DEEP LEARNING TRAINING WITH UNBALANCE SAMPLE DISTRIBUTION FOR REMOTE SENSING IMAGE SEGMENTATION

Yue XU ¹, Xiangyun HU ¹23*, Jianya GONG ¹, Xin HUANG ¹4, Jiayi LI ¹

¹ School of Remote Sensing and Information Engineering, Wuhan University – (yuexu41, huxy, gongjy, xhuang, zjerica)@whu.edu.cn
² Hubei Luojia Laboratory
³ Institute of Artificial Intelligence in Geomatics, Wuhan University
⁴ State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University

Commission III, WG III/1

KEY WORDS: Sample balance, Remote sensing image, deep learning, High spatial and spectral resolution (HSSR)

ABSTRACT:

The intelligent interpretation of remote sensing images based on deep learning has become a hot spot with the increasing satellite images acquired due to the rapid development of aerospace technology. Sufficient and reasonable distributed samples are essential for the accuracy of deep learning. The spatial distribution of natural features is inhomogeneous in the real world. When people create sample dataset, they often collect within a certain local range, which may bring problems of unbalanced distribution of samples, including the unbalance between training dataset and validation dataset, and the unbalance among different sample categories. This long-tail distribution of samples (i.e., a few classes account for most of the data, while most classes are under-represented) can lead to bias in the training model and make it difficult to ensure accuracy.

In this paper we tried to solved the above-mentioned problem in landcover classification with high spatial and spectral resolution (HSSR) remote sensing images. We first adopted an iterative stratification method for multi-label data classification to ensure that both training dataset and validation dataset contain reasonable proportion of landcover classes. Then we proposed a weighted loss algorithm to further strengthen the learning ability of the model for rare categories. Experiments on a large volume HSSR dataset shows that with our methods the accuracy of landcover classification increased by 2%.

1. INTRODUCTION

Land cover and its change are important for resource planning and monitoring, ecological environment assessment, and sustainable development. With the accelerating process of economic globalization, the demand for land cover information high spatial and temporal resolution in the region scope and even global scope is increasing (Nilsson et al., 2016) (UN, 2019). In 2015, the United Nations formulated the 2030 agenda for sustainable development for global sustainable development and proposed to achieve 17 sustainable development goals (SDGs) by 2030 (Grekovsis et al., 2015). In recent years, many of the SDGs are close-related to land cover. Thus, land cover information can effectively support the formulation and implementation of national, regional and local public policies, economic and political programs. The realization of SDGs further puts forward an urgent demand for global scope high-resolution land cover data.

At present, the most precise land cover data in the world are GlobeLand30 land cover product with 30 meters resolution developed by Chen Jun etc. (2018), FROM-GLC10 land cover with 10 meters resolution developed by Gong Peng etc. (Gong et al., 2013), and fine classification products GLC_FCS30-2015 with 30 meters resolution developed by Liu Liang-yun etc. (Zhang et al., 2019). With the development of earth observation ability, satellite remote sensing images with high spatial and temporal resolution are increasingly rapidly, which support the extraction of finer ground object in large area to product landcover data with higher resolution. However, at present, the methods and technologies of satellite remote sensing interpretation are not enough to support the large-scale production of high resolution land cover data.

The intelligent interpretation of remote sensing images based on deep learning has become a hot spot. Sufficient and reasonable distributed samples (annotated datasets) are essential for the accuracy of deep learning. This has been proved by many successful cases of deep Convolutional Neural Networks (CNNs) for visual recognition (Hinton, 2006) (Huang and Learned-Miller, 2014) with large-scale, real-world annotated datasets (Rawat and Wang, 2017) (Zhang et al., 2016). For remote sensing image interpretation, the inhomogeneous spatial distribution of natural features in the real world must be considered when creating the sample datasets. When people
create sample dataset, they often collect within a certain local range, which may bring problems of unbalanced distribution of samples, including the unbalance between training dataset and validation dataset, and the unbalance among different sample categories. This long-tail distribution of samples (i.e., a few classes account for most of the data, while most classes are under-represented) can lead to bias in the training model and make it difficult to ensure accuracy (Chen et al., 2016) (Xu et al., 2019) (Khoshgoftaar et al., 2010).

People tried to handle the problems caused by the long-tailed training data by re-sampling and re-weighting (Bengio, 2015) (Zhang et al., 2017). Cui et al. proposed a re-weighting approach to design a class-balanced loss (Cui et al., 2019).

In this paper we tried to solved the above-mentioned problem in landcover classification with high spatial and spectral resolution (HSSR) remote sensing images. We first adopted an iterative stratification method for multi-label data classification to ensure that both training dataset and validation dataset contain reasonable proportion of landcover classes. Then we proposed a weighted loss algorithm to further strengthen the learning ability of the model for rare categories. Experiments on a large volume HSSR dataset shows that with our methods the accuracy of landcover classification increased by 3%.

2. METHODS

2.1 Class balanced Loss

To address the problem of long-tailed sample distribution, typical solutions adopt class re-balancing strategies such as re-sampling and re-weighting based on the number of observations for each class. Cui et al proposed a class balanced loss method to handle this problem (Cui et al., 2019). They argued that as the number of samples increased, the additional benefit of a newly added data point would diminish. So they introduced a framework to measure data overlap by associating with each sample a small neighbouring region rather than a single point. The effective number of samples is defined as the volume of samples and can be calculated by a simple formula (1-pn)/(1-βn), where n is the number of samples and β ∈ (0, 1) is a hyperparameter. A re-weighting scheme was designed using the effective number of samples for each class to re-balance the loss, thereby yielding a class-balanced loss. We adopted this method to balance the number of samples in different landcover categories.

In this paper, we calculated the adjustment amount by using the class balance loss proposed by Cui et al. We firstly define the calculation method of the effective quantity of each category (the expected value of the sample size). Suppose that the data of a new sampling point can only be associated with the previous sampling data in two ways: either it is completely in the previous sampling data set, and the probability is recorded as p; Or completely outside the previous sampling data, and the probability is recorded as (1-p). With the increase of the number of sampled data, the probability p also increases. Using more samples in a certain category will bring diminishing marginal benefits. This is because there are inherent similarities between real-world data. With the increase of the number of samples, the newly added samples are likely to be almost duplicate of the existing samples.

The effective number of samples is called $E_n$. Where $n \in \mathbb{Z} > 0$ is the number of samples. Super parameter $\gamma$ control $E_n$ increases with the increase of $n$, then:

$$E_n = pE_{n-1} + (1-p)(E_{n-1} + 1)$$

$$= \frac{1-\gamma^n}{1-\gamma} \text{ where } \gamma \in [0,1]$$

(1)

In order to obtain the balanced loss function, we introduced a weighting factor $1/E_n$. It is in inverse proportion to the effective sample number of each class $E_n$, that is, for class $i$, $n_i$ is included in total. Adding a weighting factor to the loss function, and the sample size balance loss function can be written as:

$$L_{cb} = \frac{1}{E_n} L(o,y) = \frac{1-\gamma^n}{1-\gamma} L(o,y)$$

(2)

2.2 Iterative sample division

Generally, samples should be divided into training, validation, and test sets. The number of classes in each dataset were often unbalanced. Szymański et al. proposed an iterative stratification method for multi-label data classification (Szymański and Kajdanowicz, 2017) (Szymański and Kajdanowicz, 2019). This method considered the second-order relationship between labels so that a more balanced sample distribution can be obtained. We adopted this method to ensure both training and validation datasets contain an reasonable proportion of landcover classes.

2.3 Deep learning Model

Zheng et al. proposed a fast patch-free global learning (FPGA) framework for hyperspectral image (HSI) classification (Zheng et al., 2020). It uses an encoder-decoder based FCN to consider the global spatial information by processing the whole image. And experiments show that FPGA framework is superior to the patch-based framework in both speed and accuracy for HSI classification.
As shown in Figure 1 and Figure 2, FGPA used an encoder network and a decoder network. The encoder is responsible for computing the hierarchical convolutional feature maps over an entire input HSI. The decoder recovers the spatial dimension of the coarsest convolutional feature map progressively via lateral connection based SSF, outputting a classification probability map of the same spatial size as the input image.

### 3. EXPERIMENT AND RESULTS

#### 3.1 Original sample dataset

We used the high spatial and spectral resolution sample dataset Luojia HSSR developed by Wuhan University(2022). It is constructed based on aerial hyperspectral imagery of southern Shenyang City of Liaoning Province in China, covering area of 161 square kilometers with spatial resolution 0.75 meter 249 spectral bands as VNIR, geometric accuracy of 1.5-3 meter and corresponding to 47 categories of field-validated ground coverage. To the knowledge of this paper, it is the largest hyperspace spectral sample data set so far.

#### 3.2 Adjustment of the sample dataset

The sample size of each category in Luojia-HSSR varies greatly due to the uneven spatial distribution of landscape types in real world. For example, paddy fields account for 45.7% of the total, while roads account for only about 2%.

We adjust the original Luojia-HSSR dataset with the method proposed in previous section of this paper, including the following 3 steps:

1. Category distribution adjustment: the categories with too small sample size are deleted or merged. After adjustment, the sample categories are changed from 47 to 23, as shown in Table 1.
2. Data set division: divide the above samples into training, verification and test sets. Finally, 6438 pairs of samples are adjusted according to the method in the previous section, and the sample sizes of various types are closer. Figure 5 shows the divided training, testing and verification dataset. The sample size among the subsets is relatively equal.

#### 3.3 Result and discussion

We verified the effect of sample adjustment through experiments with the dataset described in 4.2. Frequency Weighted Intersection over Union (FWIoU) is used as the evaluation index. FWIoU is an improvement of the Mean Intersection over Union (MIoU). It sets the weight for each category according to the frequency of occurrence. Therefore, when used in the case of uneven distribution of sample size among classes, the result will be more reasonable.

\[
FWIoU = \frac{1}{\sum_{i=0}^{m} \sum_{j=0}^{n} \left(\sum_{k=0}^{l} \left(\frac{\sum_{p=0}^{m} p_i p_{ij}}{\sum_{p=0}^{m} \sum_{q=0}^{n} p_i p_{ij} p_{qj}}\right)\right)}
\]

Figure 6 is the landcover classification results with original dataset and adjusted dataset. The model is FPGA described in section 3.3. Overall, the classification accuracy (FWIoU) was improved from 0.7098 to 0.7325. In the first and second groups, the road with less pixels perform more completely in the adjusted dataset. The misclassification of natural forest, grass and dry land in the third group was also improved in the adjusted dataset.
Figure 3: Original samples with long-tail distribution

Figure 4: Samples after balancing processing

Figure 5: Distribution of training, testing, verification dataset

Figure 6: Comparison of extraction results before and after loss function optimization
Table 2 is the IoU before and after the Loss adjustment. It can be seen the accuracy of 16 categories among the 23 were improved, while the misclassification increased for river, canal and pond. River is misclassified as soil surface or sandy surface. This might be because they have similar spatial texture features.

| Category                  | Code | IoU Original | IoU Adjusted |
|---------------------------|------|--------------|--------------|
| Paddy field               | 2    | 0.9244       | 0.9283       |
| Dry farmland              | 3    | 0.7177       | 0.7496       |
| Tree-shrub orchard        | 5    | 0.0834       | 0.1000       |
| Nursery                   | 11   | 0.0899       | 0.0244       |
| Arboreal forest           | 21   | 0.4267       | 0.4702       |
| Shrubwood                 | 26   | 0.0000       | 0.0000       |
| Planted forest            | 34   | 0.2184       | 0.2339       |
| Natural grassland         | 36   | 0.1814       | 0.2017       |
| Planted grassland         | 40   | 0.5380       | 0.5890       |
| Building                  | 50   | 0.6350       | 0.6691       |
| Highway                   | 72   | 0.5716       | 0.5964       |
| Country road              | 74   | 0.3375       | 0.3737       |
| Other structure           | 80   | 0.0000       | 0.0000       |
| Open dump                 | 88   | 0.0001       | 0.0007       |
| Trampled surface          | 89   | 0.0283       | 0.0323       |
| Other hardened surface    | 90   | 0.2191       | 0.2315       |
| Greenhouse                | 110  | 0.5616       | 0.6564       |
| Dug land                  | 120  | 0.0024       | 0.0084       |
| Soil surface              | 142  | 0.0148       | 0.0012       |
| Sandy surface             | 143  | 0.0162       | 0.0211       |
| River                     | 152  | 0.7552       | 0.6474       |
| Canal                     | 153  | 0.4391       | 0.4363       |
| Pond                      | 157  | 0.4139       | 0.3986       |

Figure 7 is the confusion matrix of the classification results with original sample dataset. And Figure 8 is the confusion matrix of the classification results with adjusted sample dataset. We can also see that after adjustment, the accuracy of the model is improved, but misclassification still occur, such as dry farmland (code 3), arboreal forest (code 21), natural grassland (code 36), building (code 50), country road (code 74), trampled surface (code 89) be confused with nurse (11), Shrub-wood (26), other structure (80), open dump (88), dug land (120), Soil surface (142) and Sandy surface (143). This is because the model has insufficient discrimination ability for categories with small intra class differences. For example, dry farmland spectrum is mixed with vegetation and land. It should be classified from the aspect of spatial features. FPGA is more focus on hyperspectral feature and is insufficient for spatial features.

4. CONCLUSION

Unbalanced distribution of samples for intelligent interpretation of remote sensing image can lead to bias in the training model and make it difficult to ensure accuracy of interpretation. This paper proposed a weighted loss algorithm to strengthen the learning ability of the model for rare categories, and adopted an iterative stratification method for multi-label data classification to ensure that both training dataset and validation dataset contain reasonable proportion of landcover classes. Experiments on a large volume high spatial and spectral resolution dataset shows that the proposed methods improved the accuracy of landcover classification.
ACKNOWLEDGMENTS

This study was supported by Major Program of the National Natural Science Foundation of China (No.92038301).

REFERENCES

Bengio, S., 2015. Sharing representations for long tail computer vision problems, in: Proceedings of the 2015 ACM on International Conference on Multimodal Interaction. pp. 1–1.

Chen, F., Chen, J., Wu, H., Hou, D., Zhang, W., Zhang, J., Zhou, X., Chen, L., 2016. A landscape shape index-based sampling approach for land cover accuracy assessment. Science China Earth Sciences 59, 2263–2274. https://doi.org/10.1007/s11430-015-5280-5

Cui, Y., Jia, M., Lin, T.-Y., Song, Y., Belongie, S., 2019. Class-balanced loss based on effective number of samples, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9268–9277.

Grekousis, G., Mountrakis, G., Kavouras, M., 2015. An overview of 21 global and 43 regional land-cover mapping products. International Journal of Remote Sensing 36, 5309–5335.

Hinton, G.E., 2006. Reducing the Dimensionality of Data with Neural Networks. Science 313, 504–507. https://doi.org/10.1126/science.1127647

Wuhan University. Luojia-HSSR Dataset. https://github.com/yuexu9441/Luoija-HSSR-Dataset, 2022.

Huang, G.B., Learned-Miller, E., 2014. Labeled faces in the wild: Updates and new reporting procedures. Dept. Comput. Sci., Univ. Massachusetts Amherst, Amherst, MA, USA, Tech. Rep 14.

Khoshgoftaar, T.M., Van Hulse, J., Napolitano, A., 2010. Supervised neural network modeling: an empirical investigation into learning from imbalanced data with labeling errors. IEEE Transactions on Neural Networks 21, 813–830.

Nilsson, M., Griggs, D., Visbeck, M., 2016. Policy; map the interactions between Sustainable Development Goals. Nature 534, 320–322.

Rawat, W., Wang, Z., 2017. Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review. Neural Computation 29, 2352–2449, https://doi.org/10.1162/neco_a_00990

Szymański, P., Kajdanowicz, T., 2019. Scikit-multilearn: a scikit-based Python environment for performing multi-label classification. The Journal of Machine Learning Research 20, 209–230.

Szymański, P., Kajdanowicz, T., 2017. A network perspective on stratification of multi-label data, in: First International Workshop on Learning with Imbalanced Domains: Theory and Applications. PMLR, pp. 22–35.

UN, 2019. Sustainable development goals [WWW Document]. URL https://sustainabledevelopment.un.org/sdgs (accessed 4.1.19).

Xu, Y., Hu, X., Wei, Y., Yang, Y., Wang, D., 2019. A MACHINE LEARNING DATASET FOR LARGE-SCOPE HIGH RESOLUTION REMOTE SENSING IMAGE INTERPRETATION CONSIDERING LANDSCAPE SPATIAL HETEROGENEITY. International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences.

Zhang, Liangpei, Zhang, Lefei, Du, B., 2016. Deep Learning for Remote Sensing Data: A Technical Tutorial on the State of the Art. IEEE Geoscience and Remote Sensing Magazine 4, 22–40. https://doi.org/10.1109/MGRS.2016.2540798

Zhang, X., Fang, Z., Wen, Y., Li, Z., Qiao, Y., 2017. Range loss for deep face recognition with long-tailed training data, in: Proceedings of the IEEE International Conference on Computer Vision. pp. 5409–5418.

Zhang, X., Liu, L., Chen, X., Xie, S., Gao, Y., 2019. Fine landcover mapping in China using Landsat datacube and an operational SPECLib-based approach. Remote Sensing 11, 1056.

Zheng, Z., Zhong, Y., Ma, A., Zhang, L., 2020. FPGA: Fast patch-free global learning framework for fully end-to-end hyperspectral image classification. IEEE Transactions on Geoscience and Remote Sensing 58, 5612–5626.