西南交通大学学报

第 54 卷 第 5 期
2019 年 10 月

JOURNAL OF SOUTHWEST JIAOTONG UNIVERSITY

Vol. 54 No. 5
Oct. 2019

ISSN -0258-2724

DOI：10.35741/issn.0258-2724.54.5.23

Research article

REAL-TIME COLOR IMAGE CLASSIFICATION BASED ON DEEP LEARNING NETWORK

基于深度学习网络的实时彩色图像分类

Mohammed Hamzah Abed [1]*, Atheer Hadi Issa Al-Rammahi [2], Mustafa Jawad Radif [2]

[1] Computer Science Department, College of Computer Science and IT, University of Al-Qadisiyah, Al-Diwaniah, Iraq, mohammed.abed@qu.edu.iq
[2] Department of Computer Information System, College of Computer Science and IT, University of Al-Qadisiyah, Al-Diwaniah, Iraq, atheer.alrammahi@qu.edu.iq, mustafa.radif@qu.edu.iq

Abstract

Real-time image classification is one of the most challenging issues in understanding images and computer vision domain. Deep learning methods, especially Convolutional Neural Network (CNN), has increased and improved the performance of image processing and understanding. The performance of real-time image classification based on deep learning achieves good results because the training style, and features that are used and extracted from the input image. This work proposes an interesting model for real-time image classification architecture based on deep learning with fully connected layers to extract proper features. The classification is based on the hybrid GoogleNet pre-trained model. The datasets that are used in this work are 15 scene and UC Merced Land-Use datasets, used to test the proposed model. The proposed model achieved 92.4 and 98.8 as a higher accuracy.

Keywords: Real-Time Image, Classification, Deep Learning, Convolutional Neural Network.

摘要 在理解图像和计算机视觉领域中，实时图像分类是最具挑战性的问题之一。 深度学习方法，尤其是卷积神经网络（有线电视新闻网），已经增加并改善了图像处理和理解的性能。 由于训练风格以及从输入图像中使用和提取的特征，基于深度学习的实时图像分类的性能达到了良好的效果。 这项工作提出了一个有趣的实时图像分类架构模型，该模型基于具有完全连接的层以提取适当特征的深度学习。 分类基于混合的谷歌网预训练模型。 在这项工作中使用的数据集是 15 个场景和加州大学默塞德土地使用数据集，用于测试提出的模型。 提出的模型获得了 92.4 和 98.8 更高的精度。

关键词: 实时图像，分类，深度学习，卷积神经网络。

I. INTRODUCTION

In the last few decades, the number of users that use smart technology and intelligent devices have increased in mad fashion. This is due to the Internet in general and smart technologies like smart phones, smart watches and sensors that deal with daily challenges and are available at a cheap price [1]. Therefore, the number of raw data that is transferred over the Internet becomes huge. During the long research of machine learning, one of the grand challenges is image classification, which is the ability to classify labelled images into different groups, for example house, car and animals. Image classification is an approach of machine learning and computer vision that refer to the task of extracting features and information from each class of color images to be able to

classify other groups depending on features that are extracted. Actually, there is two types of classification: supervised and unsupervised. Recently, classification based on deep learning model has drawn significant attention [2]. There are various deep learning architectures, such as deep neural network and deep convolutional neural network, these are applied to a wide range of fields like machine learning, computer vision, and automatic image classification [3]. The basic idea of deep learning is to discover and extract the features from input images in multiple levels of representation and combine it together to make classification decision based on those features. Deep learning (convolutional neural network) gives a promising result to be used as a tool for features extraction and classification [4], [18]. Convolutional neural network architectures have used raw images as inputs, which allows the encoding of specific properties into the architecture [2]. Mainly the structure of CNN is a series of layers including a convolutional layer, a pooling layer and full connection layers, connected to each other. To be specific, a Convolutional neural network is an advance of neural network that contains one or more Convolutional layers that are responsible for extracting low features like corner edge lines in the image. Pooling layers makes the features more robust against distortion and noise. Then the third part in CNN is to produce fully connected layers that focus on the mathematical operation in order to find the summation of the weight of the previous layer of features [3].

In this study, we will use a hybrid googleNet architecture to improve the method for real-time color image classification to compare with a pre-trained convolutional neural network, AlexNet. We will test both on two different datasets: 1.) a UCMD dataset, and 2.) a 15-scene image dataset.

This paper is arranged according to the following structure. Section II discusses related work and contains a literature survey of image classification based on CNN. Section III provides an overview of the proposed method and its components. Section IV discusses the experimental result on datasets used for testing the proposed method. Section V contains the author's conclusions.

II. Related Work

During the recent decades, image processing applications and image classification based on deep learning have become an important research topic. In this part, a short description of the most widely utilized image classification techniques will be provided. In [4], M. A. Kadhim and M. H. Abed proposed a method based on pre-trained Convolutional Neural Network (CNN) for satellite image classification,by selecting the features from a fully connected layer comprising of AlexNet, GoogleNet, Resnet50 and VGG_19. They achieved 98 % accuracy when their algorithm was applied on the Merced Land dataset, as well as 95.8 % and 94.1 for SAT4 and SAT6. On the other hand, Yunlong Yu and Fuxian Liu investigated aerial scene classification, they suggest architecture model based on the experiments by using three publicly available remote sensing scene datasets—UC-Merced dataset (UCMD), Aerial Image Dataset (AID) and NWPU-RESISC45 dataset. The proposed architecture relied on the concept of feature fusion at different levels. The first approach was based on texture-coded two streams used raw RGB and planned to extract two types of features based on local binary patterns LBP. In the second proposed method, two-stream architecture are suggested to employed the saliency coded network stream as the second stream and used it with the raw RGB network stream using the same feature fusion model. The highest accuracy achieved by these authors was 98.90 % [3]. Yishu Liu and Chao Huang proposed Triplet Network for scene image classification, which uses weakly labelled images as the input. They also studied the theoretical background of the three existing loss functions for triplet networks, which allowed them to compare and analyze different underlying mechanisms for dealing with “easy” and/or “hard” triplets during training. In addition, the authors constructed four new loss functions,for triplets network to increase classification accuracy. In this work, the highest accuracy of 97.99 % was obtained when the algorithm was applied to the UCMD dataset [5]. Junhao Zhang and colleagues proposed a novel method for mutual information schema for image registration in remote sensing based on features map technique. These authors developed a new schema for remote sensing image registration between different channels to improve matching accuracy [6].

III. Proposed Architecture Model

The feature extraction and classification of the real color image methods proposed in the present study are illustrated in Figure 1. The proposed hybrid model is based on GoogleNet architecture and can be adopted for image analysis and classification because image data is represented as a set of feature vectors which can be presented

and process. The description of individual hybrid model components is given below.

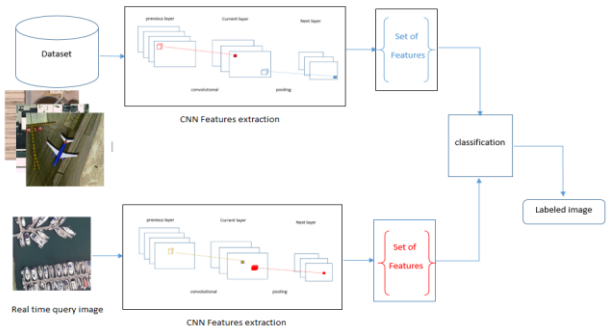

Figure 1. Proposed work architecture.

A. CNN Based on Deep Features Extraction

In this section, a general overview of the CNN architecture and configuration layers comprising the hybrid GoogleNet proposed model for real-time image classification is suggested.

B. CNN Architecture

Convolutional Neural Network is one type of feed-forward artificial neural network [4]. Its multi-layer structure consists of three parts, namely convolutional layers, pooling layers, and fully connected layers. The structure of typically CNN consists of many alternating convolutional and pooling layers, endedu87 with many fully connected layers [7]. Basically, each convolutional layer have its own weights a cross input space and the output. In general, the subset of data extracted from the previous layer will be used as an input for the current layer [8], [9]. Same as with a pooling layer conduct the behaviour of convolutional to minimize the set of output features.To avoid the complexity of entire images, the insertion of pooling layers is recommended [6], [10]. The features of the entire color image produced by convolutional and pooling layers can be considered a local feature. Therefore, the use of some of the fully connected layers is suggested to find the global features that depend on entire data. The fully connected layers are completely connected to all the previous layers, including convolutional and pooling layers, in a hierarchical fashion. This hierarchy allows the CNN model to extract the unique features from lower to higher layers.

IV. Experimental Results and Analysis

To improve the accuracy of the proposed model, it is applied for color image classification using a 15-scene image dataset, as it is challenging and the most widely used dataset. The experiments are also carried out for remote sensing image scene classification, using the familiar UC Merced Land Use dataset. For both datasets, the proposed model followed the same experimental steps to create the proposed image representations. This section provides a general overview of the experimental results and image dataset used to evaluate the accuracy of the proposed model's classification.

A. Datasets

In the proposed study, the experimental steps will be applied on two different datasets: the 15-scene image dataset and UC Merced Land Use dataset.

1) 15-Scene Image Dataset

A 15-scene image dataset is the first dataset employed in the experiments. This famous dataset contains 15 scenes comprises two categories: indoor and outdoor. Figure 2 shows the selected examples from each category. Initially, the dataset contained eight classes contributed by Oliva and Torralba [11]. Then five classes were added by Fei-Fei and Perona [12]. Finally, two additional categories were introduced by Lazebnik et al. [13]. The dataset of each class is divided into two categories: 70% training and 30% testing and validation. Table 1 shows the class number for each category, the number of images and the size of each image.

Table 1.
15 scene dataset description

Class number	Label of image	Number of images	The dimension of each image
00	Bedroom	216	267×200
01	Calsubrub	241	330×220
02	Industrial	311	345×220
03	Kitchen	210	293×220
04	Living room	289	291×220
05	MITcoast	360	256×256
06	MITforest	328	256×256
07	MIThighway	260	256×256
08	MITinsidecity	308	256×256
09	MITmountain	374	256×256
10	MITopen country	0	256×256
11	MITstreet	292	256×256
12	MITtallbuilding	356	256×256
13	PARoffice	215	352×220
14	Store	315	277×220

Figure 2. Samples of 15 scene dataset.

2) UC Merced Land-Use (UCM) Image Dataset

The second dataset in our experiments is a UCM dataset. It consists of 21 categories of land-use image data; each class contains 100 images of 256×256 dimensions. The images are extracted manually from a large dataset of images, the USGS National Map Urban Area Imagery collection. Figure (3) below shows the selected samples of the images from 21 classes [14].

Figure 3. Selected samples of UCM dataset.

V. Result and Discussion

In order to evaluate the proposed architecture for real-time color image classification, we had to comprehensively analyze and study the results for two different datasets, using different strategies for different situations. During the training phase of the 15 scene dataset and the UC Merced Land-Use dataset, the data was divided into two parts, the first 50 % for training and remaining 50 % for testing and checking results. We then increased the percentage of training to 70 % for training, and decreased the testing ratio to 30% to achieve a second set of results. Table 1 shows the accuracy results for 15 scenes using different methods, and table 2 explains the comparison to the UMD dataset in different cases.

Table 2.
Accuracy result of 15 scene dataset

Methods	Accuracy
BoVW [15]	84.11 %
CWCH [15]	88.04 %
Proposed model	92.4

Table 3.
Accuracy result of UC Merced Land - Use dataset

Methods	Training ratio	
	50 %	70 %
M. A. Kadhim and M. H. Abed [4]	95	98
Yunlong Yu and Fuxian Liu [3]	97.79	-
VGG-16-CapsNet [6]	95.33	-
Inception-v3-CapsNet [6]	97.59	-
Proposed system	97	98.8

Figures 4 and 5 show the initial accuracy and loss functions of 15 respective scene datasets with 6 Epoch and 882 iterations. The figures and tables show the accuracy for both UC Merced Land-Use and 15 scene datasets. The performance of the proposed model of the 15 scene dataset had a successful performance with more than two different strategies as shown in table 1 (92.4 % accuracy). Also, the proposed model achieved higher accuracy (98.8 %) in case of 70% training, and 97 % in the 50 % training of the dataset. The focus of the proposed model is the kind of features that can be extracted from the input images that are generated in fully connected layers of the proposed convolutional neural network.

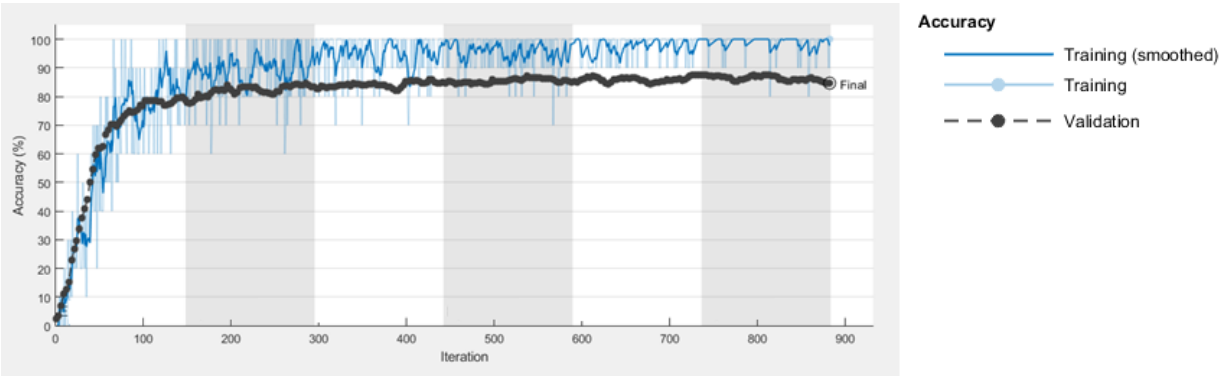

Figure 4. Accuracy of 15 scene dataset.

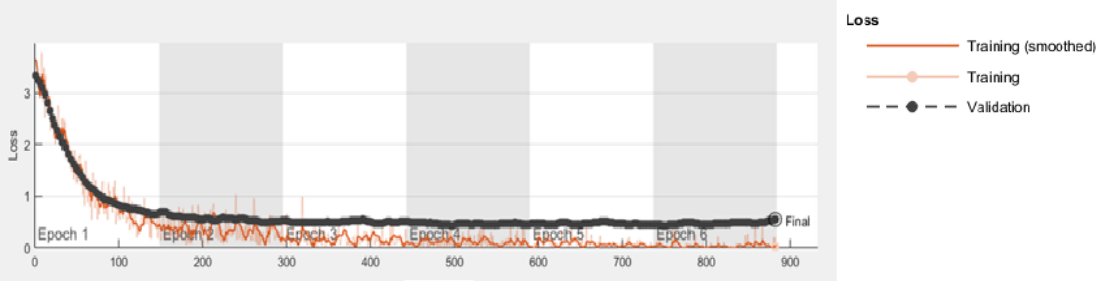

Figure 5. Loss of 15 scene dataset.

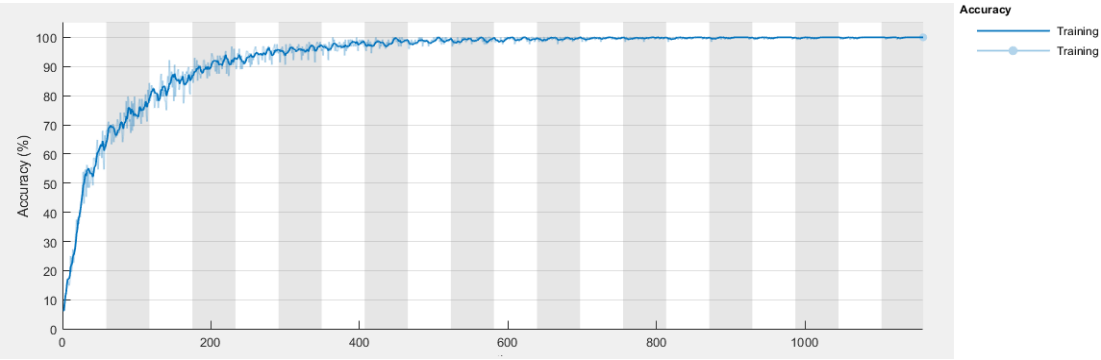

Figure 6. Accuracy of UC Merced Land-Use.

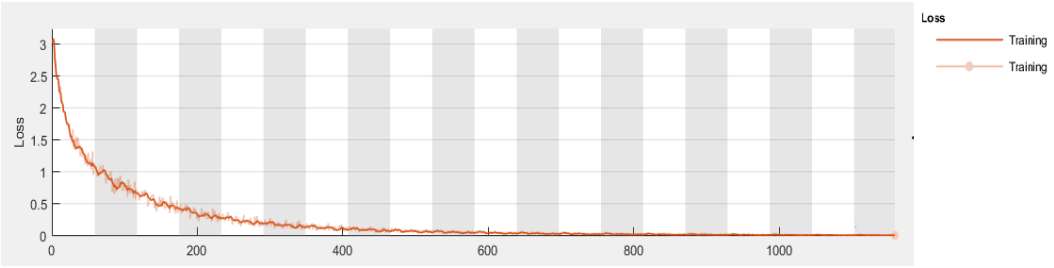

Figure 7. Loss of UC Merced Land-Use.

As figures and tables shows the accuracy of both dataset UC Merced Land-Use and 15 scene shows the performance of proposed model of 15 scene have success performance more than two different strategies as shown in table 1 is 92.4 % accuracy also the proposed model achieved higher accuracy 98.8 % in case of 70 % training and 97 % in 50 % training of dataset. In the proposed model the main point in this work is kind of features that extracted from the input images that generated in fully connected layers of the proposed convolutional neural network compare with traditional techniques for retrieval or classification [16], [17].

VI. CONCLUSION

In recent years, the number of research projects used to classify and retrieve queries from bug datasets has increased. Also in last few years, the numbers of researchers using image processing based on deep learning, especially CNN, have increased because of the good results that are being achieved. In this paper, we suggest methods based on deep learning techniques for classifying real-time images. These methods are based on the concept of features extraction from fully connected layers. The method has been tested on two different datasets, and resulted in good accuracy in comparison with results from other research. The experimental result shows this proposed architecture model achieved both a good result and better classification performance than the published results of M. A. Kadhim and M. H. Abed [4] with both 50% and 70% training ratios. In future work, this proposed classification model based on merges of more than two pre-trained CNN is expected to increase the efficiency and achieve high classification ratios of color images.

REFERENCES

[1] ACHARYA, D., YAN, W., and KHOSHELHAM, K. (2018) Real-time image-based parking occupancy detection using deep learning. *Proceedings of the 5th Annual Conference of Research@Locate*, 2087, pp. 33–40.

[2] SHAMSOLMOALI, P., JAIN, D.K., ZAREAPOOR, M., YANG, J., and ALAM, M.A. (2019) High-dimensional multimedia classification using deep CNN and extended residual units. *Multimedia Tools and Applications*, 78(17), pp. 23867-23882.

[3] YU, Y. and LIU, F. (2018) Dense Connectivity Based Two-Stream Deep Feature Fusion Framework for Aerial Scene Classification. *Remote Sensing*, 10(7), 1158.

[4] KADHIM, M.A. and ABED, M.H. (2019) Convolutional Neural Network for Satellite Image Classification. *Proceedings of the Asian Conference on Intelligent Information and Database Systems: Recent Developments*, 830, pp. 165-178.

[5] LIU, Y. and HUANG, C. (2018) Scene Classification via Triplet Networks. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(1), pp. 220-237.

[6] SHAMSOLMOALI, P., ZAREAPOOR, M., and YANG, J. (2019) Convolutional neural network in network (CNNiN): hyperspectral image classification and dimensionality reduction. *IET Image Processing*, 13(2), pp. 246-253.

[7] KÖLSCH, A., AFZAL, M.Z., EBBECKE, M., and LIWICKI, M. (2017) Real-Time Document Image Classification Using Deep CNN and Extreme Learning Machines. *Proceedings of the 14th IAPR International Conference on Document Analysis and Recognition.*

[8] WANG, J., MA, Y., ZHANG, L., GAO, R.X., and WU, D. (2018) Deep learning for smart manufacturing: Methods and applications. *Journal of Manufacturing Systems*, 48, pp. 144-156.

[9] ZAREAPOOR, M., SHAMSOLMOALI, P., and YANG, J. (2019) Learning depth super-resolution by using multi-scale convolutional neural network. *Journal of Intelligent and Fuzzy Systems*, 36(2), pp. 1773-1783.

[10] XIN, M. and WANG, Y. (2019) Research on image classification model based on deep convolution neural network. *EURASIP Journal on Image and Video Processing*, 2019, 40.

[11] OLIVA, A. and TORRALBA, A. (2001) Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision*, 42(3), pp. 145–175.

[12] FEI-FEI, L. and PERONA, P. (2005) A Bayesian hierarchical model for learning natural scene categories. *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition.*

[13] LAZEBNIK, S., SCHMID, C., and PONCE, J. (2006) Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*.

[14] YANG, Y. and NEWSAM, S. (2010) Bag-of-visual-words and spatial extensions for land-use classification. *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pp. 270-279.

[15] ZAFAR, B., ASHRAF, R., ALI, N., AHMED, M., JABBAR, S., NASEER, K., AHMAD, A., and JEON, G. (2018) Intelligent image classification-based on spatial weighted histograms of concentric circles. *Computer Science and Information Systems*, 15(3), pp. 615-633.

[16] ZHANG, W., TANG, P., and ZHAO, L. (2019) Remote Sensing Image Scene Classification Using CNN-CapsNet. *Remote Sensing*, 11(5), 494.

[17] ABED, M.H. and AL-FARTTOOSI, D.S.J. (2015) Content based image retrieval based on histogram. *International Journal of Computer Applications*, 110(3), pp. 42-47.

[18] AL-AZZAWI, D.S. (2019) Human Age and Gender Prediction Using Deep Multi-Task Convolutional Neural Network. *Journal of Southwest Jiaotong University*, 54(4).

参考文

[1] ACHARYA，D.，YAN，W.和KHOSHELHAM，K.（2018）使用深

度学习的基于图像的实时停车占用检测。第五届研究@定位年度会议论文集，2087，第 33-40 页。

[2] SHAMSOLMOALI，P.，JAIN，D.K.，ZAREAPOOR，M.，YANG，J. 和 ALAM，M.A.（2019）使用深度有线电视新闻网和扩展残差单元的高维度多媒体分类。多媒体工具和应用，78（17），第 23867-23882 页。

[3] YU，Y. 和 LIU，F.（2018）基于密集连接的两流深度特征融合框架，用于空中场景分类。遥感，10（7），1158。

[4] KADHIM，M.A. 和 ABED，M.H.（2019）用于卫星图像分类的卷积神经网络。亚洲智能信息和数据库系统会议论文集：最新进展，830，第 165-178 页。

[5] LIU Y. 和 HUANG C.（2018）通过三胞胎网络进行场景分类。电气工程师学会在应用地球观测和遥感中精选主题期刊，11（1），第 220-237 页。

[6] SHAMSOLMOALI，P.，ZAREAPOOR，M. 和 YANG，J.（2019）网络中的卷积神经网络（中国镍网）：高光谱图像分类和降维。IET 图像处理，13（2），第 246-253 页。

[7] KÖLSCH，A.，AFZAL，M.Z.，EBBBECKE，M。和 LIWICKI，M。（2017）使用深度有线电视新闻网 和极限学习机进行实时文档图像分类。国际会计师协会第十四届国际文件分析与识别会议论文集。

[8] 王建军，马耀阳，张林峰，高荣兴，吴武东（2018）智能制造的深度学习：方法和应用。制造系统杂志，48，第 144-156 页。

[9] ZAREAPOOR，M.，SHAMSOLMOALI，P. 和 YANG，J.（2019）通过使用多尺度卷积神经网络学习深度超分辨率。智能与模糊系统杂志，36（2），第 1773-1783 页。

[10] 辛敏和王 Y（2019）基于深度卷积神经网络的图像分类模型研究。EURASIP 图像和视频处理期刊，2019，40。

[11] OLIVA，A. 和 TORRALBA，A.（2001）对场景的形状进行建模：空间包络的整体表示。国际计算机视觉杂志，42（3），第 145-175 页。

[12] FEI-FEI，L。和 PERONA，P。（2005）学习自然场景类别的贝叶斯分层模型。 2005 年电气工程师学会计算机学会计算机视觉和模式识别会议论文集。

[13] LAZEBNIK，S.，SCHMID，C. 和 PONCE，J.（2006）超越功能包：用于识别自然场景类别的空间金字塔匹配。 2006 年电气工程师学会计算机协会计算机视觉和模式识别会议论文集。

[14] YANG，Y. 和 NEWSAM，S.（2010）土地用途分类的视觉词袋和空间扩展。第 18 届 SIG 空间地理信息系统发展国际会议论文集，第 270-279 页。

[15] BAF 的 ZAFAR，R。的 ASHRAF，N。的 AHMED，M.，JABBAR 的 S.，NASEER，K.，AHMAD 的 A. 和 JEON，G.（2018 年）的智能图像分类-基于同心圆的空间加权直方图。计算机科学与信息系统，15（3），第 615-633 页。

[16] ZHANG，W.，TANG，P.，和 ZHAO，L.（2019）使用美国有线电视新闻网的遥感影像场景分类。遥感，11（5），494。

[17] ABED，M.H。 D.S.J. 和 AL-FARTTOOSI（2015）基于直方图的基于内容的图像检索。国际计算机应用杂志，110（3），第 42-47 页。

[18] AL-AZZAWI，D.S.（2019）使用深度多任务卷积神经网络的人类年龄和性别预测。西南交通大学学报，54（4）。