Open-set Face Recognition with Neural Ensemble, Maximal Entropy Loss and Feature Augmentation

Rafael Henrique Vareto
Department of Computer Science
Universidade Federal de Minas Gerais
Belo Horizonte, MG, Brazil
rafaelvareto@dcc.ufmg.br

Manuel Günther
Department of Informatics
Universität Zürich
Zürich, ZH, Switzerland
guenther@ifi.uzh.ch

William Robson Schwartz
Department of Computer Science
Universidade Federal de Minas Gerais
Belo Horizonte, MG, Brazil
william@dcc.ufmg.br

Abstract—**Open-set face recognition is a scenario in which biometric systems have incomplete knowledge of all existing subjects. This arduous requirement must dismiss irrelevant faces and focus on subjects of interest only. For this reason, this work introduces a novel method that associates an ensemble of compact neural networks with data augmentation at the feature level and an entropy-based cost function. Deep neural networks pre-trained on large face datasets serve as the preliminary feature extraction module. The neural adapter ensemble consists of binary models trained on original feature representations along with negative synthetic mix-up embeddings, which are adequately handled by the designed open-set loss since they do not belong to any known identity. We carry out experiments on well-known LFW and IJB-C datasets where results show that the approach is capable of boosting closed and open-set identification accuracy.**

I. INTRODUCTION

Not only are face recognition systems sufficiently advanced nowadays to be used in social networks or photo library tagging, but also a leading mechanism to support governments [1], law enforcement agencies, and private companies [2]. Despite the significant recent progress, face recognition remains limited when facing poor image quality, a frequent condition in surveillance and CCTV environments. In addition, few researchers have devoted their efforts to solving problems that either require strong generalization or bounded open space risk, a scenario in which input samples are far from any known class and likely to represent unknown distributions.

Open-set face recognition characterizes the scenario where anonymous individuals, unseen during training and enrollment stages, only come into sight during evaluation time [3, 4]. As an illustration, one can think of immigration control at airports taking advantage of automated gates and smart face recognition. The system is expected to dismiss all law-abiding passengers and alert the security personnel whenever criminal offenders turn up. However, recent newspaper articles have shown that people being misidentified is not a hypothetical exercise but has actually occurred several times across the United States [5, 6]. To make matters worse, false alarms should be avoided by any means since a system identification error may bias the security approach and mistakenly hold up innocent people in custody [7].

Corresponding author's website: https://www.vareto.com.br

Some late face recognition systems have tackled low-quality images and major improvements have been made with the introduction of specialized loss functions [8, 9, 10, 11]. Deep neural networks (DNN) are typically trained on large datasets of public people before being applied to particular face populations [2]. For this reason, the identification task becomes intrinsically domain-adaptive as none of the individuals selected to train the network composes the gallery set, a collection that only contains subjects of interest, also referred to as a watchlist. Therefore, they often fail to distinguish whether an input face sample is enrolled in the gallery of known individuals since they cannot foresee the unknown.

Several works explore transfer learning techniques or consist of traditional machine learning algorithms fitted on deep feature representations [4, 12, 13]. Hashing functions have also been used to solve open-set face recognition tasks [14, 15, 16]. Günther *et al.* [7] take an existing pretrained deep backbone and replaces its output classification layer with a Neural Adapter Network (NAN). Other approaches rely on clustering techniques that act as a filtering barrier to unknown samples [17, 18]. Despite all contributions, the aforementioned methods present unbounded open-space risk [3] and are not very well suited for rejecting unknown individuals as generally required in the watchlist context.

Many investigators have examined the advantages of ensembles: Ma *et al.* [19] developed an adaptive-boosting classification framework but did not conduct experiments following open-set protocols. Choi *et al.* [20] combined a collection of deep neural networks with Gabor representations whereas Vareto *et al.* [18] employed a clustering technique to filter out dissimilar candidates before training a compact ensemble of binary models. The former contains a collection of deep neural networks and, in consequence, presents a high computational complexity in both training and evaluation stages. Similarly, the latter consists of an online training module that ends up making its use in real-time tasks unfeasible.

Dhamija *et al.* [21] noticed that unknown features are generally mapped near known classes and, with that in mind, proposed a novel loss function that maximizes the entropy of non-gallery samples. Data augmentation is widely adopted to prevent overfitting and strengthen the domain generalization capacity of DNNs [22]. In furtherance of modifying data in the

feature space, Verma *et al.* [23] came up with an interpolating strategy to generate new feature representations whereas Li *et al.* [24] proposed a stochastic feature augmentation procedure to perturb embeddings with Gaussian noise. None of the previously mentioned works has been evaluated on face benchmarks containing numerous identities and a limited number of samples per class and, as a consequence, they are not an accurate portrayal of realistic biometric tasks.

In this work, we propose a Neural Adapter Ensemble (NAE) of binary learners to handle unbalanced datasets. Ensembles are generally employed to reduce variance, minimize modeling bias and then decrease overfitting [25]. During inference, NAE aggregates the scores of each inner model and builds a final ranking of candidates. Moreover, we introduce a margin-based cost function called Maximal Entropy Loss (MEL) that not only produces more rigorous decision boundaries for known classes, but also increases the entropy for negative training samples. Since MEL relies upon representative negative samples, we develop an Optimized Mix-Up (OMU) feature augmentation method that synthesizes negative embeddings from feature representations of different subjects enrolled in the gallery set. The data augmentation contributes to the awareness of unknown identities since including artificial embeddings that are appropriately exploited by the cost function can improve the network's generalization performance [26].

We conduct experiments considering three pretrained face recognition networks: ARCFACE [10], VGGFACE2 [27] and AFFFE [28]. Results are obtained on two widely-explored benchmarks, namely Labeled Faces in the Wild (LFW) [29] and IARPA Janus Benchmark C (IJB-C)[30]. Seeing that LFW was initially designed for the face verification task, we adhere to the open-set protocol designed by Günther *et al.* [4]. Contrarily, IJB-C specifies an open-set identification guideline named TEST-4 that determines how face recognition algorithms must be evaluated. We optimize the hyperparameters on LFW dataset using AFFFE backbone as the feature representation module. Then, the same parameters are employed to a subsequent comparison on IJB-C with ARCFACE and VGGFACE2 as feature descriptors to check its robustness to different domains.

The major contributions of our work are: *(I)* We present a compact neural ensemble that replaces the computationally-expensive retraining of DNNs for faster ensemble learning. *(II)* We examine how the entropy-based loss and the feature augmentation mechanism enable the ensemble to better distinguish known from unknown samples. *(III)* We carry out a parameter selection evaluation to show that the very same setting can be employed in more difficult domains, considering different feature extractors and datasets.

II. PROPOSED APPROACH

A robust open-set face recognition system is expected to determine the identity of those subjects who have been previously enrolled in the gallery and reject the ones of no interest. However, when deploying biometric applications in the real world, experts must be aware that still and motion probe images very likely present low-quality captures along
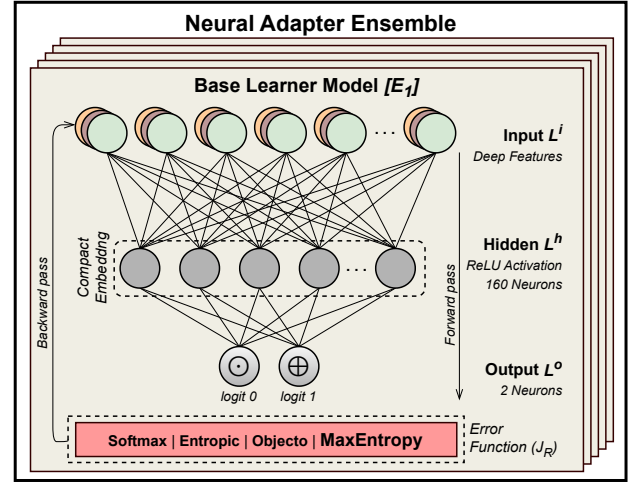

Fig. 1. NEURAL ENSEMBLE AND ITS BASE LEARNERS. Each standalone learner inputs features extracted with any deep network to learn parameters that minimize an open-set loss function and maps samples from $|G|$ individuals from the gallery set into two classes. For each learner, the ensemble keeps a record of the identities that are randomly associated with class $\odot$ (*output 0*) and class $\oplus$ (*output 1*) to later identify these people and reject unknowns.

with illumination variance and occlusion, to name a few [2]. These corrupted image samples obtained at the testing stage tend to misguide biometric systems and, therefore, end up defiling the identification process.

We design three mechanisms to address the aforestated concerns: the Neural Adapter Ensemble (NAE) encompasses binary neural networks and aims to establish a clear boundary between subjects of interest and unknown faces. The Optimized Mix-Up (OMU) augmentation synthesizes negative samples at the feature-space level by interpolating representations of different gallery-enrolled individuals. Maximal-Entropy Loss (MEL) comprehends an entropy and margin-based cost function that exploits negative samples derived from the original gallery set or extrinsic datasets. NAE is capable of boosting the predictive performance of each standalone classifier by training multiple base learners and combining their predictions [31]. OMU-made instances can enhance the generalization capability of each binary model that composes the ensemble. Additionally, MEL supports the network in handling unknown samples by maximizing the entropy of negative samples or penalizing the target class of known samples with a specified margin.

A. Feature Extraction

Given an image sample x, the feature extraction module can be defined as $\mathbf{z} = F_\Theta(x)$, a fragment of the pre-trained DNN's forward pass $\hat{y} = C_\psi \circ F_\Theta(x) = C_\psi \circ F_{\Theta^L} \circ \cdots \circ F_{\Theta^1}(x)$ with classifier C_ψ and L embedding layers in $F_\Theta(x)$. This process propagates image x forward up to the point prior to the last fully-connected layer with softmax activation and outputs the equivalent embedding z at that location. It is important to make sure that the face image x is aligned according to the requirements of the chosen DNN. Therefore, we rely on pre-determined alignment and feature extraction pipelines [32].

B. Maximal-Entropy Loss

The Maximal-Entropy Loss (MEL) is a cost function that addresses training samples in two different manners: *(i)* MEL boosts both intra-class compactness and inter-class separability among known subjects by penalizing the target classes as well as *(ii)* maximizes the entropy of negative samples by proportionately scattering their output scores among all classes. MEL encloses a soft-margin module ($\mathcal{M}$) with a margin $m \geq 0$ that makes the classification more rigorous. Then, given a deep feature representation z extracted from a face sample x, $s_c(z)$ represents the network activation (logit) for class c:

$$\mathcal{M}_c^m(z) = \frac{e^{s_c(z)-m}}{e^{s_c(z)-m} + \sum\limits_{c' \neq c} e^{s_{c'}(z)}} \tag{1}$$

The formulation of MEL ($\mathcal{J}^m$) only adds a handicapping penalty m to known classes, indicated in the first term. In favor of handling negative samples, MEL absorbs the Entropic Open-Set (EOS) loss [21], where y_i stores z_i's corresponding target class $c \in C$ and $\bar{z} \in N$ represents a negative sample:

$$\begin{aligned}
\mathcal{J}^m = &-\mathbb{E}_{(z_i, y_i) \in G} \log \mathcal{M}_{y_i}^m(z_i) \\
&- \mathbb{E}_{\bar{z} \in N} \frac{1}{|C|} \sum_{c \in C} \log \mathcal{M}_c^{m=0}(\bar{z})
\end{aligned} \tag{2}$$

MEL maximizes the uncertainty of negative instances by inducing output activations to lie uniformly distributed over all known classes $c \in C$. The insight of equalizing logit values for unknown samples lies behind not knowing anything about their corresponding class and, therefore, they must hold a similar likelihood of being assigned to any class [21]. For this reason, $\mathcal{J}^m$ is expected to propagate the entropy learned with negative data $\bar{z} \in N$ to unknown probe samples during inference.

C. Optimized Mix-Up Feature Augmentation

We introduce an augmentation strategy called Optimized Mix-Up (OMU) to build *artificial* negative samples. Differently from traditional data augmentation transforms, the designed data synthesis takes place directly at the latent space z and aims to generate tightened decision boundaries around known classes. OMU interpolates two latent embedding representations z_i and z_j into a new representation $\bar{z}$, which is assigned to the negative set N. Such embedding is generated in consonance with a mingling coefficient λ that determines the weight of each original embedding:

$$\bar{z} = \lambda \cdot z_i + (1 - \lambda) \cdot z_j$$
$$\text{s.t. } z_j = \operatorname*{arg\,max}_{(z_{i'}, g_{i'}) \in G} \cos(z_i, z_{i'}) \wedge g_i \neq g_{i'} \tag{3}$$

In summary, given feature vectors z_i and z_j, respectively associated with identities $g_i \neq g_j$, a synthetic negative feature $\bar{z}$ is manufactured in between the closest pairs of known individuals. Unlike existing works [23, 33] where different feature embeddings are randomly selected, OMU seeks the closest cosine-similar representations that, at the same time, belong to different subjects registered in gallery G.

D. Neural Ensemble Models

The ensemble is composed of multiple binary classifiers $E_n \in E$. Each classifier is trained on a different random bisecting split of gallery identities, where the task is to discern these two random groups. For training our base model E_n, we distribute the identities registered in gallery G into two equally-sized disjoint splits. The random segregation guarantees that half of the individuals are assigned to $P_n^\odot$ (partition zero) and the other fraction is allocated in $P_n^\oplus$ (partition one) so that both splits altogether, defined as $P_n = \{P_n^\odot, P_n^\oplus\}$, encompass all the subjects of interest available in the gallery. Even though all base learners share equivalent architecture and hyperparameters, each one of them is trained with an independent and identically distributed arrangement of known identities as class zero or class one.

Associating subject $g \in G$ with one of the two subsets consists of sampling from a Bernoulli distribution with probability $p = 0.5$. This partitioning P_n operates as the function $B_n : G \mapsto \{\odot, \oplus\}$ that attributes subjects g with new labels ($\odot$ or $\oplus$). Then, $G \subset \mathbb{Z}_+$ contains the original gallery identities and B_n holds the respective binary co-domain for partition P_n. Theoretically, the probability of any two subjects of interest sharing the very same sequence of binary attributions decreases as the neural ensemble size expands.

The neural ensemble E comprises the main block of the approach since it is the stage in which the feature augmentation scheme and the open-set loss act together to build a set of discriminative base models. Each base model $E_n \in E$ consists of a multi-layer *perceptron* network with fully-connected layers. In fact, E_n incorporates an input layer L^i, followed by a single hidden layer L^h and an output layer L^o. The input layer takes deep feature representations z extracted with the selected pretrained deep neural network and, consequently, its input size varies according to the DNN's feature layer dimension. As indicated in Figure 1, the hidden layer L^h employs a rectified linear unit (ReLU) activation. Layer L^o contains two neurons and outputs the corresponding activations $(a^\odot, a^\oplus)$ of the two classes. Each base learner is trained using (2), where we use the categorical loss for two classes.

E. Inference with Rank of Candidates

When given a test sample x_p, we first extract the feature embedding $z_p = F_\Theta(x_p)$ and forward these through our ensemble of binary classifiers. By construction, the activations of the base classifier E_n will be close to zero when an unknown sample is presented, and large for the corresponding partition when facing a known sample [21]. Hence, for each gallery subject g, we can simply add the activations of the partition the class g was initially assigned to:

$$\text{sim}(z_p, g) = \sum_n a_n^{B_n(g)}(z_p) \tag{4}$$

with $B_n(g) \in \{\odot, \oplus\}$. The final similarity scores can be used to identify the probe by selecting the gallery subject with the highest score, or rejecting it as unknown when the maximal score is below a certain threshold θ.

TABLE I
LFW EVALUATION. OPEN-SET ASSESSMENT TO SELECT OPTIMAL VALUES FOR PARAMETERS λ, h AND $|E|$.

Parameters	$|E|$					m					λ				
DIR/VALUES	0.10	0.30	0.50	0.75	1.00	0.10	0.20	0.30	0.40	0.50	0.55	0.65	0.75	0.85	0.95
TPIR@FPIR = 1.00	0.82	0.92	**0.94**	**0.94**	**0.94**	**0.94**	**0.94**	**0.94**	**0.94**	0.93	0.94	0.94	0.94	**0.95**	0.94
TPIR@FPIR = 0.10	0.71	0.85	0.86	**0.88**	**0.88**	0.86	0.86	**0.87**	**0.87**	**0.87**	0.87	0.87	**0.89**	**0.89**	0.87
TPIR@FPIR = 0.01	0.57	0.72	0.73	**0.75**	**0.75**	0.73	0.74	**0.77**	0.76	0.74	**0.77**	**0.77**	**0.77**	0.74	0.75

III. EXPERIMENTS

We developed the approach using PyTorch framework [34] along with other Python libraries such as Bob [32, 35] for feature extraction. The neural ensemble operates on representations obtained with SENET50-VGGFACE2 [27], RESNET50-AFFFE [28] and RESNET100-ARCFACE [10] architectures. Training has been carried out on a dedicated server running Debian Linux on a AMD EPYC 7542 32-core 128-thread CPU, 512-GB RAM, and multiple GEFORCE RTX 2080Ti GPUs.

Evaluation Metric: The open-set Receiver Operating Characteristics (O-ROC) is the canonical evaluation metric for open-set biometric systems [36]. The O-ROC plots the True Positive Identification Rate (TPIR) against False Positive Identification Rate (FPIR) by varying threshold θ. The TPIR specifies the probability that subjects from the gallery are correctly identified whereas FPIR corresponds to the number of unknown subjects mistakenly identified as someone enrolled in the gallery. An optimal open-set face identification system has TPIR of 1 at an FPIR not far from 0, while the closed-set *Rank-1* recognition rate can be obtained as TPIR @ FPIR = 1.

Datasets and Evaluation Protocols: We adopt LFW [29] as well as IJB-C [30] benchmarks. We incorporate the open-set LFW partitioning [4] for parameter selection. IJB-C provides a widely adopted open-set protocol TEST-4 that consists of two disjoint gallery sets and a probe collection holding identities from both galleries. We use IJB-C's *gallery A* for training so that probe samples corresponding to identities available in *gallery B* become unknown. IJB-C contains mostly high-quality enrollment data but low-quality probe samples, such that the application of simple open-set techniques usually does not transfer from gallery to probes [37].

Evaluated Approaches: We conduct a series of trials in the interest of verifying the enhancement provided by NAE, MEL, and OMU. We compare the ensemble with NAN [7], a compact network used for multi-class classification. We incorporate distinct cost functions in the evaluation stage: angular-based CosFace (CFL) [9], entropy-maximizing Entropic Open-set (EOS) [21], and the categorical Cross-Entropy Loss (CEL).

A. Parameter Selection

For an unbiased assessment, we select the parameters obtained on LFW to be subsequently used on IJB-C dataset. Such practice reveals whether the method is sufficiently robust to generalize across multiple domains and settings. Table I presents results achieved with AFFFE feature vectors, where we report TPIR values for various FPIR. The parametric assessment seeks to find optimal parameters $|E|$, m and λ, respectively related to the neural ensemble size, loss function penalty, and feature augmentation coefficient.

Empirically, we initially set $|E| = 0.1 \cdot |G|$, $m = 0.1$ and $\lambda = 0.55$, where $|G|$ indicates the gallery size with 610 subjects. When one of the aforementioned parameters is being modified, the two remaining stand fixed. After fixing $|E| = 0.5 \cdot |G|$ as a good compromise between speed and accuracy, we modify m and achieve better model discriminability power for $m = 0.3$. Results show that setting the augmentation coefficient $\lambda = 0.85$ (i.e. unknown samples are relatively similar to known samples) increases the ensemble's ability to distinguish subjects of interest from unknowns. Finally, we assess the optimal number of neurons for hidden layer L^h and, after ranging the number of hidden nodes from 32 to 256 in steps of 16, L^h is empirically set to 160.

B. Neural Network Evaluation

The Neural Adapter Network (NAN) [7] consists of a multi-layer perceptron with two fully-connected hidden layers. The first hidden layer encloses 512 neurons with *ReLU* activation whereas the second one holds 128 neurons. NAN is trained in a multi-class fashion for 200 epochs where the output layer size corresponds to the number of subjects enrolled in the gallery set. We train the adapter network with four distinct cost functions adhering to the best hyper-parameter specified for margin $m = 0.3$ as shown in Table I. The supplementary data is synthesized utilizing the designed OMU feature augmentation method with blending coefficient $\lambda = 0.75$.

Figure 2a demonstrates that the addition of synthetic samples holding equivalent underlying data distribution with the gallery set can significantly improve NAN's accuracy. Not only are the improved results enhanced in the closed-set evaluation (*Rank-1*) but we can observe the superior performance being propagated to the O-ROC metric when the false-positive identification rate decreases. In other words, MEL attains better identification score when FPIR= 10^{-3} than CEL or CFL lie under FPIR= 10^{-2}. The chart also demonstrates the advantage of using MEL when contrasted with EOS as the proposed loss function achieves an analogous closed-set recognition rate but outperforms EOS in the open-set evaluation.

CFL is not one of the most recent margin-based cost functions; still, two recent investigations [38, 39] demonstrated that it exceeds the results obtained with more sophisticated algorithms on IJB-C, such as ArcFace [10], CurricularFace [40] and MagFace [11]. As a result, we believe that CosFace corresponds to a good representation of most angular-margin variants of the traditional Cross-Entropy Loss. CFL demands a special parameter setting as it does not operate with probability scores. During the evaluation of CFL, we had to double the number of epochs as well as the quantity of neurons in the second hidden layer in order to obtain satisfactory results.
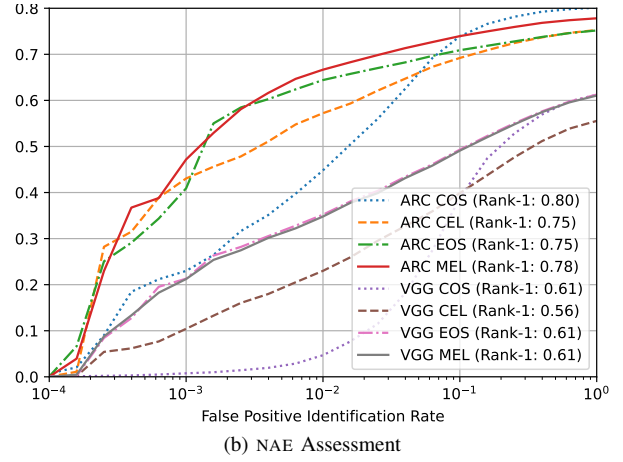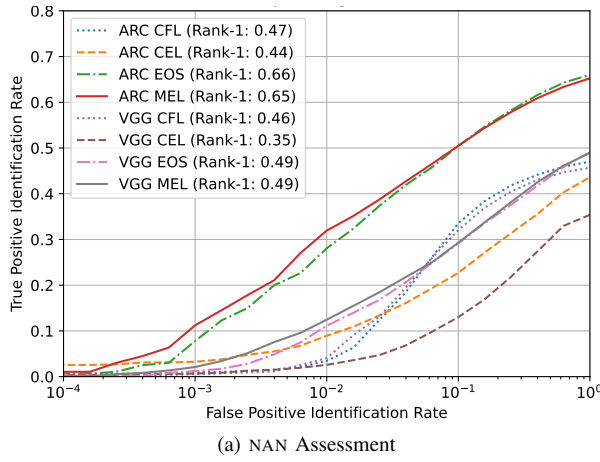
(a) NAN Assessment

(b) NAE Assessment

Fig. 2. IJB-C EVALUATION. We compare our proposed approach with several state-of-the-art approaches for open-set face recognition, using four different loss functions (CEL, CFL, EOS, MEL) and two deep feature representations ARCFACE (ARC) and VGGFACE2 (VGG). (a) indicates the improvement obtained with the Neural Adapter Network (NAN) [7] trained with MEL over long-established loss functions, whereas (b) shows the advance brought by Neural Ensemble (NAE) over multi-class NAN and the well-known cosine-similarity metric (COS) in the open-set identification task.

TABLE II
AUGMENTATION ANALYSIS. EVALUATION OF NAN TRAINED WITH CEL OR MEL ASSOCIATED WITH DIFFERENT AUGMENTATION SCHEMES ON IJB-C.

| Method | Detection and Identification Rate (TPIR@) | | | |
	FPIR=1	FPIR=0.1	FPIR=0.01	FPIR=0.001
CEL	0.44	0.23	0.09	0.03
MEL+LFW	0.58	0.23	0.10	0.03
MEL+SFA	**0.68**	**0.53**	0.31	0.05
MEL+MMU	**0.68**	0.52	0.28	0.04
MEL+OMU	0.66	0.51	**0.33**	**0.10**

Feature Augmentation Analysis: With OMU's optimal blending parameter $\lambda = 0.75$ at hand, we conduct a small set of experiments on NAN with ARCFACE in order to check how much improvement can be obtained with the novel augmentation scheme in multi-class tasks. Table II compares the proposed augmentation strategy with the Manifold Mix-Up (MMU) [23], Stochastic Feature Augmentation (SFA) [24], and the addition of original LFW samples as the negative set. Results show that CEL cannot keep up with cost functions that explore negative samples. We observe that either SFA or MMU achieves higher TPIR values under higher false-positive identifications, that is, when more unknown samples are mistakenly identified as a gallery-enrolled subject. OMU, however, is capable of achieving better detection and identification rate when FPIR drops.

C. Neural Ensemble Evaluation

Figure 2 provides a comprehensive comparison between NAN and NAE as both approaches are trained with analogous feature representations (ARCFACE and VGGFACE2) and cost functions (CEL, EOS and MEL). Results show the dominant generalization power of ensembles when compared to multi-class models. We also adopt COS, an abbreviation for the cosine-similarity computation between probe samples and the gallery of templates, as our second baseline. COS is a common similarity metric for watchlist tasks and is widely employed in the most modern face recognition applications.

As exposed in Figure 1, the ensemble consists of multiple binary models containing a single hidden layer with 160 neurons and *ReLU* activation. NAE exploits synthetic negative samples derived from the gallery set when combined with either EOS or MEL. This strategy guarantees a closer data distribution between original and artificially-made training data since OMU performs a combination of the two closest gallery samples carrying different target classes. Figure 2b shows the experimental evaluation of COS and NAE on IJB-C. The top four curves comprise experiments with ARCFACE and the bottom four refer to VGGFACE2 feature representations.

The association of NAE and MEL achieves superior open-set performance when feature representations are extracted with ARCFACE and false-positive identifications range between 10^{-1} and 10^{-4}. Moreover, MEL outperforms EOS across many FPIR ranges, which shows the importance of learning a more compact feature space through margin m. Experiments with VGGFACE2 show that EOS and MEL cost functions attain competitive results as they outperform methods without negative samples. Apparently, MEL cannot improve the performance over EOS as we presume that original VGGFACE2 representations do not include sufficient resources to handle IJB-C's low-quality probe samples and assist MEL in the training stage.

Results demonstrate that the proposed approach presents an outstanding performance regardless of the adopted network backbone. It reveals that supplementary negative data derived from the gallery set itself equips the ensemble with relevant information and boosts the algorithm's overall performance. In addition, the proposed Maximal-Entropy Loss seems capable of driving each standalone base model toward greater discriminability among known identities as well as escalating the entropy for unknown samples. The ensemble acts as an alternate mechanism to prevent the recurrent retraining of very-deep neural networks every time new individuals are enrolled in the gallery set. As a consequence, it can be attached to the penultimate layer of any pretrained DNN, which eases the maintenance of real-world biometric applications.

IV. CONCLUSION

We introduced three different approaches: a neural ensemble (NAE), a cost function (MEL), and a feature augmentation algorithm (OMU). Results show that the three methods combined provide better open-set accuracy under the presence of extensive false-positive identifications of unknown samples. In opposition to most works in the literature, NAE, MEL and OMU did not have their parameters optimized in such a way they would return favorable results on IJB-C. Instead, the hyper-parameters were selected during the evaluation of a surrogate dataset: LFW. We believe that the proposed method is likely to achieve higher performance on IJB-C dataset if we take its test set into consideration when tuning the hyper-parameters.

This work also provided an interesting insight: "transforming a gallery/training set with linear-alike perturbations may provide better generalization capability than external data". In fact, one of the experiments showed that synthesizing new samples derived from the gallery set preserves the underlying statistics of the training set and, therefore, ends up contributing more to a model generalization power than extrinsic datasets. Consequently, gallery sets with numerous identities but few available samples may not be an obscure limitation anymore when representative synthetic data can be created to assist loss functions in learning better weights.

REFERENCES

[1] J. Rio, D. Moctezuma, C. Conde, I. de Diego, and E. Cabello, "Automated border control e-gates and facial recognition systems," *Computers & Security*, 2016.
[2] H. Du, H. Shi, D. Zeng, X.-P. Zhang, and T. Mei, "The elements of end-to-end deep face recognition: A survey of recent advances," *ACM Computing Surveys*, 2022.
[3] W. Scheirer, A. Rocha, A. Sapkota, and T. E. Boult, "Toward open set recognition," *Pattern Analysis and Machine Intelligence*, 2012.
[4] M. Günther, S. Cruz, E. M. Rudd, and T. E. Boult, "Toward open-set face recognition," in *Computer Vision and Pattern Recognition Workshops*, 2017.
[5] T. Romm, "Amazon's facial-recognition tool misidentified 28 lawmakers as people arrested for a crime," *Washington Post*, July 2017.
[6] K. Hill, "Wrongfully accused by an algorithm," *New York Times*, June 2020.
[7] M. Günther, A. R. Dhamija, and T. E. Boult, "Watchlist adaptation: Protecting the innocent," in *BIOSIG*, 2020.
[8] W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, and L. Song, "Sphereface: Deep hypersphere embedding for face recognition," in *Computer Vision and Pattern Recognition*, 2017.
[9] H. Wang, Y. Wang, Z. Zhou, X. Ji, D. Gong, J. Zhou, Z. Li, and W. Liu, "CosFace: Large margin cosine loss for deep face recognition," in *Computer Vision and Pattern Recognition*, 2018.
[10] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "ArcFace: Additive angular margin loss for deep face recognition," in *Computer Vision and Pattern Recognition*, 2019.
[11] Q. Meng, S. Zhao, Z. Huang, and F. Zhou, "MagFace: A universal representation for face recognition and quality assessment," in *Computer Vision and Pattern Recognition*, 2021.
[12] P. Panareda Busto and J. Gall, "Open set domain adaptation," in *International Conference on Computer Vision*, 2017.
[13] J. Bao, D. Chen, F. Wen, H. Li, and G. Hua, "Towards open-set identity preserving face synthesis," in *Computer Vision and Pattern Recognition*, 2018.
[14] V. E. Vinzi, W. W. Chin, J. Henseler, H. Wang *et al.*, *Handbook of partial least squares*. Springer, 2010.
[15] C. E. Dos Santos and W. R. Schwartz, "Extending face identification to open-set face recognition," in *Conference on Graphics, Patterns and Images SIBGRAPI*, 2014.
[16] R. H. Vareto, S. Silva, F. Costa, and W. R. Schwartz, "Towards open-set face recognition using hashing functions," in *International Joint Conference on Biometrics*, 2017.
[17] J. Henrydoss, S. Cruz, C. Li, M. Günther, and T. E. Boult, "Enhancing open-set recognition using clustering-based extreme value machine (C-EVM)," in *Big Data*, 2020.
[18] R. H. Vareto and W. R. Schwartz, "Unconstrained face identification using ensembles trained on clustered data," in *International Joint Conference on Biometrics*, 2020.
[19] Z. Ma, Y. Liu, X. Liu, J. Ma, and K. Ren, "Lightweight privacy-preserving ensemble classification for face recognition," *IOT*, 2019.
[20] J. Y. Choi and B. Lee, "Ensemble of deep convolutional neural networks with Gabor face representations for face recognition," *Transactions on Image Processing*, 2019.
[21] A. R. Dhamija, M. Günther, and T. E. Boult, "Reducing network agnostophobia," in *Conference on Neural Information Processing Systems*, 2018.
[22] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Big Data*, 2019.
[23] V. Verma, A. Lamb, C. Beckham, A. Najafi, I. Mitliagkas, D. Lopez-Paz, and Y. Bengio, "Manifold mixup: Better representations by interpolating hidden states," in *International Conference on Machine Learning*. PMLR, 2019.
[24] P. Li, D. Li, W. Li, S. Gong, Y. Fu, and T. M. Hospedales, "A simple feature augmentation for domain generalization," in *International Conference on Computer Vision*, 2021.
[25] S. Tao, "Deep neural network ensembles," in *Conference on Machine Learning, Optimization, and Data Science*. Springer, 2019.
[26] S. Wong, A. Gatt, V. Stamatescu, and M. McDonnell, "Understanding data augmentation for classification: when to warp?" in *Conference on Digital Image Computing: Techniques and Applications*. IEEE, 2016.
[27] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman, "Vggface2: A dataset for recognising faces across pose and age," in *International Conference on Automatic Face and Gesture Recognition*, 2018.
[28] C. Li, M. Günther, and T. E. Boult, "ECLIPSE: Ensembles of centroids leveraging iteratively processed spatial eclipse clustering," in *Winter Conference on Applications of Computer Vision*, 2018.
[29] G. Huang, M. Mattar *et al.*, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," in *Real-Life Images: detection, alignment, and recognition*, 2008.
[30] B. Maze *et al.*, "IARPA Janus Benchmark - C: Face dataset and protocol," in *International Conference on Biometrics*, 2018.
[31] Z.-H. Zhou, *Ensemble learning*. Springer, 2021, pp. 181–210.
[32] T. de Freitas Pereira, D. Schmidli, Y. Linghu, X. Zhang, S. Marcel, and M. Günther, "Eight years of face recognition research: Reproducibility, achievements and open issues," *arXiv*, 2022.
[33] D.-W. Zhou, H.-J. Ye, and D.-C. Zhan, "Learning placeholders for open-set recognition," in *Computer Vision and Pattern Recognition*, 2021.
[34] A. Paszke *et al.*, "Pytorch: An imperative style, high-performance deep learning library," *Neural Information Processing Systems*, 2019.
[35] A. Anjos, L. El-Shafey, R. Wallace, M. Günther, C. McCool, and S. Marcel, "Bob: a free signal processing and machine learning toolbox for researchers," in *ACM Multimedia*, 2012.
[36] N. Poh, C. Chan, J. Kittler, J. Fierrez, and J. Galbally, "BEAT–biometrics evaluation and testing: Description of metrics for the evaluation of biometric performance," https://www.academia.edu/download/30791874/d33.pdf, Idiap, Tech. Rep., 2012.
[37] Y. Linghu, "Open-set face recognition with entropic open-set loss," Master's thesis, University of Zurich, 2022.
[38] M. Kim, A. K. Jain, and X. Liu, "AdaFace: Quality adaptive margin for face recognition," in *Computer Vision and Pattern Recognition*, 2022.
[39] M. S. E. Saadabadi, S. R. Malakshan, A. Zafari, M. Mostofa, and N. M. Nasrabadi, "A quality aware sample-to-sample comparison for face recognition," in *Winter Conference on Applications of Computer Vision*. IEEE, 2023.
[40] Y. Huang, Y. Wang, Y. Tai, X. Liu, P. Shen, S. Li, J. Li, and F. Huang, "Curricularface: adaptive curriculum learning loss for deep face recognition," in *Computer Vision and Pattern Recognition*, 2020.