Combining Deep and Manifold Learning For Nonlinear Feature Extraction in Texture Images

Cédrick Bamba Nsimba
Department of Computer Science
Federal University of São Carlos
São Carlos, SP, Brazil
cedrick.bamba@ifsp.edu.br

Alexandre L. M. Levada
Department of Computer Science
Federal University of São Carlos
São Carlos, SP, Brazil
alexandre@dc.ufscar.br

***Abstract*—This paper applies a two-step approach for texture classification by combining Manifold learning with Deep CNN feature extractors. The first step is to use CNN architecture to compute the feature vector of a given image. The second step is to apply Manifold Learning algorithms on the features computed in the first step to making a refined feature vector. Eventually, this final representation is used to train SVM classifier. In the first step, we adopted VGG-19 network trained from scratch in order to extract texture features. In the next step, we used the DIMAL (Deep Isometric Manifold Learning Using Sparse Geodesic Sampling) configuration to train a neural network to reduce the dimensionality of the feature space in a nonlinear manner for generating the refined feature vector of the input image. Our concept is that the combination of a deep-learning framework with manifold learning techniques has the potential to select discriminative texture features from a high dimensional space. Based on this idea, we adopted this combination to perform nonlinear feature extraction in texture images. The resulting learned features were then used to train SVM classifier. The experiments demonstrated that our approach achieved better accuracy in texture classification than existing models if trained from scratch.**

***Index Terms*—Texture classification, Manifold learning, Deep learning, Feature extraction, Nonlinear dimensionality reduction**

I. Introduction

The characterization of texture is of fundamental interest to the problem of most of computer vision and image processing applications. For the purpose of achieving a coherent global representation of a texture image, it is often desirable to embed the high-dimensional feature space into a space of low dimensionality while preserving only the most effective features for class separability during the upcoming classification task.

Recently, CNNs [1] have been widely applied on a bunch of image tasks due to their great image discriminative ability and consequently, significant performance. Despite its successes in image classification, CNNs alone are not very suitable for texture classification. It is known that directly applying CNNs on texture classification task results in only moderate accuracy [2], [3]. Recently, in [4] Cimpoi et al. combined CNN with Fisher Vectors (FV-CNN) to increase accuracy in texture classification. Andrearczyk et tal. [5] proposed texture CNN (T-CNN) which is a specialized version of CNN for texture classification.

However, the most common way to realize feature extraction in CNNs, even without making use of pretraining or finetuning, is to input a set of desired texture images and utilize one of the network layers (before the output one) as a feature vector. Nevertheless, in most of cases, the dimensionality of the feature vector may be higher and this can hamper the performance of the target classifier due to the well-known problem called *the curse of dimensionality*. It was Bellman who first introduced this term in 1986 [6] which refers to the fact that during the process of estimating a function of multiple variables to a given degree of accuracy, the sample size needs to grow with the number of variables.

In this paper, we propose a novel handcrafted deep convolutional networks for extracting texture features, followed by a modeling of maps, from a high-dimensional feature space generated in the previous step, that preserve geodesic distances on data manifolds. We exploit efficient deep isometric manifold learning using sparse geodesic sampling that progressively select features on the manifold by maximizing the spread of their pairwise geodesic distances. We demonstrate, in Section IV, that the resulting learned features are discriminative enough and can greatly boost the class separability of the SVM classifier.

Finally, our motivation to integrate manifold learning into deep learning based texture representation problem is threefold: (a) This approach allows to boost texture classification performance in the context of CNN based texture representation models trained from scratch; (b) Combination of deep CNN features and nonlinear dimensionality reduction techniques defining a novel approach based on deep and manifold learning and (c) Incorporation of geometric analysis into deep CNN based texture classification methods.

II. Background

A. Manifold Learning

Manifold learning consists of retrieving the low-dimensional representation of a data from a possibly a non-linear high-dimensional one. Several algorithms based on an eigendecomposition (spectral methods) provide an embedding only for given training data points, with no straightforward extension

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

for out-of-sample examples short of recomputing eigenvectors. These methods include Isomap [7], LLE [8] and Laplacian Eigenmaps [9]. Rougthly speaking, the main idea behind manifold learning techniques consists of computing the k nearest neighbors of all N data points. The next step is to construct a matrix of dimension $N \times N$ whose elements obey some geometric principle which characterizes the nature of the desired low-dimensional embedding. Therefore, a downside of spectral decomposition methods is the computational cost. As the number of data points grows, the spectral decomposition becomes large, and consequently, the computational cost is higher. In addition to this drawback, there is an out-of-sample extension of the map which is another computationally expensive task. In this paper, we overcame these issues with the use of the new framework, namely DIMAL [10].

B. CNN Based Texture Representation

In short, a standard CNN architecture consists of multiple trainable building blocks stacked on top of one another followed by a supervised classifier. Each block generally consists of three layers: a convolutional filter bank layer, a nonlinear layer, and a feature pooling layer. Many CNN-based texture representation methods have been proposed in recent years due to the high performance in many taks such as image classification achieved in [1]. A key to the success of CNNs is their ability to leverage large labeled datasets to learn high quality features [11]. The literature on CNN based texture representation methods is dominated by three categories of models: (a) the pretrained generic CNN models; (b) the finetuned CNN models and (c) handcrafted deep convolutional networks. This paper is mainly related to the last category. Among the proposed handcrafted deep convolutional networks in the literature, two of them deserve attention, namely Scattering convolution Network (ScatNet) [12] and PCANet [13]. In ScatNet, the convolutional filters are predefined being simply wavelet filters, such as Gabor or Haar wavelets, and no learning is required differently from CNN, where the convolutional filters are learned from data. PCANet, in turn, is based on trained PCA filters and LBP encoding [14] and histogramming for feature pooling. Compared with ScatNet, feature extraction in PCANet is much faster, but with weaker texture classification performance [15]. Recall that by handcrafted deep convolutional networks we mean CNNs trained from scratch.

C. Deep Isometric Manifold Learning (DIMAL)

DIMAL [10] is a nonlinear dimensionality reduction based framework which uses neural networks to model maps that preserve geodesic distances on data manifolds. The key difference from standard manifold learning techniques is that instead of optimizing over the individual coordinates of the points during MDS (multidimensional scaling) process, DIMAL optimizes over the function that generates these points by modeling this map as a neural network. This decision of modeling the isometric map with a parametric model provides a straightforward extension for out-of-sample by avoiding the need for eigenvectors recomputation. The aforementioned parametric model is in fact a simple forward pass of the network.

III. Proposed Method

Texture representation or texture feature extraction aims at transforming the input texture image into a feature vector that describes the properties of a texture, facilitating subsequent tasks such as texture classification. Within this context, this section presents a novel two-step CNN based texture representation approach for texture classification by combining Manifold learning with Deep CNN feature extractors. Our method is firstly inspired by the observation of how dimensionality reduction techniques can effectively learn relevant information from the raw data, and secondly, by the ability of CNNs to leverage large labeled datasets to learn high quality features. Based on this, we propose a straightforward solution which extracts feature vector of a texture image using VGG-19 network trained from scratch and makes a feature selection through Deep Isometric Manifold Learning to choose the most effective features for class separability.

The motivation for using the handcrafted VGG-19 network was its ability in extracting high quality texture features. Therefore, as shown in [16], its downside in texture classification task is due to the large number of the output feature vector (4096 dimensional based vector as adopted in this work), which creates the overfitting phenomenon. In this way, the combination of deep neural network based on the Siamese configuration [17], [18] and on the isometric property of such manifolds embedded into ambient space can help to improve the discriminative capacity of texture descriptors.

The proposed pipeline is sketched in Figure 1, consisting of the creation of an approach for texture classification strongly based on the application of dimensionality reduction techniques into CNN based texture representation. According to Figure 1(a), the first step of the proposed pipeline consists in training the VGG-19 network from scratch in order to extract texture features of the input image. We then input the resulting feature vector of 4096 dimension (we use the second fully connected layer, FC2, of 4096 neurons as feature vector) into DIMAL (refer to Figure 1(b)) based dimensionality reduction module for selecting the most relevant features in a nonlinear manner. In other words, in this step, feature selection is achieved DIMAL algorithm during which thresholds are used to maintain only the most effective features among all belonging to the output of the second fully connected layer of the handcrafted VGG-19 architecture. From this process, we have as an output a discriminative and reduced feature vector that contains only the best elements (the most relevant informations) of the input texture image. Finally, the selected feature vector is taken through the classification process using SVM classifier (refer to Figure 1(c)). Details about the proposed pipeline are shown in Algorithm 1.

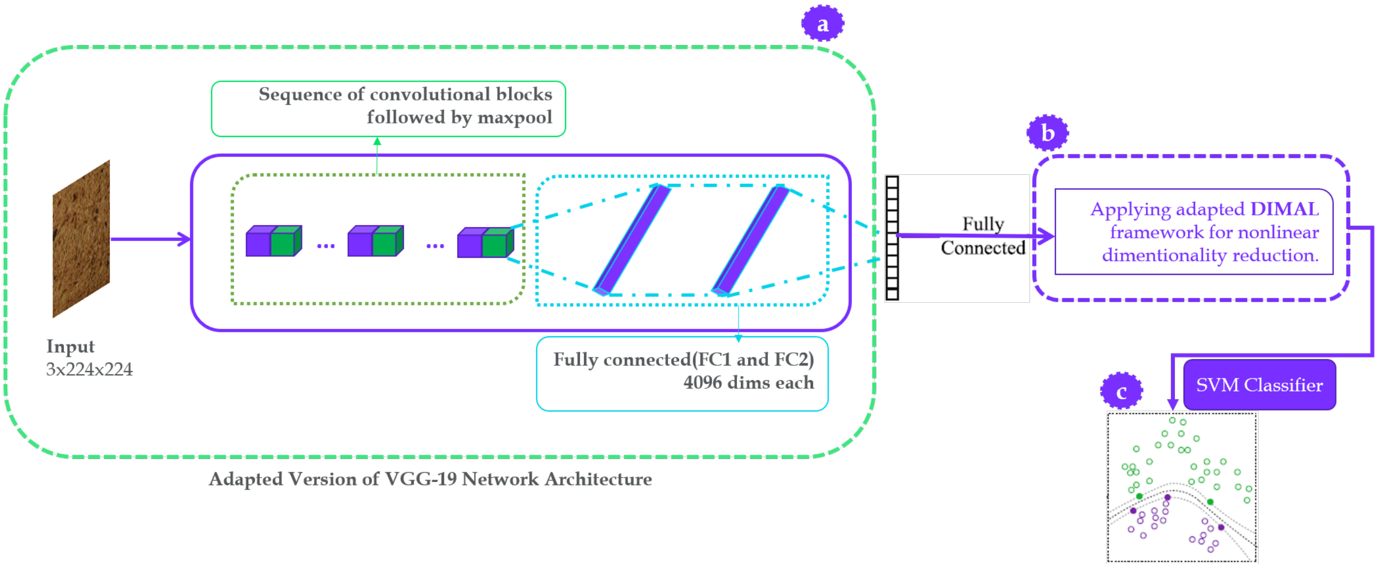

Fig. 1. The proposed pipeline.

Algorithm 1 Deep and Manifold Leaining Feature Extraction

1: From the input data $X_{m\times n}$ train VGG-19 network from scratch and consider the output of the fully connected layer (FC2) as the feature vector of dimension 4096.
2: Given the output of the previous step (for all images in the input dataset), compute the nearest-neighbor graph from the manifold data and obtain a set of K landmark points using Farthest Point Sampling Algorithm.
3: Obtain pairwise geodesic distances between landmarks using Dijkstra's or any other numerical algorithm.
4: Form a dataset of landmark pairs with corresponding geodesic distances.
5: Train network with Siamese configuration for the dataset obtained in the last step.
6: Obtain low-dimensional embedding with forward pass of the network.
7: Use SVM to classify the extracted and reduced feature vectors of the input texture images.

IV. Experiments

Datasets We used two publicly available texture datasets during our experiments, namely kth-tips2-b [19] and Salzburg [20]. The kth-tips2-b is a collection of 11 classes of images. Each class consists of 432 texture images. All images are captured varying scales, viewing angles and lighting conditions in order to study the generalization ability of material recognition methods to the new material instance. Salzburg contains a collection of 476 color texture images that have been captured around Salzburg (Austria). For each texture class (from total of 10 classes used in this paper), there were 128×128 source images, of which 80% was used for training the classifier, while the other 20% was used for testing.

Training from scratch Figure 2 shows the classification rates using the networks trained from scratch. We compared the proposed approach with two existing models considered as state-of-the-art CNN-based models for texture classification task, namely ScatNet and PCANet. Our model achieved very high classification performance on Salzburg (100.00%) and kth-tips2-b (99.5%) while ScatNet also achieved better results on both datasets, 99.7% on Salzburg and 99.4% on kth-tips2-b. PCANet, in turn, presented satisfactory performance in texture classification, 79.2% on Salzburg and 84.9% on kth-tips2-b. Compared with ScatNet and PCANet, feature extraction in our method is much faster with better invariance and texture classification performance.

V. Conclusion

This research presented an important contribution regarding the use of deep learning (CNN) in combination with dimensionality reduction techniques for texture classification task. We stated that a vector of $D = 4096$ features could be extracted directly from a VGG-19 network trained from scratch, and then these D features could be reduced to only d features (where $d < D$) that can better represent the data, achieving higher accuracies in classification task. The accuracies achieved by our model are superior to the state-of-the-art CNN-based model for texture classification task. This indicates that the combination of deep and manifold learning is capable of boosting the performance of texture descriptors.

Future studies can take into account new approaches for estimating the intrinsic dimensionality value of d for each manifold learning method, can also explore different manifold learning methods and variations of the supervised classifier's algorithms in order to produce better accuracies, as well as investigate the sub-spaces generated by such methods in order to understand better their discriminative power.

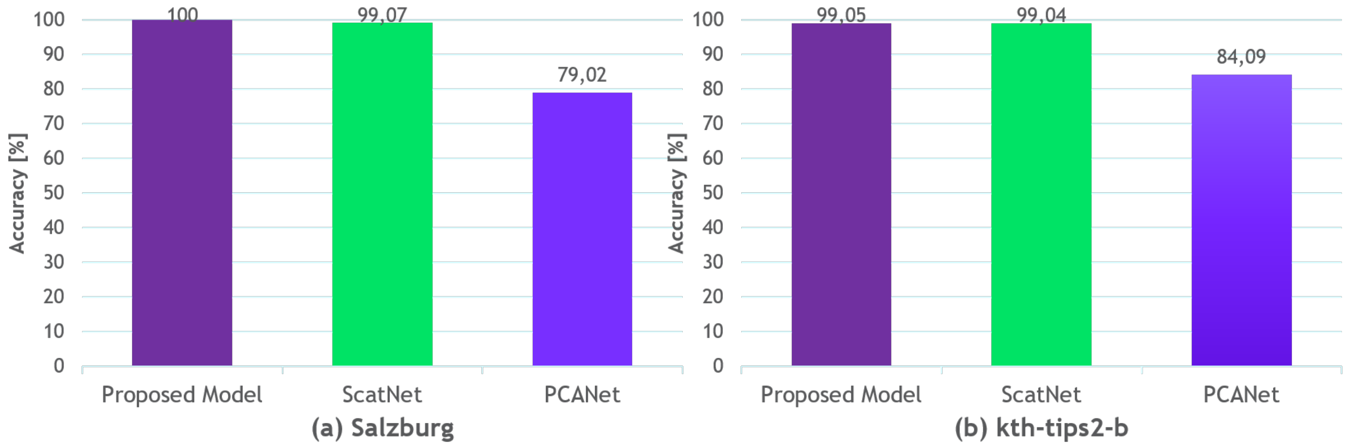

Fig. 2. Classification results of (a) Salzburg and (b) kth-tips2-b for networks trained from scratch. We compared our model with ScatNet and PCANet.

Acknowledgment

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

References

[1] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1, NIPS'12, pages 1097–1105, USA, 2012. Curran Associates Inc.

[2] S. Fujieda, K. Takayama, and T. Hachisuka. Wavelet Convolutional Neural Networks for Texture Classification. arXiv e-prints, page arXiv:1707.07394, Jul 2017.

[3] L. G. Hafemann, L. S. Oliveira, and P. Cavalin. Forest species recognition using deep convolutional neural networks. In 2014 22nd International Conference on Pattern Recognition, pages 1103–1107, Aug 2014.

[4] M. Cimpoi, S. Maji, I. Kokkinos, and A. Vedaldi. Deep filter banks for texture recognition, description, and segmentation. International Journal of Computer Vision, 118(1):65–94, May 2016.

[5] V. Andrearczyk and P. Whelan. Using filter banks in convolutional neural networks for texture classification. Pattern Recognition Letters, 84:63–69, 08 2016.

[6] R. Bellman and K. M. R. Collection. Adaptive Control Processes: A Guided Tour. 'Rand Corporation. Research studies. Princeton University Press, 1961.

[7] J. B. Tenenbaum, V. De Silva, and J. C. Langford. A global geometric framework for nonlinear dimensionality reduction. science, 290(5500):2319–2323, 2000.

[8] S. T. Roweis and L. K. Saul. Nonlinear dimensionality reduction by locally linear embedding. science, 290(5500):2323–2326, 2000.

[9] M. Belkin and P. Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. Neural computation, 15(6):1373–1396, 2003.

[10] Pai, Gautam Talmon, Ronen Bronstein, Alex Kimmel, Ron. (2018). DIMAL: Deep Isometric Manifold Learning Using Sparse Geodesic Sampling. 10.1109/WACV.2019.00092.

[11] Liu, Li Chen, Jie Fieguth, Paul Zhao, Guoying Chellappa, Rama Pietikäinen, Matti. (2018). From BoW to CNN: Two Decades of Texture Representation for Texture Classification. International Journal of Computer Vision. 10.1007/s11263-018-1125-z.

[12] Bruna, J., Mallat, S. (2013). Invariant scattering convolution networks. IEEE TPAMI, 35(8), 1872–1886.

[13] Chan, T., Jia, K., Gao, S., Lu, J., Zeng, Z., Ma, Y. (2015). PCANet: A simple deep learning baseline for image classification? IEEE Trans Image Processing, 24(12), 5017–5032.

[14] Ojala,T., Pietikäinen,M.,Maenpää,T. (2002b). Multiresolution grayscale and rotation invariant texture classification with local binary patterns. IEEE TPAMI, 24(7), 971–987.

[15] Liu, L., Fieguth, P., Guo, Y.,Wang, X., Pietikäinen, M. (2017). Local binary features for texture classification: Taxonomy and experimental study. Pattern Recognition, 62, 135–160.

[16] S. Fujieda, K. Takayama, and T. Hachisuka. Wavelet Convolutional Neural Networks for Texture Classification. arXiv e-prints, page arXiv:1707.07394, Jul 2017.

[17] J. Bromley, I. Guyon, Y. LeCun, E. S¨ackinger, and R. Shah. Signature verification using a" siamese" time delay neural network. In Advances in neural information processing systems, pages 737–744, 1994.

[18] R. Hadsell, S. Chopra, and Y. LeCun. Dimensionality reduction by learning an invariant mapping. In Computer vision and pattern recognition, 2006 IEEE computer society conference on, volume 2, pages 1735–1742. IEEE, 2006.

[19] E. Hayman, B. Caputo, M. Fritz, and J.-O. Eklundh. On the significance of real-world conditions for material classification. In T. Pajdla and J. Matas, editors, Computer Vision - ECCV 2004, pages 253–266, Berlin, Heidelberg, 2004. Springer Berlin Heidelberg.

[20] R. Kwitt and P. Meerwald. Salzburg texture image database. Available at http://www.wavelab.at/sources/STex/, last viewed February 2018.