ISSN: 1533 - 9211

DOI: 10.5281/zenodo.6879594

REMOTE SENSING SCENE CATEGORISING USING CNN

U. Deepika

Assistant Professor, Department of Electronics and Communication Engineering, Sridevi Women's Engineering College, Hyderabad, India, deepika.upadrasta@gmail.com

G. Srishma

U.G Student, Department of Electronics and Communication Engineering, Sridevi Women's Engineering College, Hyderabad, India, srishmareddi57@gmail.com

K. Ruchitha Chary

U.G Student, Department of Electronics and Communication Engineering, Sridevi Women's Engineering College, Hyderabad, India, buduruchandana@gmail.com

B. Chandana

U.G Student, Department of Electronics and Communication Engineering, Sridevi Women's Engineering College, Hyderabad, India, ruchithachary2@gmail.com

ABSTRACT: The main objective of remote sensing image scene classification is to know semantics of land covers. As CNN algorithm is used for feature representation so CNN based classification methods has been proposed for RS images. The existing system is good to capture images, which gives global information but sometimes it fails to optimise the local features. To overcome the limitation we proposed a new CNN model, for getting the local features comprehensively. It is possible due to the dual-branch structure, the input data are the image pairs that are obtained by the spatial rotation.

Considering the influence of the spatial rotation and the similarities between RS images, we developed a model to unify the salient regions and impact the RS images from the same/different semantic categories. Finally, the classification results can be obtained using the learned features. The popular RS scene datasets are selected to validate our CNN. Compared with some existing networks, the proposed method can achieve better performance. The results obtained are effective for the RS image scene classification.

KEYWORDS: Remote sensing, Scene Classification, CNN

I. INTRODUCTION

As there is increase in number RS images taken from satellites. There will be important matter present in the images. In order to organise the data RS scene classification is important. RS image scene classification is popular in many practical applications, such as agriculture, hydrology and forestry. Existing system has a two-stage classification scheme in which the support vector machine (SVM) is used to generate probability images with different handcrafted features in the first stage and the generated probability images with different features are fused in the second stage to obtain the final classification results.

SEYBOLD Report
ISSN: 1533 - 9211
DOI: 10.5281/zenodo.6879594

Taking the features of RS images in consideration, the parallel model is developed to capture the local features from the spatial /aspects. Accompanying with the global knowledge obtained by a successful CNN, the discrimination of the final features can be improved a lot. To unify different kinds of attention maps and consider the resemblance between RS images for the scene classification, we design the attention consistent model. The next major advantage is, it can also avoid the negative effects caused by the differences between attention maps of image pairs. An extensive experiment is conducted on the dataset, and the encouraging results prove that our proposed system is effective for the RS image scene classification task.

II. RELATED WORK

RS scene classification is one of the important content to understanding and performing the tasks on the RS satellite images. In recent years, with the help of CNN, the performance of RS scene classification is enhanced. Here, we roughly divide the existing CNN-based classification methods into two groups according to their architecture. The structure of the classification networks is the single-branch. In other words, these networks have only one entrance. When the RS images are input the networks, they would be mapped into the feature vectors by some operation

An old and unique algorithm known as Ad boost algorithm was also used to RS image scene classification. This Algorithm uses decision stumps as a weak classifier and combines them for continuous features and categorical features so that it becomes strong classifier but this algorithm also failed due to their low computational rates and high false rates.

III. METHODOLOGY

A. Overall Framework

The input data is takes as the image **I** and the image T(**I**) obtained by the spatial rotation as the input. There are two reasons for this kind of input data. Spatial rotation is a common data method of giving input. After the rotation, the training of data can be increased two times with visual perceptual consistency, so that the efficiency of output gets increased. By taking these kind of inputs there is also an advantage that semantic classes can be dispersed.

The consistent model that we proposed contains four stages where in each stage feature extraction will happen. In First stage the input image pairs **I** and T(**I**) are passed feature extraction model for the feature maps **X** and **X',** here basic extraction will occur which contain basic semantics information. In Second stage our proposed model explores the complex contents within the RS images from the global and local aspects. Then, the convolution representation is obtained by global average pooling (GAP) on the concatenated feature maps, which come from second stage of the model.

SEYBOLD Report
ISSN: 1533 - 9211
DOI: 10.5281/zenodo.6879594

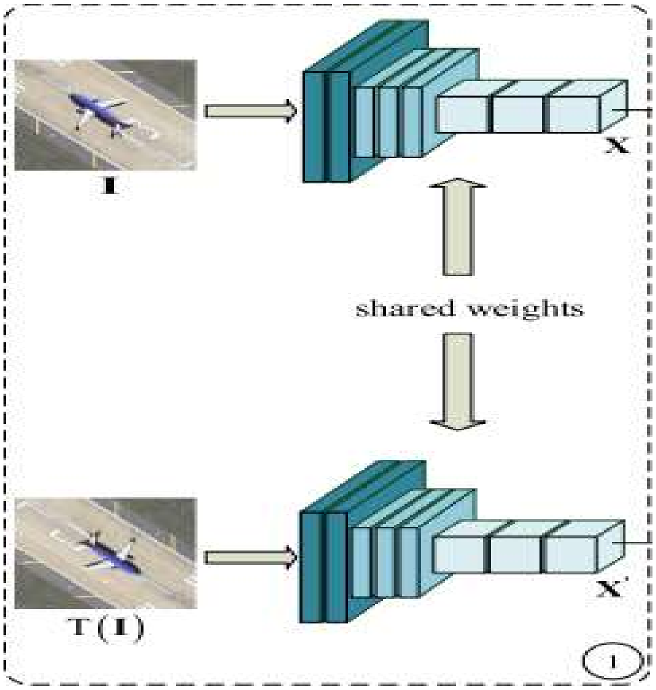


Fig1 stage one extraction

In the third stage taking the influence of spatial rotation into account, we develop to unify the attention areas of the image pairs with different angles. Also, this specific operation can compact the interclass samples and separate the intra class images

B. Second stage of feature extraction

In the second stage of feature extraction we didn't developed new CNN algorithm instead we adopted VGG16 [30] to extract the intermediate features from the RS images. Here, our intermediate feature extraction network remains convolution layers, max pooling layers, and nonlinear activation layers. VGG16 has more nonlinear layers that could improve the capacity of feature learning the sizes of convolution kernels of VGG16 are 3 × 3, which can capture images from different resolutions. The whole VGG16 network consists of 13 convolution layers with the kernel size of 3 × 3, five max pooling layers with the size of 2 × 2, five nonlinear activation layers, and three fully connected (FC) layers.

ISSN: 1533 - 9211
DOI: 10.5281/zenodo.6879594

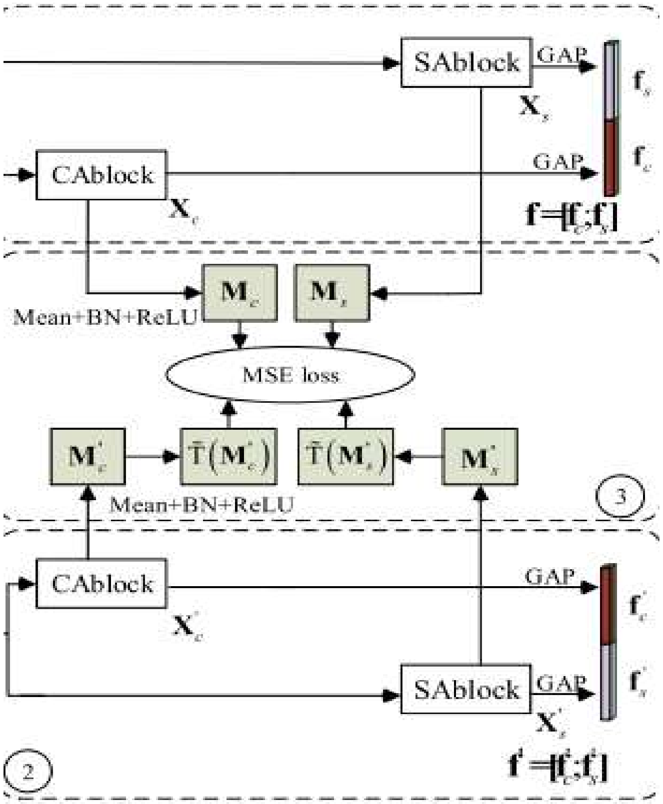


Fig2 stage 2 feature extraction

We use the pre trained weights to initialize the feature extraction model for speeding up the convergence. For the input RS image pairs **I** and T(**I**), the feature extraction networks are parameter shared. After the feature extracting mode, the feature maps Therefore, we choose VGG16 as our intermediate feature extraction network in this article.

C. Third stage of feature extraction

The intermediate feature maps extracted by VGG16 can represent the contents of RS images from the global aspect. The third stage of feature extraction model contains two blocks: channel wise attention block (CAblock) and spatial-wise attention block (SAblock). CAblock focuses on emphasizing the significant channels and suppressing the insignificant channels of **X.** SA block focuses on reshaping the block.

D. Final stage of extraction

We got the deep features **f** and **f** for the image pairs **I** and T(**I**). Taking an RS image as an example (shown in Fig. 4), when we rotate the original image by 90◦, the attention regions are changed as well for focusing on the planes to reflect the semantic of "Plane." For the original RS image, the channel-wise attention and spatial-wise attention regions are concentrated in the right part. For the rotated RS image, the channel-wise attention and spatial-wise attention regions are concentrated in the bottom part.

ISSN: 1533 - 9211

DOI: 10.5281/zenodo.6879594

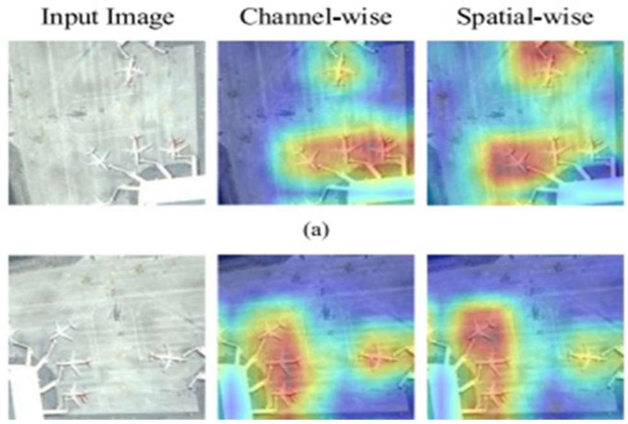


Fig3 (a) Original image and its attention maps.
(b) Rotated image (90◦) and its attention maps.

IV. EXPERIMENTS AND DISCUSSION

A. *Testing Data Introduction*

To testify the effectiveness of our Network, we select three RS image benchmarks. The first one is a small-scale aerial image dataset, which was published by the University of California Merced [50], and we name it UCM1 in this article for short.

There are 2100 aerial images in UCM that cover 20 U.S. regions, including Birmingham, New York, etc. These aerial images are divided into 21 scene classes, and each class contains 10 RS images. Their spatial resolution and sizes are one foot and 256 × 256.

Fig4 Examples of different scenes of the UCM dataset. The scene numbers and names are summarized as follows. 1-Agricultural, 2-Airplane, 3-Baseball Diamond, 4-Beach, 5-Buildings, 6-Chaparral, 7-Dense Residential, 8-Forest, 9-Freeway, 10-Golf Course, 11-Harbor, 12-Intersection, 13-Medium Density Residential, 14-Mobile Home Park, 15-Overpass, 16-Parking Lot, 17-River, 18- Runway, 19-Sparse Residential, 20-Storage Tanks, and 21-Tennis Courts.

To implement above modules author has used UCM dataset which contains 21 classes or

SEYBOLD Report
ISSN: 1533 - 9211
DOI: 10.5281/zenodo.6879594

different scenes images and those names are listed below 'agricultural', 'airplane', 'baseballdiamond', 'beach', 'buildings', 'chaparral', 'denseresidential', 'forest', 'freeway', 'golfcourse', 'harbor', 'intersection', 'mediumresidential', 'mobilehomepark', 'overpass', 'parkinglot', 'river', 'runway', 'sparseresidential', 'storagetanks', 'tenniscourt'

So by using scenes from above location we will build network model and then calculate Overall (OA) accuracy.

The next step of executing the network proposed is as follows. We need to select the data set of the trained models from our device.

In above screen dataset loaded and now click on 'Image Brightness & Rotation' button to rotate image to extract global features and to get below screen

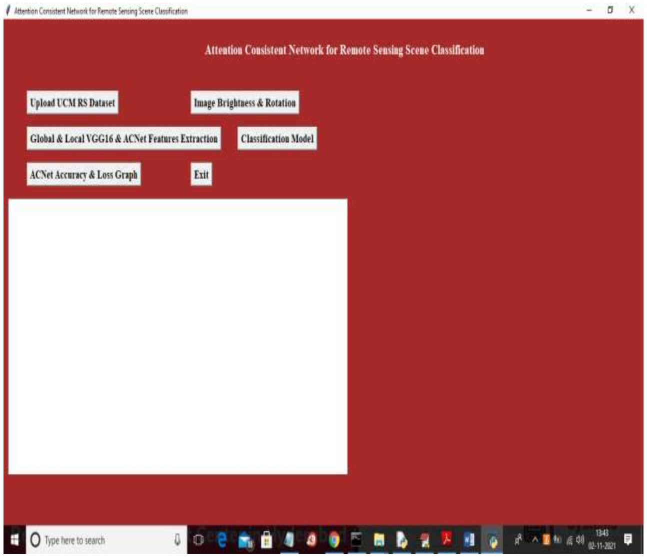


Fig5 UCM data set

In above screen click on 'Upload UCM RS Dataset' button to upload dataset and to get below screen

In above screen selecting and uploading 'UCM' dataset folder and then click on 'Select Folder' button to load dataset and to get below screen

SEYBOLD Report
ISSN: 1533 - 9211
DOI: 10.5281/zenodo.6879594

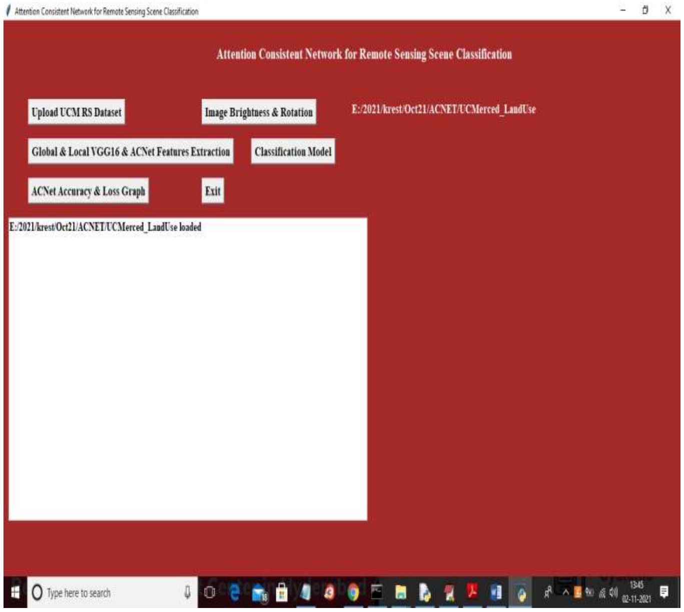


Fig 6 image brightness and rotation

In above screen dataset loaded and now click on 'Image Brightness & Rotation' button to rotate image to extract global features and to get below screen.

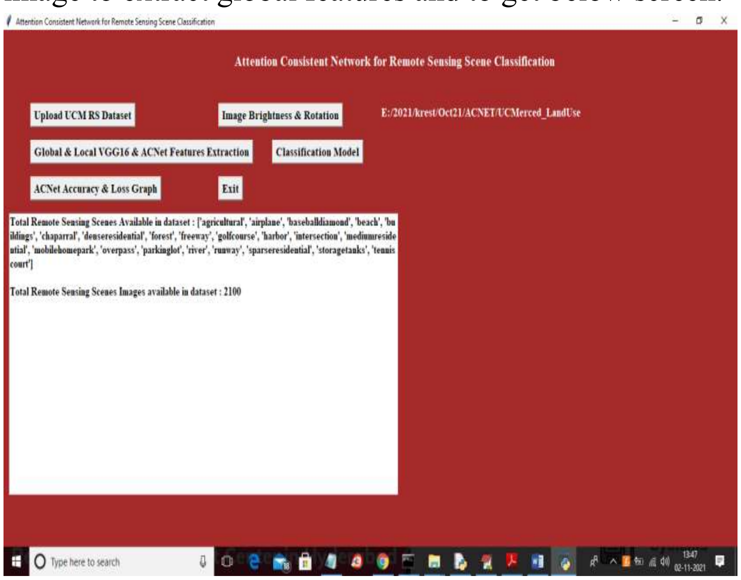


Fig 7 global feature extraction

In above screen application showing names of different scenes images and total 2100 scenes found in dataset. Now processed and image rotated dataset is ready and now click on 'Global & Local Network Features Extraction' button to train Network with both features and then will get below output

SEYBOLD Report
ISSN: 1533 - 9211
DOI: 10.5281/zenodo.6879594

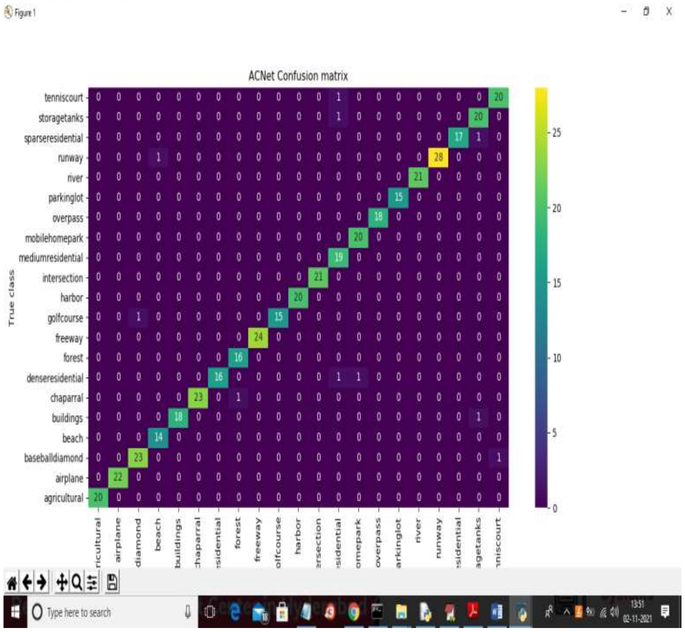
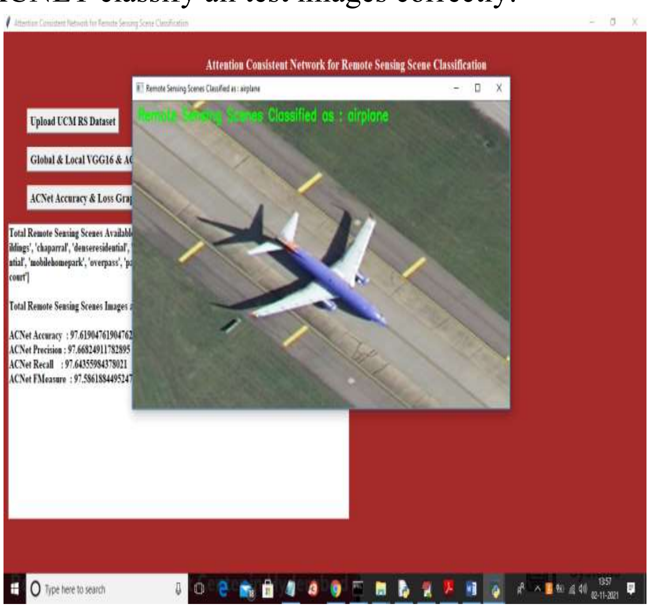


Fig 8 confusion matrix

In above screen ACNet model generated and then classify test images and we can see in above confusion matrix all images are classified correctly as all numbers values are in diagnol where you can see names of scenes on X and Y-axis are matching with predicted number of scenes so ACNET classify all test images correctly.



In above screen ACNet classified image scene as ‘Airplane’ and similarly you can upload other images and test it

SEYBOLD Report
ISSN: 1533 - 9211
DOI: 10.5281/zenodo.6879594

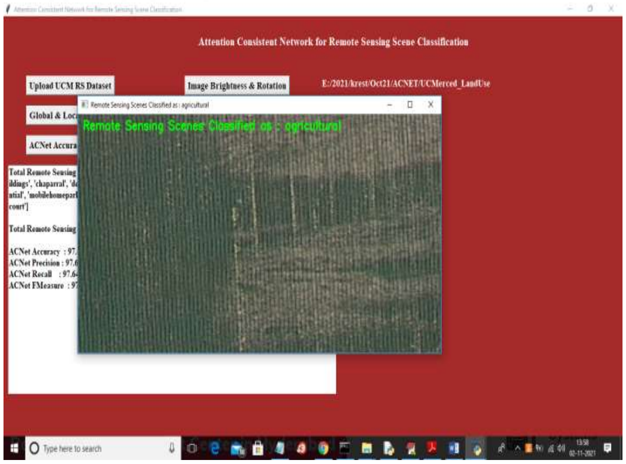


Fig 9 agriculture identification

Now click on ‘ACNet Accuracy & Loss Graph’ button to get below graph

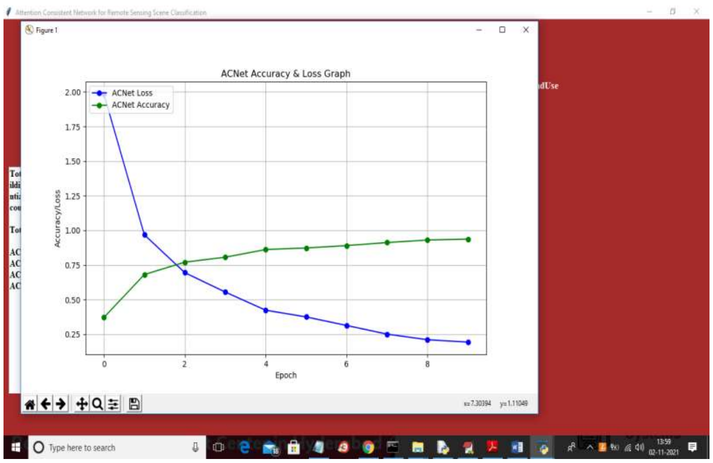


Fig 10 epoch graph

In above graph x-axis represents EPOCH and y-axis represents accuracy/loss values and in above graph green line represents accuracy and blue line represents loss values and we can see with each increasing epoch accuracy got increased and loss values get decreased. Any model with increasing accuracy and decreasing loss will be considered as accurate deep learning model.

V. CONCLUSION

From this project we proposed the network which is dual branched to classify remote sensing images. As proposed there are four feature extraction models, from four model feature extraction the first two extraction models are local feature extraction next two extraction models deals with global feature extraction finally the results obtained by the classification which have accuracy more than 99.7%.

REFERENCES

[1] N. Zhu *et al.*, “Deep learning for smart agriculture: Concepts, tools, applications, and opportunities,” *Int. J. Agricultural Biol. Eng.*, vol. 11, no. 4, pp. 32–44, 2018.

[2] J. Marçais and J.-R. de Dreuzy, “Prospective interest of deep learning for

DOI: 10.5281/zenodo.6879594
hydrological inference," *Ground Water*, vol. 55, pp. 688–692, 2017.
[3] X. Zou, M. Cheng, C. Wang, Y. Xia, and J. Li, "Tree classification in complex forest point clouds based on deep learning," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 12, pp. 2360–2364, Dec. 2017.
[4] L. Chen, W. Yang, K. Xu, and T. Xu, "Evaluation of local features for scene classification using VHR satellite images," in *Proc. IEEE Joint Urban Remote Sens. Event*, 2011, pp. 385–388.
[5] G. Sheng, W. Yang, T. Xu, and H. Sun, "High-resolution satellite scene classification using a sparse coding based multiple feature combination," *Int. J. Remote Sens.*, vol. 33, no. 8, pp. 2395–2412, 2012.
[6] L. Jiao, X. Tang, B. Hou, and S. Wang, "SAR images retrieval based on semantic classification and region-based similarity measure for earth observation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 8, pp. 3876–3891, Aug. 2015.
[7] X. Tang and L. Jiao, "Fusion similarity-based reranking for SAR image retrieval," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 2, pp. 242–246, Feb. 2017.
[8] X. Tang, L. Jiao, W. J. Emery, F. Liu, and D. Zhang, "Two-stage reranking for remote sensing image retrieval," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 10, pp. 5798–5817, Oct. 2017.
[9] X. Tang, L. Jiao, and W. J. Emery, "SAR image content retrieval based on fuzzy similarity and relevance feedback," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 5, pp. 1824–1842, May 2017.
[10] Q. Zhu, Y. Zhong, L. Zhang, and D. Li, "Scene classification based on the fully sparse semantic topic model," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 10, pp. 5525–5538, Oct. 2017.
[11] R. Xu, Y. Tao, Z. Lu, and Y. Zhong, "Attention-mechanism-containing neural networks for high-resolution remote sensing image classification," *Remote Sens.*, vol. 10, no. 10, p. 1602, 2018.
[12] Q. Zhu, Y. Zhong, L. Zhang, and D. Li, Adaptive deep sparse semantic modeling framework for high spatial resolution image scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no.10pp. 6180–6195, Oct. 2018.
[13] Q. Zhu, Y. Zhong, S. Wu, L. Zhang, and D. Li, "Scene classification based on the sparse homogeneous-heterogeneous topic feature model," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 5, pp. 2689–2703, May 2018.
[14] X. Tang, X. Zhang, F. Liu, and L. Jiao, "Unsupervised deep feature learning for remote sensing image retrieval," *Remote Sens.*, vol. 10, no. 8, p. 1243, 2018.
[15] H. Sun, S. Li, X. Zheng, and X. Lu, "Remote sensing scene classification by gated bidirectional network," *IEEE Trans. Geosci. Remote Sens.*, vol. 58,

ISSN: 1533 - 9211
DOI: 10.5281/zenodo.6879594

no. 1, pp. 82–96, Jan. 2020.
[16] X. Tang, C. Liu, J. Ma, X. Zhang, F. Liu, and L. Jiao, "Large-scale remote sensing image retrieval based on semi-supervised adversarial hashing," *Remote Sens.*, vol. 11, no. 17, p. 2055, 2019.
[17] C. Liu, J. Ma, X. Tang, F. Liu, X. Zhang, and L. Jiao, "Deep hash learning for remote sensing image retrieval," *IEEE Trans. Geosci. Remote Sens.*, to be published, doi: 10.1109/TGRS.2020.3007533.
[18] S. R. Gunn *et al.*, "Support vector machines for classification and regression," *ISIS Tech. Rep.*, vol. 14, no. 1, pp. 5–16, 1998.
[19] C. Liu and H. Wechsler, "Gabor feature based classification using the enhanced Fisher linear discriminant model for face recognition," *IEEE Trans. Image Process.*, vol. 11, no. 4, pp. 467–476, Apr. 2002.
[20] G. Csurka, C. Dance, L. Fan, J. Willamowski, and C. Bray, "Visual categorization with bags of keypoints," in *Proc. Workshop Statist. Learn. Comput. Vis.,* Prague, Czech Republic, 2004, vol. 1, pp. 1–2.
[21] A. Liaw *et al.*, "Classification and regression by randomforest," *R News*, vol. 2, no. 3, pp. 18–22, 2002.
[22] X. Lu, X. Zheng, and Y. Yuan, "Remote sensing scene classification by unsupervised representation learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 9, pp. 5148–5157, Sep. 2017