Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 07, No. 04, December 2021.

TWO-WAY METRIC LEARNING WITH MAJORITY AND MINORITY SUBSETS FOR CLASSIFICATION OF LARGE EXTREMELY IMBALANCED FACE DATASET

Ashu Kaushik[1] and Seba Susan[2]

(Received: 16-Jul.-2021, Revised: 12-Sep.-2021, Accepted: 27-Sep.-2021)

ABSTRACT

This paper proposes a new learning methodology involving deep features and two-way metric learning for large, extremely imbalanced face datasets where the number of minority classes and the imbalance ratio are both very high. The problem arises because the faces of some celebrities, being more popular, are readily available in social media and the internet, while the faces of some relatively lesser-known personalities are fewer in number. Resampling being impractical in this scenario, we propose metric learning as the tool for mitigating the class-imbalance problem prior to the classification stage. To reduce the computational overhead associated with metric learning, we separately conduct weakly supervized metric learning with majority and minority class subsets, a process that we call two-way metric learning. Transformation matrices learnt from the majority and minority subsets are used to transform the entire input space twice. The test sample in the transformed space is assigned the class of its nearest neighbor in the training set of the twice-transformed input space. Deep features derived from the state-of-the-art pre-trained deep network VGG-Face form the input space and the aggregate cosine similarity measure is used to find the closest neighbor in the training set of the twice-transformed input space. Experiments on the benchmark LFW face database having 1680 classes of celebrity faces prove that the proposed methodology is more effective than existing methods for the classification of large, extremely imbalanced face datasets. The classification accuracies of the minority classes are especially found to be boosted which is a rare accomplishment among existing methods for imbalanced learning in deep frameworks.

KEYWORDS

Face recognition, Metric learning, VGG-Face, Deep learning, Imbalanced learning, Extremely imbalanced dataset.

1. INTRODUCTION

In this computer and mobile frenzy era, almost everything is getting digitized. The large amount of data that is being generated by the use of such digital devices is creating a havoc and needs to be analyzed, sorted and stored properly and judiciously. Social media platforms, like Facebook, Instagram, Twitter and Zoom, create a large amount of data and logs based upon the usage of the account holder. Face recognition plays a crucial role in detecting and tagging the identity of a person in social media. The users who are very active on social media have a large amount of data associated with them, including images that reveal the identity of the individual. Such users constitute the majority class in the learning framework. On the other hand, users who are less active contribute to lesser data and fewer images that make automated face recognition a difficult task; such users constitute the minority class. Face recognition from such imbalanced datasets, where the difference between the volume of data between the majority and minority classes is very high, is indeed a difficult task [1]-[2]. Learning from imbalanced data is a well-researched problem in data mining [3]-[4] with various solutions proposed ranging from resampling [5] and metric learning [6] to cost-sensitive learning [7]. Selective pruning of majority and minority samples is found helpful, especially when some amount of overlapping is there between the majority and minority classes [8]. Most of the proposed solutions are effective for the binary classification problem, but classification in the multi-class scenario with a highly imbalanced class distribution is still an open research problem [9]. Resampling techniques might work on small toy datasets popular in data mining, but while dealing with a very large dataset comprising of faces derived from social media, consisting of more than a

1. A. Kaushik is a Postgraduate Student in the Department of Information Technology, Delhi Technological University, Delhi, India. Email: ashukaushik8395@gmail.com
2. S. Susan is a Professor in the Department of Information Technology, Delhi Technological University, Delhi, India. Email: seba_406@yahoo.in, ORCID: 0000-0002-6709-6591.

thousand classes and with a very high class-imbalance ratio (ratio of majority to minority population), it is not considered a feasible solution [10].

Novel learning methodologies need to be devised for extremely imbalanced large face databases in order to meet the computational overhead and at the same time improve the classification accuracy, especially for the minority classes having inadequate number of samples to learn from. This is the problem tackled in this paper and we choose metric learning [11] as the tool for transforming the entire input space in order to reduce intra-class differences and increase the inter-class differences. This is achieved by identifying two smaller subsets of the large imbalanced face dataset as the majority and minority classes and performing metric learning using these two subsets. Metric learning would be done in a weakly supervized fashion for both majority and minority subsets to learn the distance metric which can be used to transform the entire input space prior to the classification stage. The contributions made by this paper can be summarized as follows:

1. Metric learning with deep features is introduced as a viable tool for large extremely imbalanced facial datasets having more than a thousand minority classes, for which resampling is not feasible.

2. To reduce the computational overhead associated with metric learning, a weakly supervized learning scheme is devised, for which smaller-sized majority and minority class subsets are identified.

3. The entire input space is then transformed twice, once using the transformation matrix learnt from the majority class subset and likewise from the minority class subset.

4. The aggregate cosine similarity measure is eventually used for the classification of the transformed test sample by finding its closest neighbor in the training set of the twice-transformed input space.

5. Experiments on the large, extremely imbalanced LFW face database having 1680 classes, with large disparity in class populations, yield effective classification, especially for the minority classes, a rare accomplishment among existing methods for imbalanced learning in deep frameworks. The methods proposed so far mostly concentrate on the performance of majority classes only and exclude the minority classes in the learning process.

The further sections of the paper are organized as follows. Section 2 describes the motivation for our work and the proposed methodology. Section 3 analyzes the results of the experimentation and Section 4 outlines the conclusions and the future work.

2. PROPOSED METHOD

2.1 Motivation and Brief Background

Deep neural networks have been used to classify large image datasets, such as ImageNet, and have achieved excellent results [12]. However, they are computationally costly; it would take a long time to setup and train the network for accurate predictions. Pre-trained deep networks trained on large databases and fine-tuned on smaller datasets have been used successfully for complex computer vision tasks, such as face recognition and age estimation [13]. Deep neural networks have, by themselves, some inherent property of improving the scores of minority classes [14]. Pruning of the insignificant features while passing them into the deep networks is a solution to ease out on the computation part [15]. Resampling strategies prevalent in data mining are infeasible for very large, extremely imbalanced image datasets due to the high computational complexity, as is the case in our current work. We therefore, propose to use metric learning using sparse samples for transforming the input space of the large, extremely imbalanced face dataset. Our work is motivated by prior works [2], [16]-[17], [18] that have applied metric learning to mitigate class imbalance for toy datasets. Application of metric learning for large imbalanced datasets, however, requires a lot of computations and very few works have addressed the problem. In our earlier work, which is a precursor of the current work [2], we identified a majority subset of top-186 classes and learned the distance metric using a few samples of each class of the majority subset. The result was an improvement in the performance of majority classes. However, the improvement in the performance of the minority classes was only marginal. In the current work, emphasis is on improving the performance of minority classes by devising a metric learning scheme that would concentrate on the minority classes as well. Some of the earlier techniques propose oversampling of the minority class for improving the performance [19]-[20]. However, this solution is impractical in our case due to the presence of more than a thousand classes and the large size of the dataset. The process pipeline for our method is described in the following sub-sections.

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 07, No. 04, December 2021.

2.2 Deep Feature Extraction from VGG-Face Deep Pre-trained Network

VGG-Face [22], FaceNet [24], DeepFace [25] and OpenFace [39] are a few state-of-the-art deep pre-trained networks customized for face recognition. The Convolutional Neural Network (CNN) [26] is the core neural network of all these models. Because of the large number of hidden layers in their architecture, they are referred to as deep networks. Parkhi *et al.* introduced the pre-trained network VGG-Face in 2014 [22]. It is based on the VGG-16 [27] architecture, which consists of 16 convolutional layers followed by a series of pooling and activation layers. The pre-trained network originally trained on two million images is used to generate a 2622x1 feature embedding for each image in our dataset, as shown in the VGG-Face model process flow recreated in Figure 1. The 2622-dimensional feature vectors are further used, in our work, for metric learning and subsequent classification by a suitable classifier.

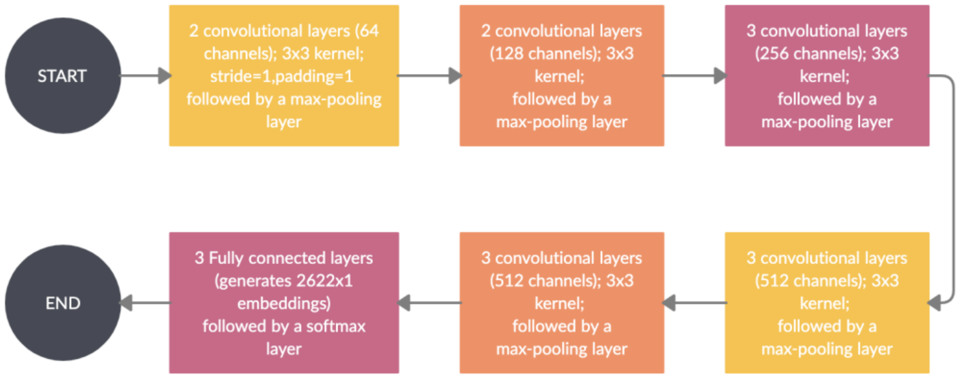


Figure 1. VGG-Face model.

2.3 Metric Learning for Transformation of the Input Space

Various distance metric learning schemes have been proposed, in the past that improved the classification performance of imbalanced datasets. The aim is to transform the input space so as to bring the samples of a class closer and push samples from different classes farther apart. Some of the most prominent distance metric learning algorithms are based on the Mahalanobis distance that is shown in Equation (1) for two feature vectors (**x, y**).

$$\boldsymbol{dm}(\boldsymbol{x},\boldsymbol{y}) = \sqrt{(\boldsymbol{x}-\boldsymbol{y})^{\boldsymbol{T}}\,\boldsymbol{M}(\boldsymbol{x}-\boldsymbol{y})} \tag{1}$$

Here, $\boldsymbol{M}$ represents the positive semidefinite matrix that is to be estimated. It is similar to the Euclidean distance in a different space or a linear projection of the distance between two points. One of the most popular algorithms using the Mahalanobis distance metric is Large Margin Nearest Neighbor (LMNN). Weinberger *et al.* developed LMNN in 2009 [23] and it has since become one of the most widely used data space modification algorithms. It is a supervized metric learning algorithm that may be used before the classification stage. Optimization problem involved is convex and simple to solve. The cost function shown in Equation (2) is the one that must be minimized.

$$\boldsymbol{costf} = (\delta)\boldsymbol{fu}_{push}(\boldsymbol{L}) + (1-\delta)\boldsymbol{fu}_{pull}(\boldsymbol{L}) \tag{2}$$

The loss function has two terms: one relates to the force that pulls samples from the same class closer together, while the other refers to the force that pushes samples from other classes apart. The cost function in (2) is a weighted sum of push and pull functions. The value of δ lies between 0 and 1. This function's transformation matrix limits the margin between k-similar samples to a minimum and maximizes the margin between samples of different classes. When the number of classes is considerable, direct application of metric learning is not recommended due to the computational complexity involved.

LMNN is based on the principle of bringing the samples belonging to the same class closer and samples belonging to different classes are moved further apart, as illustrated in Figure 2. LMNN thus brings about a global linear transformation of the input space that improves the classification of distance-based classifiers, such as *k*NN. We use the cosine similarity measure in the classification stage that follows the metric learning phase in the process pipeline. The cosine similarity measure between two feature vectors (**x, y**) is shown in Equation (3).

$$\boldsymbol{sim}(\boldsymbol{x},\boldsymbol{y}) = \frac{\sum_i \boldsymbol{x}_i \times \boldsymbol{y}_i}{\sqrt{\sum_i \boldsymbol{x}_i^2} \times \sqrt{\sum_i \boldsymbol{y}_i^2}} \quad (3)$$

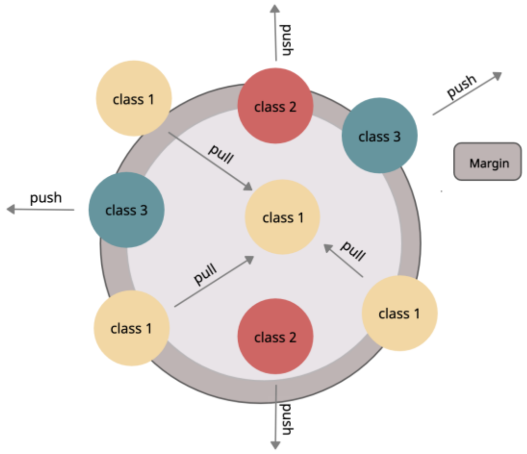


Figure 2. Diagrammatic representation of how LMNN attempts to bring similar & nearest k samples closer and moves dissimilar samples further apart (k=3, as in original paper [23]).

Other examples of metric learning are Neighborhood Component Analysis (NCA) [28], Metric Learning for Kernel Regression (MLKR) [29], Information Theoretic Metric Learning (ITML) [30], Least Squares Metric Learning (LSML) [31] and Sparse Discriminant Metric Learning (SDML) [32]. It was proved in a recent work that LMNN, NCA and MLKR yield the best performance for the *k*NN classification of toy datasets having high imbalance ratio.

2.4 Methodology

The basic outline of the methodology used in our experiments is described next. We segregate the majority and minority classes based upon the number of samples that each class contains. Figure 3 shows the class populations of the Labeled Faces in the Wild (LFW) face dataset [21] used in our experiments, that range from 530 to 2. The graph shows an extremely uneven population distribution.

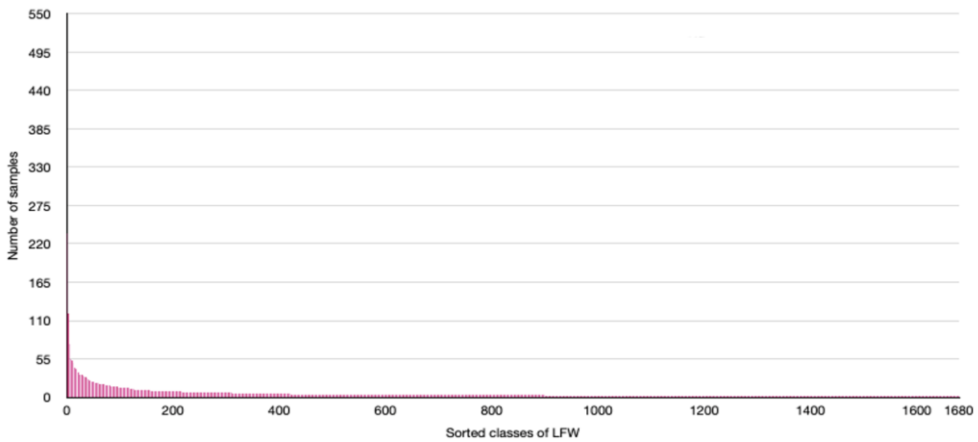
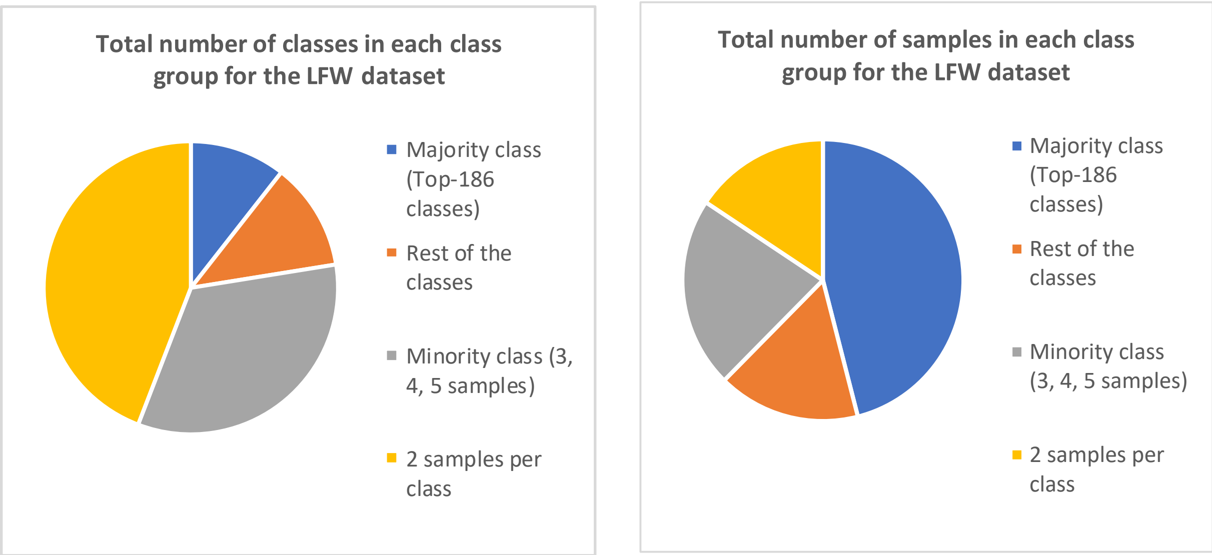


Figure 3. Sorted class populations of the LFW dataset containing 1680 classes of celebrity faces.

As the number of minority classes is very high as compared to the majority classes in the extremely imbalanced LFW dataset, we select a majority class subset and a minority class subset to perform metric learning twice, one from the perspective of the majority class and the other from the perspective of the minority class. This is the primary contribution of our work. We divided the classes based on the class populations, as shown in Figure 4. We define the group of top-186 classes as the majority class and the group of classes with 3, 4 or 5 samples per class as the minority class. To derive

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 07, No. 04, December 2021.

the top-186 classes, as per the procedure in our previous work [2], the class populations are sorted in the decreasing order of their populations and a sum-based partitioning of the sorted class populations yields the threshold as 186 as the lower boundary of the majority class in the LFW dataset. The class-wise and sample-wise groupings are shown in the pie charts in Figure 4 (a) and Figure 4 (b), respectively. It is noted from the pie chart in Figure 4 (a) that the number of minority classes having less than or equal to 5 samples is more than 1300 out of the total available 1680 classes.



(a) (b)

Figure 4. Grouping of the 1680 classes of LFW dataset into majority and minority classes (a) class-wise distribution (b) sample-wise distribution.

Due to the computational complexity involved in metric learning, we have excluded the rest of the classes having samples in the range of 6 to 8 and those having 2 samples from the metric learning computations, since the number of such classes is large and inclusion of these two groups would render metric learning computationally infeasible and impractical. The input space transformation after learning the distance metric is, however, applied to the entire training space. Also, only a few samples from each class are taken into consideration while performing metric learning to reduce the computational expense. Our learning framework is thus an instance of weakly supervized learning. The block diagram for our learning model is shown in Figure 5.

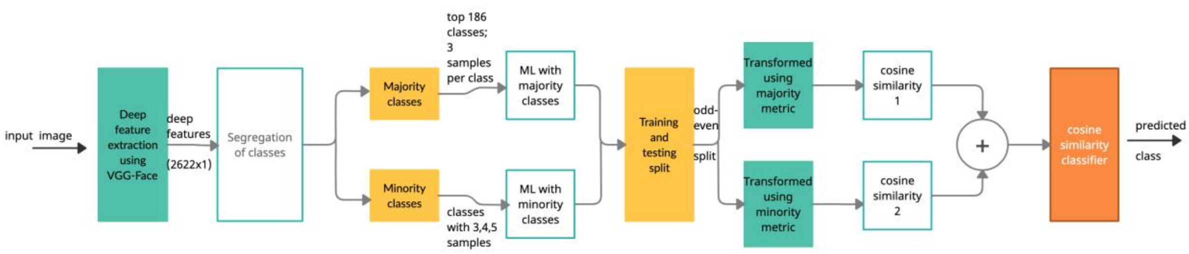


Figure 5. Proposed model.

A deep neural network VGG-Face [22] that is pre-trained on two-million facial images is used to generate the feature embeddings for the LFW face dataset. The gray input images of dimension 64x64 were given as input to the VGG-Face model and vector embeddings of dimension 2622x1 were extracted as per the guidelines in the original paper of VGG-Face [22]. We perform metric learning for both majority class and minority class subsets and get two different transformation metrics. We considered only 3 samples per class from both subsets to reduce the computational cost, since a large number of inter- and intra-class distances need to be calculated. The entire input space is divided into two parts; i.e., training and testing based on alternate sampling and after that, the entire training subset is transformed using both the minority class metric and majority class metric. We use Large Margin Nearest Neighbor (LMNN) as the metric learning technique and it is based on the nearest neighbor rule. The number of nearest neighbors is fixed as k=3, which is the same as given in the original paper of LMNN proposed by Weinberger *et al*. in 2009 [23]. For the minority classes having fewer than 3 samples in the training set, all the samples in the training set are included. The number of neighbors is to be kept small, since a large number of distances, both inter- and intra-class, need to be calculated, which is computationally expensive. The final step would be the classification stage in which the class label of the test sample has to be determined. For each test sample, we transform it using both the distance metrics and in each case, compute the cosine similarity with each sample in the training set of the transformed input space. The two cosine similarity vectors are summed up and the training sample corresponding to the maximum aggregate cosine similarity is selected as the closest neighbor in the training space; its class label is assigned to the test sample.

3. RESULTS

The experiments were performed on the publicly available dataset Labeled Faces in the Wild (LFW) developed in 2007 by Huang *et al.* [21]. It is today a benchmark in the field of facial recognition that is used for training several state-of-the-art pre-trained networks for face recognition. It is a highly imbalanced dataset consisting of 1680 celebrity classes with George W. Bush having the maximum number of samples (=530) and Michel Duclos having the minimum number of samples (=2). Only those celebs were selected who have two or more than two samples and the celebs with only one sample were discarded from our experiments. Out of the 1680 selected celebs, 1369 celebs have <=5 samples, as verified from the pie charts in Figure 4, which proves that the minority classes outnumber the majority classes in the LFW dataset.

We extracted the deep features using the pre-trained VGG-Face model, as discussed in Section 2. The gray-scale images were resized to dimensions 64x64. The pre-trained model generated the 2622-dimensional feature embeddings which were further fed to the learning module. The dataset was divided into majority and minority class subsets as explained in Section 2 and two-way metric learning was performed using these two subsets. The transformation matrices generated were used to transform the entire input space twice, separately. The cosine similarity measure was used to find the closeness of the test sample to a training sample in both the transformed spaces; this was followed by a simple summation of the cosine similarity measures.

The dataset was divided into training and testing sets by alternate sampling. In case of odd number of samples n, the training set contained (n+1)/2 samples and the test set contained (n-1)/2 samples. Cross-validation is done by swapping the training and test sets. The results - Accuracy, F1-score and AUC scores obtained from ROC curves, are compiled in Table 1 for both Validation (V) and Cross-Validation (CV). We compared the performance of our method with that of existing methods: HOG + SVM [33], HOG+ Cosine similarity [34], HOG + Metric learning with majority class [2], VGG-Face + SVM [37], VGG-Face + Cosine similarity [38] and VGG-Face + Metric learning with majority class [36]. The proposed method outperformed all existing methods in terms of accuracy, F1-score and AUC scores as observed from Table 1. The scores are overall on the lower side due to the inclusion of the entire set of 1680 classes including the 779 minority classes with only 2 samples per class of which one sample is used for training and one sample for testing. Most of the earlier experiments on LFW dataset report results only for the majority classes that have at least 10 samples following the face verification protocol in the LFW technical report in [21]. The minority classes are excluded from previous works, since they contribute to class imbalance and deteriorate the overall performance. The state-of-the-art deep networks need a large number of training samples per class for efficient classification [35].

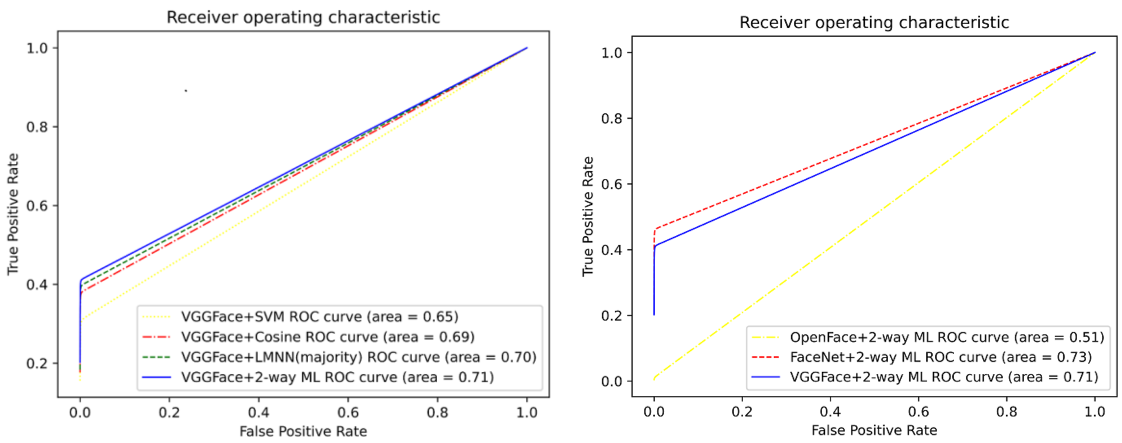
We also compared the current results with that of the weakly supervized metric learning scheme using majority classes [36] by substituting the traditional HOG features in [2] with VGG-Face deep features. We have used an Intel i5 dual core processor clocked at 2.7 GHz and python 3.7 software platform to perform the experiments. The system took half an hour each to learn the distance metrics for both majority and minority subsets. The classification took only a few minutes to execute. The deep features were found to outperform the HOG features, as observed from the classification scores in Table 1. The corresponding ROC curves are shown in Figure 6. A comparison of models that classify

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 07, No. 04, December 2021.

VGG-Face deep features from the ROC graph in Figure 6 (a) reveals that two-way metric learning prior to classification improves the performance scores. Figure 6 (b) compares our method on two other deep features other than VGG-Face; i.e., FaceNet [24] and OpenFace [39]. The FaceNet model performs better than VGG-Face for the proposed two-way metric learning scheme.

Table 1. Performance comparison of various methods on the LFW dataset.

Method	AUC		F1-score		Accuracy	
	V	CV	V	CV	V	CV
HOG + SVM [33]	0.528	0.523	0.059	0.053	28.7%	27.4%
HOG + Cosine Similarity [34]	0.544	0.541	0.078	0.076	21.5%	19.1%
HOG + LMNN metric learning with majority subset [2]	0.556	0.554	0.1006	0.097	26.8%	24.6%
VGG-Face + SVM [37]	0.654	0.635	0.287	0.263	55.4%	51.1%
VGG-Face + Cosine similarity [38]	0.689	0.675	0.342	0.329	55.4%	51.9%
VGG-Face + LMNN metric learning with majority subset [36]	0.697	0.681	0.355	0.339	57%	53.6%
VGG-Face + Two-way metric learning (proposed)	0.705	0.689	0.372	0.356	58.5%	54.8%

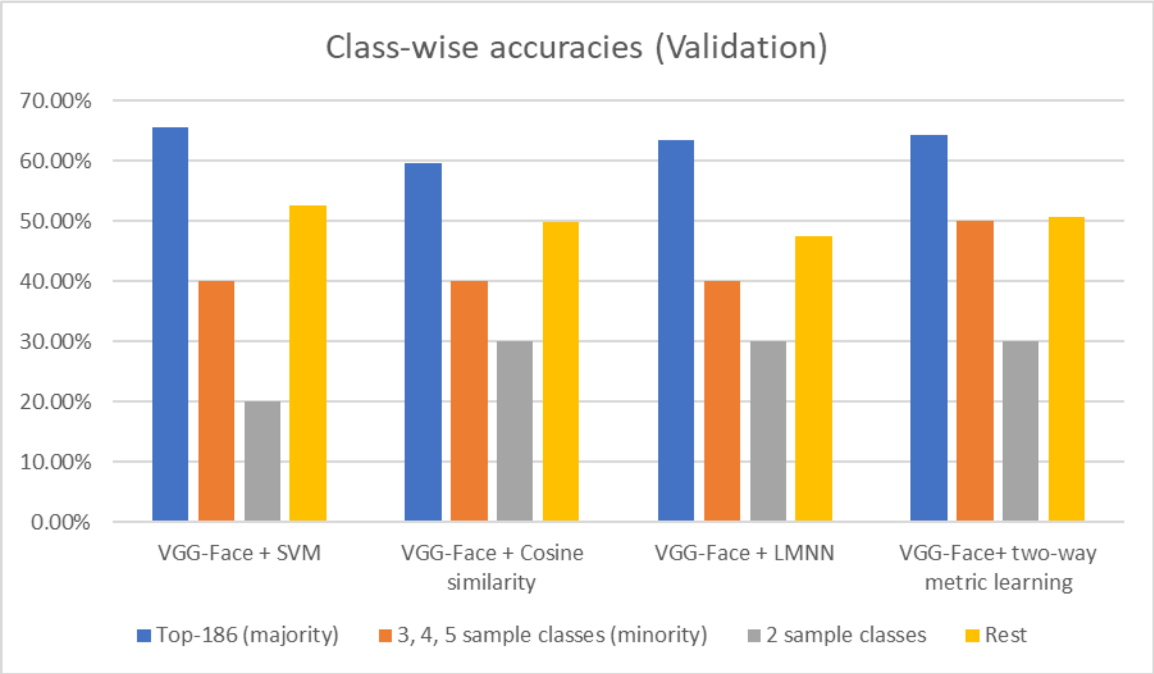


(a) (b)

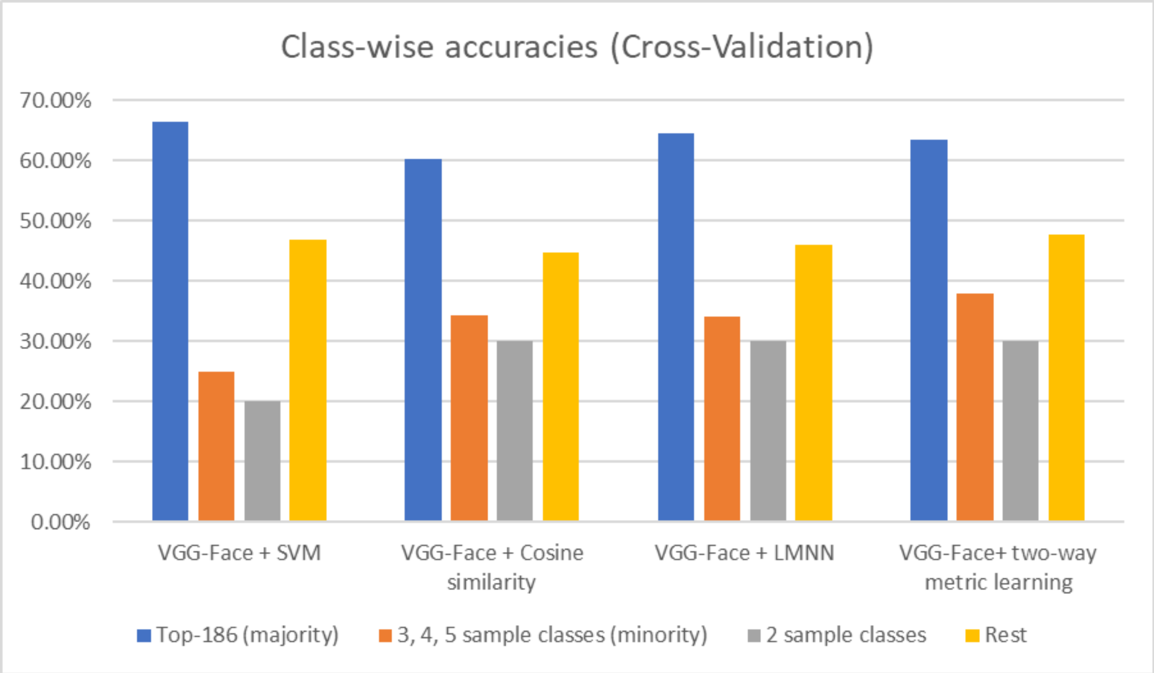
Figure 6. ROC graphs for (a) classification based on VGG-Face deep features by various methods, (b) classification based on VGG-Face, FaceNet and OpenFace deep features for the proposed two-way metric learning scheme.

Some of the class-wise accuracies are shown in Figure 7 for the majority and minority classes to understand the impact of our two-way metric learning scheme as opposed to a scenario where there is no metric learning and the deep features extracted from VGG-Face are learned directly by the classification framework. The cosine similarity measure is the classifier. As observed, Figure 7 shows a consistent performance for all majority classes for the two-way metric learning scheme, which is at par with VGG-Face + SVM. However, for the minority classes with 3, 4, 5 samples per class, a significant improvement in accuracy was recorded, with the accuracies jumping from 0% to 50% and above, for most of the minority classes.

The 2-sample classes showed a higher performance than VGG-Face+SVM. The category of classes tagged as “Rest” contain about 6 to 8 samples each. The performance of this set of classes was found improved over VGG-Face+Cosine similarity and VGG-Face+LMNN, though the performance was marginally lower than that of VGG-Face+SVM.



(a)



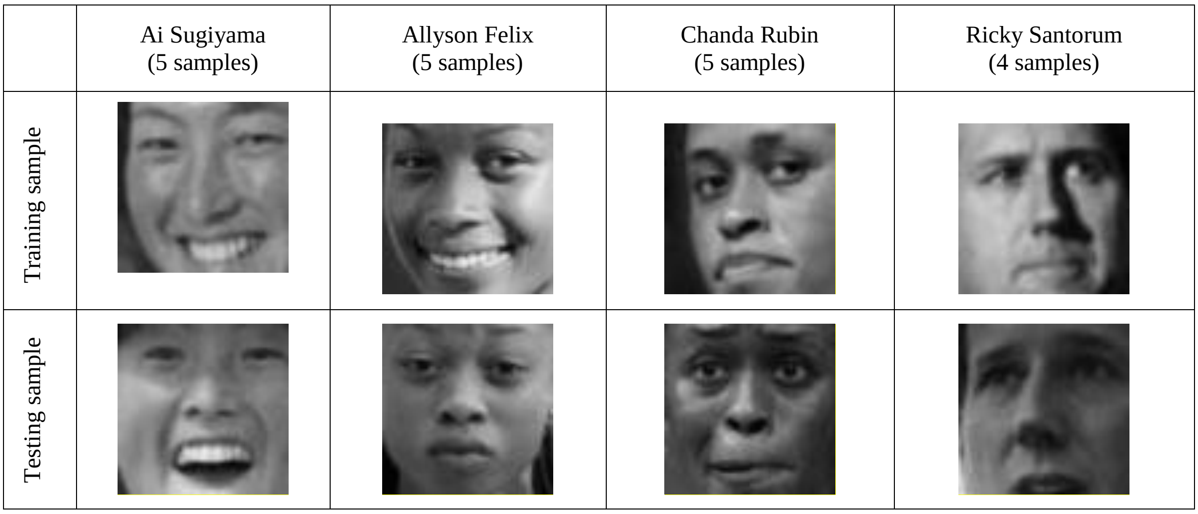
(b)

Figure 7. Performance comparison of majority class (top-186 classes), minority class (3, 4, 5 samples), 2-sample classes, rest of the classes (with samples in the range 6 to 8), with and without metric learning for (a) validation and (b) cross-validation experiments. (*LMNN was used as metric learning method).

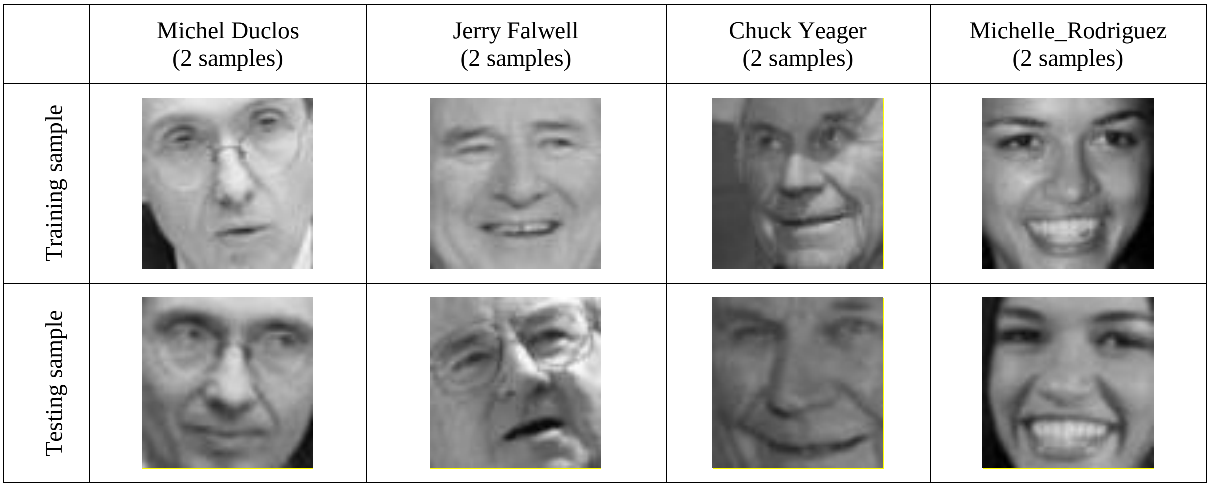
Some success cases and failure cases for the proposed method are shown in Figure 8. The success cases shown are examples when metric learning proved to be useful for the classification. The failure cases are those, which were not classified by our method. Figure 9 shows the comparison between NCA, LMNN and MLKR metric learning schemes for the proposed methodology of two-way metric learning with deep features. The classification accuracies achieved for the top-10 majority classes are shown for all three metric learning schemes.

It is noted that LMNN significantly outperforms NCA and MLKR in terms of classification accuracies. LMNN involves minimization of the distance between each training sample and its k

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 07, No. 04, December 2021.



(a)



(b)

Figure 8. (a) Some success cases of the proposed method, where VGG-Face features without metric learning could not classify the faces and (b) Some failure cases of the proposed method.

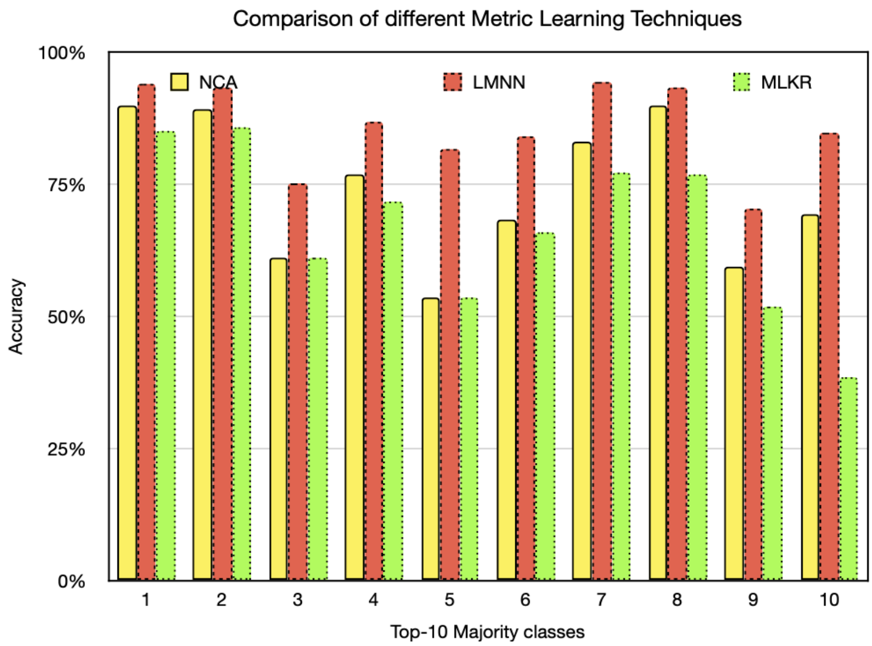


Figure 9. Performance comparison of top-10 majority classes of LFW for different metric learning techniques NCA, MLKR and LMNN.

nearest neighbors belonging to the same class while pushing the differently labelled samples farther apart. LMNN thus projects the input space into metric space in such a way that the inter-class similarities could be measured more accurately.

The primary contribution of our work as compared to our previous work and other works in literature is the improvement of accuracy of the minority classes. The challenge, here, was the existence of more than a thousand minority classes containing sparse samples, rendering metric learning a difficult task. On comparison with other methods, especially VGG-Face+SVM, we observe that though the accuracy of the majority class was comparable, the accuracy of the minority class significantly improved over all existing methods.

4. CONCLUSIONS

A novel learning methodology for large, extremely imbalanced face databases is proposed in this paper that involves deep features and two-way metric learning. LMNN is the metric learning scheme used. Deep features are extracted from the VGG-Face pre-trained model that is trained on two-million facial images. Majority and minority class subsets are identified based on the class population. Metric learning is applied twice, once for the majority subset and the second time for the minority subset. The closeness of the test sample from each training sample in the twice-transformed input space is measured using the sum of the cosine similarities computed in the two cases. The class of the closest training sample, in both the transformed spaces taken together, is assigned as the class of the test sample. Metric learning is known to transform the input space to bring samples of a class closer together. Two-way metric learning introduced in our scheme aims to improve the classification scores, especially for the minority classes, since it brings the few samples in the minority class closer together. Experiments were conducted on the LFW face dataset containing more than a thousand minority classes and the classification scores achieved indicate that the proposed learning technique is more effective than the existing methods for the classification of large, extremely imbalanced face datasets. The LFW dataset, to the best of our knowledge, is the largest extremely imbalanced dataset available for face recognition, which is the central theme of this paper. Our method can be easily adapted to other datasets having different scales of imbalance.

REFERENCES

[1] C. Huang, Y. Li, C. C. Loy and X. Tang, "Deep Imbalanced Learning for Face Recognition and Attribute Prediction," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 42, no. 11, pp. 2781-2794, 2019.

[2] S. Susan and Ashu Kaushik, "Weakly Supervized Metric Learning with Majority Classes for Large Imbalanced Image Dataset," Proceedings of the 4th International Conference on Big Data and Internet of Things, pp. 16-19, DOI: 10.1145/3421537.3421549, 2020.

[3] H. He and E. A. Garcia, "Learning from Imbalanced Data," IEEE Transactions on Knowledge and Data Engineering, vol. 21, no. 9, pp. 1263-1284, 2009.

[4] S. Xuan, G. Liu, Z. Li, L. Zheng, S. Wang and C. Jiang, "Random Forest for Credit Card Fraud Detection," Proc. of the 15th IEEE International Conference on Networking, Sensing and Control (ICNSC), pp. 1-6, Zhuhai, China, 2018.

[5] F. Zhang, G. Liu, Z. Li, C. Yan and C. Jiang, "GMM-based Undersampling and Its Application for Credit Card Fraud Detection," Proc. of the IEEE International Joint Conference on Neural Networks (IJCNN), pp. 1-8, Budapest, Hungary, 2019.

[6] S. Susan and A. Kumar, "DST-ML-EkNN: Data Space Transformation with Metric Learning and Elite K-nearest Neighbor Cluster Formation for Classification of Imbalanced Datasets," Proc. of Advances in Artificial Intelligence and Data Engineering, Part of the Advances in Intelligent Systems and Computing Book Series (AISC), vol. 1133, pp. 319-328, Springer, Singapore, 2021.

[7] H. Zhu, G. Liu, M. Zhou, Y. Xie, A. Abusorrah and Q. Kang, "Optimizing Weighted Extreme Learning Machines for Imbalanced Classification and Application to Credit Card Fraud Detection," Neurocomputing, vol. 407, pp. 50-62, DOI: 10.1016/j.neucom.2020.04.078, 2020.

[8] Z. Li, M. Huang, G. Liu and C. Jiang, "A Hybrid Method with Dynamic Weighted Entropy for Handling the Problem of Class Imbalance with Overlap in Credit Card Fraud Detection," Expert Systems with Applications, vol. 175, pp. 114750, DOI: 10.1016/j.eswa.2021.114750, 2021.

[9] S. Wang and X. Yao, "Multiclass Imbalance Problems: Analysis and Potential Solutions," IEEE Trans. on Systems, Man and Cybernetics, Part B (Cybernetics), vol. 42, no. 4, pp. 1119-1130, 2012.

[10] T. Hasanin, T. M. Khoshgoftaar, J. L. Leevy and R. A. Bauder, "Severely Imbalanced Big Data Challenges: Investigating Data Sampling Approaches," J. of Big Data, vol. 6, no. 1, pp. 1-25, 2019.

[11] B. Kulis, "Metric Learning: A Survey," Foundations and Trends in Machine Learning, vol. 5, no. 4, pp. 287-364, 2012.

[12] A. Krizhevsky, I. Sutskever and G. E. Hinton, "Imagenet Classification with Deep Convolutional Neural Networks," Advances in Neural Information Processing Systems, vol. 25, pp. 1097-1105, 2012.

[13] A. Al-Shannaq and L. Elrefaei, "Age Estimation Using Specific Domain Transfer Learning," Jordanian Journal of Computers and Information Technology (JJCIT), vol. 6, no. 2, pp. 122-139, 2020.

[14] J. M. Johnson and T. M. Khoshgoftaar, "Survey on Deep Learning with Class Imbalance," Journal of Big Data, vol. 6, no. 1, pp. 1-54, 2019.

[15] R.-C. Chen and C.-Y. Liao, "Deep Learning to Predict User Rating in Imbalance Classification Data Incorporating Ensemble Methods," Proc. of the IEEE International Conference on Applied System Invention (ICASI), pp. 200-203, Chiba, Japan, 2018.

[16] N. Wang, X. Zhao, Y. Jiang and Y. Gao, "Iterative Metric Learning for Imbalance Data Classification," Proc. of the 27th International Joint Conference on Artificial Intelligence (IJCAI-18), pp. 2805-2811, [Online], available: https://www.ijcai.org/proceedings/2018/0389.pdf, 2018.

[17] L. Gautheron, A. Habrard, E. Morvant and M. Sebban, "Metric Learning from Imbalanced Data with Generalization Guarantees," Pattern Recognition Letters, vol. 133, pp. 298-304, 2020.

[18] S. Susan and A. Kumar, "Learning Data Space Transformation Matrix from Pruned Imbalanced Datasets for Nearest Neighbor Classification," Proc. of the IEEE 21st International Conference on High Performance Computing and Communications; IEEE 17th International Conference on Smart City; IEEE 5th International Conference on Data Science and Systems (HPCC/SmartCity/DSS), pp. 2831-2838, Zhangjiajie, China, 2019.

[19] S. Barua, Md. M. Islam, X. Yao and K. Murase, "MWMOTE - Majority Weighted Minority Oversampling Technique for Imbalanced Data Set Learning," IEEE Transactions on Knowledge and Data Engineering, vol. 26, no. 2, pp. 405-425, 2012.

[20] V. Ganganwar, "An Overview of Classification Algorithms for Imbalanced Datasets," International Journal of Emerging Technology and Advanced Engineering, vol. 2, no. 4, pp. 42-47, 2012.

[21] G. B. Huang, M. Mattar, T. Berg and E. Learned-Miller, "Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments," Technical Report in Workshop on Faces in'Real-Life'Images: Detection, Alignment and Recognition, [Online], Available: http://vis-www.cs.umass.edu/lfw/lfw.pdf, 2008.

[22] O. M. Parkhi, A. Vedaldi and A. Zisserman, "Deep Face Recognition," Proc. of the British Machine Vision Conference (BMVC), pp. 41.1-41.12, [Online], Available: https://www.robots.ox.ac.uk/~vgg/ publications/2015/Parkhi15/parkhi15.pdf, Sep. 2015.

[23] K. Q. Weinberger and L. K. Saul, "Distance Metric Learning for Large Margin Nearest Neighbor Classification," Journal of Machine Learning Research, vol. 10, no. 2, pp. 207-244, 2009.

[24] F. Schroff, D. Kalenichenko and J. Philbin, "FaceNet: A Unified Embedding for Face Recognition and Clustering," Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 815-823, Boston, MA, USA, 2015.

[25] Y. Taigman, M. Yang, M. A. Ranzato and L. Wolf, "DeepFace: Closing the Gap to Human-level Performance in Face Verification," Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1701-1708, Columbus, OH, USA, 2014.

[26] Y. LeCun and Y. Bengio, "Convolutional Networks for Images, Speech and Time Series," The Handbook of Brain Theory and Neural Networks, vol. 3361, no. 10, pp. 1-14, 1995.

[27] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-scale Image Recognition," Proc. of ICLR 2015, arXiv preprint arXiv: 1409.1556, 2014.

[28] J. Goldberger, G. E. Hinton, S. T. Roweis and R. R. Salakhutdinov, "Neighborhood Components Analysis," Advances in Neural Information Processing Systems, pp. 513-520, [Online], Available: https://www.cs.toronto.edu/~hinton/absps/nca.pdf, 2005.

[29] K. Q. Weinberger and G. Tesauro, "Metric Learning for Kernel Regression," Artificial Intelligence and Statistics, pp. 612-619, [Online], Available: http://proceedings.mlr.press/v2/weinberger07a/Weinberger 07a.pdf, 2007.

[30] J. V. Davis, B. Kulis, P. Jain, S. Sra and I. S. Dhillon, "Information-theoretic Metric Learning," Proc. of the 24th Int. Conf. on Machine Learning, pp. 209-216, DOI: 10.1145/1273496.1273523, ACM, 2007.

[31] E. P. Xing, M. I. Jordan, S. J. Russell and A. Y. Ng, "Distance Metric Learning with Application to Clustering with Side-information," Proc. of the 15th International Conference on Neural Information Processing Systems (NIPS'02), pp. 521-528, 2003.

[32] G.-J. Qi, J. Tang, Z.-J. Zha, T.-S. Chua and H.-J. Zhang, "An Efficient Sparse Metric Learning in High-dimensional Space *via* l_1-penalized Log-determinant Regularization," Proc. of the 26th Annual Int. Conf. on Machine Learning, pp. 841-848, DOI: 10.1145/1553374.1553482, ACM, 2009.

[33] H. S. Dadi and G. K. M. Pillutla, "Improved Face Recognition Rate Using HOG Features and SVM Classifier," IOSR Journal of Electronics and Communication Eng., vol. 11, no. 04, pp. 34-44, 2016.

[34] D. Chen, X. Cao, F. Wen and J. Sun, "Blessing of Dimensionality: High-dimensional Feature and Its Efficient Compression for Face Verification," Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3025-3032, Portland, OR, USA, 2013.

[35] Y. C. Wong, L. J. Choi, R. S. Sarban Singh, H. Zhang and A. R. Syafeeza, "Deep Learning-based Racing Bib Number Detection and Recognition," Jordanian Journal of Computers and Information Technology (JJCIT), vol. 5, no. 3, pp. 181-194, 2019.

[36] A. Kaushik and S. Susan, "Metric Learning with Deep Features for Highly Imbalanced Face Dataset," Proc. of the International Conference on Innovative Computing and Communications, Part of the Advances in Intelligent Systems and Computing Book Series, vol. 1394, pp. 639-646, 2022.

[37] B. Knyazev, R. Shvetsov, N. Efremova and A. Kuharenko, "Leveraging Large Face Recognition Data for Emotion Classification," Proc. of the 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), pp. 692-696, Xi'an, China, 2018.

[38] S. Karahan, M. K. Yildirum, K. Kirtac, F. S. Rende, G. Butun and H. K. Ekenel, "How Image Degradations Affect Deep CNN-based Face Recognition?," Proc. of the International Conference of the Biometrics Special Interest Group (BIOSIG), pp. 1-5, Darmstadt, Germany, 2016.

[39] T. Baltrušaitis, P. Robinson and L.-P. Morency, "OpenFace: An Open Source Facial Behavior Analysis Toolkit," Proc. of the IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 1-10, Lake Placid, NY, USA, 2016.

ملخص البحث:

تقترح هذه الورقة منهجية تعلّم جديدة تتضمّن سِماتٍ عميقةٍ والتّعلُّم ذا الطّريقيْن لمجموعات البيانات المتعلّقة بتمييز الوجوه التي تتميّز بالضّخامة وشدّة عدم الاتّزان حيث عدد الأصناف الأقلية ومعدّل عدم الاتّزان كبيران جداً. وتبرز المشكلة من أنّ بعض الوجوه الأكثر شيوعاً متوافرة في وسائل التّواصل الاجتماعي والإنترنت، بينما وجوه الشّخصيات المعروفة بدرجةٍ أقلّ هي أقلّ عدداً. ولأنّ إعادة تشكيل العيّنات في ظلّ هذا السيناريو أمرٌ غير عمليّ، فإننا نقترح التّعلّم القياسي أداةً للتخفيف من مشكلة عدم الاتّزان في الأصناف قبل مرحلة التّصنيف. وللتقليل من تكلفة الحسابات المرتبطة بالتّعلّم القياسي، نقوم –بشكل منفصل- بإجراء التّعلّم القياسي المراقب بشكلٍ ضعيف على مجموعات البيانات الفرعية الأغلبية والأقلية، في عمليةٍ نسميها التّعلّم القياسي ذا الطريقْين. وقد أثبتت التجارب على مجموعة البيانات المرجعية لتمييز الوجه LFW التي تحوي (1680) صنفاً من الوجوه أنّ المنهجية المقترحة كانت أكثر فاعلية مقارنةً بالطرق القائمة. حيث انّ دقّة التّصنيف للأصناف الأقليّة ارتفعت على نحوٍ خاص، وهو إنجاز نادر في طرق تصنيف مجموعات البيانات المعتمدة على التّعلّم غير المتوازن العميق.