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 achieve the good performance using CNN.
• Pre-trained models [3] are widely adapted for simplicity and efficient 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 \( m \times n \), \( \alpha \) is a function representing the combination of convolutional layers, max-pooling layers and fully connected layers from \( I \) to last representation layer of AlexNet model, and similarly \( \varphi_{vgg} \) is a function for VGG16 model. So, the input image \( I \) is converted into \( \alpha(I) \) and \( \varphi_{vgg}(I) \), having the resolution of \( 227 \times 227 \) and \( 224 \times 224 \) for AlexNet [4] and VGG16 [5] respectively as follows,

\[
\alpha(I) = \tau(I, \beta) \circ f(I, \beta),
\]

where \( \alpha(I) \) is the transformation function to resize any image \( I \) into the resolution of \( \beta \times \beta \).

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

\[
\varphi_{alex}(I) = \alpha(I, \beta) = \tau(I, \beta) \circ f(I, \beta),
\]

\[
\varphi_{vgg}(I) = \varphi_{vgg}(I, \beta) = \tau(I, \beta) \circ f(I, \beta),
\]

where the dimensions of \( \varphi_{alex} \) and \( \varphi_{vgg} \) representation features are \( D_{alex} \) and \( D_{vgg} \) respectively as \( D_{alex} = D_{vgg} \). In the proposed work, the \( \varphi_{alex} \) and \( \varphi_{vgg} \) texture features of input image \( I \) using AlexNet and VGG16 models are fused by addition to produce a combined texture feature representation denoted as \( \varphi_{fused} \) and defined as follows,

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

where \( D \) is the dimension of the fused feature \( \varphi_{fused} \) with \( D = D_{alex} = D_{vgg} \).

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 and KTH-TIPS datasets 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.

![Figure 1: Proposed TexFusionNet architecture for texture classification.](image)

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

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


