TexFusionNet: An Ensemble of Deep CNN Feature for Texture Classification

Swalpa Kumar Roy, Shiv Ram Dubey, Bhabatosh Chanda, Bidyut B. Chaudhuri and Dipak Kumar Ghosh

swalpa@cse.jgec.ac.in, srdubey@iiits.in, {chanda,bbc}@isical.ac.in, dipak@ieee.org

Introduction

- Texture is everywhere in natural images and it describes the smoothness or roughness of image surfaces [1].
- The texture classification from images is one of the important problems in the field of pattern recognition and computer vision having many applications, including material recognition, analysis of satellite, medical images, etc.
- Designing discriminative and powerful texture features, which is robust to realistic imaging conditions is a challenging task.
- Several hand-designed methods have been proposed in last few decades for the above problem.
- Nowadays, it is observed that the convolutional neural networks (CNN) perform extremely well for the classification task mainly over object and scene images [2].
- The focus of CNN in texture classification is limited due to non-availability of large texture image datasets which is required to achieves the good performance using CNN.
- Pre-trained models [3] are widely adapted for simplicity and efficient image feature description.
- In this work, a fused CNN (TexFusionNet) is proposed for texture classification by fusing the last representation layer of widely adapted AlexNet [4] and VGG16 [5]. On the top of the fused layer, a fully connected layer is used to generate the class score.

Proposed TexFusionNet

The proposed TexFusionNet architecture for texture images classification is shown in Fig. 1 where fusing of of widely AlexNet and VGG16 is adapted at the last representation layer. On the top of the fused layer, a fully connected layer is used to generate the class score. The categorical cross entropy loss is used to train the added layer after the fusion layer. Let us consider I as the input color texture image of resolution $m \times n$, $alex$ is a function representing the combination of convolutional layers, max-pooling layers and fully connected layers from 1^{st} layer to last representation layer of AlexNet model, and similarly $vgg16$ is a function for VGG16 model. So, the input image I is converted into I_{alex} and I_{vgg} having the resolution of 227×227 and 224×224 for AlexNet [4] and VGG16 [5] respectively as follows,

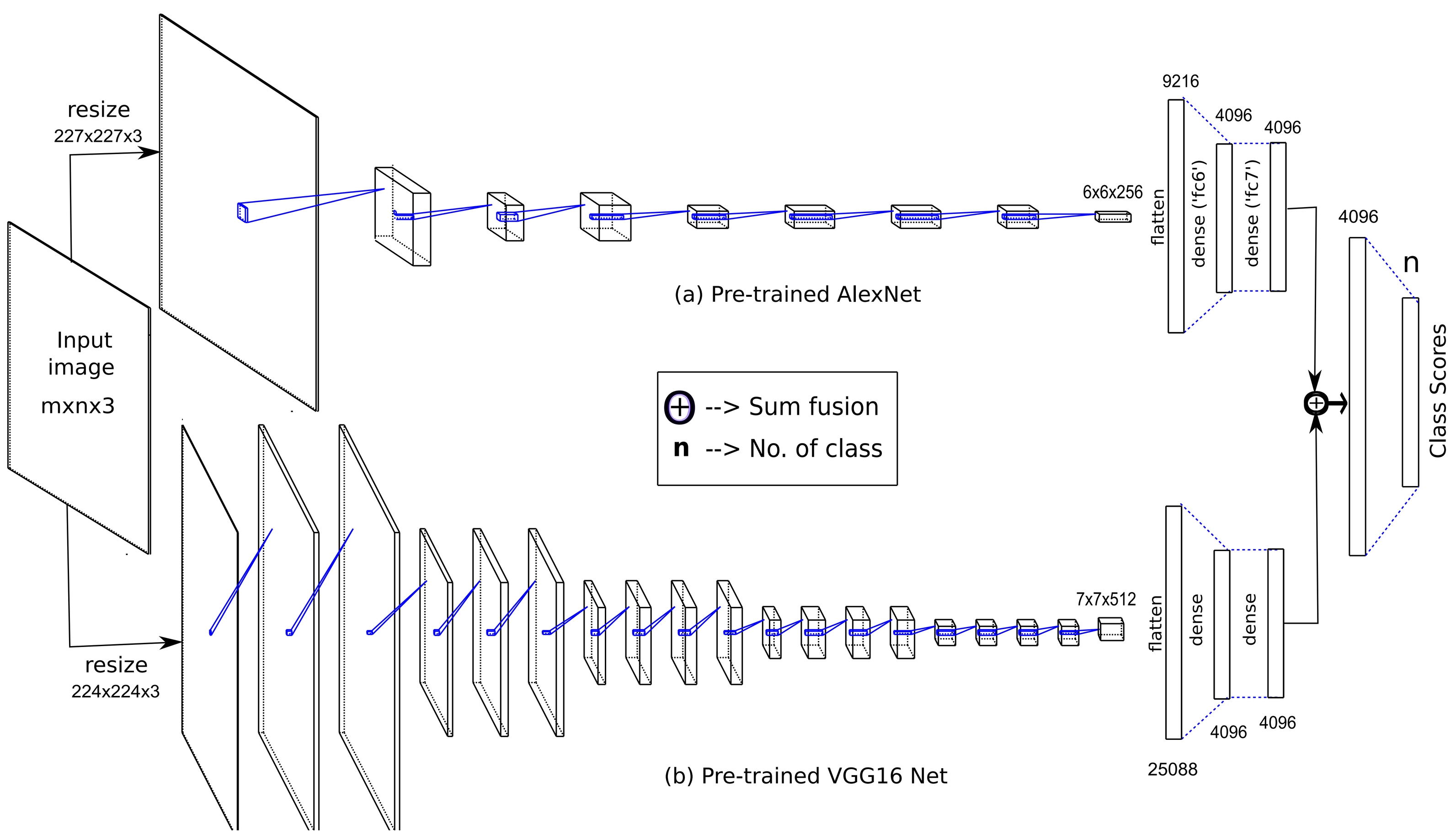

Figure 1: Proposed TexFusionNet architecture for texture classification.

$$I_{alex} = \tau(I, [227, 227]) \tag{1}$$
$$I_{vgg} = \tau(I, [224, 224]) \tag{2}$$

where $\tau(I, [\beta, \beta])$ is a transformation function to resize any image I into the resolution of $(\beta \times \beta)$.

Let us consider Φ_{alex} and Φ_{vgg} represent the features or values of last representation layer of AlexNet and VGG16 models respectively. The Φ_{alex} and Φ_{vgg} are defined as,

$$\Phi_{alex} = alex(I_{alex}) \tag{3}$$
$$\Phi_{vgg} = vgg(I_{vgg}). \tag{4}$$

where the dimensions of Φ_{alex} and Φ_{vgg} representation features are D_{alex} and D_{vgg}, respectively with $D_{alex} = D_{vgg}$.

In the proposed work, the Φ_{alex} and Φ_{vgg} texture features of input image I using AlexNet and VGG16 models are fused by addition to produce a combined texture feature representation denoted as Φ_{fused} and defined as follows,

$$\Phi_{fused}(i) = \Phi_{alex}(i) + \Phi_{vgg}(i) \quad \forall i \in [1, D] \tag{5}$$

where D is the dimension of the fused feature Φ_{fused} with $D = D_{alex} = D_{vgg}$, $\Phi_{fused}(i)$ is the i^{th} element of fused feature Φ_{fused} while $\Phi_{alex}(i)$ and $\Phi_{vgg}(i)$ are the i^{th} elements of input features Φ_{alex} and Φ_{vgg}, respectively.

Performance Evaluation

The performance comparison of proposed TexFusionNet with state-of-art hand-crafted as well as deep learning based methods are also carried out. Table 1 summarized the classification results over Brodatz, CUReT (shown in Fig. 3(a)) and KTH-TIPS databases for different methods. Though the number of training samples is very less, the proposed TexFusionNet based CNN features achieve the outstanding average classification accuracy of 98.76%, 100 % and 99.76% for Brodatz, KTH-TIPS and CUReT test suits, respectively. In addition the first plot of Fig. 2(a) shows the Loss vs. epochs whereas the second plot of Fig. 2(b) demonstrate the Accuracy vs. epochs for CUReT database.

Hand Craft Methods	Classification Rates (%)			Deep CNN Methods	Classification Rates (%)		
	Brodatz	KTH-TIPS	CUReT		Brodatz	KTH-TIPS	CUReT
LBPV	93.80	95.50	94.00	ScatNet	84.46	99.40	99.66
BRINT	98.22	97.75	97.06	PCANet	90.89	59.43	92.03
CLBP	94.80	97.19	97.40	AlexNet [4]	98.20	99.60	98.50
PRICOLBP	96.90	98.40	98.40	FV-VGGM [6]	98.60	99.80	98.70
COALBP	94.20	97.00	98.00	FV-VGGVD [6]	98.70	99.80	99.0
CDCP [7]	97.20	97.90	-	RandNet	91.14	60.67	90.87
RI-LBD	97.80	99.30	98.60	TexFusionNet	98.76	100	99.76

Table 1: Comparison of proposed TexFusionNet CNN feature with other variants of LBPs and state-of-the-art deep CNN methods in terms of the texture classification accuracy

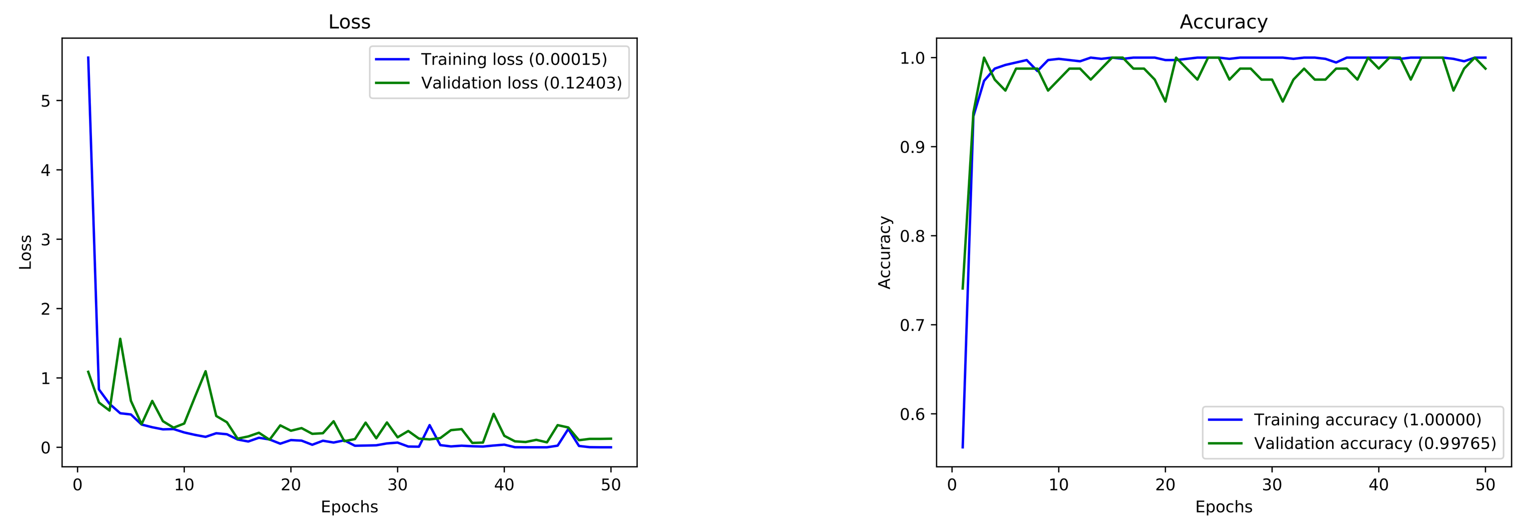

Figure 2: The performance of texture classification using the TexFusionNet method on CUReT database (a) Loss vs. epochs and (b) Accuracy vs. epochs.

To further visualize the performance of the proposed TexFusionNet in terms of the Receiver Operating Characteristics (ROC), the Area Under Curve (AUC) is measured and depicted in Fig. 3(b) over Brodatz, CUReT and KTH-TIPS databases. The true positive rate (TPR) and flase positive rate (FPR) values are plotted along y- and x-axis respectively. It is observed from Fig. 3(b) that the the performance of proposed model is reasonable over each texture database and proposed CNN feature achieves AUC values of 98.32%, 99.61% and 99.39%, respectively for Brodatz, KTH-TIPS and CUReT test suits.

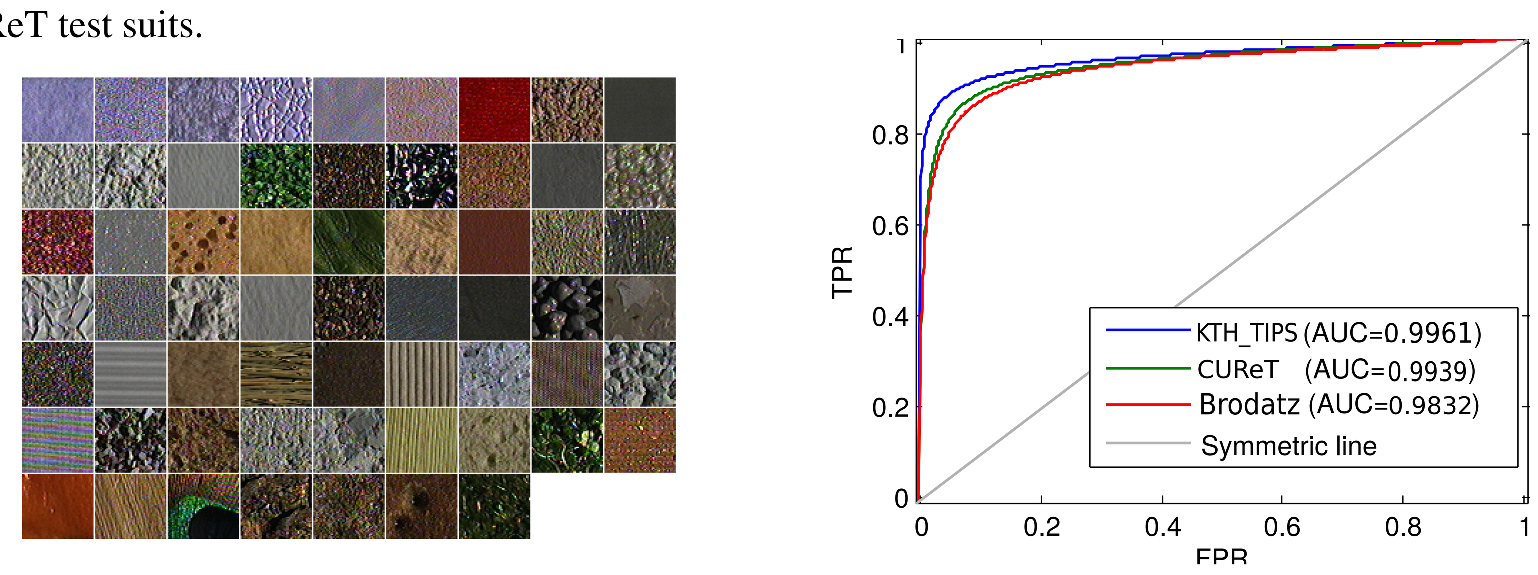

Figure 3: (a) Sample images of the ColumbiaUtrecht (CUReT) texture database having 92 samples/class labelled with 61 classes. (b) The area under the curve (AUC) indicates the probability that a model provides classification score between 0 and 1 ranks a randomly chosen true sample larger than a randomly chosen false sample.

Conclusions

- This work proposes a TexFusionNet model for the unconstrained texture classification.
- The TexFusionNet model is formed by fusing the pre-trained AlexNet and VGG16 CNN models at last representation layer.
- A fully connected layer is placed over the fusion layer to generate the class scores and the model is trained using cross-entropy loss function.
- Three benchmark Brodatz, CUReT and KTH-TIPS texture databases are used to judge the performance of proposed model.
- TexFusionNet model outperforms the hand-crafted methods and also provides a promising performance compared to other state-of-the-art deep CNN methods.

References

[1] Mihran Tuceryan, Anil K Jain, et al. Texture analysis. *Handbook of pattern recognition and computer vision*, 2:207–248, 1993.

[2] Vincent Andrearczyk and Paul F Whelan. Using filter banks in convolutional neural networks for texture classification. *Pattern Recognition Letters*, 84:63–69, 2016.

[3] Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Transactions on Knowledge & Data Engineering*, (10):1345–1359, 2009.

[4] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *NIPS*, pages 1097–1105, 2012.

[5] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[6] Mircea Cimpoi, Subhransu Maji, and Andrea Vedaldi. Deep filter banks for texture recognition and segmentation. In *CVPR*, pages 3828–3836, 2015.

[7] Swalpa K Roy, Bhabatosh Chanda, B Chaudhuri, Dipak K Ghosh, and Shiv Ram Dubey. A complete dual-cross pattern for unconstrained texture classification. In *4th Asian Conference on Pattern Recognition (ACPR 2017), Nanjing, China*, pages 741–746, 2017.

[8] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *CVPR*, pages 248–255. IEEE, 2009.