Localized Metric Learning for Large Multi-Class Extremely Imbalanced Face Database

Seba Susan* and Ashu Kaushik
Department of Information Technology, Delhi Technological University
Delhi, India - 110042
seba_406@yahoo.in

Abstract. Metric learning serves to mitigate, to a great extent, the class-imbalance problem associated with large multi-class image databases. However, the computational complexity associated with metric learning increases when the number of classes is very large. In this paper, a novel localized metric learning scheme is proposed for a large multi-class extremely imbalanced face database with an imbalance ratio as high as 265:1. The Histogram of Gradient (HOG) features are extracted from each facial image and these are given as input for metric learning. The proposed scheme involves confining the metric learning process to local subspaces that have similar class populations. The training dataset is divided into smaller subsets based on the class populations such that the class imbalance ratio within a local group does not exceed 2:1. The locally learnt distance metrics are then, one by one, used to transform the entire input space. The nearest neighbor of the test sample, in the training set, is noted for each transformation. A comparison amongst all transformations for the closest nearest neighbor in the training set establishes the class of the test sample. Experiments are conducted on the highly imbalanced benchmark Labeled Faces in the Wild (LFW) dataset containing 1680 classes of celebrity faces. All classes are retained for the experimentation including those minority classes having just two samples. The proposed localized metric learning scheme outperforms the state of the art for face classification from large multi-class extremely imbalanced face databases.

Keywords: Extremely imbalanced dataset, Face recognition, Subspace-learning, Multi-class, Metric learning, Large margin nearest neighbor.

1 Introduction

An uneven class distribution in multimedia data is a matter of concern to machine learning researchers. The reason is the bias induced while training the classification model that mis-classifies most of the minority classes whose samples are scant as compared to that of the majority classes. The class imbalance problem, as it is generally called, is a well-known highlighted issue in data mining [1]. The conventional solutions mostly revolve around resampling strategies [2] and cost-sensitive learning [3]. Resampling involves data manipulation by oversampling the minority class, undersampling the majority class, or hybrid sampling approaches [4]. Cost-sensitive learning, on the other hand, modifies the learning process such that it becomes more sensitive to the cost function pertaining to the under-represented minority class [5]. However, these solutions, that are popular in data mining, were not found easy to adapt for multimedia datasets, mainly due to the large scale of the datasets involved, the large number of classes, and the high imbalance ratio between the majority and the minority classes [6]. Oversampling the minority class created replicas that further increased the scale of the dataset, while undersampling led to loss of discriminative information. Experiments to generate new minority samples were conducted in [7] that involved affine transformations of samples from the under-represented minority class. Data augmentation using Generative Adversarial Networks (GAN) is yet another technique to correct class imbalance by creating synthetic images for the minority class [8]. Creation of synthetic samples further increases the computational overhead associated with a large dataset, with parallel computing and high-performance computing machines usually being involved [9].

Most of the solutions to the class-imbalance problem in literature, including those discussed above, are meant for binary classification problems i.e. for datasets having a single majority class and a single minority class. All these algorithms work microscopically with majority class samples and/or minority class samples as the chief particles of interest, and focus on either eradicating the redundant majority samples or replicating the minority samples. The severity of the class imbalance problem is based on the ratio of the majority class population to minority class population that is defined as the imbalance ratio (IR). IRs of 50:1, 100:1 or higher are considered as indicators of extremely imbalanced datasets. In cases where the dataset was multi-class, researchers transformed the problem into a two-class one by segregating the classes into two clusters based on the relative similarity between class populations, for the purpose of testing the utility of their sampling algorithms [10]. Fig. 1 illustrates the binarization of several popular multi-class datasets in various works in literature.

Pageblocks- 5 classes	
Classes: ‘Graphic’, ‘Vert. line, ‘Picture’	Minority class (231 instances)
Remaining classes	Majority class (5245 instances)

(a)

Yeast- 10 classes	
Classes: ‘ME2’, ‘ME3, ‘EXC’, ‘POX’, ‘ERL’, ‘VAC’	Minority class (304 instances)
Remaining classes	Majority class (1180 instances)

(b)

Yeast- 10 classes	
Class: ‘NUC’	Minority class (429 instances)
Class: ‘CYT’	Majority class (463 instances)

(c)

Glass – 6 classes	
Class:	Minority class (51 instances)
Remaining classes	Majority class (163 instances)

(d)

Fig. 1. Binarization of multi-class datasets in different works in literature: (a) *Pageblocks* by Barua et al. (2014) [11] (b) *Yeast* by Barua et al. (2014) [11] (c) *Yeast* by Tao et al. (2019) [12] (d) *Glass* by Liu et al. (2020) [13]

Apart from serving the purpose of testing and trying out new resampling strategies, the binarization of multiple classes destroys the original class information and renders the existing solutions useless for classifying the original multi-class dataset. Adaptation of existing binarized classification solutions to large multi-class multimedia datasets where the number of classes is too high and the imbalance condition is severe, is thus highly impractical. In our paper, we propose a novel solution based on distance metric learning for tackling the class-imbalance in large multi-class image datasets. Distance metric learning [14] induces data space transformation that results in an improved classification for a dataset with imbalanced class-distribution. The chief advantage over resampling strategies is the prevention of loss of information by undersampling, and the creation of redundancy and computational overhead by oversampling. Though this idea has been explored before by Feng et al. (2018) [15], Wang et al. (2018) [16], Wang et al. (2021) [17] and Susan and Kumar (2019) [10], it has not been investigated, to the best of our knowledge, for larger datasets having more than thousand classes and for a multi-class scenario with severe class-imbalance problem. The reason is the computational complexity of metric learning that attains impractical limits when the number of classes are extremely high (>1000). Specifically, we focus on the large multi-class Labeled Faces in the Wild (LFW) face database [25] containing 1680 classes, with a highly uneven class population profile ranging from 530 to 2. The researchers in [35] constrain metric learning to selected samples of the top-186 classes of LFW. However, in the current work, all the classes of the LFW dataset are involved in the metric learning process. To minimize the computational complexity, the low-dimensional HOG features are used to prove the effectiveness of the proposed localized metric learning. The contributions of our work are summarized as follows:

1. A novel localized metric learning scheme is introduced for handling severe class-imbalance in large multi-class face datasets; the LFW face dataset used in our experiments has a high imbalance ratio of 265:1.
2. The classes are first sorted based on the decreasing order of class populations, and then divided into subsets such that the maximum imbalance ratio within a subset is approximately 2:1.
3. Metric learning is performed in a distributed manner, locally within each subset. The local transformation matrices learnt are used, one by one, to transform the training and testing input space that comprise of HOG features.
4. For each transformation, the minimum Manhattan distance of the test sample from the training samples is computed. Comparison among all transformations establishes the class label of the test sample from the label of the nearest neighbor in the training space.
5. Experimental analysis has been conducted on a challenging public facial image database having 1680 classes with all the minority classes retained, and results are compared with the state of the art.
6. Our classification scores are observed to be the highest among all existing methods in terms of accuracy, AUC score and F1-score, for a very challenging train-test split of 50:50 with 2-fold cross-validation.

The organization of this paper is as follows. Section 2 reviews the metric learning concepts and the limitations of its direct application to large multi-class datasets. Section 3 describes the proposed localized metric learning scheme and section 4 describes the experimental setup and analyzes the results. The final conclusions are drawn in section 5.

2 Metric learning as an antidote for the class imbalance problem

Distance Metric Learning (ML) is the process of learning a distance metric from labeled training samples for efficient classification [14]. The process involves learning a transformation matrix that brings similar points closer and pushes dissimilar points farther apart. The result is reduced intra-class differences and increased inter-class differences. Without depleting or adding new samples, metric learning achieves improved classification for imbalanced datasets. For best results, ML is used along with distance or similarity classifiers in the classification module [18]. Weinberger et al. proposed the Large Margin Nearest Neighbor (LMNN) metric learning [19, 20] that learns a Mahalanobis distance metric for k-Nearest Neighbor (kNN) classification by semidefinite programming. We use LMNN for metric learning and data space transformation in our experiments, hence we start with a brief discussion on its procedure.
Consider an input space of *n* training samples $\boldsymbol{X} = \{\vec{\boldsymbol{x}}_{\boldsymbol{i}}\}, i = 1, 2, \ldots, n$. The corresponding class labels are denoted by $\{y_i\}, i = 1, 2, \ldots, n$. The objective of LMNN is to learn a transformation matrix $\boldsymbol{T}$ that transforms the input space such that the *k* nearest neighbors of a training sample mandatorily belongs to the same class. The squared distance between two samples $\boldsymbol{x_i}$ and $\boldsymbol{x_j}$ in the linearly transformed input space is computed as

$$\boldsymbol{D} = \left\| \boldsymbol{T}\left(\boldsymbol{x}_i - \boldsymbol{x}_j\right) \right\|^2 \tag{1}$$

The authors in [19, 20] have proposed *k*=3 to be a suitable choice, and this is adopted for our implementation of LMNN in this work. For classes with samples fewer than 3, *k* was reduced accordingly. The cost function of LMNN imposes a penalty on large distances between a training sample and its *k* nearest neighbors belonging to the same class, and another penalty on small distances between the training sample and training samples belonging to other classes. The latter penalty is responsible for the term “large margin nearest neighbor” since it ensures that a large margin is maintained between samples belonging to different classes. The Euclidean distance between samples in the transformed space is proved to be equivalent to the Mahalanobis distance between the samples in the original input space [20]. The newly transformed input space $\boldsymbol{X_{new}}$ yields improved performance for distance- or similarity-based classifiers such as kNN. The transformed input space is defined, in pure mathematical terms, by the matrix multiplication of the original input space with the transposed transformation matrix as shown in Eq. (2).

$$\boldsymbol{X}_{\boldsymbol{new}} = \boldsymbol{X} \times \boldsymbol{T}^{\boldsymbol{t}} \tag{2}$$

In case of imbalanced class-distribution for a real-world dataset, the majority samples outnumber the minority population, generating a decision bias towards the majority class. The role of metric learning in reducing the class-imbalance was explored in the works of Feng et al. (2018) [15], Wang et al. (2018) [16], Wang et al. (2021) [17] and Susan and Kumar (2019) [10]. The idea was to transform the data space and concentrate the class populations to confined spaces leading to better classification results. These methods, notably, marked a departure from the conventional solution of resampling that relied on depletion of majority samples and/or replication of minority samples for balancing the imbalanced dataset. Sampling was integrated with ML in a hybrid scheme in [21], where the distance metric was learnt from pruned datasets.

The computational complexity of metric learning constrains its use for large multi-class datasets. The real issue is the computation of the second penalizing term in the cost function that computes distance of a sample with all other samples of other classes. Hence, most of the imbalance treatment solutions mentioned above are for binary classes only with even multi-class datasets being recategorized to binary classes, as illustrated in Fig. 1. The limitations of the existing applications of ML to correct class-imbalance are summarized as follows.

1. The existing ML applications for class-imbalance treatment are limited, based on the size of the dataset and the number of classes (chosen to be two for most applications).
2. The reason is the large number of computations required to calculate inter-class distances and intra-class distances between every pair of training samples, for computing the cost function in metric learning.
3. Even if multi-class computations are somehow managed by high performance computing machines, the resulting transformation matrix may not be altogether accurate in the presence of so many classes due to spatial limitations in adjusting so many classes.
4. The transformation matrix $\boldsymbol{T}$ being a $d \times d$ matrix, the efficiency of transformation for the multi-class scenario is limited by the size of d.
5. Metric learning does not take into account the non-linear relationships between samples and class labels, instances of which are present in most of the real-world datasets.

Due to the above listed issues, ML has been rarely applied to correct class-imbalance in large multi-class multimedia databases. Some of the few works that did address this problem did so under computational constraints, and in no case, to the best of our knowledge, did the number of classes exceed 1000. In [22], the majority class was undersampled to reduce the population, followed by ML. In [23], metric learning was used to improve road traffic sign detection from a LIDAR equipped vehicle, with 112 sign categories affected by the class-imbalance problem; 20 to 80 samples were selected from each category to form a prototype set to ease the ML computations. Deep metric learning was used in [24], in an imbalanced class scenario, to augment the results of uncorrelated cost-sensitive multi-set learning (UCML) that identifies discriminative features for classification. Here, balanced subsets of *Majority class: Minority class* were created using Generative Adversarial Network (GAN), and the distance metric was learnt within each subset using an overall loss function incorporating the individual loss functions of all subsets. The datasets were converted to two-class datasets for the experiments.

In our work, we seek to address the above limitations, for an extremely imbalanced large multi-class publicly available dataset LFW containing 1680 classes of celebrity facial images, for which conventional ML is rendered impractical. Further details of our proposed technique are presented in section 3.

3 Localized metric learning – the proposed approach

Learning local metrics is considered computationally economical and efficient for large multi-class datasets rather than computing a global metric. This is the hypothesis on which we construct our idea. The concept was noted in [31] where a local similarity metric was computed for Locally Embedded Clustering (LEC), based on kd-tree partitions specific to certain localities. The objective function of SERAPH metric learning algorithm was decentralized in [32] to learn local metrics from a partitioned geographical region comprising of a finite number of nodes. With an aim to achieve parallel computation, the data in [33] is partitioned into medium-sized subsets, and results of local discriminative metric learning are aggregated into a global solution.

In our paper, we propose a localized metric learning scheme for the classification of facial images from large multi-class extremely imbalanced face databases containing more than thousand classes with numerous minority classes. Our technique, that is devoid of any sampling procedure, relies entirely on locally learnt transformation matrices. The local group formation, here, is based on the relative class sizes. LMNN is applied separately to each of the subsets, and the distance metric that is learnt from each local subspace is used, by turns, to transform the entire training and test space. For each test sample, the closest training sample in the transformed data space is determined in each case. A comparison among all transformations indicates the closest neighbor in the training space, the label of this training sample is assigned as the class label of the test sample. In summary, the work done in our paper is distinct in the following ways.

1. Our method partitions the training space into discrete subspaces based on relative class sizes and performs metric learning locally within each subspace. The task at hand is the classification of severely imbalanced large datasets having numerous classes exceeding 1000. We include all the minority classes of LFW face dataset that have at least two samples, of which one sample is applied for training and the other for testing, in a 50:50 split. To the best of our knowledge, very few works on LFW have included all the 1680 classes.

2. The locally learnt transformation matrices (7 in number for LFW) are used, in turns, to transform the entire input space including both training and test spaces.

3. The nearest (training) neighbor of the tests sample in the seven-times transformed training space provides the class label of the test sample. Therefore, a seven-fold chance is provided for each test sample to find its closest match in the training space as opposed to a single chance in the case of a global metric, the computation of which, for such a large database, is practically infeasible.

4. Our method boosts the classification accuracies, especially of the minority classes that have very few samples in the training space.

The steps are explained in greater detail in further sub-sections.

3.1 Division of dataset into subsets

The first task is the division of the dataset into smaller subsets for the purpose of localized metric learning in which local transformation matrices are learnt from each subset. Fig. 2 shows the division of the dataset into seven groups or subsets for the LFW dataset. The classes are divided such that the maximum imbalance ratio (IR) within a group is limited to approximately 2:1. This ensures that the metric learnt from a subset is not heavily biased towards one class, as is the situation with the LFW dataset, when considered as a whole. As observed from Fig. 2, *subset 1* is a majority class subset and *subset 7* contains the minority classes with least number of samples (=2 samples in each class). The distance metric is now learnt locally within each subset using the principles of LMNN explained in Eq. (1) and Eq. (2).

subset 1	***subset 2***	***subset 3***	***subset 4***	***subset 5***	***subset 6***	***subset 7***
Class size: [530, 236] IR=2.24:1	Class size: [144, 71] IR=2.02:1	Class size: [60, 30] IR=2:1	Class size: [29, 14] IR=2.07:1	Class size: [13, 6] IR=2.16:1	Class size: [5, 3] IR=1.67:1	Class size: 2 IR=1:1

Fig. 2. Division of LFW dataset into local subsets named as *subset 1, subset 2,…..,subset 7*

3.2 Localized metric learning

Within each training subset, LMNN distance metric learning, whose details were discussed in section 2, is locally applied. The overall process flow is shown illustrated in Fig. 3.

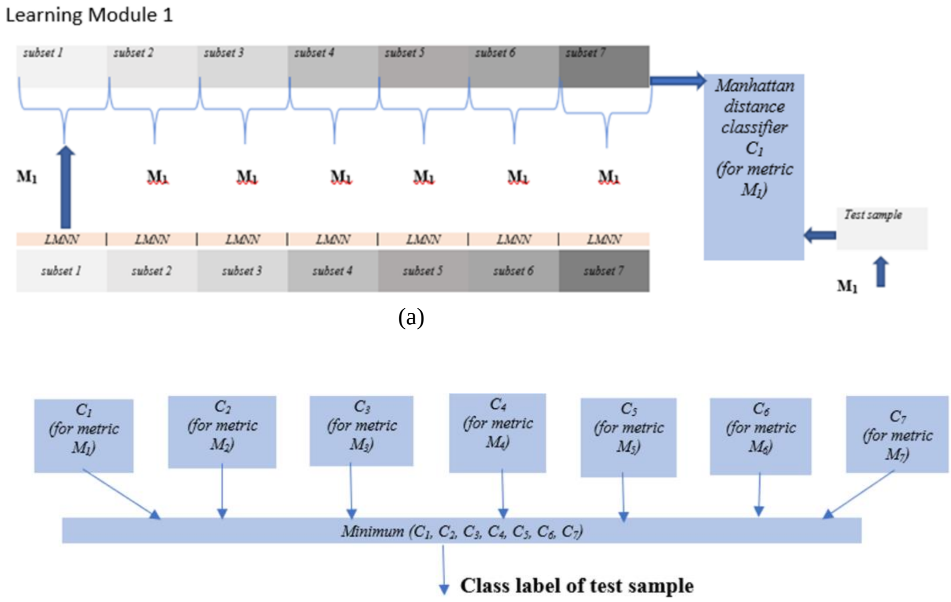


Fig. 3. Proposed scheme for localized metric learning (a) Local LMNN application and data space transformation using $\boldsymbol{M_1}$ (b) Determining the closest nearest neighbor of the test sample in the training space using the *Minimum* (·) function over all possible transformations $\boldsymbol{M_1}$, $\boldsymbol{M_2}$,…, $\boldsymbol{M_7}$

The Mahalanobis distance metric locally learnt from each subset is applied to transform the complete training and testing data, as shown in Fig. 3 (a). The figure shows Learning Module 1 that learns metric M_1 from *subset 1*, and then uses it to transform the entire input space. The other Learning Modules numbered from 2 to 7 learn similar metrics from subsets 2 to 7, respectively. The process of determining the closest neighbor of the test sample in the training space is illustrated in Fig. 3 (b). The minimum Manhattan distance of the test sample from its nearest neighbor in the training space, under all seven transformations, indicates the class of the test sample. The computations are detailed below.

The metric M_1 learnt from *subset 1* is used to transform the entire training space and also the test sample. The transformation metric in (2) is set equal to M_1 for *subset 1*.

$$\boldsymbol{T} = \boldsymbol{M}_1 \forall \boldsymbol{subset1} \tag{3}$$

The Manhattan or city-block distance of the transformed test sample from the training data transformed by M_1 is computed to determine the nearest neighbor in the transformed training space. The city-block distance between two input vectors is the sum of the absolute differences between feature values. The class of the test sample and the corresponding minimum distance are given by equations (4) and (5) respectively.

$$\boldsymbol{Class}(\boldsymbol{test}^{(\boldsymbol{M}_1)}) = \arg\min_{\forall \boldsymbol{l}} \sum_{\boldsymbol{j}} \left| \boldsymbol{X}_{\boldsymbol{new}}{}^{(\boldsymbol{j})} \left(\boldsymbol{subset1}^{(\boldsymbol{l})} \right) - \boldsymbol{X}_{\boldsymbol{new}}{}^{(\boldsymbol{j})} \left(\boldsymbol{test} \right) \right| \tag{4}$$

$$\boldsymbol{dist}(\boldsymbol{test}^{(\boldsymbol{M}_1)}) = \min_{\forall \boldsymbol{l}} \sum_{\boldsymbol{j}} \left| \boldsymbol{X}_{\boldsymbol{new}}{}^{(\boldsymbol{j})} \left(\boldsymbol{subset1}^{(\boldsymbol{l})} \right) - \boldsymbol{X}_{\boldsymbol{new}}{}^{(\boldsymbol{j})} \left(\boldsymbol{test} \right) \right| \tag{5}$$

where, *subset 1* has *l* training samples and each training sample has *j* features.

Similarly, the transformation matrices $M_2, \ldots, M_7$ are derived locally from the training subsets: *subset 2,…subset 7*, respectively, and are used to transform the entire input space, one by one. The class of the test sample is determined by finding the minimum Manhattan distance of the test sample from all the training samples in the seven-times transformed training space. This is summarized in equations (6) and (7).

$$\boldsymbol{c} = \arg\min_{\forall \boldsymbol{i}} \boldsymbol{dist}(\boldsymbol{test}^{(\boldsymbol{M}_i)}) \tag{6}$$

$$\boldsymbol{Class}(\boldsymbol{test}) = \boldsymbol{Class}(\boldsymbol{test}^{(\boldsymbol{M}_c)}) \tag{7}$$

3.3 Algorithm

The universal features considered are the Histogram of Gradient (HOG) features [26] since we wish to have a fair comparison among the classification methodologies without altering the feature space. It is understood that HOG features can be easily substituted by any of the other popular image features for the application. The algorithm for the application of localized metric learning on an extremely imbalanced face database is given below.

Algorithm Localized metric learning for extremely imbalanced face database

Input: *Training set*, *Test* sample
Output: Class label of *Test* sample
1: Find the 1784-dimensional HOG features for each facial image
2: Divide *Training set* into seven subsets based on similar class populations
3: FOR each subset *i*, DO
4: Learn the LMNN distance metric $\boldsymbol{M_i}$
5: Transform the entire input space using $\boldsymbol{M_i}$ as per Eq (2)
6: Find the minimum distance of the transformed *Test* sample from the transformed training space to determine its nearest neighbor
7: REPEAT for *i*=1, 2, …., 7
8: Find the nearest neighbor of the *Test* sample in the seven-times transformed training space
9: Assign the class of the nearest neighbor to the *Test* sample as per Eq. (7)

4 Results and discussions

The experiments were performed on an Intel dual core processor with graphics card, clocked at 2.8 GHz. The software platform is Python 3.7. The proposed approach was implemented as per the procedure outlined in section 3. The dataset for our experiments is the benchmark Labeled Faces in the Wild (LFW) dataset comprising of cropped facial images of celebrities [25]. Only classes with at least two samples were considered since at least one sample from a class is required for training purpose. The parameter *k* for LMNN was set to 3 for the subsets 1 to 5, while it was set to 2 for subset 6, and 1 for subset 7 that contained only 2 samples per class. For developing a challenging classification problem to test the robustness of our algorithm, 50% of the images in a celebrity class were assigned for training and the remaining 50% images for testing. Two-fold cross-validation (V, CV) was performed by swapping the train and test sets. Most of the current works on LFW consider subsets of LFW that contain majority classes only for efficient classification and for presenting decent scores. Very few works include the 2-sample classes (minority classes) for the experimentation, despite of their dominant presence, since they are mostly misclassified and cause the AUC and F1-scores to drop. A total of 1680 classes of celebrities were shortlisted contributing to 4857 samples in training and 4307 samples in testing. This dataset has severe class-imbalance issues, with the class population ranging from 530 (George Bush) to 2 (Michel Duclos), leading to an imbalance ratio as high as 265:1. The number of classes with the minimum number of samples (2 Nos.) is 779 out of 1680 while the number of classes having a single digit class population is 1522 out of 1680, underlining the severity of the class imbalance problem. The unevenness in class-distribution can be observed in the population statistics of LFW compiled in Fig. 4. In this scenario, the classification accuracy of a large-population class would be high and that of a low-population class would be poor due to a decision bias that favors the largely populated class with more than sufficient classes. The problem is aggravated by the presence of noise and outliers that are common characteristics of all classes in real-world datasets.

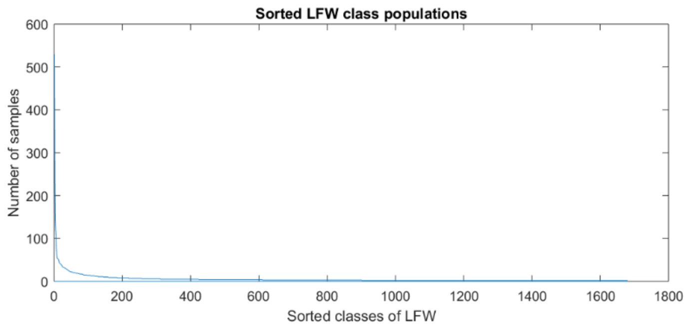


Fig. 4. LFW Class distribution sorted in decreasing order of class population ranging from 530 to 2

Metric learning transforms the data space within each local subset for efficient class separation. A demonstration of the same is shown for *subset 1* in Fig. 5 where the two classes of George Bush (label 533) versus Colin Powell (label 310) are shown for Feature 1=1762 column of HOG feature and Feature 2=1763 column of HOG feature.

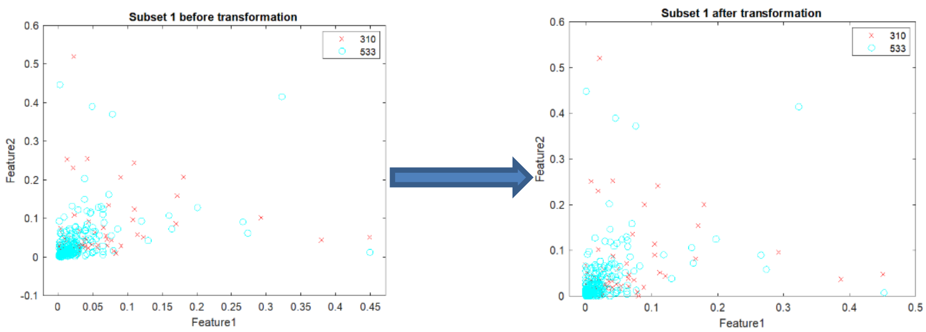


Fig. 5. Effect of local transformation on subset 1 comprising of class labels 533 (George Bush) and 310 (Colin Powell)

We compare our results with that of similar works that have tackled classification of large image datasets: HOG features with Cosine similarity measure by Chen et al. [28], HOG features with the Mahalanobis similarity metric by Fetaya and Ullman [29], HOG features with the Euclidean distance by Bhele and Mankar [30], HOG features with metric learning with majority classes [35]. The universal features considered, in all cases, are the Histogram of Oriented Gradients (HOG) features [26] since we wish to have a judicious comparison among the classification methodologies without altering the feature space. It is understood that HOG features can be easily substituted by any of the other popular image features for the application. Table 1 shows the results of our experiments in terms of AUC scores, accuracy (%) and F1-score (macro). AUC is the Area Under Curve score derived from the Receiver Operating Characteristic (ROC) curve. As observed, the proposed method outperforms the existing methods by a considerable margin. As compared to [35] which is metric learning with majority classes, the Validation (V) results of the proposed localized metric learning involving all 1680 classes are found significantly improved with regard to all three performance metrics. The Cross-Validation (CV) results were found comparable for the AUC and F1-score metrics, while the accuracy was found improved when the Manhattan distance is used as the classifier, instead of cosine similarity, as observed from Table 1.

Table 1. Results of LFW face recognition using HOG features

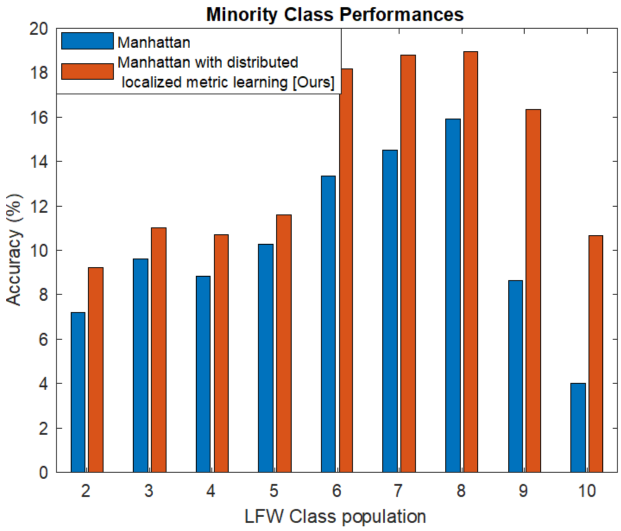
Method	AUC		F1-score		Accuracy	
	V	*CV*	*V*	*CV*	*V*	*CV*
Cosine similarity [28]	0.544	0.541	0.078	0.076	21.5%	19.1%
Manhattan [29]	0.55	0.544	0.087	0.082	22.96%	21.1%
Euclidean [30]	0.544	0.541	0.078	0.076	21.5%	19.1%
Metric learning with majority classes [35]	0.556	0.554	0.1006	0.097	26.8%	24.6%
Ours (Cosine similarity)	0.558	0.551	0.1023	0.095	26.07%	23.65%
Ours (Manhattan distance)	0.5625	0.551	0.1124	0.0962	27.72%	24.87%

It is observed from the previous works that learning the transformation matrices in metric learning consumes a lot of time (less than an hour in [35] for sparsely sampled 186 classes of LFW). The comparison of computation times for the local transformation matrices is shown in Table 2, along with class profiles. Since the computations can be conducted parallelly, in a multi-processor system with parallel computation, this would translate to a computation time of 1548.1 secs that is the time taken by the sixth subset.

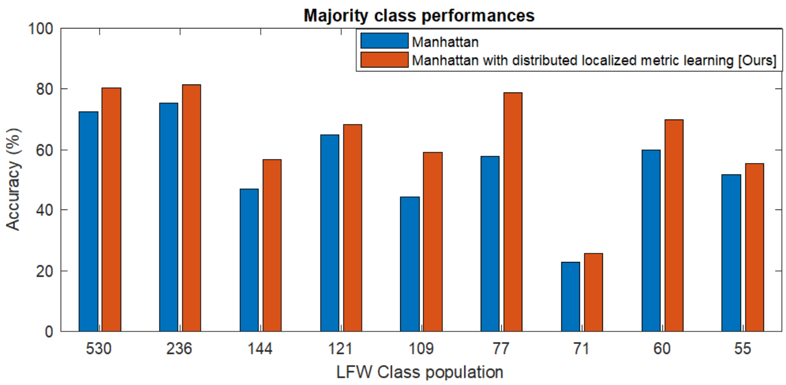
Table 2. Time taken in secs to compute the local transformation matrices for the seven subsets and the details of classes within each subset

Subset	M_1	M_2	M_3	M_4	M_5	M_6	M_7
Time (secs)	55.96	67.75	154.94	276.05	692.07	1548.1	151.34
Number of classes	2	5	27	72	205	589	779
Class size	[530, 236]	[144, 71]	[60, 30]	[29, 14]	[13, 6]	[5, 3]	[2]

Fig. 6 shows the improvement in performance using the proposed localized metric learning scheme. The results are shown for some of the minority classes with very low-class populations (ranging from 2 to 10) and some of the majority classes with high populations (ranging from 530 to 55) in Fig. 6 (a) and Fig. 6 (b), respectively. Our method thus improves classification scores for both the majority classes and the minority classes, especially for the minority classes as evident from the significant improvement in classification accuracies in Fig. 6 (a). Parallelization of the computations using a multi-GPU system is the next stage of our work.



(a)



(b)

Fig. 6. (a, b) Minority and majority class performances using the Manhattan distance classifier, with and without the proposed metric learning scheme

Overall, in our work, metric learning is introduced as an efficient solution for learning from large multi-class imbalanced datasets. This is achieved by localizing the metric learning to training subspaces having similar class populations. Our approach, notably, improved the classification performance of minority classes (Fig. 6 (b)) that are usually discarded by the current deep learning techniques [34] due to inadequate training examples in the LFW minority classes to train deep neural networks.

5 Conclusions

A novel localized metric learning is proposed in our work for learning from an extremely imbalanced face database. The aim is to learn transformation matrices separately for different class subsets having discrete groups of class populations. The data space transformation metric learnt locally from each training subset is used to transform the entire input space. The minimum Manhattan distance of the test sample from the training samples in the transformed space indicates the closest neighbor in the training space whose label is assigned to the test sample. Our localized metric learning methodology is applied successfully on a large face database LFW, with more than 1000 classes, containing a large quantity of minority classes. The results indicate a significant performance improvement for the minority classes despite of the dominating presence of the majority classes.

References

1. Krawczyk, Bartosz. "Learning from imbalanced data: open challenges and future directions." *Progress in Artificial Intelligence* 5, no. 4 (2016): 221-232.
2. Susan, Seba, and Amitesh Kumar. "SSOMaj-SMOTE-SSOMin: Three-step intelligent pruning of majority and minority samples for learning from imbalanced datasets." *Applied Soft Computing* 78 (2019): 141-149.
3. Ling, Charles X., and Victor S. Sheng. "Cost-sensitive learning and the class imbalance problem." *Encyclopedia of machine learning* 2011 (2008): 231-235.
4. Susan, Seba, and Amitesh Kumar. "The balancing trick: Optimized sampling of imbalanced datasets—A brief survey of the recent State of the Art." *Engineering Reports* 3, no. 4 (2021): e12298.
5. Mienye, Ibomoiye Domor, and Yanxia Sun. "Performance analysis of cost-sensitive learning methods with application to imbalanced medical data." *Informatics in Medicine Unlocked* 25 (2021): 100690.
6. Piras, Luca, and Giorgio Giacinto. "Synthetic pattern generation for imbalanced learning in image retrieval." *Pattern Recognition Letters* 33, no. 16 (2012): 2198-2205.
7. Saini, Manisha, and Seba Susan. "Data augmentation of minority class with transfer learning for classification of imbalanced breast cancer dataset using inception-V3." In *Iberian Conference on Pattern Recognition and Image Analysis*, pp. 409-420. Springer, Cham, 2019.
8. Rezaei, Mina, Tomoki Uemura, Janne Näppi, Hiroyuki Yoshida, Christoph Lippert, and Christoph Meinel. "Generative synthetic adversarial network for internal bias correction and handling class imbalance problem in medical image diagnosis." In *Medical Imaging 2020: Computer-Aided Diagnosis*, vol. 11314, p. 113140E. International Society for Optics and Photonics, 2020.

9. Rezaei, Mina, Haojin Yang, and Christoph Meinel. "Recurrent generative adversarial network for learning imbalanced medical image semantic segmentation." *Multimedia Tools and Applications* 79, no. 21 (2020): 15329-15348.
10. Susan, Seba, and Amitesh Kumar. "DST-ML-EkNN: data space transformation with metric learning and elite k-nearest neighbor cluster formation for classification of imbalanced datasets." In *Advances in Artificial Intelligence and Data Engineering*, pp. 319-328. Springer, Singapore, 2021.
11. Barua, Sukarna, Md Monirul Islam, Xin Yao, and Kazuyuki Murase. "MWMOTE--majority weighted minority oversampling technique for imbalanced data set learning." *IEEE Transactions on Knowledge and Data Engineering* 26, no. 2 (2012): 405-425.
12. Tao, Xinmin, Qing Li, Chao Ren, Wenjie Guo, Chenxi Li, Qing He, Rui Liu, and Junrong Zou. "Real-value negative selection over-sampling for imbalanced data set learning." *Expert Systems with Applications* 129 (2019): 118-134.
13. Liu, Tianyu, Xiubin Zhu, Witold Pedrycz, and Zhiwu Li. "A design of information granule-based under-sampling method in imbalanced data classification." *Soft Computing* 24, no. 22 (2020): 17333-17347.
14. Moutafis, Panagiotis, Mengjun Leng, and Ioannis A. Kakadiaris. "An overview and empirical comparison of distance metric learning methods." *IEEE transactions on cybernetics* 47, no. 3 (2016): 612-625.
15. Feng, Lin, Huibing Wang, Bo Jin, Haohao Li, Mingliang Xue, and Le Wang. "Learning a distance metric by balancing kl-divergence for imbalanced datasets." *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 49, no. 12 (2018): 2384-2395.
16. Wang, Nan, Xibin Zhao, Yu Jiang, and Yue Gao. "Iterative metric learning for imbalance data classification." In *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, pp. 2805-2811. 2018.
17. Wang, Cunjun, Cun Xin, and Zili Xu. "A novel deep metric learning model for imbalanced fault diagnosis and toward open-set classification." *Knowledge-Based Systems* 220 (2021): 106925.
18. Kulis, Brian. "Metric learning: A survey." *Foundations and trends in machine learning* 5, no. 4 (2012): 287-364.
19. Weinberger, Kilian Q., John Blitzer, and Lawrence K. Saul. "Distance metric learning for large margin nearest neighbor classification." In *Advances in neural information processing systems*, pp. 1473-1480. 2006.
20. Weinberger, Kilian Q., and Lawrence K. Saul. "Distance metric learning for large margin nearest neighbor classification." *Journal of Machine Learning Research* 10, no. 2 (2009).
21. Susan, Seba, and Amitesh Kumar. "Learning data space transformation matrix from pruned imbalanced datasets for nearest neighbor classification." In *2019 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. IEEE, 2019.
22. Ghanavati, Mojgan, Raymond K. Wong, Fang Chen, Yang Wang, and Chang-Shing Perng. "An effective integrated method for learning big imbalanced data." In *2014 IEEE International Congress on Big Data*, pp. 691-698. IEEE, 2014.
23. Tan, Min, Baoyuan Wang, Zhaohui Wu, Jingdong Wang, and Gang Pan. "Weakly supervised metric learning for traffic sign recognition in a LIDAR-equipped vehicle." *IEEE Transactions on Intelligent Transportation Systems* 17, no. 5 (2016): 1415-1427.
24. Jing, Xiao-Yuan, Xinyu Zhang, Xiaoke Zhu, Fei Wu, Xinge You, Yang Gao, Shiguang Shan, and Jing-Yu Yang. "Multiset feature learning for highly imbalanced data

classification." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 43, no. 1 (2019): 139-156.
25. Huang, Gary B., Marwan Mattar, Tamara Berg, and Eric Learned-Miller. "Labeled faces in the wild: A database for studying face recognition in unconstrained environments." 2008.
26. Dalal, Navneet, and Bill Triggs. "Histograms of oriented gradients for human detection." In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, vol. 1, pp. 886-893. IEEE, 2005.
27. Dadi, Harihara Santosh, and GK Mohan Pillutla. "Improved face recognition rate using HOG features and SVM classifier." *IOSR Journal of Electronics and Communication Engineering* 11, no. 04 (2016): 34-44.
28. Chen, Dong, Xudong Cao, Fang Wen, and Jian Sun. "Blessing of dimensionality: High-dimensional feature and its efficient compression for face verification." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3025-3032. 2013.
29. Abuzneid, Mohannad A., and Ausif Mahmood. "Enhanced human face recognition using LBPH descriptor, multi-KNN, and back-propagation neural network." *IEEE access* 6 (2018): 20641-20651.
30. Bhele, Sujata G., and Vijay H. Mankar. "Recognition of faces using discriminative features of LBP and HOG descriptor in varying environment." In *2015 International conference on computational intelligence and communication networks (CICN)*, pp. 426-432. IEEE, 2015.
31. Fu, Yun, Zhu Li, Thomas S. Huang, and Aggelos K. Katsaggelos. "Locally adaptive subspace and similarity metric learning for visual data clustering and retrieval." *Computer Vision and Image Understanding* 110, no. 3 (2008): 390-402.
32. Shen, Pengcheng, Xin Du, and Chunguang Li. "Distributed semi-supervised metric learning." *IEEE Access* 4 (2016): 8558-8571.
33. Li, Jun, Xun Lin, Xiaoguang Rui, Yong Rui, and Dacheng Tao. "A distributed approach toward discriminative distance metric learning." *IEEE transactions on neural networks and learning systems* 26, no. 9 (2014): 2111-2122.
34. Taigman, Yaniv, Ming Yang, Marc'Aurelio Ranzato, and Lior Wolf. "Deepface: Closing the gap to human-level performance in face verification." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1701-1708. 2014.
35. Susan, Seba, and Ashu Kaushik. "Weakly supervised metric learning with majority classes for large imbalanced image dataset." In *Proceedings of the 2020 the 4th International Conference on Big Data and Internet of Things*, pp. 16-19. 2020.