2019 15th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)

Autoencoder Based Dimensionality Reduction of Feature Vectors for Object Recognition

Reyhan Kevser Keser
Signal Processing for Computational Intelligence Group
Informatics Institute
Istanbul Technical University
Istanbul, Turkey
keserr@itu.edu.tr

Behçet Uğur Töreyin
Signal Processing for Computational Intelligence Group
Informatics Institute
Istanbul Technical University
Istanbul, Turkey
toreyin@itu.edu.tr

***Abstract*—Object recognition can be performed with high accuracy thanks to the robust feature descriptors defining the significant areas in images. However, these features suffer from high dimensional structure, in other words "curse of dimensionality" for further processes. Autoencoders (AE) are proposed in this study to solve the dimensionality reduction problem of visual features. To assess the efficacy, object recognition is performed using reduced dimensional visual features. For this purpose, dimensionalities of three well-known feature vectors, namely, HOG, SIFT and SURF, are reduced to half. Moreover, deep learning based features are also reduced. Then, reduced vectors, which are called as AE-HOG, AE-SIFT, AE-SURF and AE-DEEP are fed to object recognition task. Also, dimensionality reduction is implemented by a variant of AE, variational autoencoder (VAE) and PCA, which is the most studied unsupervised method for these features, and the results are compared. Furthermore, all experiments are repeated on noisy images. Results suggest that dimensionality reduction of these feature vectors can be accomplished successfully owing to the proposed method.**

***Index Terms*—dimensionality reduction, autoencoder, HOG, SIFT, SURF**

I. Introduction

Object recognition is a substantial problem in computer vision field and many studies are carried out on this topic. It was shown that feature based methods which extract global or local features from the image and match these, can give successful results on this problem. However, they can suffer from high dimensional structure. Hence, dimensionality reduction of feature vectors is an essential process in order to speed up the computations and reduce memory usage. Keeping the accuracy similar to the one obtained using the original sized features is crucial in this process.

The most popular feature vectors used for object recognition, are Scale Invariant Feature Transform (SIFT) vectors. These vectors are produced by the SIFT algorithm, which is a well-known method to detect interest points and form descriptors from them. This method was proposed by Lowe in 2004 [1]. It provides scale, rotation and translation invariant feature vectors for object detection and recognition. However, the local image descriptors which are produced by SIFT algorithm are 128-D vectors. Hence, many studies have proposed to reduce dimensionality of SIFT vectors using various techniques.

Some studies which aim to reduce computational load, proposed to binarize and quantize SIFT vectors [2]–[6], whereas some studies proposed linear projection methods to reduce dimensionality of SIFT vectors [7]–[9]. Furthermore, in [10], authors used autoencoder [11] to obtain feature vectors from gradient patches, which are extracted by SIFT detector. Hence, they didn't use the SIFT descriptors.

The most popular linear method used for dimensionality reduction of SIFT vectors is Principal Component Analysis (PCA) [12]–[15]. A well-known method that benefits from PCA is PCA-SIFT, which is obtained by applying PCA to normalized gradient patches instead of using smoothed weighted histograms in computation of SIFT descriptors [16]. It was shown that PCA-SIFT gives better results than SIFT in image retrieval with regards to accuracy and time. In [17], it was proposed to use PCA to reduce the dimensionality of SIFT and Speeded-Up Robust Features (SURF) [18]. The reduced vectors were named as Reduced-SIFT and Reduced-SURF. The authors stated that Reduced-SIFT was better than SIFT in image retrieval. However, Reduced-SURF could not achieve the success of SURF in image retrieval.

In [19] it is reported that supervised methods give better results than unsupervised methods. They used image class labels as feature labels for implementation of supervised methods. However, they reported that SIFT features could be found in images which belong to different classes since these features are local descriptors. In other words, a feature could not be related with one object class.

SURF is another popular technique used for feature detection and description for object recognition. It was proposed by Bay et al. and it provides local image descriptors that define significant areas on images [18]. SURF vectors are scale and rotation invariant features which are 64-D vectors.

There are many studies applying PCA on SURF and SURF based feature vectors for dimensionality reduction [14], [17], [20]. Four linear dimensionality reduction methods for SIFT and SURF vectors were studied in [8]. The results showed that Random Projection (RP), Linear Discriminant Analysis (LDA)

This work is supported in part by Istanbul Technical University (ITU) Vodafone Future Lab under Project ITUVF20180901P04 and by ITU BAP MGA-2017-40964.

978-1-7281-5686-6/19/$31.00 ©2019 IEEE
DOI 10.1109/SITIS.2019.00097

and Partial Least Squares (PLS) methods could not achieve original SURF vectors' success, while PCA outperformed original SURF vectors.

In [19], both of the supervised and unsupervised methods were used for dimensionality reduction of SIFT and SURF vectors. It is reported that unsupervised methods give acceptable results. However, supervised methods gave better results than unsupervised methods.

One way of using supervised methods can be constructing codebook vectors from feature vectors by clustering features and using cluster labels as feature labels. However, number of clusters must be defined by user and it can be different from intrinsic classes. Hence unsupervised methods are convenient for local image descriptors.

Histogram of Oriented Gradients (HOG) is another commonly used feature vector for object recognition in the literature. It is proposed in 1986 without the term HOG. However, it was not popular until that it is shown that HOG features can be used for human detection [21]. HOG features are global image descriptors which are robust to illumination changes. However, they are not rotation and scale invariant. Also, it should be noted that the size of HOG feature changes according to the image size and the features are usually very high dimensional vectors. Hence, various methods for dimensionality reduction of HOG features are proposed in the literature.

PCA is a popular dimensionality reduction technique for HOG features, too [22]–[25]. However, better results than PCA are reported using different dimensionality techniques [26]–[28]. Another popular technique for reducing HOG features is Partial Least Squares (PLS) which is a supervised method [27], [29], [30]. Moreover, some studies reported the results of Locality Preserving Projections (LPP) and its derived versions [31], [32], LDA and its derived versions [26], [28], [32] and RP [22] for dimensionality reduction of HOG features.

In our previous work, autoencoders (AE) were proposed for dimensionality reduction of SIFT vectors [33]. SIFT vectors at half and quarter size were obtained and used in vehicle logo recognition. Compared to the original vectors, the accuracy was decreased by 19% and 22% while using half and quarter sized vectors, respectively.

In this work, the study is improved and extended to SURF, HOG and deep learning based features. In addition to this, features obtained from noisy images, are included. Moreover, object recognition performances of reduced dimensional vectors are compared, which are obtained by PCA, VAE [34] and the proposed method.

Object recognition is a classification problem and in order to improve classification results, one can use methods which are aimed to maximize class separation. These methods require class labels, so they are supervised algorithms. However, class labels cannot be obtained for SIFT and SURF features, since these vectors have information about keypoints in images instead of the whole image. As in this study, object recognition can be accomplished by classifying the images which consist of one object and image classification is achieved by matching the query image with the image, which has the maximum number of similar keypoints. It should be highlighted that, this scheme demonstrates that a keypoint is not related with only one image or class.

To sum up, class labels cannot be obtained for keypoints, and, hence not for SIFT and SURF vectors, either. Thus, unsupervised methods should be used for SIFT and SURF vectors. In addition to this, HOG vectors and DEEP features are input into dimensionality reduction task applying the same method to evaluate the method's performance on both the keypoint descriptors and the image descriptors.

To achieve these goals, autoencoders, which are unsupervised neural network algorithms, are proposed for dimensionality reduction of HOG, SIFT, SURF and DEEP features. It is shown in the literature that autoencoders are useful for searching low dimensional embeddings, especially for graph structured data [35], [36]. However, here autoencoders are used for exploring useful embeddings for visual feature vectors. This method can capture nonlinear relationship in the data and provides a ready model for new input data. It yields reduced feature vectors which are named as AE-HOG, AE-SIFT, AE-SURF and AE-DEEP. The overall setup is shown in Figure 1.

Our contributions are mainly the introduction of an autoencoder based dimensionality reduction method for visual features commonly utilized in object recognition tasks, making comparison with two alternative methods and making analysis on both of the original and noisy images. Results suggest that although the data amount is reduced to half, object recognition accuracies obtained by the reduced features are comparable with the ones obtained using original sized feature vectors. Moreover, the proposed method outperforms PCA based feature reduction scheme for SIFT, SURF and DEEP feature vectors. Also, it outperforms VAE based method for HOG, SURF and DEEP features.

The remainder of this paper is organized as follows: Section II introduces the object recognition scheme using feature descriptors. Section III provides information about autoencoders, dataset and experimental setup for the implementation. Then Section IV presents the obtained results. Finally, Section V concludes this study.

II. Object recognition using feature descriptors

Object recognition using hand-crafted feature descriptors is based on vector matching and is achieved in two ways according to the feature type. Since HOG vectors are not local image descriptors like SIFT and SURF features, image matching scheme is different for HOG features. The reason is that one HOG feature is extracted per image, while hundreds of SIFT and SURF features are extracted per image. To match an image with a HOG vector, this vector is tried to match with the vectors of labelled images using a distance measure such as Euclidean distance. The image is determined in the class whose vector has the minimum distance to the query image vector. However, to match local image descriptors like SIFT and SURF, each feature vector of the queried image is compared with all of the feature vectors of labelled images. Thus each query vector selects a class whose vector has the

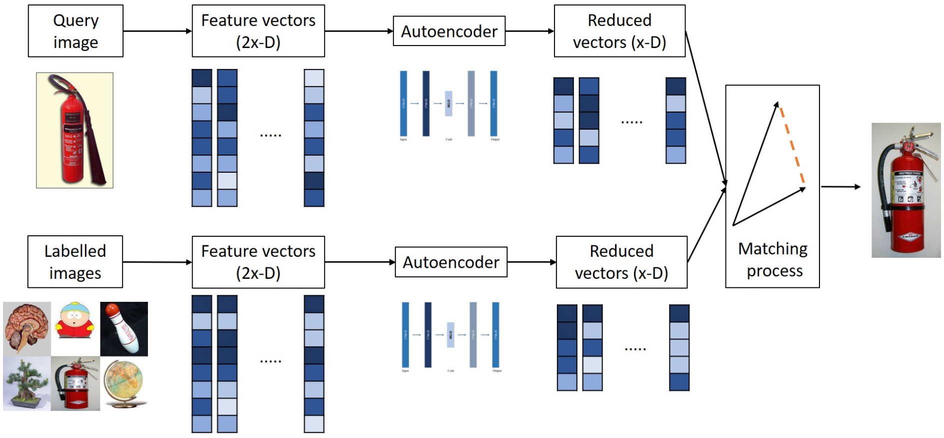

Fig. 1. The object recognition process using reduced feature vectors.

minimum distance with query vector. So, the matching result of the query image is the class which has the highest score.

Deep learning models extract some features in every layer and these features are input into other layers which make similar operations on them. In this study, the features which are output of the second to last layer, are used as DEEP features. In the experiments, the features are tried to match as in matching scheme of HOG features, instead of using some operations in the model.

Related papers of SIFT and SURF, are proposed to use threshold values for vector matching [1], [18]. This threshold is used for the ratio of distance of the closest vector to distance of the second closest vector. This second closest vector is obtained from a labelled image of another class, if there is any. If this ratio is lower than the threshold, the match is accepted; if not, the match is unaccepted.

In this study, threshold concept is used without the requirement of having the second closest vector of an image coming from another class. Different threshold values which are increased by 0.1 between 0 and 1.1, are tested for original and reduced vectors, and optimum thresholds are determined.

In order to match the feature vectors, Euclidean distance is used. To assess the performance of object recognition, accuracy metric is computed as in (1):

$$A = C/N \tag{1}$$

where A represents the accuracy, C is the number of correctly classified images and N is the number of all images in the set.

III. Method

In this section, working principle of the proposed method, designs of the used models and dataset will be explained.

A. Autoencoders

Autoencoders are neural networks whose goal is to reconstruct the input. In other words, they try to learn the identity function, in an unsupervised manner. While autoencoders try to reconstruct the input, they are limited by some restrictions. Hence, they cannot copy only the data, instead of it they must learn useful features in order to reconstruct the data. Therefore, they are useful for applications such as dimensionality reduction, data denoising and feature learning [37]. They mainly consist of two components which are encoder and decoder. The encoder maps the input to the code and the decoder maps the code to the output.

In order to reconstruct the input, they try to minimize the objective function p in (2):

$$p = L(x, g(f(x))) \tag{2}$$

where x represents the input, f and g are the encoder and decoder functions, respectively, L is a distance function, such as mean squared error.

If the code part of the autoencoder has smaller dimension than input, then the autoencoder is undercomplete. These autoencoders are limited by forcing the code being in smaller dimension and they can be used in order to learn the most useful features of the data. If the code part has greater dimension from the input, then the autoencoder is overcomplete. These autoencoders can be used with some regularization to prevent just copying the data instead of learning useful features. For example, they can be forced to learn sparse representations and to have small values of the derivatives of the representations [37].

There are a few variants of autoencoders in the literature such as denoising, sparse, variational and contractive. In this study, dimensionality reduction performances of vanilla and variational autoencoders are compared and vanilla autoencoders are proposed for this problem. Vanilla autoencoder is the simplest variant of the autoencoder family, which is only forced to have the code in smaller dimension of the input.

B. Experimental setup

In order to achieve dimensionality reduction, eight vanilla autoencoders are designed for different feature vector sets. Keras with Tensorflow backend is used for all autoencoder implementations [38], [39]. Also Scikit-learn [40], Matplotlib [41] and SciPy [42] libraries are utilized in the study. To determine the hyperparameters of the autoencoders, Bayesian optimization is performed using Kopt library [43]. The determined structures of the autoencoders are presented in Table I. Adam optimizer with recommended hyperparameters in [44], is used for optimizing the loss function which is mean squared error for all autoencoders.

C. Dataset

The experiments are conducted on three sets obtained from Caltech-256 dataset [45] which is a challenging dataset because of having high intra-class variability. Common practice of object recognition studies on this dataset is using a fixed number of training images per class and remaining for test images [46]. Numerous approaches are proposed for object recognition on this dataset, however the reported results which rely on non-Convolutional Neural Networks (CNN) and CNN methods, could not exceed the accuracy of 60% [47]–[49] and 90% [50]–[52], respectively. In this study, small sets obtained from this dataset are used, because the main purpose of this study is to demonstrate the efficacy of the proposed AE based dimensionality reduction method.

Deep features are obtained using the VGG-M model [53], which is a model that is studied on Caltech-256 dataset in [51]. Moreover, both of the HOG and DEEP features are acquired from Set 1. SIFT and SURF features are extracted from Set 2 and Set 3, respectively. Each set consists of training, validation and test sets for dimensionality reduction algorithm. For object recognition task, the test set of the dimensionality reduction algorithm is used which consists of labelled and queried images of this task. The labelled images in the test set, consist of four images per class which are used for matching with queried images. So, this part has 40 images. The other part containing queried images, consists of the remaining 70 images (7 images per class).

To form the sets, ten object classes are chosen from Caltech-256 dataset, as presented in Table II. From each class, 11 images are selected and related feature vectors are extracted from them to form the test sets. Validation sets consist of features of different 11 images per each selected class. Moreover, training sets contain vectors of the remaining images in the sets. It should be noted that the validation sets are only used to optimize hyperparameters of the proposed method.

Before computations, all images are converted to grayscale, in SIFT, SURF and HOG experiments. Moreover, Set 1 images are resized to 60×60 pixels only for HOG experiments, in order to have fixed dimensional HOG features. For DEEP feature experiments, images are resized according to the deep learning model input and then, the images are normalized by subtracting their means from them.

Moreover, to further assess the performance of the methods, noisy sets are constructed for experiments which consist of the same images with Gaussian noise with zero mean and 0.01 variance. The features obtained from the sets are named as noisy-HOG, noisy-SIFT, noisy-SURF and noisy-DEEP, respectively.

IV. Results

For dimensionality reduction of HOG, SIFT, SURF and DEEP features obtained from original images, the autoencoders are trained with the training sets of 853, 433508, 215017 and 853 vectors, respectively. The loss graphs of the models are shown in Figures 2a, 2b, 2c and 2d, respectively. After the training steps, vectors in the test sets are input into the autoencoders. The vectors obtained in the code layer are stored, which are the reduced representations of these vectors and are named as AE-HOG, AE-SIFT, AE-SURF and AE-DEEP, respectively. Then the low-dimensional vectors are used in object recognition task. To compare our results, the dimensionalities of the features are reduced to the same size by PCA as stated in [23] and [17] and VAE. For this purpose, the same training sets are used for training of these methods and reduced vectors of the Set 1, Set 2 and Set 3 are obtained by these methods, too.

Accuracy and memory usage results of object recognition for original HOG, SIFT, SURF and DEEP features, before and after the dimensionality reduction by the autoencoder (AE), VAE and PCA are given in Table III. The results show that AE-HOG and AE-SIFT vectors provide similar results to original sized vectors and vectors obtained by VAE. Moreover, half-sized SURF features could not achieve the original sized vectors' performance. However, proposed method clearly outperforms PCA and VAE on reducing SURF vectors. Besides, AE based method improved the original DEEP feature vectors' performance, in contrast to PCA and VAE.

After that, the experiments are repeated on images with noise. First, object recognition is tested on the noisy images. Then, the dimensionality of noisy features is reduced with the proposed method, VAE and PCA for comparison. Finally, object recognition with reduced features is tested. The loss graphs of the autoencoders which are trained with the training sets of 853 HOG, 486102 SIFT, 349484 SURF and 853 DEEP feature vectors, are shown in Figures 2e, 2f, 2g and 2h, respectively. The results are shown in Table IV. The results show that the proposed method preserves the performance of the original sized vectors for noisy-SIFT and DEEP features. In addition, PCA and VAE outperform the original sized noisy-SIFT vectors. Furthermore, noisy AE-HOG vectors present one of the best results among reduced vectors. Moreover, noisy AE-SURF vectors outperforms clearly the original sized noisy-SURF vectors and reduced vectors by PCA and VAE.

V. Conclusions

In this study, autoencoders are proposed for the dimensionality reduction of HOG, SIFT, SURF and deep learning based feature vectors. Firstly, sets for each feature type are

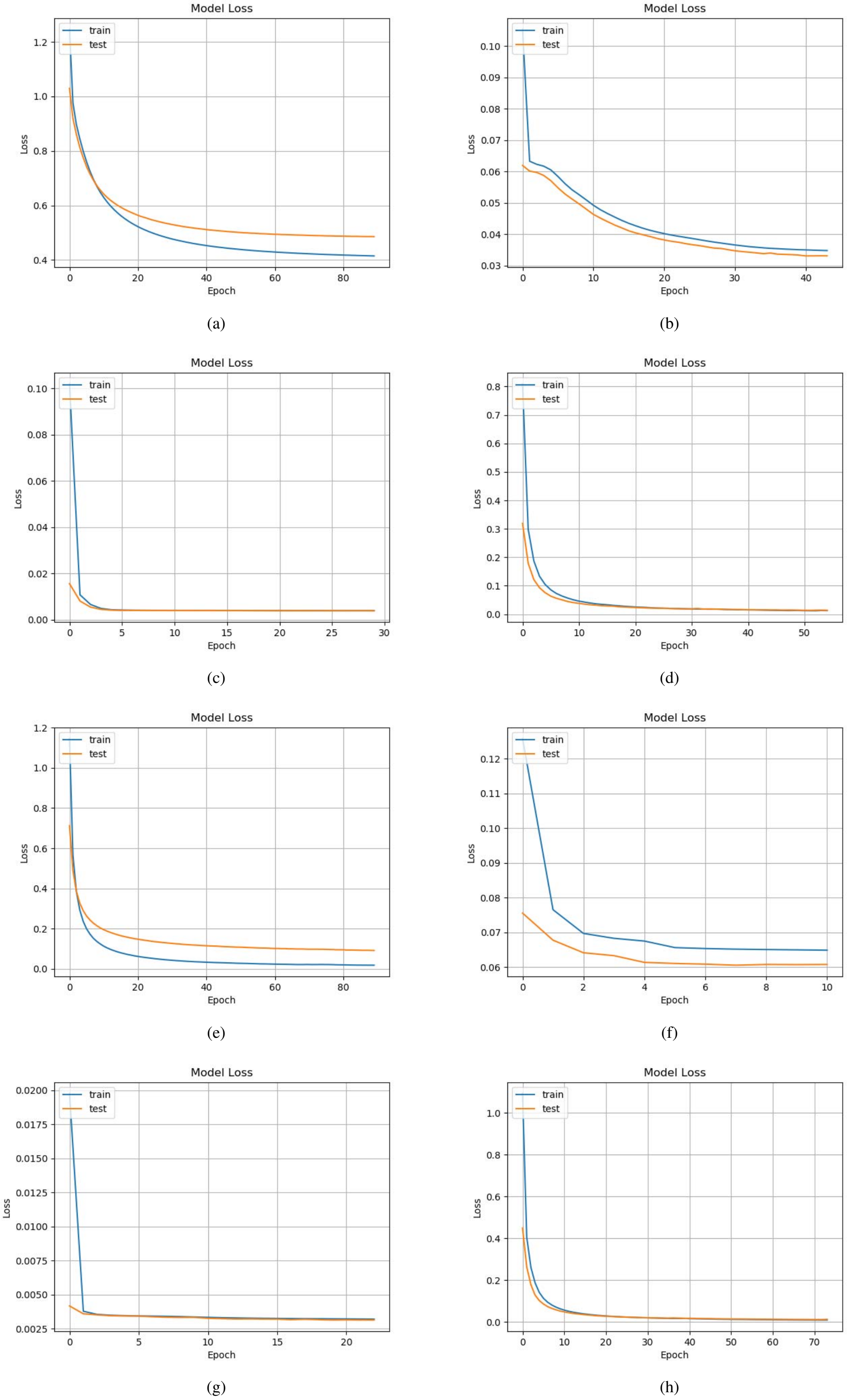

Fig. 2. Train and test losses versus epoch for a) HOG vectors, b) SIFT vectors, c) SURF vectors, d) DEEP feature vectors, e) noisy-HOG vectors, f) noisy-SIFT vectors, g) noisy-SURF vectors, h) noisy-DEEP feature vectors. Test sets consist of 110, 38051, 23895, 110, 110, 46394, 38201 and 110 vectors, respectively.

TABLE I

HYPERPARAMETERS OF THE AUTOENCODERS WHICH ARE USED FOR EIGHT TYPES OF FEATURE VECTORS. IT SHOULD BE NOTED THAT ALL OF THE LAYERS ARE DENSELY-CONNECTED LAYERS AND NOISY FEATURES DESCRIBE THE FEATURES OBTAINED FROM NOISY IMAGES

Autoencoder	**Feature type**	**#Units**	**Activation functions**	**Learning rate**	**Batch size**
AE-1	Original HOG	1296-972-648-972-1296	linear - linear linear - ReLU	2.20×10^{-4}	128
AE-2	noisy-HOG	1296-1296-648-1296-1296	linear - linear - linear - linear	4.54×10^{-4}	64
AE-3	Original SIFT	128-128-64-128-128	tanh-tanh-tanh-linear	1.25×10^{-3}	256
AE-4	noisy-SIFT	128-128-64-128-128	ReLU - ReLU - ReLU - linear	5.63×10^{-4}	128
AE-5	Original SURF	64-64-32-64-64	linear - linear - linear -linear	2.45×10^{-4}	128
AE-6	noisy-SURF	64-48-32-48-64	tanh-tanh-tanh-linear	5.95×10^{-4}	32
AE-7	Original DEEP	1000-500-500-500-1000	linear - linear - linear -linear	1.12×10^{-3}	64
AE-8	noisy-DEEP	1000-1000-500-1000-1000	linear-linear-linear-linear	9.51×10^{-4}	128

TABLE II

FOR EACH SET, 10 OBJECT CLASSES ARE SELECTED FROM CALTECH-256 DATASET

Set 1	**Set 2**	**Set 3**
Baseball glove	Baseball glove	Baseball glove
Bonsai-101	Bonsai-101	Brain-101
Bowling-pin	Brain-101	Fire extinguisher
Cartman	Calculator	French horn
Desk globe	Cartman	Frying pan
Electric guitar-101	Desk globe	Grand piano-101
Fire extinguisher	Fire extinguisher	Hamburger
Flashlight	Megaphone	House fly
French horn	Mountain bike	Megaphone
Frying pan	Paperclip	Video projector

constructed once and object recognition is tested using original sized features. Then, features are reduced to half size using the proposed method and the reduced dimensional vectors are input into object recognition. Also, results of dimensionality reduction is obtained using VAE and PCA, which is used commonly to reduce dimensionality of these feature vectors in unsupervised manner, for comparison. Moreover, noisy image sets are constructed and all experiments are repeated on these images. Finally, object recognition results are reported which belong to original sized and reduced features using the proposed method, VAE and PCA on original and noisy datasets.

The results demonstrate that using AE-SIFT and AE-HOG features on object recognition task, provides memory saving of 50% while keeping similar figures for object recognition accuracies obtained by original sized vectors. Although AE-SURF features are not as successful as the original sized features, this method outperforms VAE and PCA based feature reduction of SURF features. Moreover, this method provides improvement on object recognition using DEEP features.

The results indicate that the proposed method can preserve the object recognition accuracies of original sized noisy-SIFT and noisy-DEEP features. Besides, noisy AE-HOG vectors perform best among other reduced vectors. Moreover, this method improves the accuracy of original sized noisy-SURF features, while memory saving of 50% is achieved.

REFERENCES

[1] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.

[2] C. Strecha, A. Bronstein, M. Bronstein, and P. Fua, "LDAHash: Improved matching with smaller descriptors," *IEEE transactions on pattern analysis and machine intelligence*, vol. 34, no. 1, pp. 66–78, 2012.

[3] C. Yeo, P. Ahammad, and K. Ramchandran, "Rate-efficient visual correspondences using random projections," in *Proc. of IEEE international conference on image processing (ICIP)*, pp. 217–220, IEEE, 2008.

[4] M. Stommel and O. Herzog, "Binarising SIFT-descriptors to reduce the curse of dimensionality in histogram-based object recognition," in *Signal Processing, Image Processing and Pattern Recognition*, pp. 320–327, Springer, 2009.

[5] T. Tuytelaars and C. Schmid, "Vector quantizing feature space with a regular lattice," in *IEEE 11th International Conference on Computer Vision (ICCV)*, pp. 1–8, IEEE, 2007.

[6] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman, "Object retrieval with large vocabularies and fast spatial matching," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–8, IEEE, 2007.

[7] K. Mikolajczyk and J. Matas, "Improving descriptors for fast tree matching by optimal linear projection," in *IEEE 11th International Conference on Computer Vision (ICCV)*, pp. 1–8, IEEE, 2007.

[8] R. E. G. Valenzuela, W. R. Schwartz, and H. Pedrini, "Linear dimensionality reduction applied to scale invariant feature transformation and speeded up robust feature descriptors," *Journal of Electronic Imaging*, vol. 23, no. 3, pp. 1–13, 2014.

[9] H. Cai, K. Mikolajczyk, and J. Matas, "Learning linear discriminant projections for dimensionality reduction of image descriptors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 2, pp. 338–352, 2011.

[10] C. Zhao, A. A. Goshtasby, and S. Zhai, "An autoencoder-based image descriptor for image matching," in *Proceedings of the International Conference on Image Processing, Computer Vision, and Pattern Recognition (IPCV)*, p. 32, The Steering Committee of The World Congress in Computer Science, Computer , 2016.

[11] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *science*, vol. 313, no. 5786, pp. 504–507, 2006.

[12] N. Watcharapinchai, S. Aramvith, S. Siddhichai, and S. Marukatat, "Dimensionality reduction of SIFT using PCA for object categorization," in *International Symposium on Intelligent Signal Processing and Communications Systems (ISPACS)*, pp. 1–4, IEEE, 2009.

[13] V. Chandrasekhar, G. Takacs, D. Chen, S. S. Tsai, J. Singh, and B. Girod, "Transform coding of image feature descriptors," in *Conference on Visual Communications and Image Processing (VCIP)*, vol. 7257, International Society for Optics and Photonics, 2009.

[14] M. Asbach, P. Hosten, and M. Unger, "An evaluation of local features for face detection and localization," in *Ninth International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS)*, pp. 32–35, IEEE, 2008.

[15] M. Brown and S. Süsstrunk, "Multi-spectral SIFT for scene category recognition," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 177–184, IEEE, 2011.

[16] Y. Ke and R. Sukthankar, "PCA-SIFT: A more distinctive representation for local image descriptors," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 2, pp. II–II, IEEE, 2004.

TABLE III

Comparison of memory usage and accuracy results belonging to the vectors obtained from original images, before and after dimensionality reduction by autoencoder (AE), variational autoencoder (VAE) and PCA. Memory presents the memory space occupied by HOG, SIFT, SURF and DEEP feature vectors of test set images in Set 1, Set 2, Set 3 and Set 1, respectively. “Reduced by AE” and “Reduced by VAE” columns show the averages and standard deviations of object recognition accuracies obtained using the feature vectors on related image sets, over ten runs

Features	**Memory**		**Accuracy**			
	Original	**Reduced**	**Original**	**Reduced by AE**	**Reduced by VAE**	**Reduced by PCA**
HOG	1.14 MB	570.2 KB	65.71 %	64.29 ± 2.39 %	63.0 ± 2.07 %	**67.14 %**
SIFT	38.96 MB	19.48 MB	**62.85 %**	61.73 ± 2.11 %	62.29 ± 1.83 %	60.0 %
SURF	12.23 MB	6.12 MB	**51.43 %**	47.86 ± 3.01 %	33.86 ± 2.79 %	44.28 %
DEEP	440 KB	220 KB	97.14 %	**98.43 ± 0.45 %**	96.29 ± 1.31 %	97.14 %

TABLE IV

Comparison of memory usage and accuracy results belonging to the vectors obtained from noisy images, before and after dimensionality reduction by autoencoder (AE), variational autoencoder (VAE) and PCA. Memory presents the memory space occupied by HOG, SIFT, SURF and DEEP feature vectors of test set images in Set 1, Set 2, Set 3 and Set 1, respectively. “Reduced by AE” and “Reduced by VAE” columns show the averages and standard deviations of object recognition accuracies obtained using the feature vectors on related image sets, over ten runs

Features	**Memory**		**Accuracy**			
	Original	**Reduced**	**Original**	**Reduced by AE**	**Reduced by VAE**	**Reduced by PCA**
noisy-HOG	1.14 MB	570.2 KB	**67.14 %**	64.29 ± 1.56 %	60.29 ± 3.12 %	64.29 %
noisy-SIFT	47.43 MB	23.7 MB	50.0 %	50.28 ± 2.42 %	**55.29 ± 3.07 %**	52.86 %
noisy-SURF	19.56 MB	9.78 MB	45.71 %	**47.99 ± 2.46 %**	29.71 ± 2.62 %	45.71 %
noisy-DEEP	440 KB	220 KB	94.29 %	93.57 ± 0.75 %	92.71 ± 1.96 %	**95.71 %**

[17] R. E. G. Valenzuela, W. R. Schwartz, and H. Pedrini, “Dimensionality reduction through PCA over SIFT and SURF descriptors,” in *IEEE 11th International Conference on Cybernetic Intelligent Systems (CIS)*, pp. 58–63, IEEE, 2012.

[18] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, “Speeded-up robust features (SURF),” *Computer vision and image understanding*, vol. 110, no. 3, pp. 346–359, 2008.

[19] R. Valenzuela, H. Pedrini, and W. Schwartz, “Dimensionality reduction through LDA and bag-of-features applied to image retrieval,” in *Proc. IV ECCOMAS Thematic Conference on Computational Vision and Medical Image Processing*, pp. 31–37, 2013.

[20] Z. Boulkenafet, J. Komulainen, and A. Hadid, “Face antispoofing using speeded-up robust features and fisher vector encoding,” *IEEE Signal Processing Letters*, vol. 24, no. 2, pp. 141–145, 2017.

[21] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 1, pp. 886–893, IEEE, 2005.

[22] A. Savakis, R. Sharma, and M. Kumar, “Efficient eye detection using HOG-PCA descriptor,” in *Proc. SPIE*, vol. 9027, International Society for Optics and Photonics, 2014.

[23] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, “Object detection with discriminatively trained part-based models,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 32, no. 9, pp. 1627–1645, 2010.

[24] W.-L. Lu and J. J. Little, “Simultaneous tracking and action recognition using the PCA-HOG descriptor,” in *The 3rd Canadian Conference on Computer and Robot Vision (CRV)*, p. 6, IEEE, 2006.

[25] T. Kobayashi, A. Hidaka, and T. Kurita, “Selection of histograms of oriented gradients features for pedestrian detection,” in *International conference on neural information processing (ICONIP)*, pp. 598–607, Springer, 2007.

[26] O. Déniz, G. Bueno, J. Salido, and F. De la Torre, “Face recognition using histograms of oriented gradients,” *Pattern Recognition Letters*, vol. 32, no. 12, pp. 1598–1603, 2011.

[27] A. Misra, T. Abe, and K. Deguchi, “Hand gesture recognition using histogram of oriented gradients and partial least squares regression,” in *International Conference on Machine Vision Applications (MVA)*, pp. 479–482, 2011.

[28] D. Monzo, A. Albiol, A. Albiol, and J. M. Mossi, “Color HOG-EBGM for face recognition,” in *18th IEEE International Conference on Image Processing (ICIP)*, pp. 785–788, IEEE, 2011.

[29] S. U. Hussain and W. Triggs, “Feature sets and dimensionality reduction for visual object detection,” in *Proc. British Machine Vision Conference*, pp. 112.1–112.10, BMVA Press, 2010.

[30] W. R. Schwartz, A. Kembhavi, D. Harwood, and L. S. Davis, “Human detection using partial least squares analysis,” in *IEEE 12th international conference on Computer vision*, pp. 24–31, IEEE, 2009.

[31] Q. J. Wang and R. B. Zhang, “LPP-HOG: A new local image descriptor for fast human detection,” in *IEEE International Symposium on Knowledge Acquisition and Modeling Workshop*, pp. 640–643, IEEE, 2008.

[32] M. Mathias, R. Timofte, R. Benenson, and L. Van Gool, “Traffic sign recognitionhow far are we from the solution?,” in *The 2013 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8, IEEE, 2013.

[33] R. K. Keser, E. Ergün, and B. U. Töreyin, “Vehicle logo recognition with reduced-dimension SIFT vectors using autoencoders,” in *Multidisciplinary Digital Publishing Institute Proceedings*, vol. 2, 2018.

[34] D. P. Kingma and M. Welling, “Auto-encoding variational bayes,” *arXiv preprint arXiv:1312.6114*, 2013.

[35] W. Yu, C. Zheng, W. Cheng, C. C. Aggarwal, D. Song, B. Zong, H. Chen, and W. Wang, “Learning deep network representations with adversarially regularized autoencoders,” in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 2663–2671, ACM, 2018.

[36] S. Pan, R. Hu, G. Long, J. Jiang, L. Yao, and C. Zhang, “Adversarially regularized graph autoencoder for graph embedding,” *arXiv preprint arXiv:1802.04407*, 2018.

[37] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016. http://www.deeplearningbook.org.

[38] F. Chollet *et al.*, “Keras.” https://keras.io, 2015.

[39] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, *et al.*, “Tensorflow: a system for large-scale machine learning,” in *OSDI*, vol. 16, pp. 265–283, 2016.

[40] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, *et al.*, “Scikit-learn: Machine learning in python,” *Journal of machine learning research (JMLR)*, vol. 12, no. Oct, pp. 2825–2830, 2011.

[41] J. D. Hunter, “Matplotlib: A 2D graphics environment,” *Computing in science & engineering*, vol. 9, no. 3, pp. 90–95, 2007.

[42] E. Jones, T. Oliphant, and P. Peterson, “{SciPy}: open source scientific tools for {Python},” 2001.

[43] Ž. Avsec, “kopt - hyper-parameter optimization for keras.”

[44] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv:1412.6980*, 2014.

[45] G. Griffin, A. Holub, and P. Perona, "Caltech-256 object category dataset," 2007.

[46] Y. Pu, Z. Gan, R. Henao, X. Yuan, C. Li, A. Stevens, and L. Carin, "Variational autoencoder for deep learning of images, labels and captions," in *Advances in neural information processing systems*, pp. 2352–2360, 2016.

[47] C. Zhang, G. Zhu, C. Liang, Y. Zhang, Q. Huang, and Q. Tian, "Image class prediction by joint object, context, and background modeling," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 2, pp. 428–438, 2018.

[48] C. Zhang, C. Li, D. Lu, J. Cheng, and Q. Tian, "Birds of a feather flock together: Visual representation with scale and class consistency," *Information Sciences*, vol. 460–461, pp. 115–127, 2018.

[49] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong, "Locality-constrained linear coding for image classification," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3360–3367, IEEE, 2010.

[50] L. Zheng, Y. Zhao, S. Wang, J. Wang, and Q. Tian, "Good practice in CNN feature transfer," *arXiv:1604.00133*, 2016.

[51] Q. Li, Q. Peng, and C. Yan, "Multiple VLAD encoding of CNNs for image classification," *Computing in Science & Engineering*, vol. 20, no. 2, pp. 52–63, 2018.

[52] C. Zhang, J. Cheng, and Q. Tian, "Multiview label sharing for visual representations and classifications," *IEEE Transactions on Multimedia*, vol. 20, no. 4, pp. 903–913, 2017.

[53] K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman, "Return of the devil in the details: Delving deep into convolutional nets," in *Proceedings of the British Machine Vision Conference*, BMVA Press, 2014.