2018 24th International Conference on Pattern Recognition (ICPR)
Beijing, China, August 20-24, 2018

Texture Segmentation using Siamese Network and Hierarchical Region Merging

Ryusuke Yamada, Hidenori Ide, Novanto Yudistira, and Takio Kurita
Department of Information Engineering
Graduate School of Engineering
Hiroshima University

***Abstract*—This paper proposes an texture segmentation algorithm. In the proposed texture segmentation algorithm, the feature vectors at each pixel of an input image are extracted by using the deep neural networks such as the deep convolutional network (CNN) or the Siamese Network. Then they are used as input of the hierarchical region merging. Unlike the semantic segmentation such as fully connected network (FCN) or U-Net which are based on the supervised learning, the proposed algorithm can correctly segment the texture regions whose texture is taken from the other types of the texture. The effectiveness of the proposed texture segmentation algorithm is experimentally confirmed by using the famous texture images taken from book by P.Brodatz.**

I. Introduction

Image segmentation is one of the oldest and most widely studied problems in computer vision [1]. The task of image segmentation is to find groups of similar pixels. Image segmentation in pathology involves the partitioning the image into histological primitives or biologically related tissue or cellular [2]. This allows the extraction and quantification of tissue or cell-specific features that can be used for diagnosis or prognosis. In statistics, this problem is known as cluster analysis. In image segmentation, the constraint between pixels must be taken into account because only adjacent regions can be merged into a group.

Brice and Fennema used a dual grid for representing boundaries between pixels and merged regions based on their relative boundary lengths and the strength of the visible edges at the boundaries [3]. In the simplest version of pixel-based region merging, adjacent regions are combined when the average color difference is below a given threshold [4]. Regions obtained by this method are known as ”superpixels” which does not have semantically meaning but they can be useful for pre-processing of stereo matching and optical flow.

While many region merging algorithms simply apply a fixed rule that groups pixels and regions together, Kurita proposed an efficient agglomerative clustering algorithm for region merging [5]. In this algorithm, the constraint between the regions is represented as the adjacency graph and only best adjacent pair is merged in the each step of hierarchical clustering. To speed up the search of the best pair of regions which is merged into one region, dissimilarity values of all possible pairs of regions are stored in a heap. Then the best pair can be found as the element of the root node of the binary tree corresponding to the heap. The constraint of neighboring relations of the regions is maintained by using the sorted linked lists. Then they reduced computational cost drastically for updating the dissimilarity values and neighboring relations which are influenced by the merging of the best pair. Similar approach in which adjacency graph is used to represent the constraint on the neighboring regions has been also proposed by Felzenszwalb and Huttenlocher [6].

Recently the convolutional neural network (CNN) becomes very popular for image classification and several algorithms for image segmentation which use the CNN have been already proposed in the literature. For example, Long et al. showed that semantic segmentation by fully convolutional networks (FCNs) trained for pixel-wise prediction can outperform the previous best results [7]. Noh et al. [8] also proposed a novel semantic segmentation algorithm by learning a deep deconvolution network. The deconvolution network which contains deconvolution and unpooling layers are learned on top of the convolution layers adopted from VGG 16-layer net. The pixel-wise class labels are identified by the trained deconvolution network. U-Net [9], which has achieved state of the art result on biomedical image segmentation, has skip connections over various resolutions of both down sampling and up sampling network making it robust to localization within context. It is not only accurate but also low computational cost which is fast on single GPU.

In these image segmentation algorithms, supervised learning is used. This means that there is a difficulty for segmentation of images taken from the other type of images. To overcome this drawback of the semantic segmentation, Fukushima et al. [10] proposed a image segmentation method which uses hypercolumn feature vectors extracted from the trained deep CNN [11] and performs hierarchical region merging based on the extracted hypercolumn feature vectors. The hypercolumn is a method to extract pixel-wise feature vector from the trained CNN and is defined at a given input location as the outputs of all units above that location from all layers of the CNN and stacked into one vector. In their experiments, the hypercolumn feature vectors are extracted from the trained AlexNet [12] and the dimension reduced feature vectors by Principal Component Analysis (PCA) [13] are used as inputs for the hierarchical region merging algorithm proposed by Kurita [5].

Texture is one of the important cues for image segmentation, especially for the applications to medical images or biological images. Texture regions in a image are generally defined as a statistical spatial distribution of pixel intensities [14]. There are many similar but different textures in real images. When the patterns of texture are the same but the density of the pattern is different at different scales and frequencies, the classification of such texture becomes difficult.

In many of the classical texture analysis, features are extracted by manually designed filters bank and they are used for clustering or classification [15], [16], [17], [18], [19]. A co-occurrence matrix proposed by Haralick et al. [14] and local spectral histograms proposed by Liu et al. [15] are examples of such handcrafted features. The CNN is also suitable for the texture analysis [20], [21]. We can use the CNN as a feature extractor in which the filters bank are trained with a powerful learning algorithm. For example, Siamese Network has two CNNs as subnetworks and can extract the dimension reduced discriminative feature vector. The weights of the two CNNs are shared. It was successfully applied to face verification, face recognition, and face clustering [22], [23]. Similarly Triplet Network which consists of three subnetworks was proposed as the extension of the Siamese Network [24].

In this paper we propose a texture segmentation algorithm in which feature vectors at each pixel are extracted by using the Siamese Network and hierarchical region merging is performed based on the extracted feature vectors. The effectiveness of the proposed texture segmentation algorithm is experimentally confirmed by using the famous texture images taken from book by P.Brodatz [25].

This paper is organized as follows. Section II explains related works such as the Siamese Network and the hierarchical region merging algorithm proposed by Kurita [5]. Section III explains the proposed texture segmentation algorithms. Section IV shows the experiments and results. Section V concludes this paper.

II. Related Works

A. Siamese Network

Siamese Network is a deep network model with two deep CNNs as subnetworks and is used to extract the dimension reduced discriminative feature vectors. Chopra et al. applied the Siamese Network for face verification [22]. The Siamese Networks was also applied by Schro et al. for face recognition and face clustering [23]. The goal of the Siamese Network is to construct a mapping from a given input image to the dimension reduced feature vector such that the distances between a pair of images with the same label are close while they are far for a pair of image with the different labels. To train the parameters of the network by using pairs of the training samples, two deep CNNs with the same weights as shown in the middle of the Figure 1 are used in the training phase. The objective function for optimization to estimate the parameters of the Siamese Network is defined by using the outputs of the two subnetworks. The standard loss function of the Siamese Network is the contrastive loss [22].

Consider a pair of input images p_1 and p_2. Let $\boldsymbol{f}(p_1)$ and $\boldsymbol{f}(p_2)$ be the output vectors of the two subnetworks for the pair of the input images. The Euclidean distance between the output vectors of the two subnetwork of the Siamese Network for the pair of the input images is given as

$$D(\boldsymbol{f}(p_1), \boldsymbol{f}(p_2)) = ||\boldsymbol{f}(p_1) - \boldsymbol{f}(p_2)||_2. \quad (1)$$

Then the contrastive loss is defined for the pair of the input images as

$$L(p_1, p_2) = \alpha(1-y)D(\boldsymbol{f}(p_1), \boldsymbol{f}(p_2))^2 + \beta y \max(0, m - D(\boldsymbol{f}(p_1), \boldsymbol{f}(p_2)))^2 \quad (2)$$

where y is a binary indicator that indicates whether two images belong to the same class or not. This loss function has a margin parameter m and two more tuning parameters α and β. In the learning process, this objective function is optimized to obtain the optimum weights of the subnetworks by using the error back-propagation learning algorithm.

B. Hierarchical Region Merging

The main task of image segmentation is to partition a digital image into multiple groups of similar pixels. In the merging process of image segmentation, only adjacent pairs of regions can be merged with the region adjacency constraints. Brice and Fennema proposed an image segmentation method in which a dual grid for representing boundaries between pixels is used and adjacent regions are merged based on their relative boundary lengths and the strength of the visible edges at the boundaries [26].

In the simplest version of pixel-based region merging, adjacent regions are merged into one region when the average color difference is below a given threshold [27].

Kurita proposed an efficient agglomerative clustering algorithm for region merging [5] . In his algorithm, the constraints between the regions are represented as the region adjacency graph (RAG) and only the best adjacent pair is merged in the each step of hierarchical clustering.

To speed up the search of the best pair of regions, dissimilarity values of all possible pairs of regions are stored in a heap. Then the best pair can be found as the element of the root node of the binary tree corresponding to the heap. The constraints on the each pair of neighbouring regions are maintained by using the sorted linked lists. Then we can drastically reduce the computation to update the dissimilarity values and the neighboring relations which are influenced by the merging of the best pair.

An efficient agglomerative clustering algorithm for region merging starts with the initial N regions in which each region includes only one pixel. Then the total number of regions is gradually reduced by iteratively merging the best pair among all possible pairs of regions in terms of a given criterion. This merging process is repeated until the required number of regions is obtained.

This algorithm is able to apply not only image but also feature vector. This means that this algorithm can perform the clustering of feature vector from deep neural network.

III. Proposed Method

A. Outline of the Proposed Texture Segmentation Algorithm

In the proposed texture segmentation algorithm, feature vectors at each pixel of the input image are extracted by the deep neural network such as the deep CNN or the Siamese Network. Each pixel of the input image is corresponded to each rectangle region of given images. This means that we can classify each pixel if each rectangle region of given images

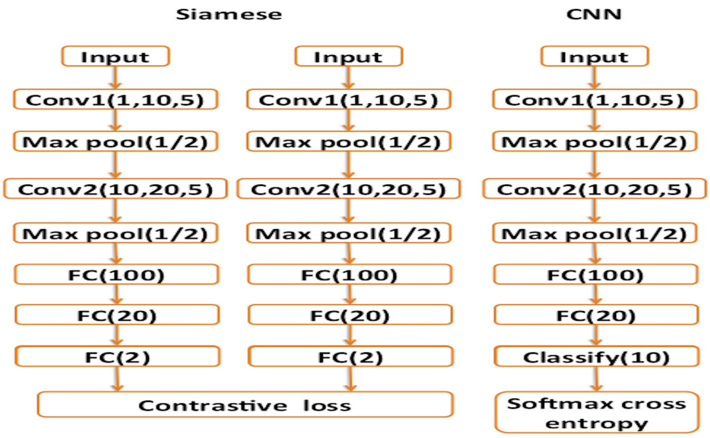

Fig. 1. The network architecture considered in this paper. Left: the Siamese Network. Right: the standard CNN.

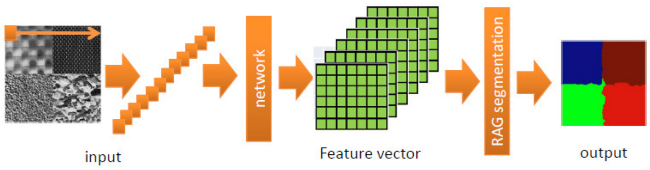

Fig. 2. Flow diagram of the proposed hierarchical region merging algorithm based on the feature vectors extracted by the deep neural network. At first the feature vectors at each pixel of the input image are extracted. Then they are used as the input of the hierarchical region merging.

is classified. Deep CNN has classifier and it is trained to classify such rectangle region with training texture class labels in the supervised learning. The Siamese Network is trained to generate feature map with the labels that means input pair paths are same training texture class or not. Then the hierarchical region merging algorithm explained in the section II-B is applied for texture segmentation based on the feature vectors of each pixel from trained network.

Figure 2 shows the flow diagram of the proposed texture segmentation algorithm. This algorithm is designed as an extension of the algorithm proposed by Fukushima et al. [10].

B. *Feature Extraction by the Deep Neural Networks*

The neighboring rectangular local regions of each pixel of a given image are clipped from the image and they are used as the inputs of the deep neural network to extract the features vectors at each pixel. In this paper, we tried two deep neural networks to extract the feature vectors. They are the deep CNN or the Siamese Network.

To train the deep neural networks, many small patches are clipped from 10 different texture images. The 1st small patch (32×32) is extracted from the (32×32) rectangular local region of given image and the 2nd small path patch (32×32) is extracted from the (32×32) rectangular local region of 1 pixel neighboring 1st patch's region on the given image. Other small patches are extracted by the same scheme. Then the weights of the networks are trained by using these patches.

1) The Deep CNN: The network architecture of the deep CNN is shown in the right of Figure 1. The size of the input image of the network (rectangular local region on the given image) is a 32×32 pixel. The network is composed of two pairs of the convolution and the pooling layers, two fully-connected layers, and classification layer. In this network, the size of the convolution filter is 5×5 pixels and the size of the pool is 2×2 pixels. The number of convolution filters in the first layer is set to be 10 and the number convolution filters of second layer is set to be 20. Relu functuion $(\max(0, x))$ is used as activation function of neurons of each layer. The number of neurons of the first fully-connected layer is set to be 100 and the number of neurons of the second fully-connected layer is set to be 20. The soft-max function is used in the classification layer for classification. The soft-max cross entropy loss is used as the objective function of this network and the weights of the network is obtained by using the stochastic gradient descent (SGD) optimization algorithm. This network is trained to classify input image (each rectangular local region) at the training phase. After the training we can extract the feature vectors from the hidden layers at each pixel of the input image by feeding the neighboring rectangular local regions of each pixel to the trained deep CNN.

2) The Siamese Network: The network architecture of the Siamese Network is also shown in the middle of Figure 1. The size of the input images of this network (rectangular local region on the given image) is also a 32×32 pixel. Configuration of the two subnetworks is set to be the same as the deep CNN. The contrastive loss is used as the objective function for learning. Adam optimization [28] is used as the optimizer and the margin parameter of Adam is set to be 3. This network is trained to generate the feature vectors from the pairs of input images (rectangular local regions) at the training phase. The feature vectors at each pixel of the input image can be extracted by feeding the neighboring rectangular local regions of each pixel to one of the subnetworks of the trained Siamese Network.

C. *Hierarchical Region Merging*

After the feature vectors at all pixels in the given input image are extracted by using the deep CNN or the Siamese

Network, they are used as the input of hierarchical region merging.

At first, each pixel is set to be an initial region, and four neighboring pixels are regarded as adjacent regions. A region adjacency graph is created and dissimilarities between adjacent regions are calculated using the extracted feature vectors of each pixel. Then the hierarchical region merging algorithm described in the section II-B is applied to obtain the segmentation results.

In the following experiments we use the Ward's method [29] which minimizes the sum of squared errors to select the best pair of regions at the each step in the hierarchical region merging.

IV. EXPERIMENT

A. Dataset and Evaluation

To confirm the effectiveness of the proposed texture segmentation algorithm, we have performed experiments using texture images taken from book by P.Brodatz [25]. Examples of the textures are shown in Figure 3. From these texture images, we used 10 class of 10 images for the training and evaluating the parameters of network and hierarchical region merging as the training and test images. We called as the known texture and the rests are used as the test images (other texture). The other texture are taken from the other types of texture. The texture patterns of other texture are not included in known texture.

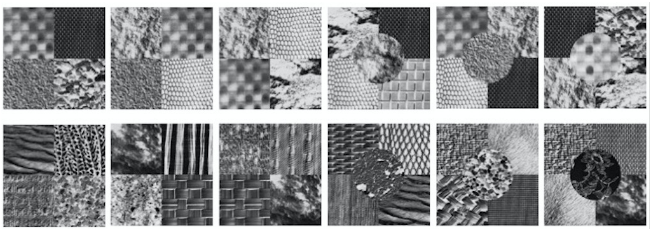

Fig. 3. Examples of evaluation images. Top images are created from the texture images in the training images while bottom images are created by using the textures in the test images.

To evaluate the segmentation results, we created texture images for evaluation in which several textures are pasted together as shown in Figure 3. Upper 6 big images in the Figure 3 were created from the textures in the training images. Here these are called known class images. On the other hand, lower 6 big images were created from the test images. These are called unknown class images. We used these texture images as ground truth to evaluate the correctness of the result of segmentation.

The deep CNN and the Siamese Network were trained by using the local 32×32 pixels regions in the training images of known texture. The deep CNN is trained to classify the local regions of training images as the supervised learning with the training images of the known texture. The Siamese Network was trained to generate feature vectors with the labels that means the pairs of input images (the training images of known texture) are same class or not. After training each networks, we extracted feature vector from hidden layer of the deep CNN or the outputs of the Siamese Network. Then we applied the hierarchical region merging to the feature vectors. The parameters of the hierarchical region merging algorithm were determined by the segmentation of the training images of known texture. And then we applied hierarchical region merging to the feature vectors of other texture images with the parameter obtained by the known texture.

To evaluate the texture segmentation results numerically, the ratio of the number of pixels which are correctly segmented to the number of pixels in the ground truth is calculated as the accuracy of the segmentation.

B. Comparison with other methods

The proposed segmentation algorithm is compared with the standard semantic segmentation using the deep CNN. For the semantic segmentation, the deep CNN was also trained by using the local 32×32 pixels regions in the training images of the known texture. In the proposed segmentation algorithm, the deep CNN was used to extract the feature vectors of each pixels. The deep CNN was trained with the training images of known texture. And then we applied the hierarchical region merging to the feature vectors of images (the test images of known texture and other texture images) from the hidden layer of the deep CNN. The results of the segmentation accuracies for the known textures and other textures are shown in Table I. Examples of the segmentation results are shown in Figure 4.

From the Table I, the accuracy of the proposed algorithm is about 20% better than the semantic segmentation especially for other textures. From the these results, it can be said that the proposed texture segmentation algorithm is effective not only for known texture but also for other textures.

TABLE I. ACCURACIES FOR THE KNOWN TEXTURE IMAGES AND THE OTHER TEXTURE IMAGES. CNN-RGA DENOTES THE PROPOSED TEXTURE SEGMENTATION ALGORITHM.

	method(Feature vector)	Average accuracy
known texture	semantic segmentation	88.04%
	CNN-RAG(FC20)	**90.97%**
other texture	semantic segmentation	57.96%
	CNN-RAG(FC20)	**77.11%**

Image to image estimation algorithm such as U-Net is also applicable to the semantic segmentation. Especially, the U-Net is one of standard algorithm for semantic segmentation. It gains attention since its successful application on biomedical image segmentation [9]. The architecture of U-Net is downsampling-upsampling autoencoder network with skip connection for every resolutions. It is known that the coarse to fine (multiple layers) analysis of U-Net is helpful to obtain clear segmentation for the known textures. In this experiment we use the implementation of U-Net that uses Generative Adversarial Network (GAN) [30] for regularization. This is the extension of U-Net. Examples of texture segmentation with U-Net for unknown texture images are shown in Figure 5. Figure 4 and 5 showed that these semantic segmentation algorithms do not work well for these other texture images and our proposed is better than these semantic segmentation algorithms.

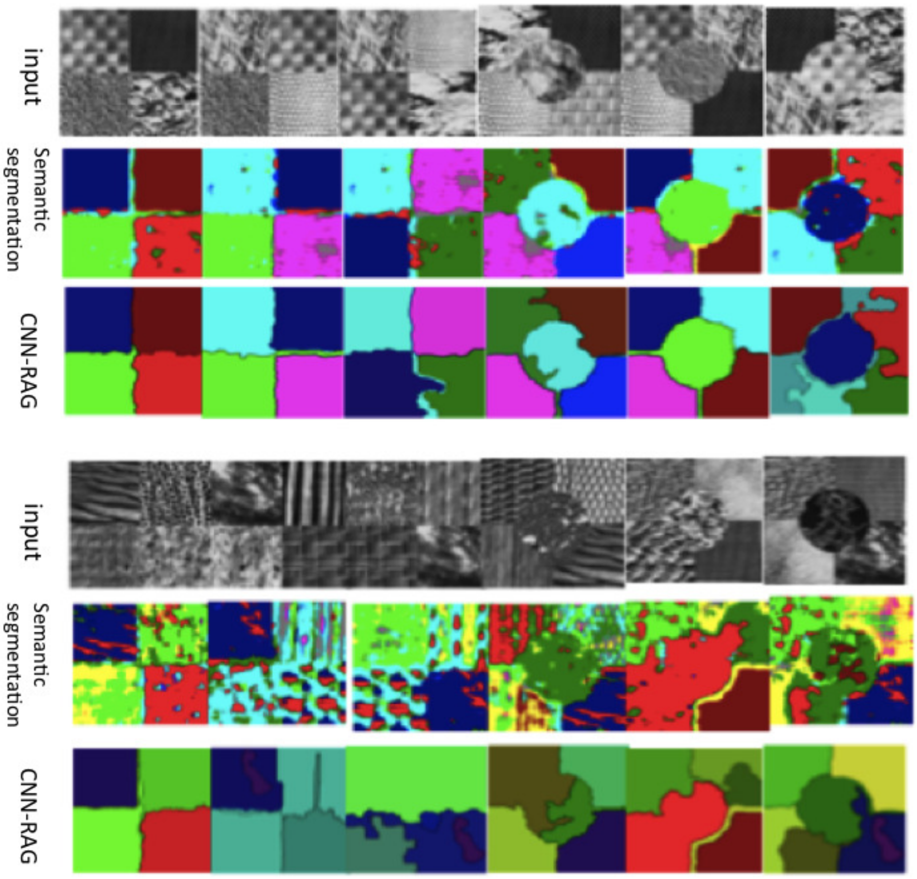

Fig. 4. Examples of the texture segmentation results. The upper half is the results for the known texture images and the lower half is the results for the other texture images. CNN-RAG denotes the proposed segmentation algorithm.

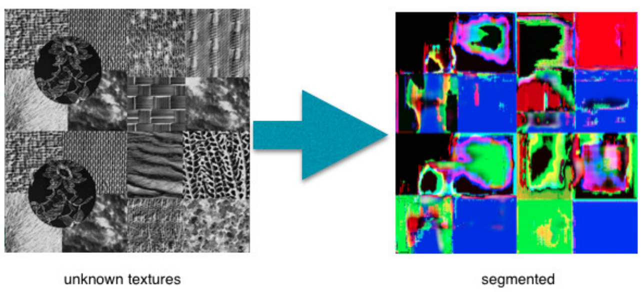

Fig. 5. Examples of texture segmentation by U-Net for unknown texture images.

Experimental results clearly show that U-Net gives high accuracy for region segmentation for learned images (known texture images), but good accuracy can not be obtained for the regions of other type of texture (other texture images).

C. Comparison of Feature Extraction Methods

Finally, we have performed the experiments to compare the feature extraction methods including the method which use hypercolumn to extract the feature vector from the trained deep CNN [10]. This method is similar to our methods. The difference is that we used the feature vector from the hidden layer or output layer of network while the hypercolumns extracted from thte trained network are used in this method. The accuracies of the texture segmentations are shown in Table II.

From Table II, the best accuracy is obtained for both of the known and other texture images when the two dimensional feature vectors are extracted by the Siamese Network denoted as Siamese-RAG (FC2). These results showed that hierarchical region merging should be applied to the feature vector from the output of Siamese Network.

V. CONCLUSION

This paper proposed texture segmentation algorithms which combine the feature extraction by using the deep neural networks such as the deep CNN or the Siamese Network and the hierarchical region merging with regions adjacency graph

TABLE II. ACCURACIES OF SEGMENTATION WITH DIFFERENT FEATURE EXTRACTION METHODS.

	method(Feature vector)	Average accuracy
known texture	Hyper-RAG (conv1+conv2)	75.64%
	Hyper-RAG (conv2)	83.50%
	CNN-RAG (clasify10)	90.27%
	CNN-RAG (FC20)	90.97%
	Siamese-RAG (FC20)	91.52%
	Siamese-RAG (FC2)	**93.88%**
other texture	Hyper-RAG (conv1+conv2)	65.63%
	Hyper-RAG (conv2)	65.65%
	CNN-RAG (clasify10)	76.33%
	CNN-RAG (FC20)	77.11%
	Siamese-RAG (FC20)	66.56%
	Siamese-RAG (FC2)	**81.68%**

constraints. The effectiveness of the proposed algorithms, especially for the other textures are experimentally shown by using the famous texture images of taken from book by P.Brodatz. Especially, hierarchical region merging should be applied to the feature vector from the output layer of the Siamese Network.

As future works, we would like to apply the proposed algorithm for the segmentation of the medical images and the bioimages.

ACKNOWLEDGMENT

This work was partly supported by JSPS KAKENHI Grant Number 16K00239 and 16H01430.

REFERENCES

[1] R. Szeliski, "Computer Vision – Algorithms and Applications –", Springer, 2011.

[2] S.Kevin Zhou, Hayit Greenspan, and Dinggang Shuen, *Deep Learning for Medical Image Analysis*, Academic Press, 2017.

[3] C. R. Brice and C. L. Fennema, "Scene analysis using regions," Artificial Intelligence, Vol.1, No.3-4, pp.205-226, 1970.

[4] G. Mori, X. Ren, A. Efros, and J. Malik, "Recovering human body configurations: Combining segmentation and recognition," Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR2004), pp.326-333, 2004.

[5] T. Kurita, "An efficient clustering algorithm for region merging," . IEICE Trans. on Information and Systems, Vol.78-D, No.12, pp.1546-1551, 1995.

[6] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient graph-based image segmentation," International Journal of Computer Vision, Vol.59, No.2, pp.167-181, 2004.

[7] J. Long, E. Shelhame and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR2015), pp.3431-3440, 2015.

[8] H. Noh, S. Hong and B. Han, "Learning Deconvolution Network for Semantic Segmentation," Proc. of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR2016), pp.1520-1528, 2016.

[9] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. InInternational Conference on Medical Image Computing and Computer-Assisted Intervention 2015 Oct 5 (pp. 234-241). Springer, Cham.

[10] Y.Fukushima,T.Kurita,Hierarchical Region Merging with Region Adjacency Graph Constraints using Hypercolumns extracted from the trained CNN(FCV2017),2017.

[11] B. Hariharan, P. Arbelez, R. Girshick, and J. Malik, Hypercolumns for object segmentation and finegrained localization, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 447-456, 2015.

[12] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks, Proc. of he Neural Information Processing Systems Conference (NIPS2012), 2012.

[13] Takio Kurita, Principal Component Analysis (PCA), Computer Vision, A Reference Guide, pp.636-639, 2014.

[14] R. M. Haralick, K. Shanmugam,et al., Textural features for image classification, IEEE Transactions on systems, man, and cybernetics, vol. 3, no. 6, pp. 610-621, 1973.

[15] X. Liu and D. Wang, Image and texture segmentation using local spectral histograms, IEEE Transactions on Image Processing, vol. 15, no. 10, pp. 3066-3077, 2006.

[16] N. Mevenkamp and B. Berkels, Variational multi-phase segmentation using high-dimensional local features, in 2016 IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 19, IEEE,2016.

[17] J. Yuan, D. Wang, and A. M. Cheriyadat, Factorization-based texture segmentation, IEEE Transactions on Image Processing, vol. 24, no. 11, pp. 3488-3497, 2015.

[18] J. Shi and J. Malik, Normalized cuts and image segmentation, IEEE Transactions on pattern analysis and machine intelligence, vol. 22, no. 8, pp. 888-905, 2000.

[19] N. Paragios and R. Deriche, Geodesic active regions and level set methods for supervised texture segmentation, International Journal of Computer Vision, vol. 46, no. 3, pp. 223-247, 2002.

[20] V. Andrearczyk and P. F. Whelan, Using Filter Banks in Convolutional Neural Networks for Texture Classification, Pattern Recognition Letters , vol. 84, pp. 63-69, 2016.

[21] M. Cimpoi, S. Maji, I. Kokkinos, and A. Vedaldi, Deep filter banks for texture recognition, description, and segmentation, International Journal of Computer Vision , vol. 118, no. 1, pp. 65-94, 2016.

[22] S. Chopra, R. Hadsell, Y. LeCun, Learning a similarity metric discriminatively, with application to face verification, (CVPR2005), 2005, pp. 539-546.

[23] F. Schro, D. Kalenichenko, J. Philbin, Facenet: A unified embedding for face recognition and clustering, in: CVPR, 2015, pp. 815-823.

[24] J.Wang, Y. Song, T. Leung, C. Rosenberg, J.Wang, J. Philbin, B. Chen, and Y. Wu, Learning fine-grained image similarity with deep ranking, in Proc. of IEEE Computer Vision and Pattern Recognition, 2014, pp. 1386-1393.

[25] P.Brodaz, "Textures," Dover,1966.

[26] C. R. Brice and C. L. Fennema, Scene analysis using regions, ArtificialIntelligence, Vol.1, No.3-4, pp.205-226, 1970.

[27] G. Mori, X. Ren, A. Efros, and J. Malik, Recovering human body configurations: Combining segmentation and recognition, Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR2004), pp.326-333, 2004.

[28] D. P. Kingma, and J. Ba, "Adam: A method for stochastic optimization," International Conference for Learning Representations (ICLR2015), 2015.

[29] J. H., Jr. Ward, Hierarchical grouping to optimize an objective function, Journal of the American Statistical Association, Vol.58, pp.236-244, 1963.

[30] Isola P, Zhu JY, Zhou T, Efros AA. Image-to-image translation with conditional adversarial networks. arXiv preprint arXiv:1611.07004. 2016 Nov 21.