Open-set Face Recognition for Small Galleries Using Siamese Networks

Gabriel Salomon[1], Alceu Britto Jr.[2], Rafael H. Vareto[3], William R. Schwartz[3], David Menotti[1]

[1]Vision, Robotics and Imaging Laboratory, Universidade Federal do Paraná, 82590300, Brazil
[2]PPGIA, Pontifícia Universidade Católica do Paraná, 80215901, Brazil
[3]Smart Sense Laboratory, Department of Computer Science, Universidade Federal de Minas Gerais, 31270901, Brazil
gsaniceto@inf.ufpr.br

Abstract—Face recognition has been one of the most relevant and explored fields of Biometrics. In real-world applications, face recognition methods usually must deal with scenarios where not all probe individuals were seen during the training phase (open-set scenarios). Therefore, open-set face recognition is a subject of increasing interest as it deals with identifying individuals in a space where not all faces are known in advance. This is useful in several applications, such as access authentication, on which only a few individuals that have been previously enrolled in a gallery are allowed. The present work introduces a novel approach towards open-set face recognition focusing on small galleries and in enrollment detection, not identity retrieval. A Siamese Network architecture is proposed to learn a model to detect if a face probe is enrolled in the gallery based on a verification-like approach. Promising results were achieved for small galleries on experiments carried out on Pubfig83, FRGCv1 and LFW datasets. State-of-the-art methods like HFCN and HPLS were outperformed on FRGCv1. Besides, a new evaluation protocol is introduced for experiments in small galleries on LFW.

Keywords—open-set face recognition, siamese networks, face recognition, small galleries.

I. INTRODUCTION

Face recognition is a real-world necessity, mainly for human identification and surveillance. It presents many advantages over other biometrics (e.g.,fingerprint and iris), as face images capture is non-intrusive and can be done at a distance [1]. Moreover the availability of face samples on the Internet.

A common definition of face recognition assumes the existence of three tasks, as follows: verification, identification and watch list recognition [2]. Face verification is defined as a binary problem that consists of comparing two face images and then determining if both of them belong to the same individual. Closed-set face identification, or just identification, focus on determining to which gallery member a face image sample belongs. Watch list recognition (or open-set face recognition), works similarly to the latter by also comparing the face sample to the gallery, but with two major distinctions: i) the individual may not belong to the gallery; ii) the face corresponding identity is only relevant if it was previously enrolled in the gallery. More precisely, an open-set face recognition task addresses the problem of determining whether a face image is enrolled in a previously defined gallery and, in that case, retrieving the sample's corresponding enrolled identity.
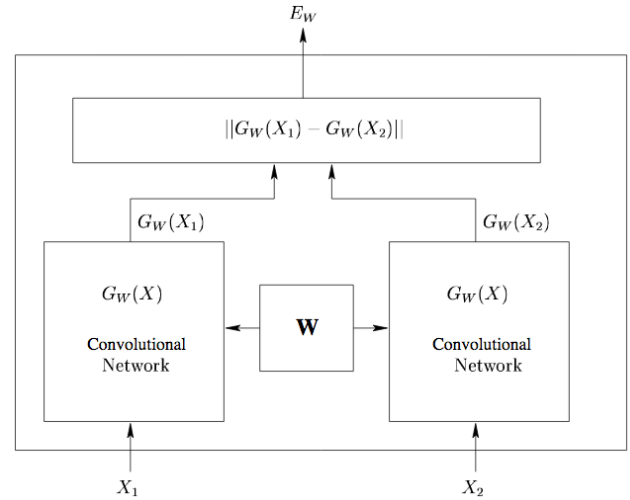
Fig. 1: Representation of a Siamese Network. It has two input images X_1, X_2 for each network (convolutional networks in this example). The architecture (or function G) is the same for both and the weights W are shared. The outputs are joined at the end using Euclidean distance. The final output is E_W. Source: Chopra et al. [3].

Machine learning methods usually learn a model that uses training data to map feature vectors to known classes [4], [5]. Testing these methods consists in dealing with unknown samples of known classes. However, dealing with unknown classes on testing is not trivial. Open-set face recognition methods have to deal with an "unlimited" number of unknown classes (subjects' identities) when compared to the countable quantity of individual identities enrolled in the gallery set [6].

It is important to stress that this work aims to determine whether a given face sample belongs to any individual registered in the gallery set, and not at retrieving its genuine identity. In that direction, our proposal is directly comparable to works in the literature [7]. Our main hypothesis is that Siamese Neural Networks (SNNs) can be effectively employed in open-set recognition. A SNN is composed of two independent sub-networks with the same architecture and shared weights with their outputs joined at the end using a distance metric [8]. Each independent sub-network has its own input, and the whole network outputs a numeric distance value, indicating the

similarity between the two inputs. Although the SNN performs face verification, the proposed architecture uses it to compute scores and perform open-set recognition. Fig. 1 illustrates the process.

Small galleries are the focus of the present work due to a scalability limitation of the present method on the experiments performed. To the best of our knowledge, this is the first time that SNN are considered in open-set face recognition scenarios focusing on small galleries. The term *small gallery* refers to sets containing no more than 20 identities, but with no restrictions on the number of samples per identity. Many business establishments require visual authentication of small personnel, rarely exceeding fifteen or twenty employees. The lack of sufficient classes and samples turns the small gallery recognition task into a difficult problem, especially due to overfitting. Similarly, scarce face samples may reduce the models ability to learn general facial characteristics.

The proposed method can be divided into two stages: first, an SNN is trained on generated pairs of face samples that belong to the gallery. It is trained for the usual task (verification): learning to differentiate the members of the gallery. In the second stage, the trained SNN is used to calculate the minimum Euclidean distance between a probe image and training samples from all members of the gallery. It is expected that the SNN learns a representation afterward expressed in terms of distance of a pair of images. These representations should allow establishing a separation between the probe images that belong to the gallery and the ones that do not.

Our main contribution is twofold. A new method using SNN for performing open-set recognition on small galleries; and a new experimental protocol for comparison of methods devoted to open-set small galleries. The method addresses a real necessity of algorithms that can perform well in constrained data (few examples and few identities known) and the protocol was defined to support the development of such methods.

Experiments performed on Experiment 4 of FRGCv1 dataset (Experiment 4) [9] achieved state-of-the-art results for small galleries in comparison with published work [7].

II. RELATED WORK

A face verification-like approach towards open-set recognition is proposed in the current work, using CNN and SNN. In this respect, research regarding deep learning methods, open-set recognition and Siamese Networks for face recognition are reviewed.

One of the main focus of the face recognition research community in the last years has been the improvement of the feature extraction quality, which is measured by its ability to represent and differentiate face samples. The most commonly used methods are based on Convolutional Neural Networks (CNNs).

However, CNNs only became the most used feature extraction and learning techniques within the last decade. Combined with GPU and large datasets, face recognition methods are capable of surpassing human performance using deep network architectures, achieving state-of-the-art performance on many surveillance and biometrics tasks [10]–[15].

One of the most well-established networks with a focus on face recognition is the VGGFace [11], introduced in 2015. VG-GFace is pre-trained on 2.6 million face images and achieves 98.95% accuracy on the Labeled Faces in the Wild (LFW) dataset [16] under the face verification task. Conversely to most works in the literature, we design an open-set face recognition approach following a face-verification strategy.

According to Kemelmacher et al. [17], recent identification and verification face recognition approaches are not very promising when the gallery set is upscaled, that is, when the set of known individuals escalates towards thousands or millions of subjects. For that reason, scalability is drawing the attention of several researchers [14], [18]–[20]. Vareto et al. [7] concentrate on the open-set face recognition problem as they assess both Partial Least Squares (PLS) and Multi-Layer Perceptron (MLP) classification models in pursuance of an algorithm that is not directly dependent on the gallery set size. In fact, the authors build a collection of either PLS or MLP binary models, defined as hashing functions, and a voting system scheme (candidate list) to determine whether the queried subject is known or unknown. If a subject stands out from the other individuals of the candidate list, the subject is recognized; otherwise, the subject is discarded since the method considers s/he is not enrolled in the gallery set.

Siamese Networks are used for metric learning and were originally developed for verifying signatures using Neural Networks [8]. A few years later, they were employed in face verification for the first time, using a new loss function based on Energy-Based Models (EBM), called Contrastive Loss [3]. Since the boom of CNNs in face recognition, Siamese Networks have been widely used for face verification in large datasets, obtaining great performance over other methods [10], [11], [21]. As Siamese Networks are a metric learning method, they make use of similarity scores to perform recognition. Several works have also proposed approaches based on thresholding similarity-like scores, but focusing generally on open-set face recognition [13], [14], [20], [22]–[26].

III. PROPOSED METHOD

The proposed method consists of 4 steps: i) face alignment; ii) feature extraction; iii) training on pairs; iv) recognition. The two first steps are straightforward approaches. On the third step, negative and positive pairs are generated from the face gallery samples to train the SNN to learn how to differentiate between the gallery members. Using the trained SNN, in the last step, a test sample is compared to selected training samples from each gallery member and the recognition is performed based on the similarity score. Each stage is detailed in the next sections and illustrated in Fig. 2.

A. Face Alignment

Alignment is conducted in all face images using Joint Face Detection and Alignment using Multi-task Cascaded CNNs [27]. The Multi-task Cascaded CNN can perform both
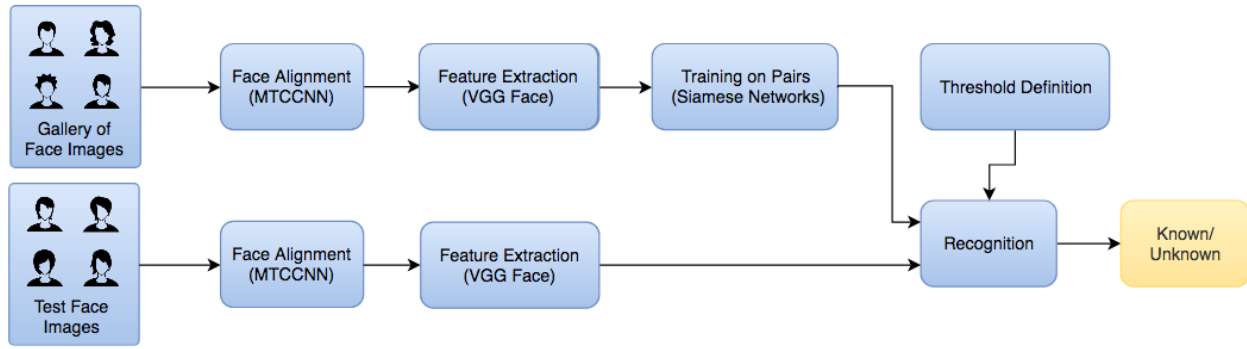
Fig. 2: Pipeline for proposed method. The gallery is used for training the Siamese Networks, after face alignment/crop and feature extraction. At the end, a threshold is defined. The test face images are aligned/crop and their features are extracted to perform recognition. The recognition stage assigns a binary label to each test sample: "Known" or "Unknown".

face detection and alignment at the same time. There are three stages to perform both tasks:

i. a shallow CNN produces a large set of windows in the image that might be a face or not;
ii. a more complex CNN reject most of the windows, which are unlikely to contain faces;
iii. an even more complex CNN analyzes the results refining it and outputs landmarks positions for aligned faces.

This state-of-the-art method was used in the pre-processing stage, as face cropping and alignment may improve feature extraction discrimination capability.

B. Feature Extraction

The extraction of features from the images for training and testing was made using a Deep Convolutional Network: VGG Face [11]. The network has 21 layers (18 for convolution and pooling and 3 for classification, respectively) plus a Softmax function at the end for prediction purposes. VGG is pre-trained on a celebrity database that contained 2.6 million images of 2,622 individual identities. Aiming only feature extraction, the Softmax function is removed, and the resulting feature vector for every image has 2,622 dimensions.

C. Siamese Networks

The SNN architecture is composed of two identical 3-layer sub-networks, each layer is fully connected and composed of 2,048 neurons. The architecture and the number of neurons were obtained empirically. The input layer has size 2,622, which is the exact size of the output of the VGG Face feature extractor. Fig. 3 presents the network architecture.

1) Pair Generation: The maximum possible number of training samples pairings is the combination two at a time of the n samples, i.e., $\binom{n}{2} = O(n^2)$. This can be costly, considering that few iterations on training over almost n^2 pairs can take very long, as neural networks are expensive to train. Subsampling is needed to reduce the number of pairs used in training.

Also, an evaluation on how the pairs choice impacts the performance is desirable. Therefore, two methods of pair subsampling were evaluated. These specific techniques were selected empirically on initial experiments.
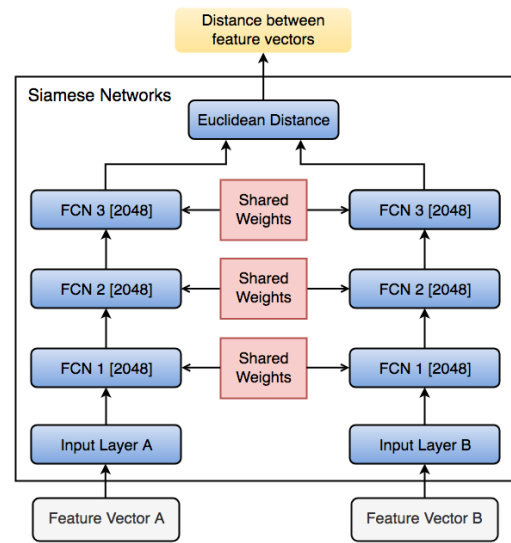

Fig. 3: SNN architecture. The weights of each FCN are shared among the two separated internal pipelines of the network, this means that the weights are the same (not that they are connected). In the end, the output is the euclidean distance.

For explanation purposes, they are denominated: Algorithms P1 and P2. The training samples set is denoted as $\mathcal{S}$, and the set of identities I. The function $id()$ returns the identity associated with the sample supplied as parameter. The procedure $rand\_sample()$ receives a set of identities and returns a random sample that belongs to one of the identities provided. In the end, each algorithm return two sets of pairs, PP and PN, which correspond to the set of pairs of the samples from the same (positive) and different persons(negative), respectively. Algorithm P1 is described below and illustrated in Fig. 4a.

Both algorithms generate equal numbers of positives and negatives pairs, for balancing purposes. **P1** generates pairs using the individual samples *versus* every other subject in the gallery. For each training sample, pairings with $z << n$ random samples from every other individual in the gallery are generated. In cases where z is greater than the number of samples of each individual, repeated pairs will occur. On the

Algorithm 1 Pairing 1

```
1:  PN = [ ]
2:  PP = [ ]
3:  for x in S do
4:      for i in I − id(x) do
5:          c = 0
6:          while c < z do
7:              PN+ = [x, rand_sample(i)]
8:              PP+ = [x, rand_sample(id(x))]
9:              c+ = 1
```

other hand, **P2** generates fewer pairs, only pairing samples with a few other individuals randomly chosen. Every training sample is paired with z random samples from other random identities. Algorithm P2 is presented below and illustrated in Fig. 4b.

Algorithm 2 Pairing 2

```
1:  PN = [ ]
2:  PP = [ ]
3:  for x in S do
4:      c = 0
5:      while c < z do
6:          PN+ = [x, rand_sample(I − id(x))]
7:          PP+ = [x, rand_sample(id(x))]
8:          c+ = 1
```

Observe that for P1, $n \times 2(k-1) \times z$ pairs are generated, where $k = |I|$, while for P2 only $n \times 2 \times z$ pairs are generated and 2 stands for the "positive" and "negative" sets.

2) Loss and Distance Functions: The loss function chosen is the Contrastive Loss [3], [28]. Contrastive Loss takes pairs of samples and maps them into a space where intra-class samples reside in neighbor areas and inter-class samples are far away.

The weights (W) of the network are the parameters of the mapping function (called G), that reduces the dimensionality by mapping two feature vectors $\vec{X}_1$ and $\vec{X}_2$ into $G_W(\vec{X}_1)$ and $G_W(\vec{X}_2)$ and computing the distance measure $D_W(\vec{X}_1, \vec{X}_2)$, i.e., $D_W(\vec{X}_1, \vec{X}_2) = ||G_W(\vec{X}_1) - G_W(\vec{X}_2)||$. For simplicity $D_W(\vec{X}_1, \vec{X}_2)$ is denoted D_W and the final loss function [3] is defined as

$$L(W, Y, \vec{X}_1, \vec{X}_2) = \frac{(1-Y)(D_W)^2 + Y\{\max(0, m - D_W)\}^2}{2} \quad (1)$$

where Y is 0 if the feature vectors belong to the same person and 1 if they represent different identities, and the variable m is a predefined margin around $G_W(\vec{X}_1)$ or $G_W(\vec{X}_2)$ that should be greater than 0. In other words, if $\vec{X}_1$ and $\vec{X}_2$ belong to different individuals and $D_W(\vec{X}_1, \vec{X}_2)$ is smaller than the m margin, the loss function improves their separation, otherwise, no margin changes are made.

The Siamese Network's distance chosen is the euclidean distance. Other commonly used distances are cosine similarity and Manhattan distance, but in the experiments performed with Contrastive Loss, euclidean distance outperformed both.

IV. EXPERIMENTS

In this section, the experiments executed to validate the proposed approach are described. For such, the evaluation metrics, datasets employed and the experimental protocols are described. For literature comparison, we chose the work of Vareto et al. [7], taking into account the clarity on the evaluation protocol which assists reproducibility of experiments, enabling a fair comparison.

A. Metrics

AUC (Area Under Curve) from the ROC (Receiver Operating Characteristic) curve is the metric selected for method evaluation. It is a well-known technique as it describes the algorithm performance on various thresholds. ROC curve shows the relation between TPR (True Positive Rate) and FPR (False Positive Rate). ROC curves with AUC closer to 1 indicate better performance, while the ones with AUC close to 0.5 indicate poor performance. Values below 0.5 are worst than random guessing.

B. Datasets

One of the most challenging and used datasets for Face Recognition is certainly LFW [16]. However, as previously stated, there is no universal protocol for open-set evaluation and for literature comparison. The first two datasets described below are used for literature comparison, and, the third one is used for evaluating the proposed (new) protocol.

PubFig83: Pubfig83 is a subset of the original PubFig [29]. It has only 83 subjects and 13,838 uncontrolled images (average of 166 samples per subject). Ranges from 89 to 367 samples available per subject.

FRGCv1: This is the dataset provided for the Face Recognition Grand Challenge v1.0 [9]. It will be used to compare with Vareto et al. experiments. It has a total of 50,000 images and 6 experiments. Tests were only conducted on experiment 1 and 4. The first experiment is composed of 760 controlled images and the fourth has 152 controlled photos + 1064 uncontrolled ones. Both experiments have 152 distinct subjects.

LFW: LFW is a well-known dataset among face recognition researchers, in fact, one of the most used for face recognition. It consists of 13,233 uncontrolled images of 5,749 people. Many subjects in this dataset have only 1 sample available. Some have more than 15 images.

C. Experimental Protocols

The main issue in choosing the experimental protocol (EP) lies upon the lack of consensus among authors on which is the best protocol for open-set recognition evaluation. Also, the protocol is usually not clearly stated making reported results unable to be reproduced and compared fairly.

As we are comparing our proposal to the one by Vareto et al. [7], and their protocol is well described, we used their protocol, which we refer to as *Experimental Protocol I*.

We also propose a slight modification on EP-I focusing on *small galleries*, thus generating the so-called *Experimental Protocol II*. The difference between these two protocols is that
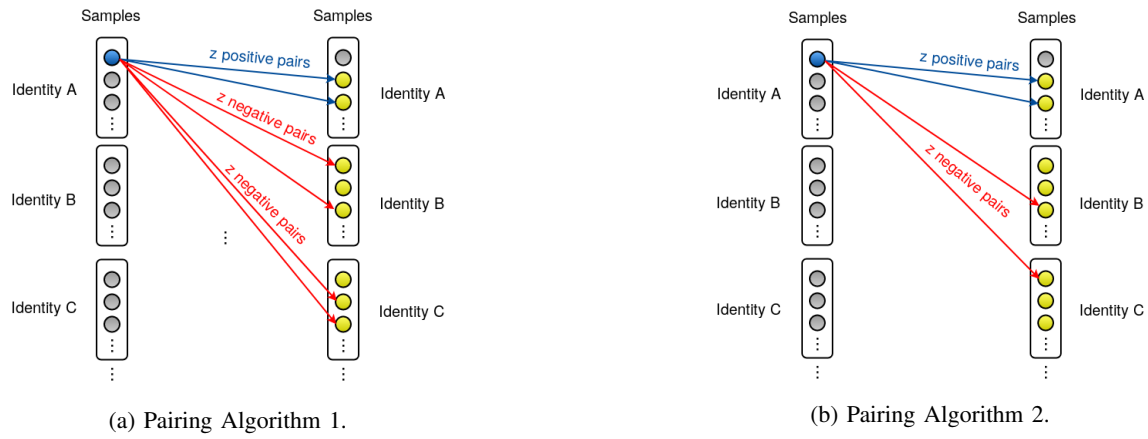
(a) Pairing Algorithm 1.

(b) Pairing Algorithm 2.

Fig. 4: Comparison of P1 and P2. All "Samples" columns represent the same samples. Note that in P1, the z pairs are formed between the same identity and z pairs for each one of the other identities. P2, in contrast, generate z pairs choosing randomly samples for any other identity (generating less pairs).

in the EP-I, we choose a *percentage* of subjects, i.e., 10%, 50% and 90% of identities in the dataset to be part of the known individual, while in the EP-II an *absolute* number of subjects (e.g., 5, 10, 15 and 20) is chosen.

In both protocols, the database is split into training and testing data. The training data has only known individuals (individuals that should be recognized by the method). The testing data has known individuals unseen samples (samples of the individuals that the method has never seen before) and unknown individuals (samples of people the method has never seen before).

The whole dataset is first divided between known/unknown individuals. The known subjects are evaluated in different scenarios, i.e., 10%, 50% and 90% of all dataset identities for EP-I and 5, 10, 15, and 20 subjects for EP-II in the dataset and the remainder is considered as unknown and is only used for testing.

For each known identity, the samples are split in half: 50% for training and 50% for testing. Each experiment on each database is repeated 10 times by randomly splitting the data as aforementioned, executing the method and, in the end, plotting the ROC curves with AUC. The mean AUC is computed, along with the standard deviation.

D. Experimental Results

In this section, we report results yielded by our proposed method using: i) the EP-I (proposed by Vareto et al. [7]) on the Pubfig83 [18] and FRGCv1 [9] (exp. 1 and 4) datasets (Table I); ii) The new proposed EP-II on the LFW dataset [16] (Table II); iii) the EP-I on the FRGCv1 dataset - exp. 4 for comparison with the literature (Table III).

PubFig83 (EP-I): The proposed method achieved great results in Pubfig83 for small galleries (few subjects known). Although with the increase of the number of subjects, the AUC dropped significantly in 90% known. Statistically, there is no difference between the pairings P1 and P2.

TABLE I: Results for experiments on Pubfig83 and FRGCv1 (Experiments 1 & 4) dataset using EP-I.

Pair	Known Rate	Datasets		
		Pubfig83	FRGCv1 Exp. 1	FRGCv1 Exp. 4
		AUC	AUC	AUC
P1	10%	.981 ± .008	.996 ± .002	.938 ± .027
P2		.981 ± .007	.996 ± .003	.904 ± .032
P1	50%	.922 ± .015	.986 ± .004	.868 ± .013
P2		.916 ± .018	.986 ± .003	.816 ± .019
P1	90%	.856 ± .017	.972 ± .018	.797 ± .030
P2		.854 ± .025	.974 ± .008	.770 ± .021

FRGCv1 - Experiment 1 (EP-I): In experiment 1 from FRGCv1 dataset outstanding results are obtained as well. We believe that this is due to the fact that there are only controlled images on this dataset. There is no great challenge in discriminating controlled images. There is no difference in the results concerning the type of pairing used in training. The increase in gallery size did not influence dramatically the AUC, the difference is not statistically significant.

FRGCv1 - Experiment 4 (EP-I): When using small size gallery better results than the bigger galleries are obtained, as in other experiments. There is a significant discrepancy in the AUC rate for a small gallery between P1 and P2, where P1 shows better performance.

LFW (EP-II): The method performed well for small galleries, taking into account that LFW is unbalanced in the number of samples per subject and uncontrolled. There is no discrepancy in the AUC rates for small galleries between P1 and P2.

Literature Comparison: As shown in Table III, for a small number of identities (small galleries) the proposed method with SNN can surpass state-of-the-art methods in experiment 4 on FRGCv1. Note that we report the best result obtained by P1 (.938 ± .027) – from Table I. As there are approximately 150 identities, there are 15 subjects in the gallery. When there are 50% known people in the gallery (75 identities), no statistical

TABLE II: Results of the experiments on LFW using EP-II.

Pair	Known	ROC (AUC)
P1	5	0.971 ± 0.019
P2		0.969 ± 0.019
P1	10	0.967 ± 0.028
P2		0.967 ± 0.030
P1	15	0.976 ± 0.011
P2		0.946 ± 0.006
P1	20	0.972 ± 0.007
P2		0.977 ± 0.008

difference between the methods occurs.

TABLE III: Comparison with the work of Vareto et al. on FRGCv1 - experiment 4.

Known Individuals		AUC		
		10%	50%	90%
SN (proposed)	mean	0.938	0.868	0.783
	std	0.027	0.013	0.029
HPLS [7]	mean	0.794	0.850	0.856
	std	0.078	0.009	0.022
HFCN [7]	mean	0.872	0.877	0.856
	std	0.015	0.022	0.014

V. CONCLUSIONS

We introduced a recognition system, with well-defined pipelines, that employs a SNN for open-set face recognition, a task that still has not been extensively researched. We showed that SNNs outperform state-of-the-art methods like HPLS and HFCN on small galleries, being useful in applications on which reliability is more relevant than scalability. Another contribution is the Experimental Protocol II proposed for small galleries.

For future work, our goals are improving pairing selection, testing new loss functions and also combining SNNs with other metric learning methods to enhance performance.

ACKNOWLEDGMENTS

REFERENCES

[1] Z. L. Stan and K. J. Anil, *Handbook of face recognition*. Springer, 2011.

[2] R. Chellappa, P. Sinha, and P. J. Phillips, "Face recognition by computers and humans," *Computer*, vol. 43, no. 2, 2010.

[3] S. Chopra, R. Hadsell, and Y. LeCun, "Learning a similarity metric discriminatively, with application to face verification," in *CVPR*, vol. 1, pp. 539–546, IEEE, 2005.

[4] G. Salomon, R. Laroca, and D. Menotti, "Deep learning for image-based automatic dial meter reading: dataset and baselines," in *International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8, 2020.

[5] D. V. Ruiz, G. Salomon, and E. Todt, "Can giraffes become birds? An evaluation of image-to-image translation for data generation," *COTB*, 2020.

[6] W. J. Scheirer, A. de Rezende Rocha, A. Sapkota, and T. E. Boult, "Toward open set recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 7, pp. 1757–1772, 2013.

[7] R. H. Vareto, S. Silva, F. Costa, and W. R. Schwartz, "Towards open-set face recognition using hashing functions," in *Biometrics (IJCB), 2017 IEEE International Joint Conference on*, pp. 634–641, IEEE, 2017.

[8] J. Bromley, I. Guyon, Y. LeCun, E. Säckinger, and R. Shah, "Signature verification using a" siamese" time delay neural network," in *Advances in neural information processing systems*, pp. 737–744, 1994.

[9] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, *et al.*, "Overview of the face recognition grand challenge," in *CVPR*, vol. 1, pp. 947–954, IEEE, 2005.

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

[11] O. M. Parkhi, A. Vedaldi, A. Zisserman, *et al.*, "Deep face recognition.," in *BMVC*, p. 6, 2015.

[12] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in *CVPR*, pp. 815–823, 2015.

[13] Y. Sun, D. Liang, X. Wang, and X. Tang, "DeepID3: Face recognition with very deep neural networks," *arXiv preprint arXiv:1502.00873*, 2015.

[14] W. AbdAlmageed, Y. Wu, S. Rawls, S. Harel, T. Hassner, I. Masi, J. Choi, J. Lekust, J. Kim, P. Natarajan, *et al.*, "Face recognition using deep multi-pose representations," in *WACV*, pp. 1–9, IEEE, 2016.

[15] R. H. Vareto, S. S. da Silva, F. D. O. Costa, and W. R. Schwartz, "Face verification based on relational disparity features and partial least squares models," in *SIBGRAPI*, pp. 209–215, IEEE, 2017.

[16] G. B. Huang, M. Mattar, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database forstudying face recognition in unconstrained environments," in *Workshop on faces in 'Real-Life' Images: detection, alignment, and recognition*, 2008.

[17] I. Kemelmacher-Shlizerman, S. M. Seitz, D. Miller, and E. Brossard, "The megaface benchmark: 1 million faces for recognition at scale," in *CVPR*, pp. 4873–4882, 2016.

[18] N. Pinto, Z. Stone, T. Zickler, and D. Cox, "Scaling up biologically-inspired computer vision: A case study in unconstrained face recognition on facebook," in *CVPR Workshops*, pp. 35–42, IEEE, 2011.

[19] G. de Paulo Carlos, H. Pedrini, and W. R. Schwartz, "Classification schemes based on partial least squares for face identification," *Journal of Visual Communication and Image Representation*, vol. 32, pp. 170–179, 2015.

[20] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "Web-scale training for face identification," in *CVPR*, pp. 2746–2754, 2015.

[21] J. Yang, P. Ren, D. Zhang, D. Chen, F. Wen, H. Li, and G. Hua, "Neural aggregation network for video face recognition," in *CVPR*, p. 7, 2017.

[22] F. Li and H. Wechsler, "Open set face recognition using transduction," *IEEE transactions on pattern analysis and machine intelligence*, vol. 27, no. 11, pp. 1686–1697, 2005.

[23] N. Pinto, D. D. Cox, and J. J. DiCarlo, "Why is real-world visual object recognition hard?," *PLoS computational biology*, vol. 4, no. 1, p. e27, 2008.

[24] C. E. dos Santos Junior and W. R. Schwartz, "Extending face identification to open-set face recognition," in *SIBGRAPI*, pp. 188–195, IEEE, 2014.

[25] M. Günther, S. Cruz, E. M. Rudd, and T. E. Boult, "Toward open-set face recognition," in *CVPR Workshops*, 2017.

[26] W. Liu, Y. Wen, Z. Yu, M. Li, *et al.*, "Sphereface: Deep hypersphere embedding for face recognition," in *CVPR*, pp. 212–220, 2017.

[27] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," *IEEE SPL*, vol. 23, no. 10, pp. 1499–1503, 2016.

[28] R. Hadsell, S. Chopra, and Y. LeCun, "Dimensionality reduction by learning an invariant mapping," in *CVPR*, IEEE, 2006.

[29] N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar, "Attribute and simile classifiers for face verification," in *Computer Vision, 2009 IEEE 12th International Conference on*, pp. 365–372, IEEE, 2009.