SPD Hashing Network for Fast Image Set Classification and Retrieval

Xiaxin Wang, Lixuan Zong, Xiaobo Shen

***Abstract*—Symmetric Positive Definite (SPD) manifold has powerful capability of modeling image set, and achieves good image set classification performance. Hashing has been widely used for large-scale image retrieval due to its superiority of computation and storage. In this work, we propose SPD Hashing Network (SPDHNet) that employs deep manifold neural network and hashing for fast image set classification and retrieval. Specifically, the proposed SPDHNet includes BiMap, ReEig, and LogEig Layers to extract nonlinear discriminative features of SPD manifold, and then uses hashing block to generate hash code. The experiments demonstrate the proposed method can reduce time and storage while achieving competitive accuracies.**

***Keywords*—Riemannian manifold, hashing, deep learning, image set classification**

I. Introduction

In recent years, due to the rapid development of multimedia technology, there has been a surge in the amount of images captured by cameras or surveillance videos. An image set consists of multiple images belonging to a specific object, and offers richer variability information about the object than a single-shot image. Image set tasks, e.g., classification, retrieval have attracted increasing attention, and have showed their great potential. However, the appearance of image sets often varies a lot due to different shooting conditions, e.g., illumination, pose, and thus makes image set tasks challenging.

Riemannian manifolds have been widely explored as image set modeling technique due to their powerful nonlinear representation capability. Among them, covariance matrix has achieved remarkable success, and some works have been proposed. For instance, [1] embeds SPD manifolds into reproducing kernel Hilbert space (RKHS) using well-established kernel function, and [2], [3] directly learn similarity Riemannian metrics of SPD matrix. To further enhance modeling capability, SPDNet [4] is proposed to learn deep SPD manifold feature. However, most existing methods mainly learn continuous representation, which cannot be efficiently applied for large-scale tasks.

Hashing [5] encodes data into low-dimensional hash code while preserving similarity information among data, which brings great improvement in terms of computational and storage cost. It motivates us to leverage advantages of hashing to enhance the efficiency of image set tasks. To this end, we propose SPD Hashing Network (SPDHNet) by combing the advantages of hashing, SPD manifold, and deep neural network. As illustrated in Fig. 1, the proposed SPDHNet includes SPD manifold feature block and hashing block. The experimental results on two image set benchmarks demonstrate that the proposed method achieves competitive accuracies while obtaining much higher efficiency.

This work was supported by the National Natural Science Foundation of China under Grant No. 62101268, the Natural Science Foundation of Jiangsu Province, China under Grant No. BK20230095. *(Corresponding author: Xiaobo Shen.)*

X. Wang, X. Shen are with School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China (e-mail: wangxiaxin@njust.edu.cn).

L. Zong is with Department of Electrical and Electronic Engineering, The Hong Kong Polytechnic University, Hong Kong SAR.

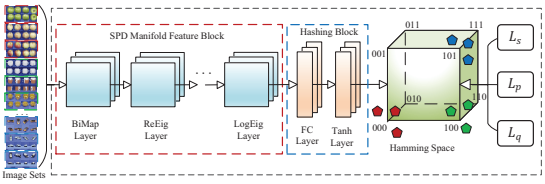

Fig. 1. The illustration of the proposed SPDHNet.

II. Methodology

A. SPD Manifold Feature Block

The SPD manifold feature block [4] consists of three layers, i.e., BiMap, ReEig, LogEig layers to extract nonlinear features of SPD manifold. We first assume $X_{k-1} \in \mathbb{R}^{d_{k-1} \times d_{k-1}}$ is the input of the k-th layer, and define the three layers as follows:

BiMap Layer: This layer projects each input SPD matrix into a new one via $X_k = f_b^{(k)}(W_k, X_{k-1}) = W_k^\top X_{k-1} W_k$, where f_b denotes a bilinear mapping, $W_k \in \mathbb{R}^{d_{k-1} \times d_k} (d_k < d_{k-1})$ denotes its corresponding transformation matrix.

ReEig Layer: This layer injects nonlinearity by tuning-up the small positive eigenvalues of each input SPD matrix via $X_k = f_r^{(k)}(X_{k-1}) = U \max(\epsilon I, \Sigma) U^\top$, where f_r is a nonlinear rectification function, ϵ is a small activation threshold, and eigenvalue decomposition of X_{k-1} is denoted as $X_{k-1} = U \Sigma U^\top$.

LogEig Layer: This layer embeds the SPD matrix into a flat space with the following mathematical form $X_k = f_l^{(k)}(X_{k-1}) = U \log(\Sigma) U^\top$.

B. Hashing Block

The hashing block first employs a FC layer to obtain r-dimensional continuous features, and further uses a Tanh layer to constrain feature in the range of $[-1, 1]$. The proposed

979-8-3503-5409-6/24/$31.00 ©2024 IEEE
DOI 10.1109/CAI59869.2024.00234

SPDHNet minimizes cross-entropy loss L_s to ensure the learned hash code achieves good classification performance:

$$L_s = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{c} y_{ij}\log\hat{y}_{ij} \tag{1}$$

where c and n are the class number and batch size respectively, y_i and $\hat{y}_i$ denote ground-truth and predicted label of the i-th image set respectively.

In addition, the proposed SPDHNet further employs pair-wise cross entropy loss to preserve the semantic similarity of each hash code pairs. Specifically, given two hash codes b_i and b_j, we have the following loss:

$$L_p = \sum_{s_{ij}\in S}\Big(\log(1+\exp(b_i^\top b_j)) - s_{ij}b_i^\top b_j\Big) \tag{2}$$

where $S = \{s_{ij}\}$ denotes the semantic similarity matrix generated by labels. We further define a quantization loss to enable the continuous feature to approach to hash code:

$$L_q = \sum_{i=1}^{n}\Big(\big\||b_i| - 1\big\|_1\Big) \tag{3}$$

To this end, the final objective loss is formulated as:

$$L = L_s + \alpha L_p + \beta L_q \tag{4}$$

where α and β are two regularization parameters to balance the three losses.

III. EXPERIMENTS

A. Experiment Setting

The two large-scale image set datasets are used for experiment. **AFEW** collects 1345 videos of actors in movies, each video is classified into one of seven expressions. We follow [4], and select 80% and 20% clips for training and testing. **BBT** includes 4,667 video clips of 15 characters. We randomly select 3268 image sets for training and the rest for testing.

We compare the proposed SPDHNet with the following baselines: CDL [1], RMML [3], SPDML-AIM [2], SPDML-Stein [2], ITQ [6], SDH [7], and SPDNet [4]. Among, our SPDHNet is an improved method based on SPDNet. The source codes of the baselines are kindly provided by the authors. The network of the proposed SPDHNet is constructed as $X_0 \rightarrow f_b^1 \rightarrow f_r^2 \rightarrow f_b^3 \rightarrow f_r^4 \rightarrow f_b^5 \rightarrow f_l^6 \rightarrow f_f^7$, where f_b, f_r, f_l, f_f denote BiMap, ReEig, LogEig, FC layers respectively. The learning rate, rectification threshold, batch size, hash code length are set to 0.01, 10^{-4}, 30, 64 respectively. The trade-off parameters α, β are set to 10^{-2}, 10^{-4} respectively.

B. Results

The accuracies and mAPs of all the methods on image set classification and retrieval tasks are summarized in Table I. From the table, we see the proposed SPDHNet outperforms all the baselines on AFEW, and takes the second place on BBT. Specifically, the proposed method improves accuracies of SPDNet by 7.27%, 0.44% on AFEW and BBT respectively, and improves mAPs of SPDNet by 9.18%, 1.14% on AFEW and BBT respectively.

TABLE I
THE ACCURACIES AND MAPS (IN PERCENTAGES) OF ALL THE METHODS. '-' DENOTES NO AVAILABLE RESULT IN THIS CASE AS RUNNING TIME OF THE METHOD EXCEEDS ONE WEEK.

Method	AFEW		BBT	
	Acc	mAP	Acc	mAP
CDL	29.65	19.96	87.56	88.49
RMML	15.90	17.07	**97.78**	50.93
SPDML-AIM	15.36	16.79	-	-
SPDML-Stein	16.98	16.84	-	-
ITQ	19.14	25.30	93.14	92.32
SDH	31.27	32.51	95.00	**94.80**
SPDNet	25.34	25.10	97.20	93.45
SPDHNet	**32.61**	**34.28**	97.64	94.59

TABLE II
STORAGE AND RUNNING TIME OF DISTANCE CALCULATION OF SOME REPRESENTATIVE METHODS.

Method	Storage		Running Time	
	AFEW	BBT	AFEW	BBT
CDL	28.23MB	116.36MB	0.25s	1.85s
RMML/SPDML	2.52GB	9.11GB	4847.73s	14364.62s
SPDNet	8.27MB	18.23MB	0.28s	1.97s
Hashing	132.38KB	291.69KB	0.01s	0.02s

Table II reports storage of the features learned by some representative methods, and the running time of distance calculation in testing stage. It can be seen that the storage of hash code is significantly lower than those of the other representations. In addition, the running time of Hamming distance calculation is also much lower than that of Euclidean distance calculation. The empirical results verify the superiority of hashing in terms of efficiency in image set tasks.

IV. CONCLUSION

This paper proposes SPD Manifold Hashing Network (SPDHNet) for large-scale image set classification and retrieval. The SPDHNet employs SPD manifold feature block to extract nonlinear feature of SPD manifold, and employs hashing block to generate hash code that can achieve efficient storage and computation. The experimental results validate the effectiveness and efficiency of the proposed method.

REFERENCES

[1] R. Wang, H. Guo, L. S. Davis, and Q. Dai, "Covariance discriminative learning: A natural and efficient approach to image set classification," in *CVPR*, 2012, pp. 2496–2503.

[2] M. Harandi, M. Salzmann, and R. Hartley, "Joint dimensionality reduction and metric learning: A geometric take," in *ICML*, 2017, pp. 1404–1413.

[3] P. Zhu, H. Cheng, Q. Hu, Q. Wang, and C. Zhang, "Towards generalized and efficient metric learning on riemannian manifold." in *IJCAI*, 2018, pp. 3235–3241.

[4] Z. Huang and L. Van Gool, "A riemannian network for spd matrix learning," in *AAAI*, vol. 31, no. 1, 2017.

[5] J. Wang, T. Zhang, j. song, N. Sebe, and H. T. Shen, "A survey on learning to hash," *IEEE TPAMI*, vol. 40, no. 4, pp. 769–790, 2018.

[6] Y. Gong, S. Lazebnik, A. Gordo, and F. Perronnin, "Iterative quantization: A procrustean approach to learning binary codes for large-scale image retrieval," *IEEE TPAMI*, vol. 35, no. 12, pp. 2916–2929, 2012.

[7] F. Shen, C. Shen, W. Liu, and H. Tao Shen, "Supervised discrete hashing," in *CVPR*, 2015, pp. 37–45.