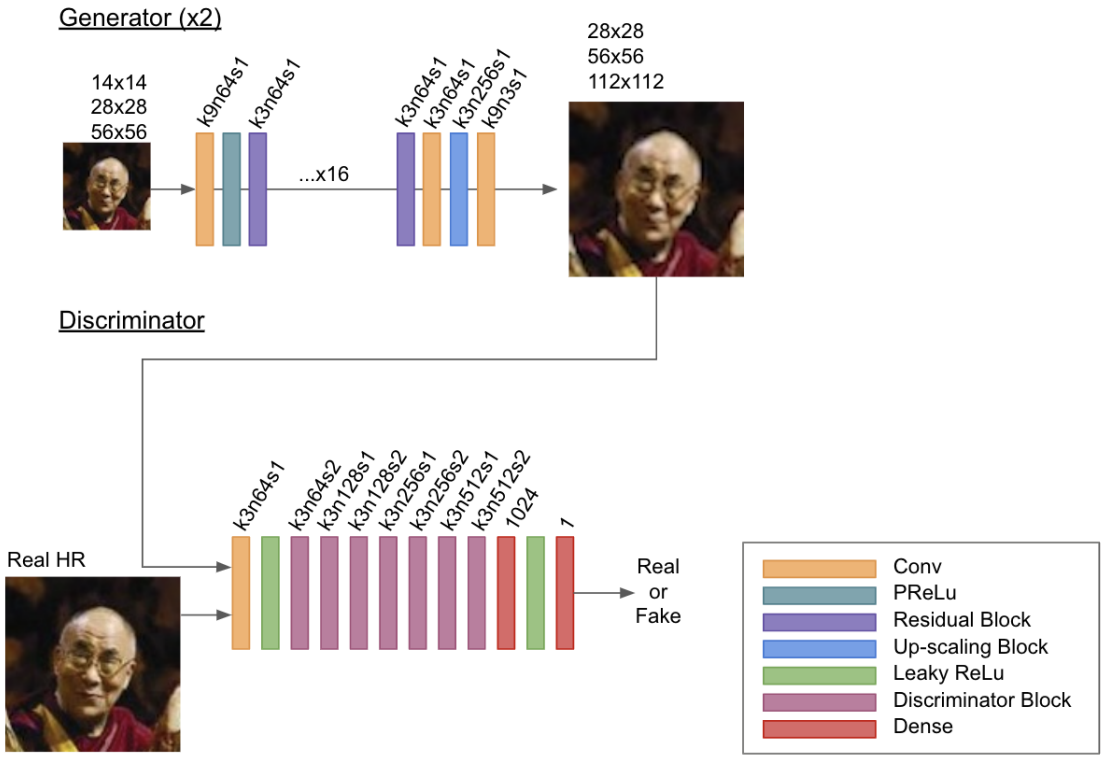


FIGURE 3.5. Visual representation of the created GAN network.

The created GAN does a two times scaling by each pass of the image. For this work, three different GANs where trained with the same architecture: 1) from 14x14 to 28x28, 2) from 28x28 to 56x56 and 3) from 56x56 to 112x112. The Discriminator network was trained using binary cross-entropy as a loss function. The Generator Network was trained with VGG Loss function as describe in (Ledig et al., 2017). There is one exception, for the network with input image 14x14 binary cross-entropy was used instead since such a lower resolution couldn't be fed to VGG Loss function.

Every GAN was trained with Adam optimizer, $learningrate = 1E-4, beta_1 = 0.9, beta_2 = 0.999, epsilon = 1e-08$. The training was done in a GTX 1080 TI for 150 epochs. Sample images of the results can be found in figures 3.6, 3.7 and 3.8 for 14x14, 28x28 and 56x56 respectively.

FIGURE 3.6. Sample images from the test set for 14x14 pixels after passing through three trained generator networks, as to achieve 3 times a x2 scales. The final resolution is 112x112 pixels.

FIGURE 3.7. Sample images from the test set for 28x28 pixels after passing through two trained generator networks, as to achieve 2 times a x2 scales. The final resolution is 112x112 pixels.

We can observe that images for 56x56 (figure 3.8) scale-up look like standard faces with some blurriness and no visual artifacts can be easily seen. In the samples for the 28x28 set (figure 3.7), we can see some artifacts causing minor deformities in the faces and an amount of blurriness. For the 14x14 set (figure 3.6) the appearance of a face is not easily seen and there is a significant amount of blurriness in the images. It should be noticed each pass through one of the trained GANs outputs an image slightly lighter than the original one. For the 28x28 and 14x14 sets it becomes noticeable since they need to pass through a GAN 2 and 3 times respectively.

FIGURE 3.8. Sample images from the test set for 56x56 pixels after passing through a trained generator network, as to achieve a x2 scale. The final resolution is 112x112 pixels.

3.2.2.2. Trained Siamese

For feature extraction we built a Siamese network where the columns are based on the VGG Face Architecture (Parkhi et al., 2015). A visual representation of a column architecture can be seen in figure 3.9. It should be noted that each convolution had a ReLu activation.

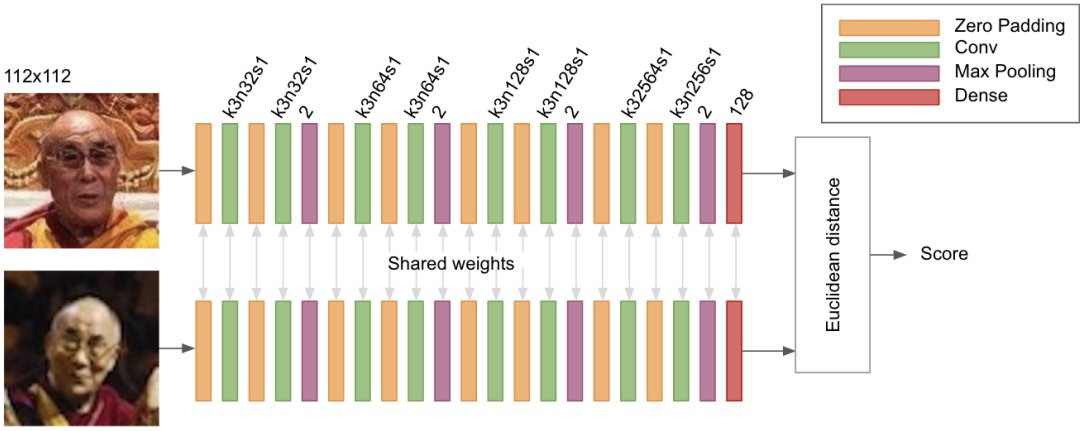


FIGURE 3.9. Visual representation of the trained Siamese network.

For each dataset one column receive the real 112x112 image, while the second received a up-scaled image. This Siamese architecture was trained using the 14x14, 28x28 and 56x56 pair sets obtain from VGGFace 2 dataset as described in 5.1. The up-scaled method used to training separately where: the trained GAN model as explained in 3.2.2.1, area interpolation, bicubic interpolation, Lanczos interpolation over an 8x8 neighbor and

nearest neighbor interpolation generating each one model. Each model was trained until the validation accuracy showed no improvements for at three epochs, saving the best epoch weights as the final weights. The training was done in a GTX 1080 TI with a RMSprop optimizer, $learningrate = 1E - 5$ and contrastive loss.

4. PROPOSED DATASET PROTOCOL

Within low resolution face recognition finding a good dataset is a challenge of its own. Having a common dataset is key for research comparison but not enough. For claiming the state-of-the-art methods requires using the same testing set to obtain comparable results, meaning for each work that is not on a known dataset needs to be replicate and obtain results again in a standard testing set.

It is important when choosing a dataset to keep in mind: 1) Is there enough images and subjects to train heavy models?, 2) Are the images of reasonable quality and sharp enough?, 3) Do the images represent the problem you are looking to solve? (i.e, if we are training an algorithm to recognize people outdoors, are the images in my dataset outdoors or indoors?), 4) Are the images real low/high resolution or a big downscale/upscale needed?

The first step for creating the protocol was to choose the resolutions for the dataset. For high resolution images, previous work has been seen to use 224x224, for example (Schroff et al., 2015; Szegedy et al., 2015) or 112x112 pixels, for example (Deng et al., 2019). We decide to use the latter to reduce the gap between the number of pixels in the high and low resolutions.

As for low resolutions, several articles have recognized that a face with a tight bounding box smaller than 32x32 pixels begin to present significant accuracy challenge to face recognition systems (Boom, Beumer, Spreeuwers, & Veldhuis, 2006). According to (Harmon, 1973), 16x16 is very close to the minimum resolution that allows for facial recognition. For the protocol a 14x14 pixels resolution is a good fit, this way the up-scaling algorithms can be set to do a times 8 augmentation instead of a decimal one. We created a second low resolution group of 28x28 pixels, in some datasets it is fair to consider it low resolution if the images have a reasonable amount of background to account for. Finally, for a third group, faces of 56x56 pixels is recommended since some up-scaling algorithm would do a x2 three times instead of one augmentation of x8. Images depicting faces in the different resolutions can be seen in figure 4.1.

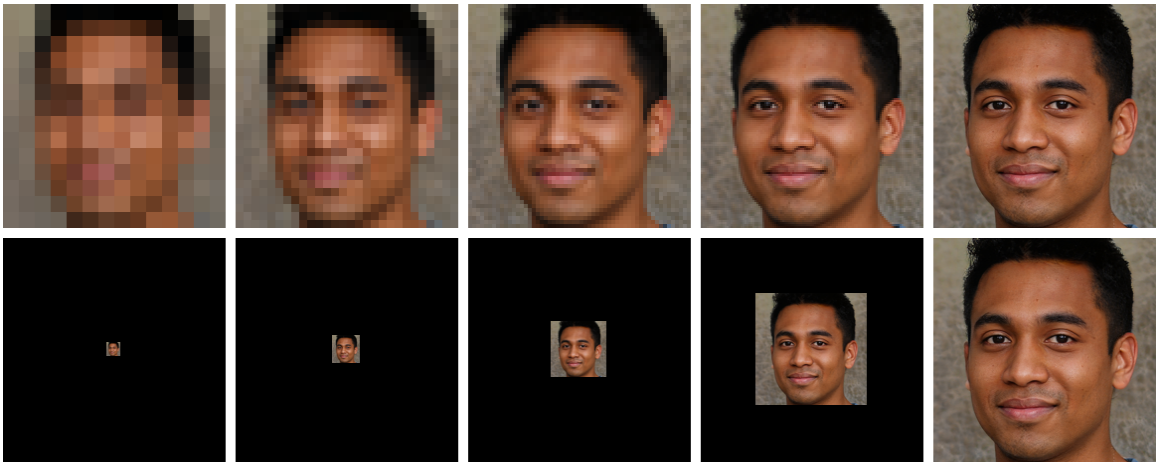


FIGURE 4.1. In the first row we can see the same image which has been down-scale and then up-scale to different resolutions. In the second row we can observe the same image which have been only down-scale to different resolutions. The resolutions are 14x14, 28x28, 56x56, 112x112 and 224x224 from left to right. It is clear how difficult is to recognize a face of 14x14 pixels.

If any interpolation is needed, bicubic interpolation should be used. Since inorganic information is created when up-scaling an image special care needs to be dedicated so never to do a bicubic interpolation from a lower resolution to a higher one. Also, big down-scales are not recommended since real low resolution images have more types of challenges than just low resolution (i.e, blurriness, artifact due to low-rate compression, acquisition conditions that add noise to the image) (Li et al., 2018).

For the pair of high resolution and low resolution images, two algorithm where done: one to create pairs between two different resolutions (Algorithm 1) and a second one for within the same resolution (Algorithm 2) in case it is needed for experiments. The algorithms ensure that there is an equal amount of positive and negative pairs for training and that each image from the lower resolution is at least on one pair.

For pair creation between two different resolutions, we start by creating a dictionary where each key is a subject id from the upper resolution set and the values are the names for the images each subject has. Then, we start going over every subject in the lower resolution set and check if the subject exists in the upper resolution set. If so, we pick a random image from this subject and save it as a positive pair. Afterwards we compute the negative pair or

Algorithm 1: Pair creation between two different resolutions

```
Result: File with each pair and label as a line
Input: Path to the directory with all the images
Let baseImages be a dictionary where the key is the subject id and the value is an array of
all the images from that subject in the higher resolution;
for each image in the lower resolution do
    Get the subject id of the image;
    if the subject exists as key in baseImages then
        Choose randomly between the images on the value of baseImages[subject id];
        Write this a a positive pair;
        if It is a training pair creation then
            Pick a random subject different to the initial one and a random image within
            baseImages[random subject id];
            Write this as a negative pair;
        else
            for each different subject as key in baseImages do
                Pick a random image within baseImages[different subject id];
                Write this as a negative pair;
            end
        end
    else
        Skip to the next image;
    end
end
```

pairs depending if it is a training set or testing set respectively. For training, we proceed to choose a subject from the upper resolution and image of that subject randomly (the subject needs to be different from the lower resolution) and save it as a negative pair. For testing, for each subject in the upper resolution a random image is picked and saved as a negative pair.

For pair creation within the same resolution, we also start by creating a dictionary with all the subjects and images per subject of the resolution set. Then for each subject in the dictionary, if the subject has at least two images, we iterate through all the subject images picking another random image from the subject and saving it as a positive pair. Then we compute the negative pair or pairs depending if it is a training (positive and negative should have the same amount) or testing respectively. For training, a random subject and image from that subject is picked randomly and saved as a negative pair. For testing, we iterate for

Algorithm 2: Pair creation within the same resolution

```
Result: File with each pair and label as a line
Input: Path to the directory with all the images
Let baseImages be a dictionary where the key is the subject id and the value is an array of
all the images from that subject in that resolution set;
for each subject id/key on baseImages do
    if the subject has at least two images then
        for each image within baseImages[subject id] do
            Pick random image from the same subject (different than itself);
            Write this as the positive pair;
            if It is a training pair creation then
                Pick random subject different to the initial one and image within
                baseImages;
                Write this as negative pair;
            else
                for each different subject as key in baseImages that different to the initial
                subject do
                    Pick a random image within baseImages[different subject id];
                    Write this as a negative pair;
                end
            end
        end
    else
        Pass to the next subject;
    end
end
```

each different subject in the dictionary picking a random image and saving it as a negative pair.

It should be noted that this protocol was thought for datasets with a significant amount of subjects and images per subjects. So the pair lists are constructed with the use of randomly picking images and subject in order to avoid having an unmanageable amount of pairs. If the protocol wants to be applied for smaller datasets, instead of using randoms, all possible combination can be computed, always keeping in mind that the amount of positive and negative pairs should be the same for training.

On the other hand, for big dataset computing all the possible subjects for negative pair combination can cause a unmanageable amount of pairs for testing. In cases like this, we recommend obtaining one negative pair for each positive pair (same as training).

Afterwards we need to prove the pair set obtained is representative of the testing set, several pair set should be obtained and observe how much the results vary with different pair sets. If the results variance is little to none then we can consider the pair sets representative.

New datasets are constantly being created for specific needs. We believe each dataset owner should use this protocol as a guideline for creating their pairs for training, validation and testing sets and publish them together with the dataset itself.

5. EXPERIMENTAL RESULTS

We begin this section by describing how the dataset. protocol can be applied to VGG Face 2 to create two sets of genuine and impostor HR-LR pairs for training and testing. As explained in section 3.1, a experiment should have four parts to comply with the protocol. In section3.2.1 we described our choice of algorithms for up-scaling, feature extraction and feature comparison in a baseline and 12 other methods. In subsection 5.2 we will describe the results obtained with the different methods computing the genuine/impostor curve with it's d', ROC curve with the area under the curve, the FNMR and FMR as described in 3.1.4. We close this section with some global analysis and conclusions.

5.1. DATASET PROTOCOL WITH VGG FACE 2

In this section, we will explain how we applied the dataset protocol to VGG Face 2 dataset released in 2018 (Cao, Shen, Xie, Parkhi, & Zisserman, 2018). It contains more than 9000 identities with an average of 362 images per identity and a total of more than 3.3 million faces. VGG face 2 images are centered on the face allowing some background to be present. It is important to know that more than half of this images have a resolution lower than 100 pixels with a significant amount under 50 pixels. All the faces are captured in the wild with variations in pose, emotions, lighting and occlusion situation.

First we noticed most of the images were not square but had enough background to crop. To keep as much of the image as possible we grabbed the difference between height and width and cropped the largest one. The difference was divided into two and then the image cropped in both sides, keeping the face centered and the image squared in the final result.

To correctly divide the images in the four resolutions (14x14, 28x28, 56x56 and 112x 112 pixels), we first ordered them by height and counted the amount of images under 112 pixels (a total of 490.692 for the training set and 26.371 images for the testing set). Then we divided the total number of images by three (163.564 images for training, 8.791 images for testing) and started down-scaling to the lower resolution by order (14x14, 28x28, 56x56)

until each set had as many as 163.564 images for training and 8.791 images for testing. Such that the smallest 163.564 images where down-scaled to 14x14 for the training set and so on. This way the minimum incorrectness in real low resolution was added as possible. Afterwards, the same number of images was picked by resolution order of a resolution higher than 112 so that this set was also composed of 163.564 images for training and 8.791 images in testing.

It is important to notice, that no up-scale was done while creating this sets, therefore no new information was created in this step. For all down-scaling we used a bicubic interpolation. Samples of the images in each group before manipulation can be found on figures 5.1, 5.2 and 5.3 for 14x14, 28x28 and 56x56 sets respectively.

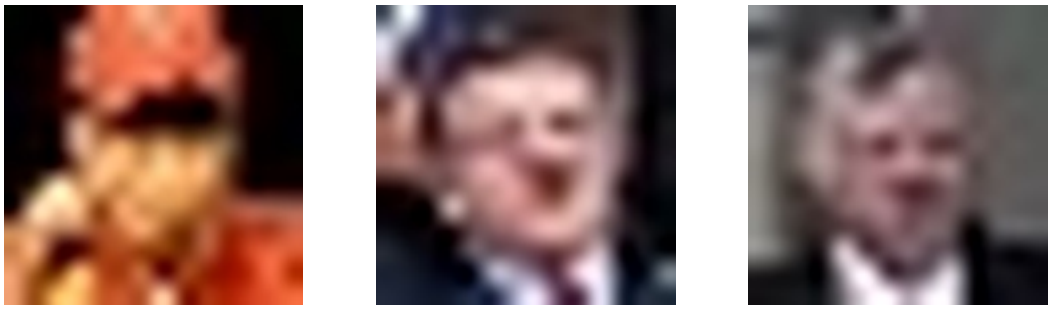


FIGURE 5.1. Sample images for 14x14 set with VGG Face 2 dataset (images have been upscale with bicubic interpolation for viewing purposes).

FIGURE 5.2. Sample images for 28x28 set with VGG Face 2 dataset (images have been upscale with bicubic interpolation for viewing purposes).

Pairs for the following sets matches were created (14x14, 112x112), (28x28, 112x112) and (56x56, 112x112) using algorithm 1 and (112x112, 112x112) using algorithm 2 for the experiments in this work. Given the size of the dataset and the amount of the images

FIGURE 5.3. Sample images for 56x56 set with VGG Face 2 dataset (images have been upscale with bicubic interpolation for viewing purposes).

not all negative pairs for testing where created as specified in the algorithms in section 4. The testings pairs were created with the same method as training pairs, a random subject id different to the current one with a random image was picked for the negative pair giving the same amount of positive and negative pairs. Since this is a smaller subset of the testing pairs the pair creation was repeated several times checking for variations in results, there where only variation on the second decimal number not making any change in which experiment was better or worse.

5.2. EXPERIMENTS WITH HR-MAPPING PROTOCOL

It is very important for evaluation that we use a testing set to match the characteristics of the real world problem we are aiming to solve. In table 5.1 we have made a comparison of the most common dataset that can be used for LR Face Re-identification: VGG Face 2 (Cao et al., 2018), SCFace (Grgic, Delac, & Grgic, 2011), ICB-RW (Neves & Proença, 2016) and LFW (Huang, Mattar, Berg, & Learned-Miller, 2008). We can see that both ICB-RW and VGG Face have more real world characteristics. Also, VGG Face 2 has significantly more identities but lacks an existing protocol making it harder for researchers to compare their results. In this section we will detail the results with the testing protocol constructed as explained in section 5.1. Subsections will detail the results and analysis for the developed experiments as explained in 3.2.

TABLE 5.1. Comparison of commonly used datasets in LR Face Re-identification.

	VGG Face 2	SCFace	ICB-RW	LFW
Number of identities in testing set	500	130	90	697
Captured in the wild or closed set	Wild	Closed	Wild	Wild
Some low resolution	Yes	Yes	Yes	No
Within subject variance in pose	Yes	Yes	Yes	Yes
Within subject variance in age	Yes	No	No	No
Variance in lightning	Yes	No	Yes	Yes
Surveillance Quality	No	Yes	Yes	No
Publishing year	2018	2011	2015	2007
Existing protocol	No	Yes	Yes	Yes

5.2.1. Specific Results

In the following subsections we will display the experimental results and analysis for all the methods described in 3.2.1.

5.2.1.1. Baseline

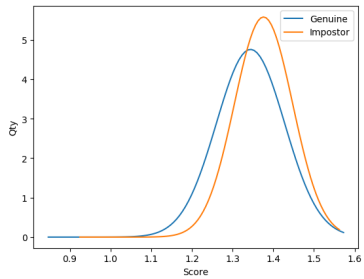
A genuine/impostor and ROC curve for each resolution can be seen at figures 5.4 and 5.5 respectively. We can see the baseline performance in the 56x56 set is quite good but as the resolution gets lower performance drops significantly.

5.2.1.2. Bicubic + Facenet

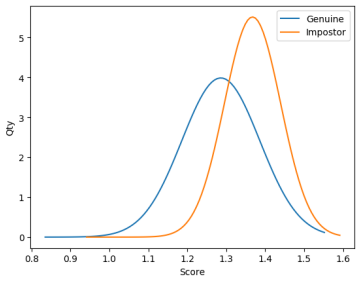
A genuine/impostor and ROC curve for each resolution can be seen at figures 5.6 and 5.7 respectively. We can see significant improvement from the baseline for the 28x28 and 56x56 sets. For the 14x14 set we can see the results get have some improvement in comparison to the baseline.

5.2.1.3. Trained GAN + Arcface

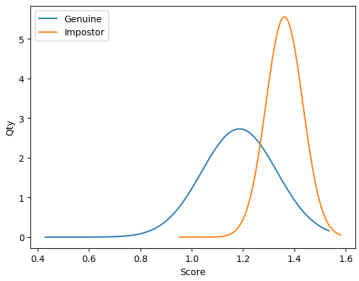
A genuine/impostor and ROC curve for each resolution can be seen at figures 5.8 and 5.9 respectively. We can see some improvement from the baseline for 56x56 set and a



(a) Genuine/Impostor curve for Baseline with LR 14x14 dataset ($d' = 0.411$).

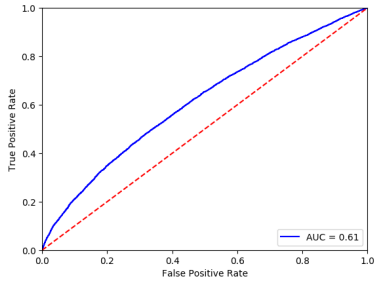


(b) Genuine/Impostor curve for Baseline with LR 28x28 dataset ($d' = 0.933$).

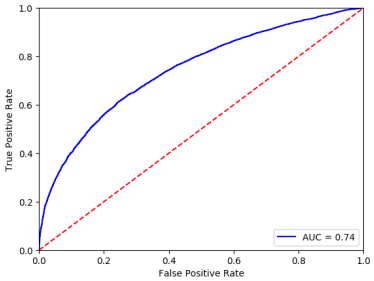


(c) Genuine/Impostor curve for Baseline with LR 56x56 dataset ($d' = 1.523$).

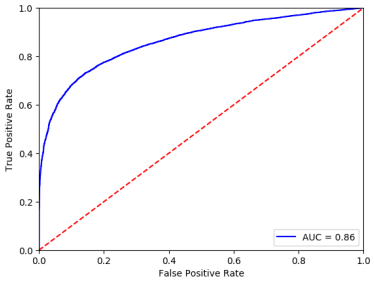
FIGURE 5.4. Genuine/Impostor curves for Baseline for LR 14x14, 28x28 and 56x56 respectively and HR 112x112.



(a) ROC curve for Baseline with LR 14x14 dataset ($AUC = 0.61$).

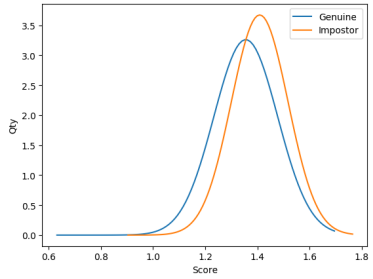


(b) ROC curve for Baseline with LR 28x28 dataset ($AUC = 0.74$).

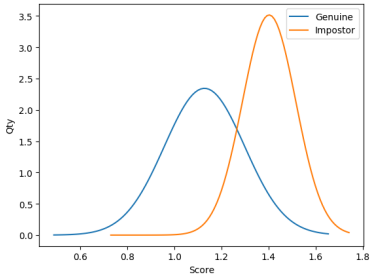


(c) ROC curve for Baseline with LR 56x56 dataset ($AUC = 0.86$).

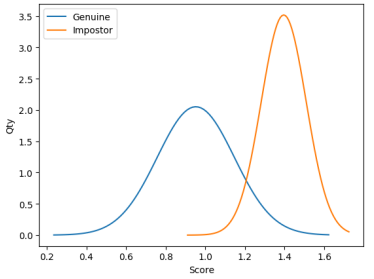
FIGURE 5.5. ROC curves for Baseline for LR 14x14, 28x28 and 56x56 respectively and HR 112x112.



(a) Genuine/Impostor curve for Bicubic + Facenet with LR 14x14 dataset ($d' = 0.462$).

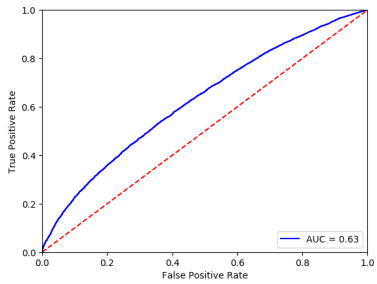


(b) Genuine/Impostor curve for Bicubic + Facenet with LR 28x28 dataset ($d' = 1.900$).

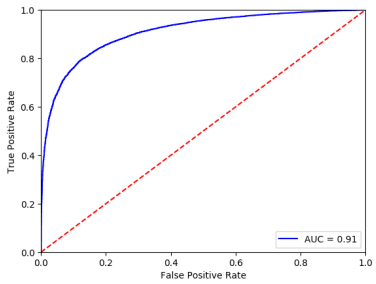


(c) Genuine/Impostor curve for Bicubic + Facenet with LR 56x56 dataset ($d' = 2.787$).

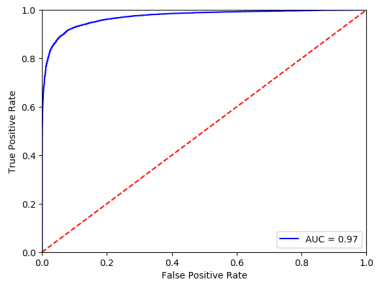
FIGURE 5.6. Genuine/Impostor curves for Bicubic + Facenet for LR 14x14, 28x28 and 56x56 respectively and HR 112x112.



(a) ROC curve for Bicubic + Facenet with LR 14x14 dataset ($AUC = 0.63$).



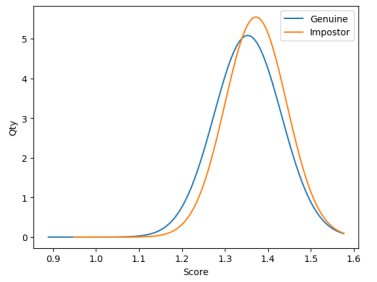
(b) ROC curve for Bicubic + Facenet with LR 28x28 dataset ($AUC = 0.91$).



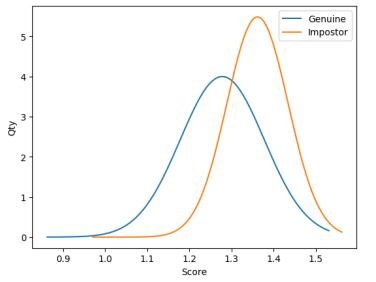
(c) ROC curve for Bicubic + Facenet with LR 56x56 dataset ($AUC = 0.97$).

FIGURE 5.7. ROC curves for Bicubic + Facenet for LR 14x14, 28x28 and 56x56 respectively and HR 112x112.

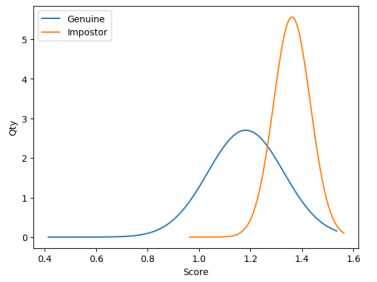
smaller improvement to the 28x28 set. For the 14x14 set we can see the results get worse than baseline.



(a) Genuine/Impostor curve for Trained GAN + Arcface with LR 14x14 dataset ($d' = 0.411$).



(b) Genuine/Impostor curve for Trained GAN + Arcface with LR 28x28 dataset ($d' = 0.933$).

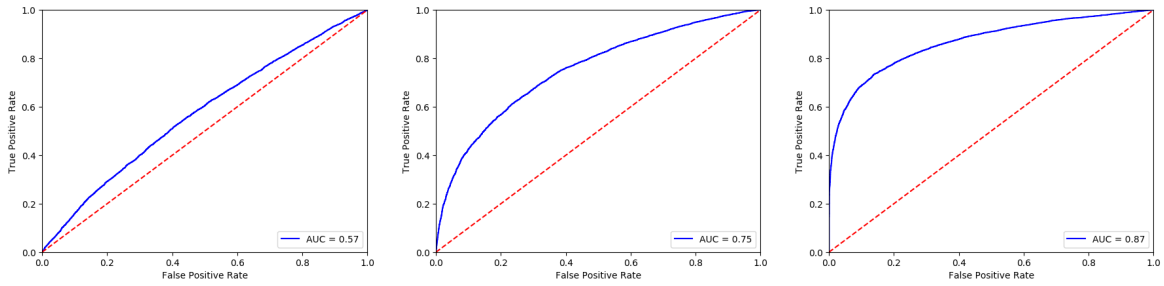


(c) Genuine/Impostor curve for Trained GAN + Arcface with LR 56x56 dataset ($d' = 1.523$).

FIGURE 5.8. Genuine/Impostor curves for Trained GAN + Arcface for LR 14x14, 28x28 and 56x56 respectively and HR 112x112.

5.2.1.4. Trained GAN + GAN Trained Siamese

A genuine/impostor and ROC curve for each resolution can be seen at figures 5.10 and 5.11 respectively. We can observe that performance for the 14x14 and 28x28 improves significantly in contrast to baseline results. The 56x56 set shows a small decay in performance in contrast to baseline.

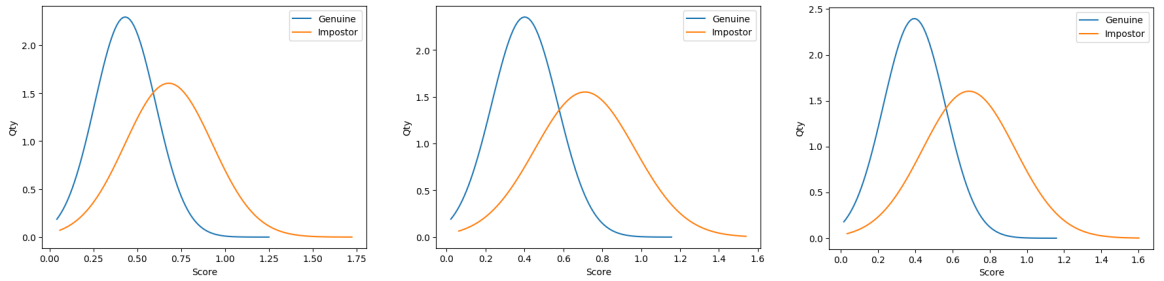


(a) ROC curve for Trained GAN + Arcface with LR 14x14 dataset ($AUC = 0.61$).

(b) ROC curve for Trained GAN + Arcface with LR 28x28 dataset ($AUC = 0.74$).

(c) ROC curve for Trained GAN + Arcface with LR 56x56 dataset ($AUC = 0.86$).

FIGURE 5.9. ROC curves for Trained GAN + Arcface for LR 14x14, 28x28 and 56x56 respectively and HR 112x112.



(a) Genuine/Impostor curve for Trained GAN + GAN Trained Siamese with LR 14x14 dataset ($d' = 0.411$).

(b) Genuine/Impostor curve for Trained GAN + GAN Trained Siamese with LR 28x28 dataset ($d' = 0.933$).

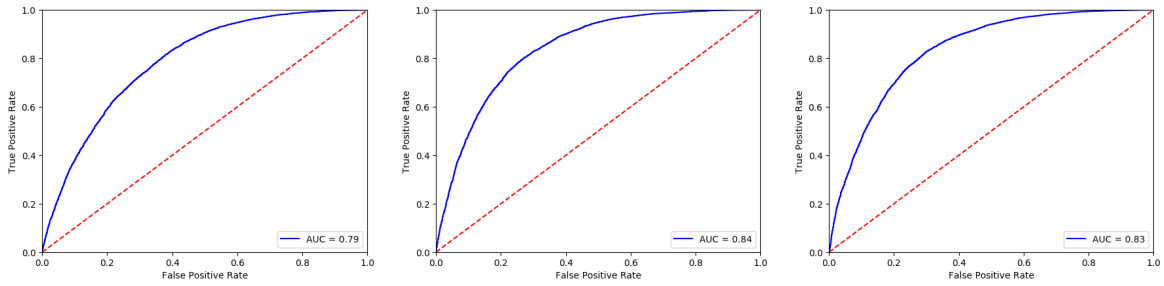
(c) Genuine/Impostor curve for Trained GAN + GAN Trained Siamese with LR 56x56 dataset ($d' = 1.523$).

FIGURE 5.10. Genuine/Impostor curves for Trained GAN + GAN Trained Siamese for LR 14x14, 28x28 and 56x56 respectively and HR 112x112.

5.2.1.5. Interpolations + GAN Trained Siamese

In the following paragraphs we will show the results for the four experiments done with a handcrafted interpolation for the up-scale and used the GAN Trained Siamese, as mentioned in 3.2.2.2, for feature extraction. In general, we can see a small performance increase for the 14x14 set when comparing to the baseline. Performance for the 28x28 and 56x56 sets drops significantly.

Area Interpolation: a genuine/impostor and ROC curve for each resolution can be seen at figures 5.12 and 5.13 respectively.

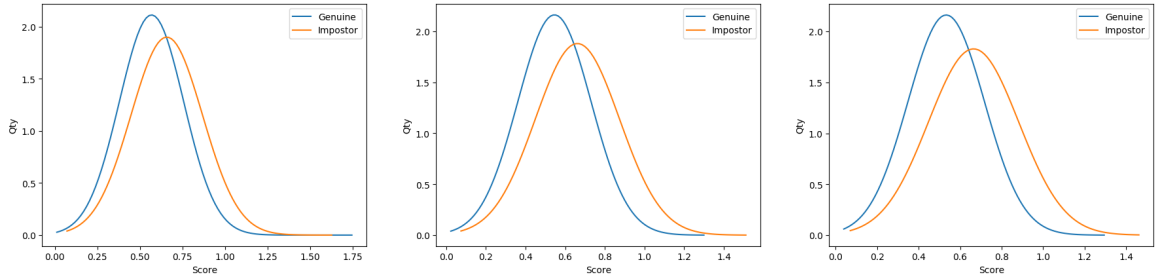


(a) ROC curve for Trained GAN + GAN Trained Siamese with LR 14x14 dataset ($AUC = 0.61$).

(b) ROC curve for Trained GAN + GAN Trained Siamese with LR 28x28 dataset ($AUC = 0.74$).

(c) ROC curve for Trained GAN + GAN Trained Siamese with LR 56x56 dataset ($AUC = 0.86$).

FIGURE 5.11. ROC curves for Trained GAN + GAN Trained Siamese for LR 14x14, 28x28 and 56x56 respectively and HR 112x112.



(a) Genuine/Impostor curve for Area + GAN Trained Siamese with LR 14x14 dataset ($d' = 1.202$).

(b) Genuine/Impostor curve for Area + GAN Trained Siamese with LR 28x28 dataset ($d' = 1.400$).

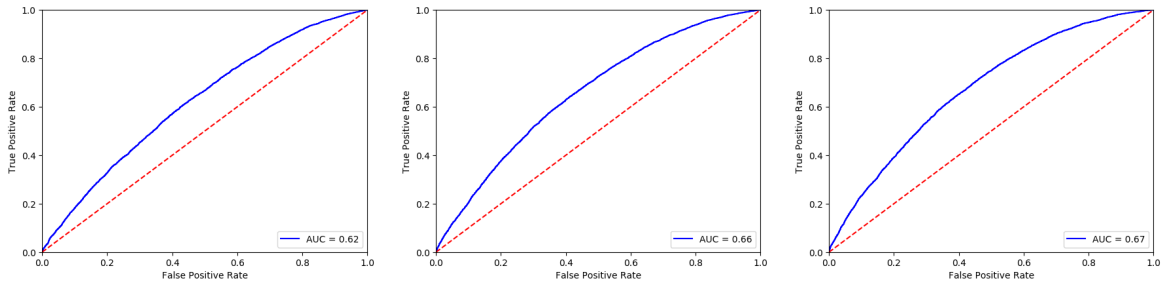
(c) Genuine/Impostor curve for Area + GAN Trained Siamese with LR 56x56 dataset ($d' = 1.420$).

FIGURE 5.12. Genuine/Impostor curves for Area + GAN Trained Siamese for LR 14x14, 28x28 and 56x56 respectively and HR 112x112.

Bicubic Interpolation: a genuine/impostor and ROC curve for each resolution can be seen at figures 5.14 and 5.15 respectively.

Lanczos Interpolation: a genuine/impostor and ROC curve for each resolution can be seen at figures 5.16 and 5.17 respectively.

Nearest Neighbor Interpolation: a genuine/impostor and ROC curve for each resolution can be seen at figures 5.18 and 5.19 respectively.

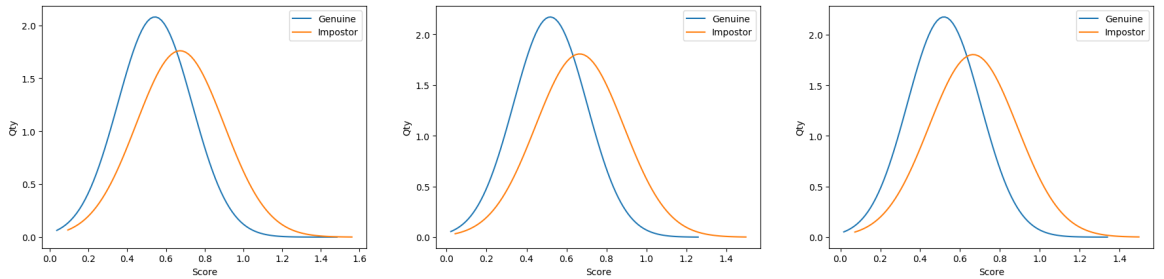


(a) ROC curve for Area + GAN Trained Siamese with LR 14x14 dataset ($AUC = 0.80$).

(b) ROC curve for Area + GAN Trained Siamese with LR 28x28 dataset ($AUC = 0.84$).

(c) ROC curve for Area + GAN Trained Siamese with LR 56x56 dataset ($AUC = 0.84$).

FIGURE 5.13. ROC curves for Area + GAN Trained Siamese for LR 14x14, 28x28 and 56x56 respectively and HR 112x112.



(a) Genuine/Impostor curve for Bicubic + GAN Trained Siamese with LR 14x14 dataset ($d' = 0.619$).

(b) Genuine/Impostor curve for Bicubic + GAN Trained Siamese with LR 28x28 dataset ($d' = 0.724$).

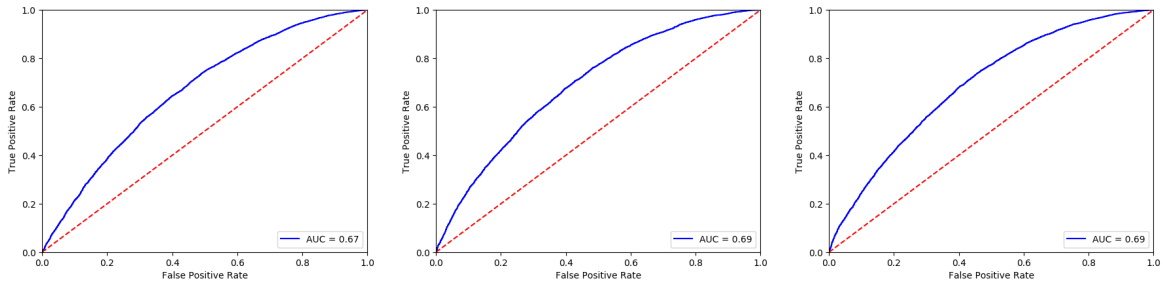
(c) Genuine/Impostor curve for Bicubic + GAN Trained Siamese with LR 56x56 dataset ($d' = 0.720$).

FIGURE 5.14. Genuine/Impostor curves for Bicubic + GAN Trained Siamese for LR 14x14, 28x28 and 56x56 respectively and HR 112x112.

5.2.1.6. Interpolations + Interpolations Trained Siamese

In the following paragraphs we will show the results for the four experiments done with a handcrafted interpolation for the up-scale and used the interpolation Trained Siamese, as mentioned in 3.2.2.2, for feature extraction. In general, we can see a significant improvement in the 14x14 and 28x28 sets when comparing with the baseline. Also, a small performance drop in the 56x56 set.

Area Interpolation: a genuine/impostor and ROC curve for each resolution can be seen at figures 5.20 and 5.21 respectively.

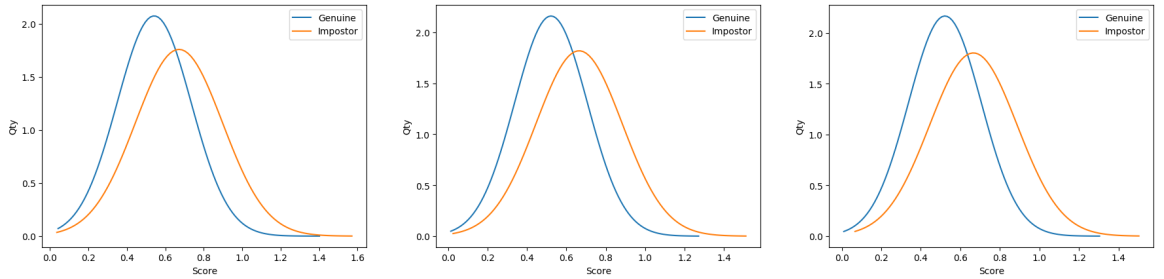


(a) ROC curve for Bicubic + GAN Trained Siamese with LR 14x14 dataset ($AUC = 0.67$).

(b) ROC curve for Bicubic + GAN Trained Siamese with LR 28x28 dataset ($AUC = 0.69$).

(c) ROC curve for Bicubic + GAN Trained Siamese with LR 56x56 dataset ($AUC = 0.69$).

FIGURE 5.15. ROC curves for Bicubic + GAN Trained Siamese for LR 14x14, 28x28 and 56x56 respectively and HR 112x112.



(a) Genuine/Impostor curve for Lanczos + GAN Trained Siamese with LR 14x14 dataset ($d' = 0.613$).

(b) Genuine/Impostor curve for Lanczos + GAN Trained Siamese with LR 28x28 dataset ($d' = 0.704$).

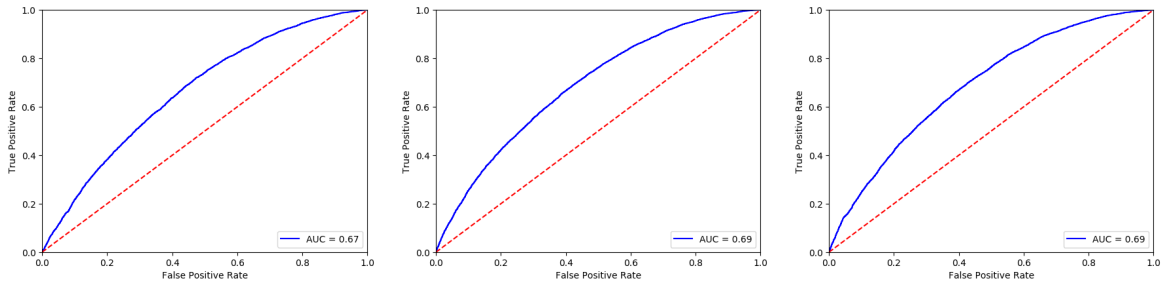
(c) Genuine/Impostor curve for Lanczos + GAN Trained Siamese with LR 56x56 dataset ($d' = 0.708$).

FIGURE 5.16. Genuine/Impostor curves for Lanczos + GAN Trained Siamese for LR 14x14, 28x28 and 56x56 respectively and HR 112x112.

Bicubic Interpolation: a genuine/impostor and ROC curve for each resolution can be seen at figures 5.22 and 5.23 respectively.

Lanczos Interpolation: a genuine/impostor and ROC curve for each resolution can be seen at figures 5.24 and 5.25 respectively.

Nearest Neighbor Interpolation: a genuine/impostor and ROC curve for each resolution can be seen at figures 5.26 and 5.27 respectively.

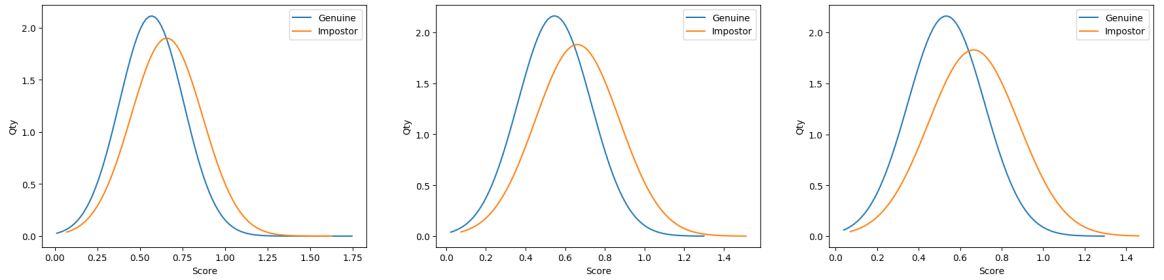


(a) ROC curve for Lanczos + GAN Trained Siamese with LR 14x14 dataset ($AUC = 0.67$).

(b) ROC curve for Lanczos + GAN Trained Siamese with LR 28x28 dataset ($AUC = 0.69$).

(c) ROC curve for Lanczos + GAN Trained Siamese with LR 56x56 dataset ($AUC = 0.69$).

FIGURE 5.17. ROC curves for Lanczos + GAN Trained Siamese for LR 14x14, 28x28 and 56x56 respectively and HR 112x112.



(a) Genuine/Impostor curve for Nearest + GAN Trained Siamese with LR 14x14 dataset ($d' = 0.448$).

(b) Genuine/Impostor curve for Nearest + GAN Trained Siamese with LR 28x28 dataset ($d' = 0.582$).

(c) Genuine/Impostor curve for Nearest + GAN Trained Siamese with LR 56x56 dataset ($d' = 0.653$).

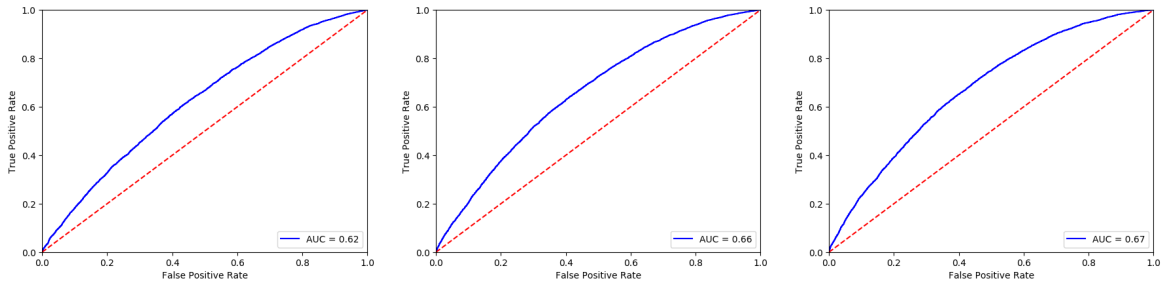
FIGURE 5.18. Genuine/Impostor curves for Nearest + GAN Trained Siamese for LR 14x14, 28x28 and 56x56 respectively and HR 112x112.

5.2.1.7. ESRGAN + Trained Siamese

A genuine/impostor and ROC curve for each resolution can be seen at figures 5.28 and 5.29 respectively. We can see a small increase in performance for set 14x14, a small decrease in performance for set 28x28 and a significant decrease in performance for thee 56x56 set.

5.2.2. Global Results and Analysis

For an easier comparison we have summarized the results in the following table 5.2.

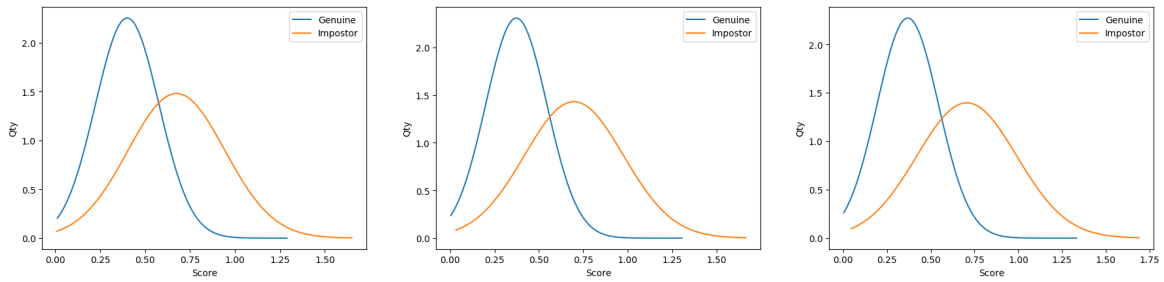


(a) ROC curve for Nearest + GAN Trained Siamese with LR 14x14 dataset ($AUC = 0.62$).

(b) ROC curve for Nearest + GAN Trained Siamese with LR 28x28 dataset ($AUC = 0.66$).

(c) ROC curve for Nearest + GAN Trained Siamese with LR 56x56 dataset ($AUC = 0.67$).

FIGURE 5.19. ROC curves for Nearest + GAN Trained Siamese for LR 14x14, 28x28 and 56x56 respectively and HR 112x112.



(a) Genuine/Impostor curve for Area + Area Trained Siamese with LR 14x14 dataset ($d' = 0.448$).

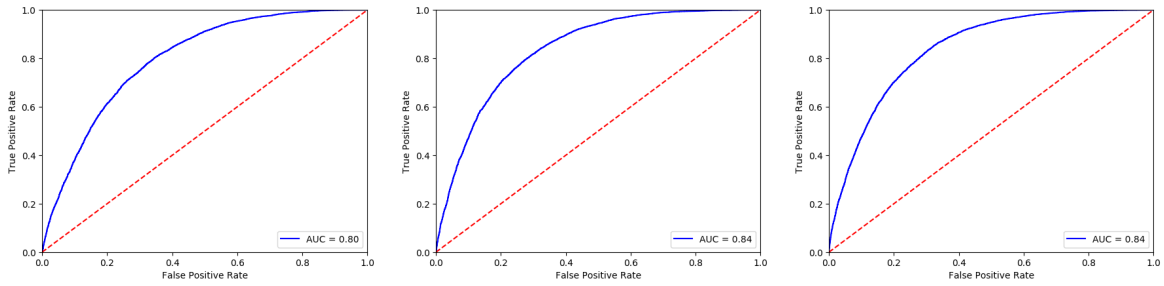
(b) Genuine/Impostor curve for Area + Area Trained Siamese with LR 28x28 dataset ($d' = 0.582$).

(c) Genuine/Impostor curve for Area + Area Trained Siamese with LR 56x56 dataset ($d' = 0.653$).

FIGURE 5.20. Genuine/Impostor curves for Area + Area Trained Siamese for LR 14x14, 28x28 and 56x56 respectively and HR 112x112.

Some conclusions can extracted from this results:

- As seen from experiments 3.2.1.5 and 3.2.1.6 it is best for the Siamese network to be trained by the same up-scaling algorithms.
- It is a known fact that training neuronal networks can be quite time and resources consuming. Adding to this the performance the experiment 3.2.1.6 had, it is possible to say that interpolations offer a good solution when resources are scarce.
- For the 56x56 pixels we can see that feature extraction models trained for higher resolution tend to work better than those train with lower resolutions.

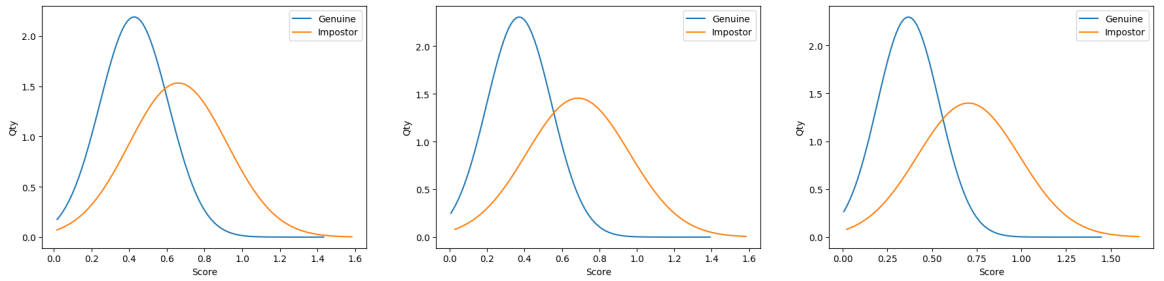


(a) ROC curve for Area + Area Trained Siamese with LR 14x14 dataset ($AUC = 0.62$).

(b) ROC curve for Area + Area Trained Siamese with LR 28x28 dataset ($AUC = 0.66$).

(c) ROC curve for Area + Area Trained Siamese with LR 56x56 dataset ($AUC = 0.67$).

FIGURE 5.21. ROC curves for Area + Area Trained Siamese for LR 14x14, 28x28 and 56x56 respectively and HR 112x112.



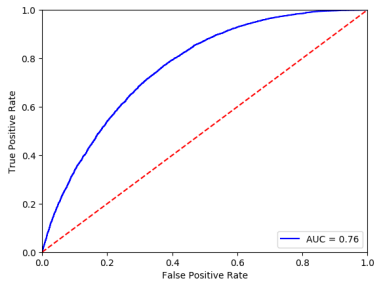
(a) Genuine/Impostor curve for Bicubic + Bicubic Trained Siamese with LR 14x14 dataset ($d' = 1.041$).

(b) Genuine/Impostor curve for Bicubic + Bicubic Trained Siamese with LR 28x28 dataset ($d' = 1.373$).

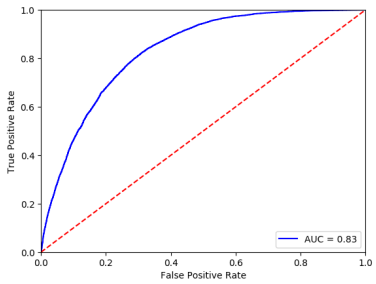
(c) Genuine/Impostor curve for Bicubic + Bicubic Trained Siamese with LR 56x56 dataset ($d' = 1.424$).

FIGURE 5.22. Genuine/Impostor curves for Bicubic + Bicubic Trained Siamese for LR 14x14, 28x28 and 56x56 respectively and HR 112x112.

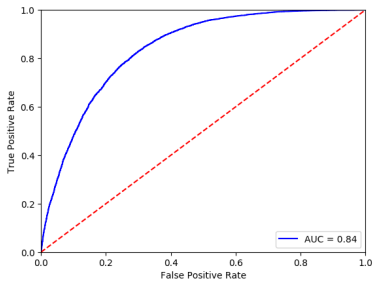
- As future work, a pre-trained GAN network could be added to the feature extraction network and trained as one whole model. It is our intuition that this will create images that do not visually correct but maximize the classification, hence removing a limitation that the current 3.2.1.4 model has. It should be noted that such model would not belong to the HR-mapping category but the unified space representation, since we are not trying to up-scale the image but to obtain a feature for each image in a unified space.



(a) ROC curve for Bicubic + Bicubic Trained Siamese with LR 14x14 dataset ($AUC = 0.76$).

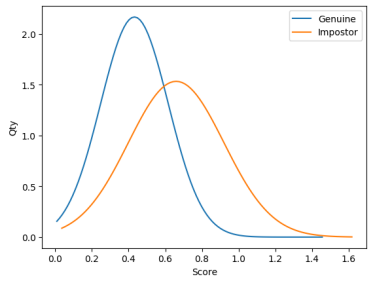


(b) ROC curve for Bicubic + Bicubic Trained Siamese with LR 28x28 dataset ($AUC = 0.83$).

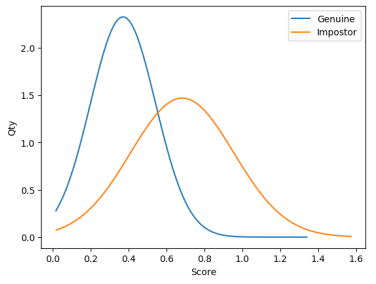


(c) ROC curve for Bicubic + Bicubic Trained Siamese with LR 56x56 dataset ($AUC = 0.84$).

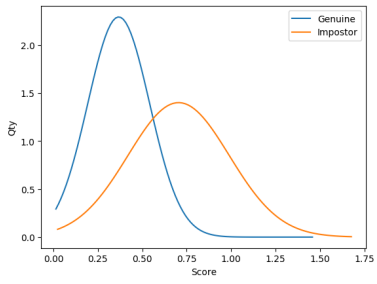
FIGURE 5.23. ROC curves for Bicubic + Bicubic Trained Siamese for LR 14x14, 28x28 and 56x56 respectively and HR 112x112.



(a) Genuine/Impostor curve for Lanczos + Lanczos Trained Siamese with LR 14x14 dataset ($d' = 1.007$).

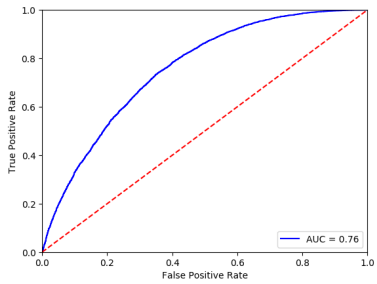


(b) Genuine/Impostor curve for Lanczos + Lanczos Trained Siamese with LR 28x28 dataset ($d' = 1.371$).

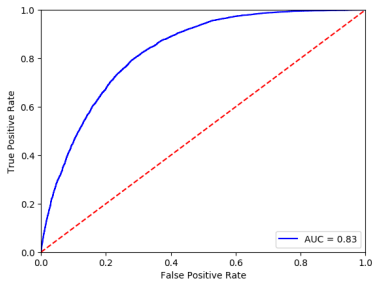


(c) Genuine/Impostor curve for Lanczos + Lanczos Trained Siamese with LR 56x56 dataset ($d' = 1.426$).

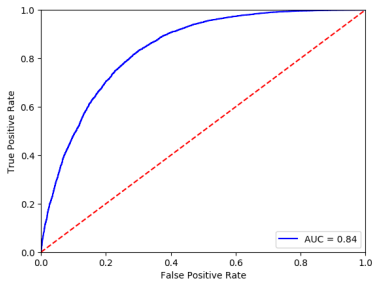
FIGURE 5.24. Genuine/Impostor curves for Lanczos + Lanczos Trained Siamese for LR 14x14, 28x28 and 56x56 respectively and HR 112x112.



(a) ROC curve for Lanczos + Lanczos Trained Siamese with LR 14x14 dataset ($AUC = 0.76$).

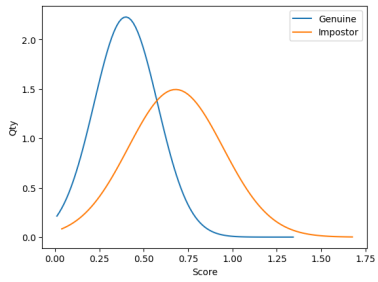


(b) ROC curve for Lanczos + Lanczos Trained Siamese with LR 28x28 dataset ($AUC = 0.83$).

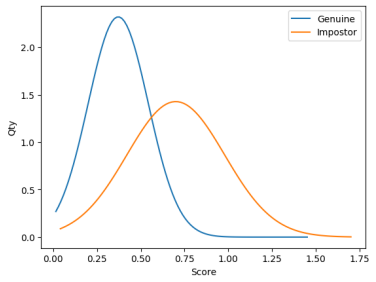


(c) ROC curve for Lanczos + Lanczos Trained Siamese with LR 56x56 dataset ($AUC = 0.84$).

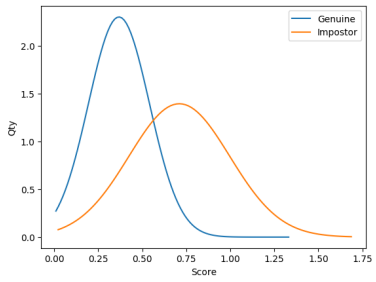
FIGURE 5.25. ROC curves for Lanczos + Lanczos Trained Siamese for LR 14x14, 28x28 and 56x56 respectively and HR 112x112.



(a) Genuine/Impostor curve for Nearest + Nearest Trained Siamese with LR 14x14 dataset ($d' = 1.236$).

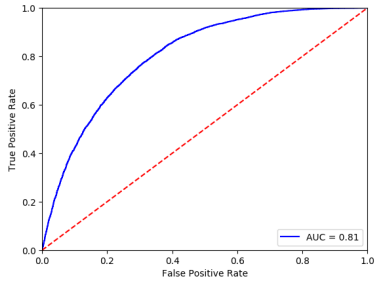


(b) Genuine/Impostor curve for Nearest + Nearest Trained Siamese with LR 28x28 dataset ($d' = 1.416$).

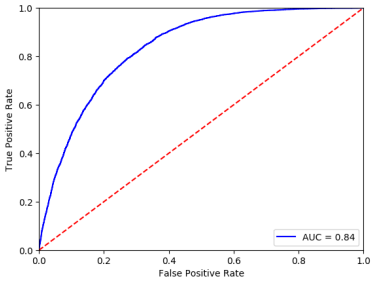


(c) Genuine/Impostor curve for Nearest + Nearest Trained Siamese with LR 56x56 dataset ($d' = 1.449$).

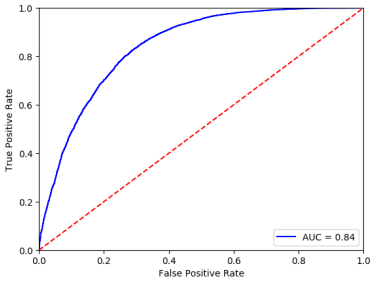
FIGURE 5.26. Genuine/Impostor curves for Nearest + Nearest Trained Siamese for LR 14x14, 28x28 and 56x56 respectively and HR 112x112.



(a) ROC curve for Nearest + Nearest Trained Siamese with LR 14x14 dataset ($AUC = 0.81$).

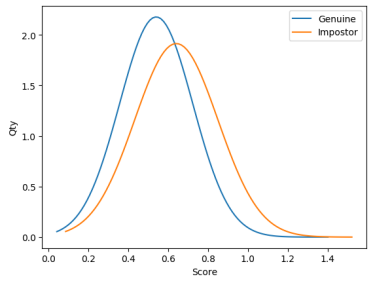


(b) ROC curve for Nearest + Nearest Trained Siamese with LR 28x28 dataset ($AUC = 0.84$).

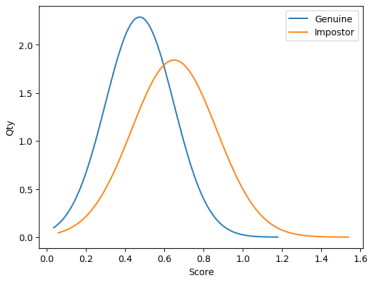


(c) ROC curve for Nearest + Nearest Trained Siamese with LR 56x56 dataset ($AUC = 0.84$).

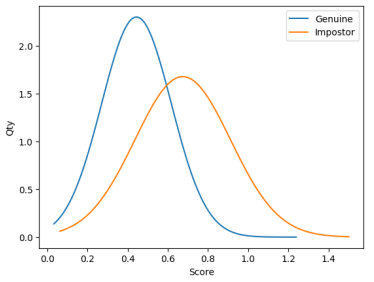
FIGURE 5.27. ROC curves for Nearest + Nearest Trained Siamese for LR 14x14, 28x28 and 56x56 respectively and HR 112x112.



(a) Genuine/Impostor curve for ESRGAN + GAN Trained Siamese with LR 14x14 dataset ($d' = 0.514$).

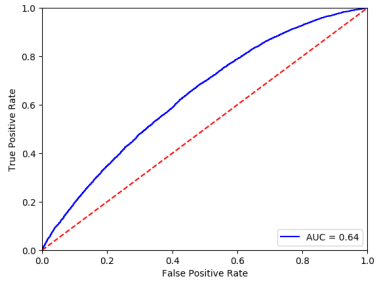


(b) Genuine/Impostor curve for ESRGAN + GAN Trained Siamese with LR 28x28 dataset ($d' = 0.894$).

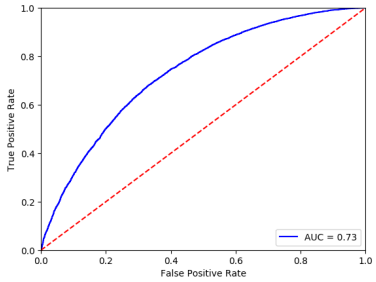


(c) Genuine/Impostor curve for ESRGAN + GAN Trained Siamese with LR 56x56 dataset ($d' = 1.106$).

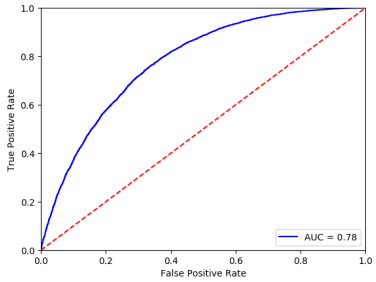
FIGURE 5.28. Genuine/Impostor curves for ESRGAN + GAN Trained Siamese for LR 14x14, 28x28 and 56x56 respectively and HR 112x112.



(a) ROC curve for ESRGAN + GAN Trained Siamese with LR 14x14 dataset ($AUC = 0.64$).



(b) ROC curve for ESRGAN + GAN Trained Siamese with LR 28x28 dataset ($AUC = 0.73$).



(c) ROC curve for ESRGAN + GAN Trained Siamese with LR 56x56 dataset ($AUC = 0.78$).

FIGURE 5.29. ROC curves for ESRGAN + GAN Trained Siamese for LR 14x14, 28x28 and 56x56 respectively and HR 112x112.

TABLE 5.2. Summary of all the experiments results (three best results by resolution in bold).

	Exp. name	14x14		28x28		56x56		Average	
		d'	AUC	d'	AUC	d'	AUC	d'	AUC
1	Baseline	0.411	0.61	0.933	0.74	**1.523**	**0.86**	0.956	0.74
2	Bicubic + Facenet	0.462	0.63	**1.900**	**0.91**	**2.787**	**0.97**	**1.716**	**0.84**
3	T. GAN + Arcface	0.253	0.57	0.959	0.75	**1.547**	**0.87**	0.920	0.73
4	T. GAN + GAN T. S.	**1.156**	**0.79**	**1.421**	**0.84**	1.388	0.83	1.322	0.82
5	Area + GAN T. S.	0.448	0.62	0.582	0.66	0.653	0.67	0.561	0.65
6	Bicubic + GAN T. S.	0.619	0.67	0.724	0.69	0.720	0.69	0.688	0.68
7	Lanc. + GAN T. S.	0.613	0.67	0.704	0.69	0.708	0.69	0.675	0.68
8	Nearest + GAN T. S.	0.448	0.62	0.582	0.66	0.653	0.67	0.561	0.65
9	Area + Area T. S.	**1.202**	**0.80**	1.400	0.84	1.420	0.84	**1.341**	**0.83**
10	Bicubic + Bicubic T. S.	1.041	0.76	1.373	0.83	1.424	0.84	1.279	0.81
11	Lanc. + Lanc. T. S.	1.007	0.76	1.371	0.83	1.426	0.84	1.268	0.81
12	Nearest + Nearest T. S.	**1.236**	**0.81**	**1.416**	**0.84**	1.449	0.84	**1.367**	**0.83**
13	ESRGAN + GAN T. S.	0.514	0.64	0.894	0.73	1.106	0.78	0.838	0.72

6. CONCLUSIONS

Over the past years we have witnessed remarkable advancement in face recognition. With some applications that were once considered science-fiction. It is clear that there is still more work to be done with low resolution images, especially in low resolution face re-identification whether it is done by humans or machines. In this work, the attempt was to establish a useful protocol that could aid the advancement in this field. Protocols do not place a burden on research, but rather are a agreed set of guidelines for writing understandable work that can be used by other researchers. Protocols are the stepping stone for reproducibility, the difference between science and a simple opinion.

We have detailed the first protocol for HR-mapping and a dataset protocol for face re-identification. To validate these protocols we have used them as a guideline to create a set of training and testing pairs for face re-identification using VGG Face 2 dataset. We have also put together a baseline and 12 experiments to validate the experimental protocol. Obtaining a $d' = 1.236$ and $AUC = 0.81$ for a 14x14 LR pixel size set, $d' = 1.900$ and $AUC = 0.91$ for a 28x28 LR pixel size set and $d' = 2.787$ and $AUC = 0.97$ for a 56x56 LR pixel size set, surpassing the baseline.

We hope that this protocols serve a basis for future researchers as they build their own research, making it comparable and replicable. As future work, protocols for the other categories (LR and HR robust features and Unified space representation) can be created, further standardizing the research in this field. It should be noted that the same dataset protocol could be used, making the comparison between categories seamless and consistent.

References

Al-Maadeed, S., Bourif, M., Bouridane, A., & Jiang, R. (2016). Low-quality facial biometric verification via dictionary-based random pooling. *Pattern Recognition*, *52*, 238–248.

Boom, B., Beumer, G., Spreeuwers, L. J., & Veldhuis, R. N. (2006). The effect of image resolution on the performance of a face recognition system. In *2006 9th International Conference on Control, Automation, Robotics and Vision* (pp. 1–6).

Bradski, G. (2000). The OpenCV Library. *Dr. Dobb's Journal of Software Tools*.

Cao, Q., Shen, L., Xie, W., Parkhi, O. M., & Zisserman, A. (2018). VGGFace2: A dataset for recognising faces across pose and age. In *International Conference on Automatic Face and Gesture Recognition.*

Cheng, Z., Dong, Q., Gong, S., & Zhu, X. (2020). Inter-task association critic for cross-resolution person re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 2605–2615).

Deng, J., Guo, J., Niannan, X., & Zafeiriou, S. (2019). Arcface: Additive angular margin loss for deep face recognition. In *CVPR*.

Grgic, M., Delac, K., & Grgic, S. (2011). SCface–surveillance cameras face database. *Multimedia tools and applications*, *51*(3), 863–879.

Harmon, L. D. (1973). The recognition of faces. *Scientific American*, *229*(5), 70–83.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770–778).

Heinsohn, D., Villalobos, E., Prieto, L., & Mery, D. (2019). Face recognition in low-quality images using adaptive sparse representations. *Image and Vision Computing*, *85*, 46–58.

Hernández-Durán, M., Cheplygina, V., & Plasencia-Calaña, Y. (2015). Dissimilarity representations for low-resolution face recognition. In *International Workshop on Similarity-Based Pattern Recognition* (pp. 70–83).

Huang, G. B., Mattar, M., Berg, T., & Learned-Miller, E. (2008). Labeled faces in the wild: A database forstudying face recognition in unconstrained environments..

Jia, G., Li, X., Zhuo, L., & Liu, L. (2016). Recognition oriented feature hallucination for low resolution face images. In *Pacific Rim Conference on Multimedia* (pp. 275–284).

Jiang, J., Ma, J., Chen, C., Jiang, X., & Wang, Z. (2016). Noise robust face image super-resolution through smooth sparse representation. *IEEE Transactions on Cybernetics*, *47*(11), 3991–4002.

Jiang, J., Yu, Y., Hu, J., Tang, S., & Ma, J. (2018). Deep CNN denoiser and multi-layer neighbor component embedding for face hallucination. *arXiv preprint arXiv:1806.10726*.

Jiang, J., Yu, Y., Tang, S., Ma, J., Qi, G.-J., & Aizawa, A. (2017). Context-patch based face hallucination via thresholding locality-constrained representation and reproducing learning. In *2017 IEEE International Conference on Multimedia and Expo (ICME)* (pp. 469–474).

Keys, R. (1981). Cubic convolution interpolation for digital image processing. *IEEE Transactions on Acoustics, Speech, and Signal processing*, *29*(6), 1153–1160.

Kim, D., Kim, M., Kwon, G., & Kim, D.-S. (2019). Progressive face super-resolution via attention to facial landmark. *arXiv preprint arXiv:1908.08239*.

Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., … others (2017). Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE Conference on Computer vision and Pattern Recognition* (pp. 4681–4690).

Li, P., Prieto, L., Mery, D., & Flynn, P. (2018). Face recognition in low quality images: a survey. *arXiv preprint arXiv:1805.11519*.

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

Li, P., Prieto, M. L., Flynn, P. J., & Mery, D. (2017). Learning face similarity for re-identification from real surveillance video: A deep metric solution. In *2017 IEEE International Joint Conference on Biometrics (IJCB)* (pp. 243–252).

Lu, C., & Tang, X. (2014). Surpassing human-level face verification performance on lfw with gaussianface. *arXiv preprint arXiv:1404.3840*.

Lu, T., Xiong, Z., Zhang, Y., Wang, B., & Lu, T. (2017). Robust face super-resolution via locality-constrained low-rank representation. *IEEE Access*, *5*, 13103–13117.

Lu, Z., Jiang, X., & Kot, A. (2018). Deep coupled resnet for low-resolution face recognition. *IEEE Signal Processing Letters*, *25*(4), 526–530.

Mudunuri, S. P., Sanyal, S., & Biswas, S. (2018). 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)* (pp. 602–60209).

Neves, J., & Proença, H. (2016). ICB-RW 2016: International challenge on biometric recognition in the wild. In *2016 International Conference on Biometrics (ICB)* (pp. 1–6).

Parkhi, O. M., Vedaldi, A., & Zisserman, A. (2015). Deep face recognition.

Saha, P., & Das, A. (2019). NFGS enabled face re-identification for efficient surveillance in low quality video. In *2019 Fifth International Conference on Image Information Processing (ICIIP)* (pp. 114–118).

Schroff, F., Kalenichenko, D., & Philbin, J. (2015). Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 815–823).

Selwood, D. L. (2019). Commentary: Reproducible science, why protocols matter. *Chemical Biology & Drug Design*, *93*(6), 975–978.

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Sønderby, C. K., Caballero, J., Theis, L., Shi, W., & Huszár, F. (2016). Amortised map inference for image super-resolution. *arXiv preprint arXiv:1610.04490*.

Sun, Y., Liang, D., Wang, X., & Tang, X. (2015). Deepid3: Face recognition with very deep neural networks. *arXiv preprint arXiv:1502.00873*.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1–9).

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 2818–2826).

Tang, Y., Yang, X., Wang, N., Song, B., & Gao, X. (2020). Person re-identification with feature pyramid optimization and gradual background suppression. *Neural Networks*, *124*, 223–232.

Viola, P., & Jones, M. J. (2004). Robust real-time face detection. *International Journal of Computer Vision*, *57*(2), 137–154.

Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., ... Change Loy, C. (2018). Esrgan: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European Conference on Computer Vision (ECCV)* (pp. 0–0).

Wang, Y., Perazzi, F., McWilliams, B., Sorkine-Hornung, A., Sorkine-Hornung, O., & Schroers, C. (2018). A fully progressive approach to single-image super-resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (pp. 864–873).

Wang, Z., Liu, D., Yang, J., Han, W., & Huang, T. (2015). Deep networks for image super-resolution with sparse prior. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 370–378).

Wang, Z., Yang, W., & Ben, X. (2015). Low-resolution degradation face recognition over long distance based on CCA. *Neural Computing and Applications*, *26*(7), 1645–1652.

Yu, X., Fernando, B., Hartley, R., & Porikli, F. (2018). Super-resolving very low-resolution face images with supplementary attributes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 908–917).

Yu, X., & Porikli, F. (2017). Face hallucination with tiny unaligned images by transformative discriminative neural networks. In *Thirty-First AAAI Conference on Artificial Intelligence.*

Zhao, K., Xu, J., & Cheng, M.-M. (2019). Regularface: Deep face recognition via exclusive regularization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1136–1144).

Zhao, L., Bai, H., Liang, J., Zeng, B., Wang, A., & Zhao, Y. (2019). Simultaneous color-depth super-resolution with conditional generative adversarial networks. *Pattern Recognition*, *88*, 356–369.