INCREMENTAL ZERO-SHOT LEARNING BASED ON ATTRIBUTES FOR IMAGE CLASSIFICATION

Nan Xue[1] *Yi Wang*[1,2] *Xin Fan*[*1,2] *Maomao Min*[1]

[1] School of Software, Dalian University of Technology, Dalian, China
[2] Key Laboratory for Ubiquitous Network
and Service Software of Liaoning Province, Dalian, China
xin.fan@dlut.edu.cn

ABSTRACT

Instead of assuming a closed-world environment comprising a fixed number of objects, modern pattern recognition systems need to recognize outliers, identify anomalies, or discover entirely new objects, which is known as zero-shot object recognition. However, many existing zero-shot learning methods are not efficient enough to incrementally update themselves with new samples mixed with known or novel class labels. In this paper, we propose an incremental zero-shot learning framework (IIAP/QR) based on indirect-attribute-prediction (IAP) model. Firstly, a fast incremental classifier based on null space based linear discriminant analysis with QR-updating (NLDA/QR) is put forward, which can solve small-sample-size (SSS) problem and unequal-sample-size (USS) problem that usually occur in incremental learning using the centroid of each class as input. Then with the probabilistic inference of Class-Attribute layer and Attribute-Zero shot classification layer, IIAP/QR model can efficiently update itself for the insertion of both new samples to the existing class and totally novel classes with comparable recognition accuracy for zero-shot object recognition.

Index Terms— zero-shot, incremental learning, NLDA/QR, image classification

1. INTRODUCTION

For zero-shot identification [1, 2], human can learn completely unseen objects without any training samples from a list of high-level semantically meaningful properties which are called “attributes” [3, 4]. Lampert *et al.* [5, 6] proposed the zero-shot classifiers (direct-attribute-prediction (DAP) and IAP using attributes as an intermediate layer to recognize object classes which had no a single training example. And both DAP and IAP are based on multi-class classification and probabilistic inference. Several existing methods [7, 8, 9, 10] improved performance of zero-shot recognition by finding the better mapping between classes. But these methods are not incremental learners that can adapt themselves to the change of environment.

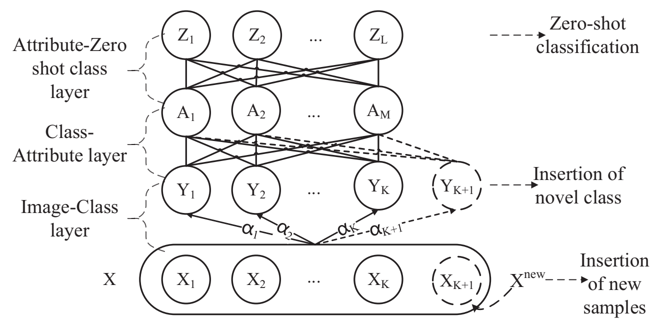


Fig. 1: IIAP/QR learning model. Referring to IAP, for each training class Y and test class Z with $Y \cap Z = \emptyset$, we employ our incremental algorithm in Image-Class layer to estimate $p(Y|x)$. Attribute representation A^Y and $A^Z \in A$ are available beforehand, then we can transfer information between Y and Z through attribute layer A.

In incremental learning and changed scenarios, the changes may lie in the following three situations: (a) Attributes updating. A certain of attributes may be added incrementally. (b) Samples updating. More samples for training is available. (c) Objects categories updating. Some novel objects categories (having training samples) are added by human in an interactive environment. As to the first situation, self-organizing and incremental neural networks (SOINN) [11] can learn new attributes and update existing attributes in an online incremental manner while retaining the accuracy. Yet in this paper, we mainly focus on the latter two cases. As we all known, multi-classifier plays an important role in the incremental learning, especially for quickly updating need, which inevitably exists SSS and USS problems in zero-shot classes and objects categories updating. In original IAP [6], kernel non-linear support vector machine (SVM) is used as classification mechanism. But the regression and SVM are not suitable for these circumstances, for SVM suffers from training a mass of support vectors and the error accumulation [12, 13], especially when considering incremental learning in a dynamic environment.

In this paper, we propose an incremental attributed-based zero-shot learning framework (IIAP/QR) as shown in Fig. 1.

Algorithm 1 NLDA/QR

1. **Input**: data $X \in \mathbf{R}^{d\times n}$, the number of all samples n, number of each class n_i, and cluster number K
2. **Output**: the optimal projection matrix G
3. **The first QR decomposition**
4. Get $X = U_1 R$ through economic QR factorization of X, where $U_1 \in \mathbf{R}^{d\times n}$, $R \in \mathbf{R}^{n\times n}$
5. **The second QR decomposition**
6. Set P a permutation matrix which is transformed by exchanging i-th column and $\sum_{j=1}^{i-1} n_j + 1$-th column of the $n\times n$ identity matrix, $i = 2,...,K$.
7. Denote $e_i = [1\cdots 1]^T \in \mathbf{R}^{n_i\times 1}$, W and W_i is the Householder Transformation of vector $\left[\sqrt{n_1}\sqrt{n_2}\cdots\sqrt{n_K}\right]^T$ and $e_i,\ i = 1,...,K$
8. Set $\begin{bmatrix} R_1 & R_2 & R_3 \end{bmatrix} = R\begin{bmatrix} W_1 & & \\ & \ddots & \\ & & W_K \end{bmatrix} P\begin{bmatrix} W & \\ & I \end{bmatrix}$, where $R_1 \in \mathbf{R}^{n\times 1}$, $R_2 \in \mathbf{R}^{n\times(K-1)}$, $R_3 \in \mathbf{R}^{n\times(n-K)}$
9. *Set* $q = rank(R_3)$, $\gamma = rank\begin{bmatrix} R_2 & R_3 \end{bmatrix}$, $\begin{bmatrix} R_3 & R_2 \end{bmatrix} = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix}\begin{bmatrix} R_{12} & R_{13} \\ 0 & R_{23} \end{bmatrix}$, where $Q_1 \in \mathbf{R}^{n\times q}, Q_2 \in \mathbf{R}^{n\times(\gamma-q)}, R_{12} \in \mathbf{R}^{q\times(n-K)}, R_{23} \in \mathbf{R}^{(\gamma-q)\times(K-1)}$
10. $G = U_1 Q_2$
11. **return** G

We firstly put forward a fast incremental classifier based on null space based linear discriminant analysis(NLDA/QR) [14] with QR-updating for Image-Class Layer. Particularly, in order to reduce computational complexity and solve the problems of SSS and USS of classes in incremental learning, we take the centroid of each class as input in IIAP/QR. Then in Class-Attribute layer and Attribute-Zero shot classification layer, IIAP/QR model can update itself for the insertion of both new samples and novel classes with comparable recognition accuracy and computation performance by probabilistic inference.

2. INCREMENTAL ZERO-SHOT LEARNING

2.1. Null Space Based Classification

Null space-based linear discriminant analysis (NLDA) has great potentials in pattern recognition as a powerful feature extraction tool due to its ability tackling SSS problem [15]. The main task of NLDA is to find the optimal projection matrix $G^T : x \in R^{d\times 1} \to y = G^T x \in R^{l\times 1}$, where $l \ll d$, d is the dimensionality of data, so that the distance between-class is the largest, and the within-class spacing is the smallest after projection. NLDA/QR [14] employs Al.1 to obtain the optimal projection matrix G.

We propose an incremental implementation of Al. 1 as the classifier of Image-Class layer, which takes centroid matrix $C = [c_1,...,c_K] \in \mathbf{R}^{d\times K}$ (K is the cluster number) instead of all data X as input. It is an innovation of improving computational and recognition performance. In next section, we introduce the two updating cases of IIAP/QR in details.

2.2. Updating for Novel Classes

As shown in Fig. 1, the number of training classes Y increases in this case. Due to the centroid matrix as input of IIAP/QR model, we can insert a few instances of some novel classes at one time. We update centroid matrix $\widetilde{C} = [C, C_h]$, where h is the number of novel classes, and C_h is the set of center of h novel classes. We will update Image-Class layer for getting the new optimal projection matrix $\widetilde{G}$ using QR-updating. As follows:

$$\widetilde{C} \equiv (C, C_h) = (U_1 R, C_h) = (U_1, C_h)\begin{pmatrix} R & 0 \\ 0^T & 1 \end{pmatrix} \tag{1}$$

Here, U_1 and R have been calculated before. Further, applying Gram-Schmidt orthogonalization [16], we can obtain such that

$$(U_1, C_h) = (U_1, q)\begin{pmatrix} I & r \\ 0 & \rho \end{pmatrix},\ {U_1}^T q = 0, \|q\| = 1.$$

Therefore, we can present $\widetilde{C}$ as

$$\widetilde{C} = (U_1, q)\begin{pmatrix} R & r \\ 0^T & \rho \end{pmatrix} \equiv \bar{U}_1 \bar{R} \tag{2}$$

Then, we employ Givens rotation matrix $\left(W_{K,K+1}, \ldots, W_{K+h-1,K+h}\right)$, we can get

$$W\bar{R} \equiv W_{K,K+1}\cdots W_{K+h-1,K+h}\bar{R} \equiv \widetilde{R} \tag{3}$$

$$\bar{U}_1 W^T = \bar{U}_1 W_{K+h-1,K+h}\cdots W_{K,K+1} \equiv \widetilde{U}_1 \tag{4}$$

Thus,

$$\widetilde{C} = \widetilde{U}_1 \widetilde{R} \tag{5}$$

It can be seen that we re-express the first QR factorization of $\widetilde{C}$ instead of recalculating it by utilizing relevant parameters calculated. Subsequently, do the second economic QR factorization of a $(K+h)\times(K+h)$ matrix with column pivoting described by Al.1 to get the optimal projection matrix $\widetilde{G} = \widetilde{U_1}\widetilde{Q}_2$. Afterwards, we apply K-Nearest-Neighbor (KNN) to renew Image-Class layer $\widetilde{p}(Y|x)$. At present, the change occurs at the number of known classes from K to $K+h$. Attribute-Class and Attribute-Zero shot class layer refers to probability inference of zero-shot class of IAP to be renewed. So

$$\widetilde{p}\,(A|x) \quad = \prod_{m=1}^{M}\left(\sum_{k=1}^{K+h} p\left(A_m|Y_k\right)\widetilde{p}\left(Y_k|x\right)\right) \tag{6}$$

$$\widetilde{p}\,(\mathrm{Z}|\mathrm{x}) \quad = \frac{p(Z)}{p\left(A^Z\right)}\prod_{m=1}^{M}\left(\sum_{k=1}^{K+h} p\left(A_m^Z|Y_k\right)\widetilde{p}\left(Y_k|x\right)\right) \tag{7}$$

$$\widetilde{f}\,(x) \quad = \underset{l=1,\ldots,L}{\arg\max}\,\widetilde{p}(Z|x) \tag{8}$$

The whole process is summarized as following Al.2.

Algorithm 2 The insertion of novel classes

1: **Input**: the original centroid matrix C and the centroid matrix of h novel classes C_h
2: **Output**: classification accuracy of zero-shot classes
3: Call Eq.(1) to Eq.(5) to get $\widetilde{C} = \widetilde{U}_1 \widetilde{R}$.
4: Refer to Al.1 to get the optimal projection matrix $\widetilde{G} = \widetilde{U}_1 \widetilde{Q}_2$.
5: Update $\widetilde{p}(Y|x)$
6: Call Eq.(6) to Eq.(8) to renovate probability of zero-shot classes.
7: **return** classification accuracy of zero-shot classes

Algorithm 3 The insertion of new samples to existing classes

1: **Input**: the original centroid matrix C, the sample number n_i of each class, new samples X^i belong to each classes
2: **Output**:classification accuracy of zero-shot classes
3: Call Eq.(9) to renew centroid matrix $\widetilde{C}$
4: Refer to Al.1 to get the optimal projection matrix $\widetilde{G} = \widetilde{U}_1 \widetilde{Q}_2$.
5: Update $\widetilde{p}(Y|x)$.
6: Call Eq.(10) to Eq.(12) to renovate probability of zero-shot classes.
7: **return** classification accuracy of zero-shot classes

2.3. Updating for New Samples to the Existing Classes

Under this circumstance, some new instances $X^{(i)}$ belonging to each known category X_i added into the trained model, we replace the centroid matrix C with $\widetilde{C} = [\widetilde{c}_1, ..., \widetilde{c}_K]$, where

$$\widetilde{c}_i = (n_i c_i + X^{(i)}) \Big/ \widetilde{n}_i, \tag{9}$$

$\widetilde{n}_i$ is the new number of category X_i after inserting new samples. Then we employ Al.1 to get $\widetilde{G} = \widetilde{U}_1 \widetilde{Q}_2$. Hereto, we renew the responding parameters $\widetilde{p}(Y|x)$ in Image-Class layer.

In the second stage, we need to re-express the probability associated with Attribute-Class and Attribute-Zero shot class layer in IIAP/QR model. But obviously the overall structure of them is not changed. So we can obtain the probability of each test sample x belonging to each test class Z according to the following representation.

$$\widetilde{p}\,(A|x) = \prod_{m=1}^{M} p\,(A_m|x) = \prod_{m=1}^{M} \sum_{k=1}^{K} p\,(A_m|Y_k)\,\widetilde{p}\,(Y_k|x) \tag{10}$$

so

$$\widetilde{p}(Z|x) = p(Z|x) \tag{11}$$

$$\widetilde{f}(x) = f(x) \tag{12}$$

The following Al.3 sums up the all implementations of this incremental learning.

3. EXPERIMENTS

In this section, we evaluate the efficiency of proposed incremental zero-shot model in terms of the execution time and classification accuracy for SSS problem. So we set the dimension d of feature is far more than the number n of samples ($d \gg n$). Due to using centroid matrix as input, our method avoids the problem caused by unequal sample size of all classes. Our experiments are conducted on AWA [5], which includes 50 different kinds of animals with eight kinds of features. We select three feature representations (cq, decaf and vgg19) in our experiments. The other dataset is aPascal-aYahoo dataset [17], from which we draw aPascal dataset including 20 classes with 64 attributes. The details of training and test set of two datasets are shown in Tab. 1. In incremental learning, as to original IAP (using SVM as classifier) and NLDA/QR, we make them learn new samples in batch mode. We compare the execution time and classification accuracy for original IAP, NLDA/QR, IDR/QR [18], ILDA/SSS [19] and IIAP/QR (our) in two incremental recognition tasks. KNN(with K=5) is used for classification purpose. All experiments are the average of five times of test, performed in MATLAB 2016a and ran on an Intel (R) Core (TM) *i*7 PC with 3.40 GHz CPU and 8 GB RAM.

3.1. Insertion of novel classes

In this section, for AWA, we divide 40 training classes into two parts, *i.e.*, initial training set S_1 including 30 classes, and 10 classes in incremental set S_2. And we set 10 classes for zero-shot class S_3. About aPascal, the corresponding numbers of each part are 15, 10, 5, 5. The samples number of each class is referred to N_1 and N_2 in Tab. 1. Then we make novel classes arrive one by one as new insertion. The comparison of execution time and accuracy of four algorithms are shown in Fig. 2 and Fig. 4(a).

Compared to original IAP and IDR/QR, the accuracy of IIAP/QR approximately raises by 8%, 10-15%, 30-35%, and 12-15% for feature cq, decaf, vgg19 and aPascal respectively. IIAP/QR improves about 5% or is closed to NLDA/QR. In performance of execution time, IIAP/QR improves 2-3 orders of magnitude in contrast to IAP and NLDA/QR. Although IDR/QR achieves time reduce on incremental learning, IIAP/QR is still faster 4-5 times than it.

3.2. Insertion of new samples to the existing classes

In this case, we take some samples from each class for initial training, seen N_3 in Tab. 1. We make new samples (from the rest of each training class) arrive one by one. Fig. 3 and Fig. 4 (b) shows the results of five algorithms.

The computation time of original IAP, NLDA/QR and ILDA/SSS gradually increases with the increasing of new samples. However, IIAP/QR has been around a lower value in

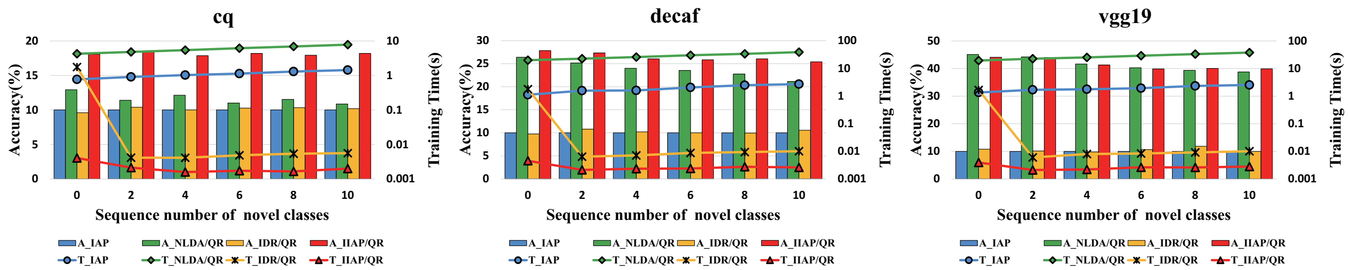


Fig. 2: The execution time and recognition accuracy of inserting novel classes successively. From left to right in each bar of accuracy and line charts of execution time with ◦, ⋄, ∗, and △ respectively represent original IAP, NLDA/QR+IAP, IDR/QR and IIAP/QR (our). And the vertical axis on the left in the graph indicates the recognition accuracy, as well it is execution time on the right.

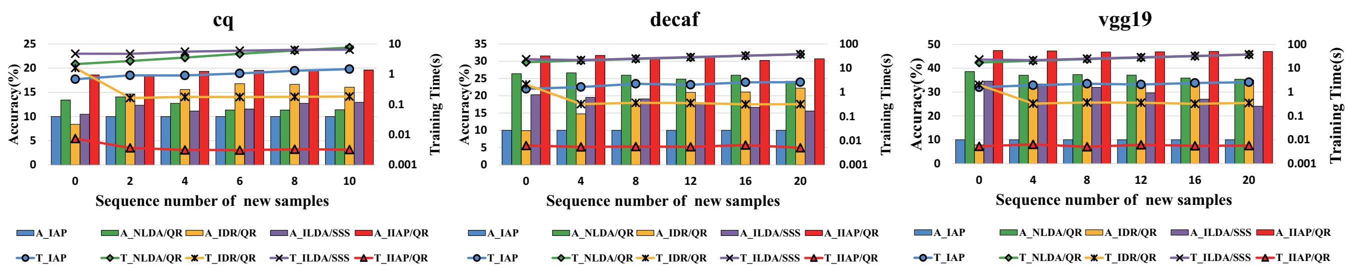
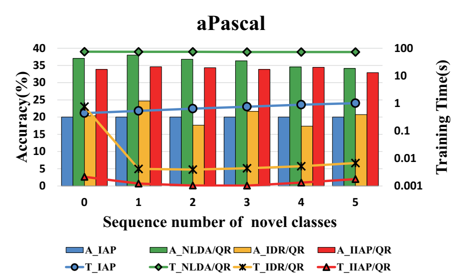


Fig. 3: The training time and recognition accuracy of inserting new samples one by one.

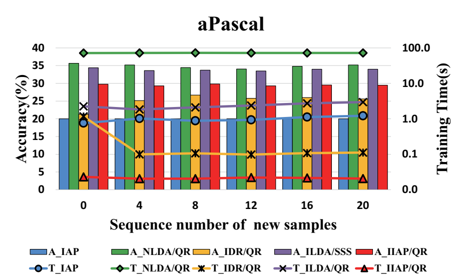
Table 1: The parameters of experiments for two datasets. N_1 is the samples number of each training class, N_2 is the samples number of each novel class as test, N_3 is the sample number of each initial training class, and d is the dimension of each feature.

	cq	decaf	vgg19	aPascal
d	2688	4096	4096	9751
N_1	40	70	70	80
N_2	30	40	40	70
N_3	20	50	50	60

terms of time and avoid unequal sample size problem for stable identification performance, owing to choosing the centroid matrix as input of the model. In the recognition accuracy, IIAP/QR is higher 3-12% than other four algorithms on AWA, or closed to NLDA/QR and ILDA/SSS on aPascal, while it is lower 1-4 orders of magnitude in terms of time.



(a)



(b)

Fig. 4: The trianing time and recognition accuracy of inserting new samples one by one.

4. CONCLUSION

In this work, we have proposed IIAP/QR for fast incremental zero-shot learning based on attributes. Integrating our fast incremental algorithm and zero-shot model IAP, it can get superior performance for insertion of new samples to existing classes or novel classes in terms of recognition accuracy and execution time in zero-shot incremental learning.

5. ACKNOWLEDGMENTS

This research is sponsored in part by the National Natural Science Foundation of China (No.61402072, No.61272371, No. 61572096, and No. 61432003).

6. REFERENCES

[1] Hugo Larochelle, Dumitru Erhan, and Yoshua Bengio, "Zero-data learning of new tasks.," in *AAAI*, 2008, p. 3.

[2] Marcus Rohrbach, Michael Stark, and Bernt Schiele, "Evaluating knowledge transfer and zero-shot learning in a large-scale setting," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE, 2011, pp. 1641–1648.

[3] Xiaodong Yu and Yiannis Aloimonos, "Attribute-based transfer learning for object categorization with zero/one training example," in *European conference on computer vision*. Springer, 2010, pp. 127–140.

[4] Devi Parikh, Adriana Kovashka, Amar Parkash, and Kristen Grauman, "Relative attributes for enhanced human-machine communication.," in *AAAI*, 2012.

[5] Christoph H Lampert, Hannes Nickisch, and Stefan Harmeling, "Learning to detect unseen object classes by between-class attribute transfer," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. IEEE, 2009, pp. 951–958.

[6] Christoph H Lampert, Hannes Nickisch, and Stefan Harmeling, "Attribute-based classification for zero-shot visual object categorization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 3, pp. 453–465, 2014.

[7] Ziming Zhang and Venkatesh Saligrama, "Zero-shot learning via semantic similarity embedding," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 4166–4174.

[8] Ziming Zhang and Venkatesh Saligrama, "Zero-shot learning via joint latent similarity embedding," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 6034–6042.

[9] Maxime Bucher, Stéphane Herbin, and Frédéric Jurie, "Improving semantic embedding consistency by metric learning for zero-shot classiffication," in *European Conference on Computer Vision*. Springer, 2016, pp. 730–746.

[10] Soravit Changpinyo, Wei-Lun Chao, Boqing Gong, and Fei Sha, "Synthesized classifiers for zero-shot learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 5327–5336.

[11] Pichai Kankuekul, Aram Kawewong, Sirinart Tangruamsub, and Osamu Hasegawa, "Online incremental attribute-based zero-shot learning," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 3657–3664.

[12] Xiaoou Li and Wen Yu, "Data stream classification for structural health monitoring via on-line support vector machines," in *Big data computing service and applications (bigdataservice), 2015 IEEE first international conference on*. IEEE, 2015, pp. 400–405.

[13] Youlu Xing, Furao Shen, Chaomin Luo, and Jinxi Zhao, "l^3-svm: A lifelong learning method for svm," in *Neural Networks (IJCNN), 2015 International Joint Conference on*. IEEE, 2015, pp. 1–8.

[14] Delin Chu and Goh Siong Thye, "A new and fast implementation for null space based linear discriminant analysis," *Pattern Recognition*, vol. 43, no. 4, pp. 1373–1379, 2010.

[15] Rui Huang, Qingshan Liu, Hanqing Lu, and Songde Ma, "Solving the small sample size problem of lda," in *Pattern Recognition, 2002. Proceedings. 16th International Conference on*. IEEE, 2002, vol. 3, pp. 29–32.

[16] James W Daniel, Walter Bill Gragg, Linda Kaufman, and GW Stewart, "Reorthogonalization and stable algorithms for updating the gram-schmidt qr factorization," *Mathematics of Computation*, vol. 30, no. 136, pp. 772–795, 1976.

[17] Ali Farhadi, Ian Endres, Derek Hoiem, and David Forsyth, "Describing objects by their attributes," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. IEEE, 2009, pp. 1778–1785.

[18] Jieping Ye, Qi Li, Hui Xiong, Haesun Park, Ravi Janardan, and Vipin Kumar, "Idr/qr: An incremental dimension reduction algorithm via qr decomposition," *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 9, pp. 1208–1222, 2005.

[19] Tae-Kyun Kim, Shu-Fai Wong, Bjorn Stenger, Josef Kittler, and Roberto Cipolla, "Incremental linear discriminant analysis using sufficient spanning set approximations," in *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*. IEEE, 2007, pp. 1–8.