A MACHINE LEARNING DATASET FOR LARGE-SCOPE HIGH RESOLUTION REMOTE SENSING IMAGE INTERPRETATION CONSIDERING LANDSCAPE SPATIAL HETEROGENEITY

Yue Xu¹, Xiangyun Hu^{1,*}, Yujun Wei¹, Ying Yang², Donghua Wang²

1. Wuhan University, Wuhan, China (email address: yuexu41@whu.edu.cn, huxy@whu.edu.cn, 2013302590157@whu.edu.cn) 2.National Geomatics of China, Beijing, China (email address: yangying@ngcc.cn, donghua@ngcc.cn)

KEY WORDS: Deep Learning, Sample Dataset, Sample distribution, Land cover and Land use, Landscape spatial Heterogeneity

ABSTRACT:

The demand for timely information about earth's surface such as land cover and land use (LC/LU), is consistently increasing. Machine learning method shows its advantage on collecting such information from remotely sensed images while requiring sufficient training sample. For satellite remote sensing image, however, sample datasets covering large scope are still limited. Most existing sample datasets for satellite remote sensing image built based on a few frames of image located on a local area. For large scope (national level) view, choosing a sufficient unbiased sampling method is crucial for constructing balanced training sample dataset. Dependable spatial sample locations considering spatial heterogeneity of land cover are needed for choosing sample images. This paper introduces an ongoing work on establishing a national scope sample dataset for high spatial-resolution satellite remote sensing image processing. Sample sites been chosen sufficiently using spatial sampling method, and divided sample patches been grouped using clustering method for further uses. The neural network model for road detection trained our dataset subset shows an increased performance on both completeness and accuracy, comparing to two widely used public dataset.

1. INTRODUCTION

The demand for timely information about the earth's surface, such as land cover and land use (LC/LU), is consistently increasing, because such information provides a base for many local, regional or global scope applications, e.g. resources monitoring, change monitoring, and environmental studies (Grekousis et al., 2015; Jun et al., 2014). Especially, recently large scope (national or global range) information extraction is in urgent demand, under the SDG30 goals set by UN for global sustainable development (UN, n.d.).

Reliable methods are needed for extracting geospatial information. Geospatial information with high spatial-temporal resolution and reliability are essential for monitoring global environmental and resources conditions. With the rapid development of space technology, remote sensing images have become main resources to acquire such large scope geospatial information (Song et al., 2018), while the efficiency of information extraction has been one of the main bottlenecks (Townshend and Justice, 1981).

On the other hand, Machine learning method shows its advantage on collecting such information from remotely sensed images while requiring sufficient training sample (Zhang et al., 2016). Artificial intelligence and computer vision technologies can help to mine the information from remote sensing data (Rawat and Wang, 2017; Zhang et al., 2016). Deep neural network based approach has been widely used on remote sensing image interpretation for various purpose, especially on land use and land cover monitoring (Zhang et al., 2016). Sample dataset is an important basis for deep learning applications. The well-known ImageNet (Deng et al., n.d.) for ground view images has more than ten thousand categories of objects which can be used to train

networks. With networks pertained on it, features with corresponding labels can be learnt, then further application like object detection (Shu et al., 2018) and semantic segmentation (Yujun et al., 2018) can be done more efficiently.



Figure 1. Main flow for building the sample dataset

For satellite remote sensing image, however, sample datasets covering large scope (such as national or regional scope) are still limited. Most existing datasets for satellite remote sensing image were built based on a few frames of image located on a local area (Demir et al., 2018; Zhang, 2017). This might cause the low generalization ability for trained models, because small sample size could be easily overfitted. Moreover, current sample datasets on satellite remote sensing images usually annotate pixels into only around ten categories (Demir et al., 2018; Xia et al., 2017;

^{*} Corresponding author

Zhang, 2017). The coarse classification might also cause huge inclass differences, which will make the trained models fragile while testing on larger data volume.

For constructing remote sensing image sample dataset with large scope (national level, for example) view, choosing a sufficient sampling method is also crucial for selecting sample images. When training dataset is imbalanced, i.e. the training samples are unevenly distributed among classes, Machine learning models often perform biased on the majority class, where tends to misclassify the minority class (Khoshgoftaar et al., 2010; Murphey et al., 2004). Dependable spatial sample locations considering spatial heterogeneity of land cover are needed (Chen et al., 2016) to obtain optimized sample distributions, sufficient sample numbers for rare classes, and adequate sample size for each region of geographical space.

To insure sufficient sample distribution, our approach considers unbalanced data problem while constructing the dataset. While previously, some researches deal with unevenly distributed training data by developing algorisms (Geng et al., 2016) or reshape dataset using resampling strategies (Estabrooks et al., 2004; Kubat and Matwin, 1997).



Figure 2. Sample location for choosing remote sensing image

This paper introduces an ongoing work on establishing a national scope sample dataset for high spatial-resolution satellite remote sensing image processing. We consider avoiding unbalanced sample distribution of LC/LU categories through different scale from national to the patch groups.

Figure 1 shows the main flow of construction the sample dataset: (1) To insure sample balance in national scope level, we select sample points with the method which takes landscape spatial heterogeneity into consideration;

(2)Based on the location of each sample point, find the corresponding image frame from the high-spatial resolution datasets with fully annotated pixel-level ground truth.

(3) Divide each image frame into patches. Then based on fully annotated pixel-level labels or image feature pattern, grouping small patches into specific clusters using unsupervised approach.(4) Every patch can be considered as a group of specific scene, by training based on the clustered scenes, specific interpretation model can learn features corresponding to each scene.

(5) Further data process method could be applied to our dataset based on clustering results, to insure sample balance among scope of scenes (patches). E.g. as a preparation for the uses like semantic segmentation for multiple classes, users can adjusting sample distribution based on clustering information.

Thus, this dataset would serve the use of large scale productions on land use and land cover monitoring. Based on grouped scenes, we can train multiple models as an ensemble optimization to improve the overall performance of LU/LC segmentation. In addition, image features extracted from each category of scene can be considered as one 'bag of features, allowing for similar image searching not only inside the dataset, also outside the dataset to expend the dataset scope to a global level.

Two Specific machine learning tasks could be set. (1) The grouped scene can serve as a multi-label scene classification task. (2) A semantic segmentation task for remote sensing image with all objects or specific objects. Besides, the pre-trained networks for scene classification can serve as a baseline for further using on other information extraction methods of remotely sensed images.

From a preliminary experiment, the neural network model for road detection trained our grouped dataset shows an increased performance on both completeness and accuracy, comparing to two widely used public dataset.

Section 2 of this paper introduces related methods, where the third section shows some results as well as some discussion. The conclusion and future works are introduced at the end of this paper.

2. METHOD

The main flow for constructing our dataset introduces in section 1 and shown in figure 1. To avoid unbalanced sample distribution through different scale from national to the scope within a patch, we adopted a spatial sampling method provided by Chen et al.(Chen et al., 2016) for national scope sample distribution and use a clustering method which is prepared for adjusting sample distribution among patches for further applications. Here we briefly introduce the related methods.

2.1 Sampling method

Sample points been selected from the 30-meter resolution land cover datasets (Jun et al., 2014) with the method proposed by Chen et.al. which taking into consideration of landscape spatial heterogeneity. Then, based on location of each sample point, we can find the corresponding image frame from the high-spatial resolution (higher than 1 meter) datasets established by the Chinese National Geographic Conditions Survey Program with fully annotated pixel-level ground truth (NGCC, n.d.).

The landscape index (LSI) is a quantitative measurement of landscape heterogeneity and is based on the ratio of the perimeter to area(Patton, 1975; Rutledge, 2003). In general, higher LSI is caused by more complex landscapes, as such landscapes have bigger borders for a given zone (Plexida et al., 2014; Zhou et al., 2014).

For given regions, Chen's (Chen et al., 2016) method can determine corresponding sample sizes and their spatial distributions according to landscapes classes. They use LSI as three levels to characterize the spatial heterogeneity: rLSI for a region, cLSI for land cover classes under such region and uLSI for each geographic sampling unit.

Specifically, for raster data, the basic LSI (Chen et al., 2012) is:

$$LSI = \frac{\sum b_j}{4\sqrt{qs^2}} = 0.25 \times \sum \frac{b_j}{\sqrt{q}}$$
(1)

Where q is the number of pixel and b is the number of edges for a pixel j. s is the spatial resolution.

Then for each region we can calculate its LSI called rLSI. It is used to calculate sample sizes N_i for each target region:

$$N_i = \frac{rLSI_i \times A_i}{\sum_{i=1}^{j=n} rLSI_j \times A_j} \times N, (i = 1, ..., n)$$
(2)

 A_i and A_j are the areas of region *i* and *j*, respectively. *N* is the total sample size. Thus, region that is more heterogeneous will have a higher sample density leading to a larger sample sizes.

The cLSIs are used to allocate sample numbers to land type classes. For class k of a given region i, the sample number is:

$$cN_{i,k} = N_i \times \frac{cLSI_{i,k} \times W_{i,k}}{\sum_{k=1}^{k=M} cLSI_{i,k} \times W_{i,k}}, (k = 1, ..., m)$$
(3)

Where, N_i is the sample size, $W_{i,k}$ is proportion of area, k. is the cLSI for a class

The uLSIs been used for adaptively choosing optimal geographic units for sample sites location according to spatial heterogeneity. Suppose that region *i* is composed by $R \times L$ geographical units.

$$\int_{k}^{i} uLSI_{r,l}$$
(4)
Is the uLSI of class k in row r and column l.

Sample site location can be selected with uLSI curves. uLSI curves are derived by descending rank of the geographical units based on the uLSIs for each class.

2.2 Clustering method

We also want to prepare samples groups for further applications. With derived sample points from method described in section 2.1, remote sensing image frames can be found based on locations. After that, we divide each image frame into 512*512 patches. Then based on fully annotated pixel-level labels, we further divided patches into specific clusters using unsupervised meanshift clustering method, considering categories appearance or image texture pattern depends on different application purpose.

Since the number of clustered scenes as a priori is unknown, clustering the patches with parametric strategies like k-means are difficult. We choose to use the mean-shift procedure (Comaniciu and Meer, 2002), a non-parametric density based method. A feature vector is generated for each divided patch, which is formed with a 10 dimensional histogram: 10 bins for 10 first-classes of land cover categories form ground truth image. Image texture as further cues can be added by extending the vector, 22 bins for LBP features derived from corresponding remote sensing image. Uniform LBPri (Ojala et al ., 2002) is used for describing small-scale appearance (textures) of the image and which is rotation invariance with uniform patterns. This kind of cue is prepared as a clustering option for further uses of our dataset. The main flow for clustering is:

- 1) Derive image features using uniform LBPri
- 2) Histogram for image texture and semantic label
- 3) Mean-shift (Comaniciu and Meer, 2002) for scene clustering

With these steps, we categorized pre-divided sample patches into multiple scenes for overall ensemble model for large scale image processing. Each scene can be considered as one 'bag of features', allowing for similar image searching among the whole dataset and for further training specific interpretation models.

3. RESULT AND DISCUSSION

With the above approach, we generated 1602 random sampled points using the tool GLCVal (NGCC, n.d.) developed by the National Geomatic Center of China from Global Landcover30 dataset with ten categories (Grasslands, Forrest, Cultivated land, Bareland, Shrublands, Artificial surfaces, Water bodies and Permanent snow/ice). Based on these sample points, about 1000 high-resolution image frames (Multi data source, mainly 1 meter resolution image from Chinese Gaofen 2 Satellite (CRESDA, n.d.) and 0.5 meter aerial image. with supplement of 2 meter image from Chinese ZY-3 Satellite (Deren, 2016) and other image) were found from the Chinese National Geographic Conditions Survey Dataset, and more than 1 million patches of 512*512 images were generated. Considering the semantic labels' ontology components and their corresponding image texture using uniform LBP, we categorized these patches into multiple scenes for further specific model training.

30m Land cover point type	Count	
Artificial surfaces	161	
Barren land	199	
Cultivated land	216	
Forests	227	
Grass lands	234	
Permanent snow and ice	126	
Shrub lands	185	
Water bodies	154	
Wetland	100	
Grand Total	1602	

A.		
		<u>(</u>

Table 1. Number of sample location for each Land cover type

Figure 3. Divided patches and ground truth

The resulted sample site locations followed the spatial heterogeneity across China, which insures sufficient number of sample for each land cover and land use category has been chosen. Figure 2 shows sample locations for choosing remote sensing image. As can be seen from the image, the distribution of sample locations followed the distribution pattern of land use and land cover type cross the whole China. Artificial surfaces, cultivated land and forests are denser in the eastern region than western region, while barren land and shrub land sample points occurred more in the western area of China. Sufficient sample numbers for each class have been generated. Table 1 shows the number of

sample location for each land cover type under the 30 meter global land use and land cover production.

The divided patches show a high quality of annotation result, which make this production a perfect source for feeding deep neural networks. Figure 3 shows some divided patches and corresponding ground truth. With fully annotated pixel-level ground truth generated by the Chinese National Geographic Conditions Survey Program, which has been validated by visual interpretation and field verification. In this dataset the land cover and land use objects have been classified into 10 first-class, 59 second-class and 143 third-class categories ("Content and Evaluation Indicators for Chinese National Geographic Conditions Survey Program," 2017). Appendix 1 shows one of the first class category, cultivated land, with its second and third categories.



Figure 4. Cluster result, upper side: labels, lower side: images.

Since the overall accuracy has been validated via field survey, this production becomes a perfect source of sample making for deep learning. The accuracy is within 5 pixels, or 10 pixels under shadows or shaded condition("Regulations on Inspection, Acceptance and Quality Assessment for Chinese National Geographic Conditions Survey Program," 2013).

Every patch can be considered as a specific scene, by training based on the categories of each scene clusters, networks can learn features corresponding to specific objects in scene. Figure 4 shows the clustering result using the mean-shift clustering method considering the semantic labels' appearance and image texture.

Dataset	Completeness	Accuracy
Massachusetts Road dataset	39.6%	90.7%
Massachusetts Road dataset +	69.0%	69.0%
deepglobe		
ours	71.1%	80.1%

Table 2. Results of different datasets

As a preliminary experiment, we implemented a deep learning approach proposed by Wei et al. (Yujun et al., 2018) for road centreline extraction. Then test the result completeness and accuracy (Heipke et al., n.d.) on a same frame of ZY3 satellite image.

$$Completeness = \frac{TP}{TP + FN}$$
(5)
$$Correctness = \frac{TP}{TP + FP}$$
(6)

Where TP, FN and FP is the length of correctly extracted roads, missing roads and falsely extracted roads, respectively.

We trained three models for comparison. First, trained one network on the Massachusetts Road dataset (Mnih, 2013), then trained one on the combination of two datasets: Massachusetts Road dataset and deepglobe dataset (Demir et al., 2018). Finally we trained the network on our dataset. As shown in table 2. For same model on same remote sensing image, the completeness

increased from 39.6% to 71% for Massachusetts Road dataset and our dataset. Where the accuracy increased from 69% to 80.15% for the combined dataset and ours. The accuracy on the Massachusetts road dataset is not comparable due to its low completeness. Figure 5 shows a part of extracted result of the road centreline. Result from our road category (Figure 5.c) shows an obvious improvement on completeness comparing to Massachusetts Road dataset (Figure 5.a). while the completeness difference of Figure 5.b and Figure 5.c is not obvious, from Table 2, the improvement of overall accuracy calculated from the whole frame of remote sensing is still significant. The improvement on completeness and accuracy indicating that our dataset included sufficient variance of images for such category.



Figure 5. Part of interpreting result of road category. a.: result from Massachusetts Road dataset; b.: result from Massachusetts Road dataset + deep global; c: result from out dataset; d: GT

4. CONCLUSION

This paper introduces an ongoing work on establishing a national scope sample dataset for high spatial-resolution satellite remote sensing image processing. The neural network model for road-centreline detection trained our dataset obtained a performance increase on both completeness and accuracy, comparing to two widely used public dataset.

Two Specific machine learning tasks could been set. (1) The grouped scene can serve as a multi-label scene classification task. (2) A semantic segmentation task for remote sensing image with all objects or specific objects.

For future works, feature information extracted from categories of our grouped scenes can be used for overall ensemble model for large scale production. Each category of scene can be considered as one 'bag of features, allowing for further training interpretation models and for similar image searching within the dataset or searching outside the dataset to expending the scope to a global level,

Further assessments on improvement of method performance using our dataset will conduct in the future. E.g. semantic segmentation for multi-objects should be conducted as well as its overall performance assessment. The dataset or pre-trained model could possibly become public for research purpose.

ACKNOWLEDGEMENTS

This research is sponsored by National Key Research and Development Projects (No.2016YFB0501400).

REFERENCES

- 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
- Chen, X., Yamaguchi, Y., Chen, J., 2012. Weighted misclassification rate: a new measure of classification error designed for landscape pattern index. Remote sensing letters 3, 57–65.
- Comaniciu, D., Meer, P., 2002. Mean shift: a robust approach toward feature space analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence 24, 603– 619. https://doi.org/10.1109/34.1000236
- Content and Evaluation Indicators for Chinese National Geographic Conditions Survey Program, 2017.
- CRESDA, n.d. China Centre For Resources Satellite Data and Application, www.cresda.com/CN/index.shtml (accessed 4.1.19).
- Demir, I., Koperski, K., Lindenbaum, D., Pang, G., Huang, J., Basu, S., Hughes, F., Tuia, D., Raska, R., 2018. Deepglobe 2018: A challenge to parse the earth through satellite images, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). IEEE, pp. 172– 17209.
- Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L., n.d. ImageNet: A Large-Scale Hierarchical Image Database 8.
- Deren, L.I., 2016. China's first civilian three-line-array stereo mapping satellite: ZY-3. Acta Geodaetica et Cartographica Sinica 41, 317–322.
- Estabrooks, A., Jo, T., Japkowicz, N., 2004. A multiple resampling method for learning from imbalanced data sets. Computational intelligence 20, 18–36.
- Geng, M., Wang, Y., Tian, Y., Huang, T., 2016. CNUSVM: Hybrid CNN-Uneven SVM Model for Imbalanced Visual Learning, in: 2016 IEEE Second International Conference on Multimedia Big Data (BigMM). Presented at the 2016 IEEE Second International Conference on Multimedia Big Data (BigMM), IEEE, Taipei, Taiwan, pp. 186–193. https://doi.org/10.1109/BigMM.2016.19
- 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.
- Heipke, C., Mayer, H., Wiedemann, C., n.d. EVALUATION OF AUTOMATIC ROAD EXTRACTION 10.
- Jun, C., Ban, Y., Li, S., 2014. China: Open access to Earth landcover map. Nature 514, 434.
- 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.
- Kubat, M., Matwin, S., 1997. Addressing the curse of imbalanced training sets: one-sided selection, in: Icml. Nashville, USA, pp. 179–186.
- Mnih, V., 2013. Machine learning for aerial image labeling. University of Toronto (Canada).

- Murphey, Y.L., Guo, H., Feldkamp, L.A., 2004. Neural learning from unbalanced data. Applied Intelligence 21, 117– 128.
- NGCC, n.d. Chinese National Geographic Conditions Survey , http://chzt.sbsm.gov.cn/article/zxgz/dlgqjc/ (accessed 4.1.19a).
- NGCC, n.d. GLCVal, URL glcval.geo-compass.com (accessed 4.1.19b).
- Ojala, T., Pietikainen, M., Maenpaa, T., 2002. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence 24, 971– 987. https://doi.org/10.1109/TPAMI.2002.1017623
- Patton, D.R., 1975. A diversity index for quantifying habitat" edge". Wildlife Society Bulletin (1973-2006) 3, 171– 173.
- Plexida, S.G., Sfougaris, A.I., Ispikoudis, I.P., Papanastasis, V.P., 2014. Selecting landscape metrics as indicators of spatial heterogeneity—A comparison among Greek landscapes. International Journal of Applied Earth Observation and Geoinformation 26, 26–35.
- 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
- Regulations on Inspection, Acceptance and Quality Assessment for Chinese National Geographic Conditions Survey Program, 2013.
- Rutledge, D.T., 2003. Landscape indices as measures of the effects of fragmentation: can pattern reflect process?
- Shu, Z., Hu, X., Sun, J., 2018. Center-Point-Guided Proposal Generation for Detection of Small and Dense Buildings in Aerial Imagery. IEEE Geoscience and Remote Sensing Letters 15, 1100–1104. https://doi.org/10.1109/LGRS.2018.2822760
- Song, X.-P., Hansen, M.C., Stehman, S.V., Potapov, P.V., Tyukavina, A., Vermote, E.F., Townshend, J.R., 2018. Global land change from 1982 to 2016. Nature 560, 639.
- Townshend, J., Justice, C., 1981. Information extraction from remotely sensed data. A user view. Remote Sensing 2, 313–329.
- UN, n.d. Sustainable development goals, https://sustainabledevelopment.un.org/sdgs (accessed 4.1.19).
- Xia, G.-S., Bai, X., Ding, J., Zhu, Z., Belongie, S., Luo, J., Datcu, M., Pelillo, M., Zhang, L., 2017. DOTA: A Large-scale Dataset for Object Detection in Aerial Images. arXiv:1711.10398 [cs].
- Yujun, W., Xiangyun, H., Jinqi, G., 2018. End-to-End Road Centerline Extraction via Learning a Confidence Map, in: 2018 10th IAPR Workshop on Pattern Recognition in Remote Sensing (PRRS). IEEE, pp. 1–5.
- 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, M., 2017. Learning Dual Multi-Scale Manifold Ranking for Semantic Segmentation of High-Resolution Images. Remote Sensing 9.
- Zhou, W., Cadenasso, M., Schwarz, K., Pickett, S., 2014. Quantifying spatial heterogeneity in urban landscapes: integrating visual interpretation and object-based classification. Remote Sensing 6, 3369–3386.

Id	First class	Second class	Third	Mini
			class	mum
				area
0100	Cultivated			400 m
	land			
0110		paddy field		400 m
0120		Dry land		400 m
0130		orchard		400 m
0131			Shrub	400 m
0132			Liana	400 m
0133			Herbal	400 m
0140		Tea garden		400 m
0150		mulberry field		400 m
0160		rubber plantation		400 m
0170		nursery garden		400 m
0180		flower nursery		400 m
0190		Other Economic		400 m
		seedlings		
0191			Shrub	400 m
0192			Liana	400 m
0193			Herbal	400 m

APPENDIX

Appendix 1: Example of categories from Chinese National Geographic Conditions Survey Program