A MULTI-TASK ARCHITECTURE FOR REMOTE SENSING BY JOINT SCENE CLASSIFICATION AND IMAGE QUALITY ASSESSMENT

Cong Zhang[1], *Qi Wang*[1*], *Xuelong Li*[1]

[1]School of Computer Science and Center for OPTical IMagery Analysis and Learning (OPTIMAL),
Northwestern Polytechnical University, Xi'an 710072, Shaanxi, P.R. China.

ABSTRACT

In this work, we propose a compact multi-task architecture based on deep learning for remote sensing scene classification and image quality assessment (IQA) simultaneously. The model can be trained in an end-to-end manner, and the robustness of classification is improved in our method. More importantly, by exploiting IQA and super-resolution, the accurate classification results can be obtained even if the images are distorted or with low quality. To the best of our knowledge, it is the first successful attempt to associate IQA with scene classification in a unified multi-task architecture. Our method is evaluated on the expanded UC Merced Land-Use dataset after data augmentation. In comparison with some other methods, the experimental results show that the proposed structure makes a great improvement on both classification and IQA.

Index Terms— Remote sensing, scene classification, image quality assessment, image super-resolution, multi-task learning, deep learning

1. INTRODUCTION

Recently, remote sensing scene classification tends to attracting more attention due to its wide range of applications, such as disaster relief, land-cover/land-use classification, and urban planning. However, intricate structures involved in remote sensing images make it a challenging task to extract valuable features from the images. To remedy this problem, great effort has been made for research of remote sensing scene classification algorithms [1, 2, 3, 4, 5]. In spite of this, there are still some issues to deal with, for example, it is difficult to obtain accurate and robust results from remote sensing images with very low quality.

Generally in practice, remote sensing scene images suffer from various degradation not only from the unstable imaging system but also from harsh environments like extreme weather, illumination and atmosphere [6]. Therefore, although these existing methods can achieve high accuracy in some datasets of scene images classification, there is a still certain limitation: they are mainly for clear and simple images but not for distorted poor quality ones. However, a robust system for remote sensing scene classification should also work well and produce accurate results even in complex conditions.

With the development of image quality assessment in remote sensing [6, 7], it is becoming more and more suitable to break through the above limitation. Image Quality Assessment (IQA) can be divided into two types: subjective IQA and objective IQA. The former approach fully considers the visual experience of the human eyes, which is very labor-intensive and time-consuming. The latter aims to calculate the parameters representing the image quality according to a predetermined standard algorithm and finally obtain the image quality scores, which ideally meets our requirements. Further, objective IQA can be divided into Full-Reference IQA (FR-IQA), Reduced-Reference IQA (RR-IQA) and Non-Reference IQA (NR-IQA). FR-IQA contains the pristine reference image and the distorted one, while NR-IQA only operates on the distorted ones. Due to its widely applications, NR-IQA is becoming more important. In the early works, handcrafted features are extensively used for remote sensing IQA, such as structural similarity [8]. However, as this task becomes more difficult, these existing methods are not suitable for our complex application scenarios. Inspired by IQA-CNN [9], we explore applying Convolutional Neural Networks (CNN) for remote sensing IQA. In addition, scene classification and IQA are integrated in this work, since multi-task learning has demonstrated its ability on discriminative classifiers [10].

After evaluating image quality, in order to further contribute to other tasks, the images are expected to be post-processed in the next step, such as image super-resolution. Image Super-Resolution (SR) refers to the construction of corresponding high-resolution images from the observed low-resolution ones. It has been widely used in remote sensing and medical images. Recently thanks to the outstanding achievements of CNN in other fields, CNN is increasingly applied to image SR. Among them, DRCN [11] is a representative algorithm based on single-resolution reconstruction, which ob-

[*]Corresponding Author.

tains competitive results.

To address the problems depicted above, in this paper a novel architecture is proposed, which can accurately classify the remote sensing scene images with low quality. The contributions of this work can be summarized as follows:

(1) We design a robust and end-to-end multi-task architecture. This architecture can adaptively process images based on their quality scores. With the work of image quality assessment (IQA), it is able to obtain categories and IQA scores of input images simultaneously. To the best of our knowledge, it is the first successful attempt to associate image quality assessment and remote sensing scene classification in a unified multi-task architecture.

(2) Image super-resolution [11] is added after IQA, and this module improves the resolution of low qulity images, which can contribute to the performance of scene classification.

2. OUR METHOD

The overview of our proposed architecture is presented in Fig. 1. The main components of our algorithm are as follows. First of all, the parameters sharing CNN is used to extract feature from an input image. Then two branches of several fully connected layers after pooling operations are exploited to accomplish two different but related tasks. One is remote sensing scene classification, the other is image quality assessment (IQA). Meanwhile the IQA module generates a quality score to represent the credibility of the predicted classification result. If not confident, the input image will be preprocessed by the image super-resolution module, and its output will be used as a new input to the multi-task framework. Along with the IQA and super-resolution stages, the scene images are classified more accurately. Moreover, the proposed algorithm is robust to distorted images with low quality, and it can be trained in an end-to-end manner easily.

2.1. Shared CNN for Multi-task Learning

For better performance, remote sensing scene classification usually operates at feature levels, as is IQA. Therefore, the CNN module is designed for feature extraction. As illustrated in Fig. 1, feature maps will be obtained immediately from the input images through CNN without any preprocessing operations. More importantly, we apply the idea of multi-task learning here, so that the parameters in this module are shared by two tasks. The feature map can be directly used for scene classification and IQA, which avoids repeated feature extraction, greatly reducing computational resource and time consumption.

Even if the CNN are shared, there are still millions of parameters to be adjusted. However, the dataset of remote sensing scene classification usually is far from satisfied which only consists of several thousands images. So it should be noted that the dataset used in our method requires data augmentation, for example, adding noise, blurring images and degrading resolution. Another purpose of these operations is to provide effective training samples for the IQA module. Our implementation of CNN is based on the modified VGGNet-16.

2.2. Joint Scene Classification and IQA

In this section, we briefly present the design of multi-task module as described in Fig. 1. It is composed of two branches: remote sensing scene classification and IQA. Then we introduce the two different loss functions and the total loss function applied to these tasks respectively.

In the classification branch, we design three fully connected layers of 512 nodes each after a max pooling operation. The part of fully connected layers is a $512-512-512-n$ structure, where n denotes the total number of categories to classify. Then its output will be fed to softmax layer, and the probability of each class is obtained. Hence one with the highest probability is selected to represent the class of the input image. While in the quality assessment branch, inspired by [9], we proposed a network composed of four layers. As shown in Fig .1, the pooling operations reduce each feature map come by shared CNN to one max and one average. Two fully connected layers of 800 nodes each follow the pooling. Finally, the last layer is a simple linear regression with a one dimensional output that gives the quality score [9]. In this paper, the range of IQA scores is fixed to $[0, 10]$.

The design of loss functions is essential for multi-task learning, which contributes to the performance directly. In our proposed architecture, the total loss function can be formulated as

$$L = \lambda_c L_c + \lambda_i L_i, \tag{1}$$

where L_c and L_i represents the losses for scene classification and image quality assessment, respectively, and λ_c and λ_i weighs the importance between these two losses. In our experiment, we set both λ_c and λ_i to 1.

In the stage of remote sensing scene classification, we use softmax cross entropy loss function given by

$$L_c = \sum_j \ell(\hat{y_j}, y_j), \tag{2}$$

$$\ell(\hat{y}, y) = -y^T \log \hat{y}, \tag{3}$$

where $\hat{y}$ and y represent the prediction (i.e., the output of softmax layer) and its ground truth, respectively. The softmax function is denoted as

$$\hat{y_j} = softmax(z_j) = \frac{e^{z_j}}{\Sigma_j e^{z_j}}. \tag{4}$$

For the loss of IQA task, L_i is the ℓ_1 norm of the prediction error, as defined in [9].

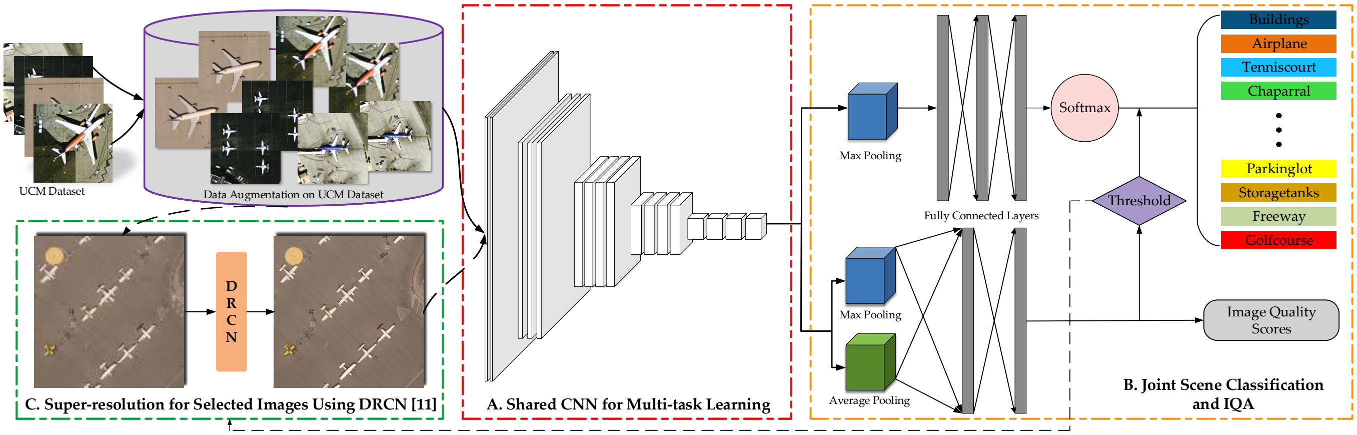

Fig. 1. Overview of the components of our proposed architecture

2.3. Super-resolution for Selected Images Using DRCN [11]

If the quality of an image is low, the predicted score from IQA module will be low. When it is above the threshold η, our algorithm will normally output the classification result and its quality score. Nevertheless, if the predicted score is below η, the input image will be selected and then sent to the image super-resolution module automatically. In the experiment, we set η to 8.0.

It has been shown that after super-resolution, images were more easily classified correctly. Moreover, there are many proposed related methods based on CNN. They can ideally meet our requirements. For this reason, we apply the algorithm DRCN proposed in [11] to our model. DRCN outperforms many other algorithms [11], and we directly adapt it for low quality images selected by our IQA module. It is worth noting that super-resolution will change the size of images, which affects the fully connected layers. To resolve this issue, we use global pooling instead of normal pooling in the pooling layer. To the best of our knowledge, this strategy that adapt super-resolution after IQA is proposed in the first time.

3. EXPERIMENTS

To evaluate our proposed model, we trained and tested it on a public remote sensing scene classification dataset named UC Merced Land-Use (UCM) dataset [12]. This dataset contains 2100 scene images, which are divided into 21 typical land-use scene classes.

However, UCM dataset can only be used to train classification network, but there is no scene image dataset which contains image quality scores. The labels of quality scores is indispensable since IQA model should be trained simultaneously. Therefore, we apply data augmentation to this dataset. With operations adding white noise on images, blurring and degrading images, and reducing resolution of images, the total number of images for training and testing overall framework is expanded from 2100 to 21000. Data augmentation not only improves classification performance, but more importantly, provides the training data for IQA module which requires low-quality and low-resolution images. Further, we generate quality scores as ground truths using a full-reference IQA method named FSIM [13]. As shown in Fig. 2, the original images in UCM dataset are used as reference images, while the ones obtained by data augmentation are used as the distorted. The label distribution histogram of IQA module can be seen in Fig .2 clearly. With the above approach, our model can be trained in an end-to-end manner easily.

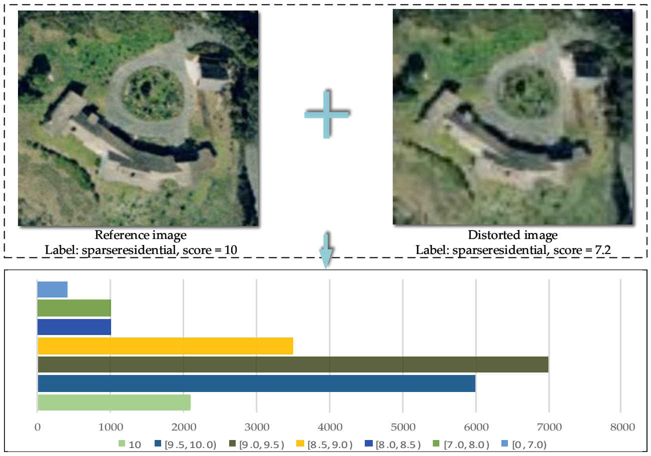

Fig. 2. The sample and image quality label distribution histogram of UC Merced Land-Use dataset after data augmentation.

Stochastic Gradient Decent (SGD) is used to train our model. The parameters are set as follows: the learning rate is set to 0.001 and the batch size is 100. Dropout and regularization are both adopted to prevent overfitting in the experiment. In general, for the UCM dataset, we choose a common ratios using 80% samples selected randomly for training. Overall accuracy (OA) is used as evaluation measurement for scene classification. Besides, we follow the same protocol as in [9] to use Linear Correlation Coefficient (LCC) and Spearman Rank Order Correlation Coefficient (SROCC) to evaluate the performance of the proposed IQA branch. TensorFlow is chosen as deep learning framework to implement our model.

Table 1. TESTING CLASSIFICSATION RESULTS OF DIFFERENCE METHODS

Methods	Overall Accurary (%)
BoVW [3]	76.81
Pyramid of Spatial Relations [1]	89.10
Gradient Boosting Random CNNs [4]	94.53
Fine-tuning GoogLeNet [3]	97.10
Our Proposed Method	**98.57**

We compare our method with some advanced ones for scene classification like BoVW [3], Pyramid of Spatial Relations [1], Gradient Boosting Random CNNs [4] and Fine-tuning GoogLeNet [3]. As illustrated in Table. 1, traditional methods achieve poor performance, while CNN-based approaches are very effective which improve performance significantly. The proposed method in this paper achieves the best scene classification performance, which indicates the active influence of both IQA module and image super-resolution module.

Table 2. PERFORMANCE OF QUALITY ASSESSMENT ON UCM DATASET AFTER DATA AUGMENTASTION

Methods	LCC	SROCC
IQA-CNN [9]	0.834	0.810
Our Proposed Method	**0.891**	**0.850**

In order to evaluate the performance of our proposed IQA module, we compare it with IQA-CNN [9] as shown in Table. 2. Experiments on IQA performance evaluation were conducted on the UCM dataset after data augmentation. As can be seen in Table. 2, our method outperforms IQA-CNN algorithm [9], which shows that the proposed multi-task architecture is reliable.

4. CONCLUSIONS

In this paper, a novel multi-task architecture based on CNN is proposed for scene classification and image quality assessment (IQA). The robustness and performance of classification are both improved by applying IQA and image super-resolution. The experimental results on UC Merced Land-Use dataset after data augmentation demonstrate its effectiveness of our proposed method.

5. ACKNOWLEDGEMENT

This work was supported by the National Key R&D Program of China under Grant 2018YFB1107403, National Natural Science Foundation of China under Grant U1864204 and 61773316, Natural Science Foundation of Shaanxi Province under Grant 2018KJXX-024, and Project of Special Zone for National Defense Science and Technology Innovation.

6. REFERENCES

[1] Shizhi Chen and YingLi Tian, “Pyramid of spatial relatons for scene-level land use classification,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 4, pp. 1947–1957, 2015.

[2] Qi Wang, Shaoteng Liu, Jocelyn Chanussot, and Xuelong Li, “Scene classification with recurrent attention of vhr remote sensing images,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 2, pp. 1155–1167, 2019.

[3] Marco Castelluccio, Giovanni Poggi, Carlo Sansone, and Luisa Verdoliva, “Land use classification in remote sensing images by convolutional neural networks,” *arXiv preprint arXiv:1508.00092*, 2015.

[4] Fan Zhang, Bo Du, and Liangpei Zhang, “Scene classification via a gradient boosting random convolutional network framework,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 3, pp. 1793–1802, 2016.

[5] Qi Wang, Xiang He, and Xuelong Li, “Locality and structure regularized low rank representation for hyperspectral image classification,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 2, pp. 911–923, 2019.

[6] Yatong Xia and Zhenzhong Chen, “Quality assessment for remote sensing images: approaches and applications,” in *Proceedings of the IEEE Conference on Systems, Man, and Cybernetics*, 2015, pp. 1029–1034.

[7] Jicheng Wang, Yuanxin Ye, Li Shen, Zhipeng Li, and Songbo Wu, “Research on relationship between remote sensing image quality and performance of interest point detection,” in *Geoscience and Remote Sensing Symposium, 2015 IEEE International*. IEEE, 2015, pp. 545–548.

[8] Di Liu, Yingchun Li, and Shaojun Chen, “No-reference remote sensing image quality assessment based on the region of interest and structural similarity,” in *Proceedings of the 2nd International Conference on Advances in Image Processing*, 2018, pp. 64–67.

[9] Le Kang, Peng Ye, Yi Li, and David Doermann, “Convolutional neural networks for no-reference image quality assessment,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1733–1740.

[10] Jun Yu, Baopeng Zhang, Zhengzhong Kuang, Dan Lin, and Jianping Fan, “iPrivacy: image privacy protection by identifying sensitive objects via deep multi-task learning,” *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 5, pp. 1005–1016, 2017.

[11] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee, “Deeply-recursive convolutional network for image super-resolution,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 1637–1645.

[12] Yi Yang and Shawn Newsam, “Bag-of-visual-words and spatial extensions for land-use classification,” in *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 2010, pp. 270–279.

[13] Lin Zhang, Lei Zhang, Xuanqin Mou, David Zhang, et al., “Fsim: a feature similarity index for image quality assessment,” *IEEE Transactions on Image Processing*, vol. 20, no. 8, pp. 2378–2386, 2011.