Unsupervised Superpixel-Driven Parcel Segmentation of Remote Sensing Images Using Graph Convolutional Network

Fulin Huang*†
University of Toronto
Toronto, ON, Canada

Zhicheng Yang*
PAII Inc.
Palo Alto, CA, USA

Hang Zhou
PAII Inc.
Palo Alto, CA, USA

Chen Du
PAII Inc.
Palo Alto, CA, USA

Andy J.Y. Wong
PAII Inc.
Palo Alto, CA, USA

Yuchuan Gou
PAII Inc.
Palo Alto, CA, USA

Mei Han
PAII Inc.
Palo Alto, CA, USA

Jui-Hsin Lai‡
PAII Inc.
Palo Alto, CA, USA

ABSTRACT

Accurate parcel segmentation of remote sensing images plays an important role in ensuring various downstream tasks. Traditionally, parcel segmentation is based on supervised learning using precise parcel-level ground truth information, which is difficult to obtain. In this paper, we propose an end-to-end unsupervised Graph Convolutional Network (GCN)-based framework for superpixel-driven parcel segmentation of remote sensing images. The key component is a novel graph-based superpixel aggregation model, which effectively learns superpixels' latent affinities and better aggregates similar ones in spatial and spectral spaces. We construct a multi-temporal multi-location testing dataset using Sentinel-2 images and the ground truth annotations in four different regions. Extensive experiments are conducted to demonstrate the efficacy and robustness of our proposed model. The best performance is achieved by our model compared with the competing methods.

CCS CONCEPTS

• **Computing methodologies** → **Image segmentation**; **Unsupervised learning**; • **Mathematics of computing** → **Graph algorithms**.

KEYWORDS

parcel segmentation, unsupervised learning, graph convolutional network, remote sensing images

*Fulin Huang and Zhicheng Yang equally contributed to this work.
†This work was done during Fulin Huang's internship at PAII Inc., USA.
‡Corresponding Author: Jui-Hsin Lai (email: Juihsin.lai@gmail.com)

WWW '22 Companion, April 25–29, 2022, Virtual Event, Lyon, France

ACM ISBN 978-1-4503-9130-6/22/04…$15.00
https://doi.org/10.1145/3487553.3524716

ACM Reference Format:
Fulin Huang, Zhicheng Yang, Hang Zhou, Chen Du, Andy J.Y. Wong, Yuchuan Gou, Mei Han, and Jui-Hsin Lai. 2022. Unsupervised Superpixel-Driven Parcel Segmentation of Remote Sensing Images Using Graph Convolutional Network. In *Companion Proceedings of the Web Conference 2022 (WWW '22 Companion), April 25–29, 2022, Virtual Event, Lyon, France.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3487553.3524716

1 INTRODUCTION

Parcel segmentation is a building block of many environmental remote sensing applications, such as crop classification and growth monitoring [12, 20, 27], land use change detection [26], etc. These applications inform governance and business decisions related to food security, climate change, and environmental protection. The vast majority of existing parcel segmentation tasks are based on supervised learning methods, which require precise parcel-level ground truth annotation in the target area [2, 8, 9]. This requirement has almost become indispensable in the era of deep learning. While satellite images provide a wealth of spatial, temporal and spectral information of the earth surface, annotating parcel-level reference is time-consuming and labor-intensive. As a result, those supervised learning-based algorithms suffer from unsatisfactory generalization in other regions. Some existing datasets are constructed by per-pixel classification [28], which inherently have inevitable salt-and-pepper noise, thus hindering their usability.

Unsupervised learning-based segmentation methods, on the other hand, do not need expensive ground truth information during the learning process. These methods purely rely on image content to accomplish a segmentation task instead, leading to much better generalization capacity [3, 5]. Specifically, *superpixel* is widely used in remote sensing segmentation tasks, which is a group of pixels that share similar properties [22]. A superpixel-level result can facilitate image processing and significantly eliminate the salt-and-pepper noise. With the vigorous development of deep learning nowadays, a superpixel output is commonly used as an intermediate result or guidance to achieve better performance in supervised learning-based segmentation tasks in remote sensing [16, 17, 19]. However, their proposed methods still require ground truth training data.

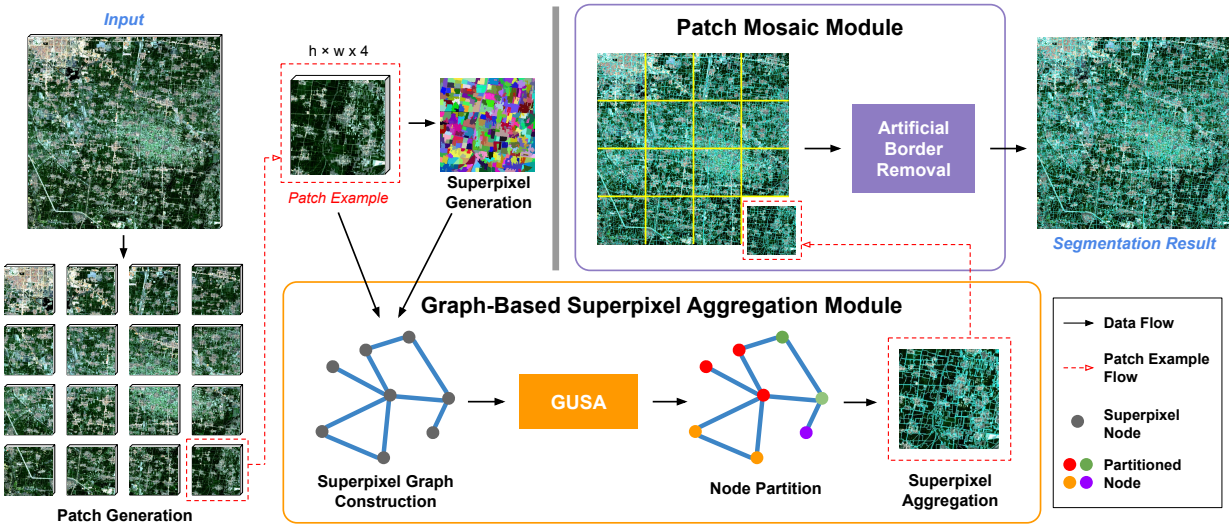

Figure 1: The proposed framework.

Recently, Graph Convolutional Networks (GCNs) [11] have empowered numerous applications in the Web and social good, computer vision, and natural language processing [31]. Among various graph problems, *graph partitioning* aims to divide the vertex set under constraints, such that the edge cut across the partitions is minimized [4]. Since a superpixel can be transformed into a node in a graph, it is possible to leverage GCNs to learn the latent relationship among superpixels and partition them into a few larger segments without ground truth. These larger segments, which are the aggregations of superpixels, are visually the segmentation result of an image.

In this work, we propose an unsupervised GCN-based framework of superpixel-driven parcel segmentation. We use Sentinel-2 data as our image source because of their high spatial resolution (10m) and high temporal resolution (5-day revisiting interval) as well as free accessibility [7]. Our key contributions are listed below:

- To the best of our knowledge, this is the first end-to-end unsupervised GCN-based framework for superpixel-driven parcel segmentation of remote sensing images. It incorporates the powerful graph-learning capacity of GCNs and the great generalization of superpixels.
- In our framework, we design GUSA (Graph-based Unsupervised Superpixel-Aggregation) by modifying the network structure and loss function of a GCN-based model, dedicated to effectively learning the latent affinity relationship among superpixels and better aggregating similar ones in spatial and spectral spaces.
- We conduct extensive experiments on our multi-temporal multi-location Sentinel-2 image dataset to demonstrate the efficacy, robustness, and generalization of GUSA. In particular, GUSA achieves best performance compared with the competing methods. The newly defined hyper-parameters in GUSA are also validated by ablation studies.

The rest of this paper is organized as follows. Related studies are reviewed in Sec. 2. We elaborate on the proposed method in Sec. 3. The experiment setup and results are presented in Sec. 4. We conclude this paper in Sec. 5.

2 RELATED WORK

Conventional unsupervised learning algorithms for superpixel aggregation are mainly based on the idea of Normalized Cut [21] on a graph, which calculates the cut cost as a fraction of all nodes' edge connections, and further based on a bipartite graph [15, 25, 29]. However, those traditional methods still suffer from heavy computational complexity when the number of superpixels increases [30], which are not quite feasible for large-scale model deployments.

Many deep learning-based approaches incorporate superpixels in their proposed frameworks [6, 16, 17, 19, 23]. An affinity loss is designed to improve the superpixel segmentation [23], which is also adopted in remote sensing tasks [17]. Nevertheless, the objectives of these usages are to either enhance the generation of superpixel itself or further boost the entire supervised learning tasks. Recently, some superpixel-guided unsupervised frameworks are proposed for image segmentation such as Unsupervised Image Segmentation by Backpropagation (UISB) [10] and Deep Image Clustering (DIC) [32]. They utilize Convolutional Neural Networks (CNNs) to learn the spectral features and calculate the iterative refinement loss guided by a superpixel segmentation result, but they do not well emphasize the subtle spatial affinity among superpixels.

To solve the graph partitioning problem, different from supervised GCN-based learning methods [31], an unsupervised GCN-based graph partitioning framework Generalizable Approximate Graph (GAP) is presented in [18], nonetheless, its usability for image segmentation tasks is not well studied.

3 METHODOLOGY

3.1 Framework Overview

Fig. 1 briefly illustrates our proposed framework. A large remote sensing image input is first cropped into smaller patches for efficient

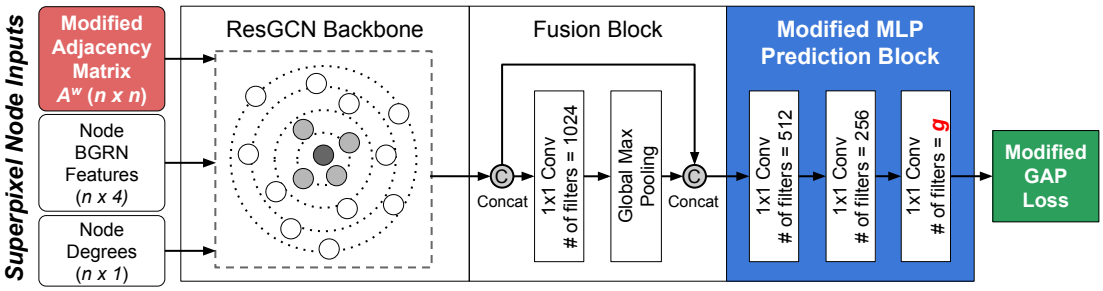

Figure 2: GUSA architecture.

processing. Superpixel generation is next performed on each patch. An image patch and its superpixel result together construct a *superpixel graph* that is fed into the Graph-Based Superpixel Aggregation Module, where the superpixel graph is well learned and partitioned by our designed GUSA. The partition result is equivalent to the superpixel aggregation of the image patch. In the Patch Mosaic Module, every patch is mosaicked back into the whole image size. The next artificial border removal procedure is able to eliminate fake resultant boundaries at the patch mosaicking borders. The processed output is the final parcel segmentation result of the entire input image. We will detail each step in the following sub-sections.

3.2 Patch Generation and Superpixel Generation

As a single Sentinel-2 image has 10,980×10,980 pixels, it is rare to directly train it on a deep learning model due to hardware limitation and expensive computational overhead. Instead, we crop a whole image into smaller patches. In our case using Sentinel-2 images, a cropped patch has a fixed size of height and width with 4 channels (blue, green, red, and near-infrared bands (BGRN)). Each patch is then processed individually before the patch mosaic module. Regarding superpixel generation, we adopt the Simple Non-Iterative Clustering (SNIC) algorithm because of its overall satisfaction in terms of visual quality and compactness [1].

3.3 Graph-Based Superpixel Aggregation Module

In this module, a superpixel graph is first constructed by the original image patch and its superpixel result, where each node of the graph represents a superpixel. The mean BGRN values of all pixels inside a superpixel contribute to its 4-dimensional node features. Recently, GAP is proposed to partition a graph in an unsupervised manner [18]. It accepts three inputs (node degree, node features, and adjacency matrix), and has two modules for graph embedding and partitioning, respectively. A trainable multi-task loss function is designed for minimizing a continuous relaxation format of normalized cut and a new balanced cut without ground truth. However, it is not quite suitable for *superpixel graph partitioning* due to the following issues:

(1) The GCN in the graph embedding module of GAP suffers from the vanishing gradient problem, limiting itself to shallow models. Another limitation is that the graph edges in GCN are fixed so that the relationship of a superpixel node and its neighbors are not dynamically learned during training.

(2) The graph partitioning module of GAP comprises fully connected layers, which is not the ideal structure to maintain spatial information when reducing the length of channels.

(3) The adjacency matrix of superpixel nodes should include both spatial and spectral affinities.

Therefore, we design GUSA, by modifying the architecture and the loss function of the GAP model to effectively consider the specificity of a superpixel graph.

Modification to the architecture. We leverage DeepGCN [14] inside the GUSA to overcome the issues listed above as shown in Fig. 2. The DeepGCN exploits the ResGCN backbone, which adds residual connections between the input and output layers, to alleviate the vanishing gradient problem. By using a Dilated K-nearest-neighbors (KNN) function, the DeepGCN can dynamically change neighbors in the GCN to mitigate the over-smoothing issue and learn better graph representations. This is an advantage over the GCN in which only vertex features are updated at each iteration. The fusion block fuses the global features as well as local features from the ResGCN backbone. Suppose that a superpixel graph has n nodes, and the expected number of aggregated partitions after GUSA is given as g. The modified MLP prediction block comprises several 1×1 convolutional layers to maintain the spatial information and assign n nodes to g partitions. Consequently, the graph partitioning module of GAP is not necessary, since DeepGCN is regarded as an improved holistic combination of the graph embedding and partitioning modules of GAP.

Modification to the loss function. As shown in Fig. 2, we design the new adjacency matrix to consider both spatial and spectral affinities, and define two new hyper-parameters to adjust the dominant terms in the loss function. We modify the original loss function in GAP as follows:

$$L = \overbrace{\sum_{\text{reduce-sum}} (Y \oslash \Gamma)(\mathbf{1} - Y)^T \odot A^w}^{\text{normalized cut loss}} + \sigma \overbrace{\sum_{\text{reduce-sum}} (\mathbf{1}^T Y - \frac{n}{g})^2}^{\text{balanced cut loss}}, \tag{1}$$

where $\sigma > 0$ denotes one new hyper-parameter to investigate the importance of the balanced cut loss, and $A^w = (A^w_{ij})_{n \times n}$ represents the weighted adjacency matrix. All other terms have the same definitions as the original version in GAP [18]. $Y \in \mathbb{R}^{nxg}$ is a probability matrix that represents the probability of a node

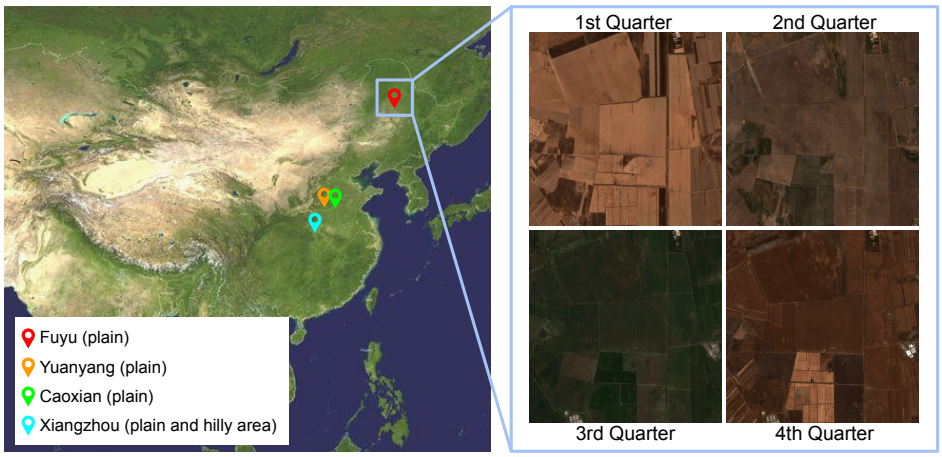

Figure 3: Physical locations and terrain types of counties in our dataset and example of land cover visualization in four quarters.

belonging to a partition. $\Gamma = Y^T D$ calculates the expected value of node degrees on each partition. $\oslash$ means element-wise division, and $\odot$ means element-wise multiplication. The matrix element A^w_{ij} of A^w is defined as:

$$A^w_{ij} = \delta c_{ij} + (1-\delta)e^{-\beta d_{ij}}, \tag{2}$$

where c_{ij} is 1 if superpixel nodes i and j are spatially adjacent, otherwise 0; $e^{-\beta d_{ij}}$ represents the weighted similarity in spectral space; d_{ij} is the Euclidean distance between the average BGRN spectral values of two nodes i and j, and $\beta > 0$ is a weight to control the significance of d_{ij}; $\delta \in [0, 1]$ denotes the other new hyper-parameter to adjust the balance of spatial and spectral affinities.

When the training process is completed, the output $n \times g$ matrix indicates every superpixel's partition class. Visually, the adjacent superpixels with the same partition label appear aggregated together (i.e. the "superpixel aggregation" result in Fig. 1).

3.4 Patch Mosaic Module

Once all patches are processed, they are fed into the Patch Mosaic Module and stitched back together into the input image size. However, due to the individual patch-based result, artificial borders are present on the patch edges, leading to fake segmentation boundaries there. The yellow lines in the Patch Mosaic Module in Fig. 1 illustrate this effect. Therefore, we design a procedure of artificial border removal to appropriately eliminate those artifacts and merge the similar segments at a shared patch border across the adjacent patches. The *full lambda schedule algorithm* [13] is utilized to calculate the merging cost of two segments S_i and S_j, which is defined as:

$$C(S_i, S_j) = \frac{\frac{a_i \cdot a_j}{a_i + a_j} d^2_{S_i S_j}}{\text{len}(\partial(S_i, S_j))} < \lambda \tag{3}$$

where a_i denotes the area of S_i (i.e. the number of pixels of S_i); $d_{S_i S_j}$ is the spectral Euclidean distance between S_i and S_j; $\text{len}(\partial(S_i, S_j))$ refers to the length of the shared border of the segments S_i and S_j. If the merging cost $C(S_i, S_j)$ is less than the pre-defined threshold λ, S_i and S_j are merged and their shared border is removed. If the segments have a large common border or small Euclidean distance value, they have a higher chance to merge. To facilitate this process, the patches are first merged horizontally to remove vertical artificial borders, and a vertical merging is then processed to remove horizontal ones.

4 EXPERIMENTS

4.1 Dataset, Evaluation Metric and Implementation Details

Our testing dataset is built using Sentinel-2 images over four county regions which have parcel-level ground truth labels. The four counties are located in the areas of major grain production in China and cover a total area of ~4600 km^2, including Fuyu, Yuanyang, Caoxian, and Xiangzhou. The left figure of Fig. 3 shows their geographic locations. Since the ground truth reference was annotated in 2019, we acquire and decloud all valid images in 2019 from Sentinel-2 data, excluding the defective/corrupted ones or those that have dense clouds. Every image is cropped into patches with the size of 512×512. There are total ~9,800 patches to cover the four testing area in 2019. The right figure of Fig. 3 presents an image patch example of land cover variations at Fuyu county throughout four quarters.

We adopt the commonly used Probabilistic Rand Index (PRI) [24] as our segmentation evaluation metric, which quantifies the partition similarity between the segmentation result and the ground truth, ranging from 0 to 1. A higher PRI value means a better segmentation result. We calculate the average PRI value over multi-temporal images to evaluate the performance of a model.

We modify the SNIC implementation to support 4-channel image patch input. In the DeepGCN model of GUSA, we adopt 28 GCN layers. The maximum number of neighbors of a node is set to 8. We assign 30 to the weight β of spectral Euclidean distance. The learning rate, the dropout rate, and the decay rate are set to 0.001, 0.3, and 0.5, respectively. The merging cost λ is set to 30,000. The newly defined hyper-parameters will be evaluated in Sec. 4.3. All

(a) Original image (b) Ground truth (c) DIC [32]

(d) UISB [10] (e) GAP [18] (f) GUSA (ours)

Figure 4: Visualized results of our proposed GUSA and competing methods on an image patch example in Yuanyang County in 2019.

our codes are implemented in Python 3.9.4 and PyTorch 1.9.0, and sped up with 2 NVIDIA Tesla V100 32GB GPUs.

4.2 Comparison of Different Methods

To perform a timely and fair comparison, we include the recent DIC [32], UISB [10], and GAP [18] as the competing methods, since they are unsupervised, deep learning-based, and superpixel-involved models. Every model uses the same superpixel generation input and the number of partitions g. Table 1 lists the average PRI values of parcel segmentation results on our dataset using the four models. As we can see, while graph-based models have better performance, GUSA achieves the best performance compared with other models across all counties. For each county, Fuyu has the highest PRI values thanks to its simpler parcel layout and lower urban density. On the contrary, the hilly topography in Xiangzhou indeed impacts on the segmentation results of all four models due to the more irregular parcel shapes and distribution. Yuanyang and Caoxian counties are located very close to each other, so their terrains and parcel layouts are comparable, leading to similar PRI results.

Table 1: Performance (PRI) comparison of our proposed GUSA with competing methods using our testing dataset.

Method	Fuyu	Yuanyang	Caoxian	Xiangzhou
DIC [32]	0.7071	0.6381	0.6238	0.6013
UISB [10]	0.7784	0.7522	0.7344	0.6890
GAP [18]	0.8509	0.8056	0.8022	0.7560
GUSA (ours)	**0.8826**	**0.8439**	**0.8415**	**0.8137**

Fig. 4 provides visualized results of an image patch example in Yuanyang county. Particularly, both DIC and UISB maintain the boundaries between urban and cropland areas, but DIC undergoes the effect of rapid convergence to over-aggregate cropland parcels, and UISB have lots of tiny segments. Since the spatial affinity among superpixels is not learned well in these models, the cropland parcel boundaries are obviously ignored. For graph-based models, GAP obtains clearer parcel boundaries but is unable to separate the croplands from the urban areas well. Many incorrect segments can be thus found at the actual urban boundaries. GUSA achieves better

Table 2: Performance (PRI) of different hyper-parameter settings in the loss function of GUSA in Yuanyang county.

σ (in Eq. 1)	δ=0	δ=0.3	δ=0.7	δ=1
0.1	0.6817	0.7608	0.8121	0.7872
1	0.8203	0.8328	**0.8439**	0.8374
10	0.8100	0.8125	0.8138	0.8126

Table 3: Performance of GUSA in different quarters and counties.

Quarter	Fuyu	Yuanyang	Caoxian	Xiangzhou
First	**0.8847**	0.8472	**0.8448**	0.8062
Second	0.8816	**0.8497**	0.8435	0.8104
Third	0.8803	0.8415	0.8406	**0.8225**
Fourth	0.8837	0.8373	0.8371	0.8157

parcel segmentation results as well as the clear boundaries of urban and cropland areas. We believe that the reasons are threefold. First, DeepGCN learns the graph embedding better than the conventional GCN by mitigating the gradient vanishing problem. Second, the fully connected layers inside GAP are not desirable to preserve superpixels' spatial information. Third, we improve the adjacency matrix to consider superpixels' affinities in spatial and spectral spaces.

4.3 Ablation Studies

Analysis of hyper-parameters in the GUSA loss function. For efficient ablation experiments, we fix one representative county Yuanyang because it has a relatively balanced percentage of urban and agricultural areas, and the model performance in this county is intermediate. We set different orders of magnitude for σ and create intervals for δ. Table 2 shows the average PRI results of GUSA under various combinations of the parameter values. The best result is achieved when σ and δ are 1 and 0.7, respectively. The settings of σ and δ indicate that (i) the normalized cut loss and balanced cut loss synergistically contribute to the learning process; (ii) significantly decreasing the dominance of balance cut loss (σ=0.1) notably degrades the performance; (iii) in an adjacency matrix, similarity in the spectral space is valuable and even more dominant but the spatial superpixel adjacency is not ignored; (iv) purely relying on similarity in either spectral space (δ=0) or spatial space (δ=1) is not ideal.

Quarters in different counties. As shown in the right figure of Fig. 3, images even in the same region can vary a lot in different seasons due to environmental change and plant growth. We evaluate the performance of GUSA in four individual quarters on our dataset. The PRI values in Table 3 demonstrate the overall robustness of GUSA across the different seasons. We also find that the performance is most stable in Fuyu but fluctuates in Xiangzhou. We believe that one of the reasons is the topographic simplicity and low urbanization in Fuyu. Another observation is that Fuyu has relatively better performance in cold seasons, Yuanyang and Caoxian enjoy outstanding results in the first half-year, and Xiangzhou achieves the best PRI values in the third quarter. These differences are potentially caused by intra-annual climate and vegetation variability.

5 CONCLUSIONS AND FUTURE WORK

In this paper, we propose the first end-to-end unsupervised GCN-based framework for superpixel-driven parcel segmentation of remote sensing images. A dedicated model GUSA is designed to effectively learn the latent affinity among superpixels and better aggregate similar ones in spatial and spectral spaces. Extensive experiments are conducted on our multi-temporal multi-location Sentinel-2 image dataset to demonstrate the outstanding performance and robustness of our proposed framework. The newly defined hyper-parameters in GUSA are also validated using ablation studies. We are currently annotating more regions to expand our dataset size. We will also investigate the effectiveness of some edge-enhanced approaches to improve the performance of our proposed method.

ACKNOWLEDGMENTS

We thank Mr. Saurabh Dash from Georgia Institute of Technology, USA for his comments on the code implementation.

REFERENCES

[1] Radhakrishna Achanta and Sabine Susstrunk. 2017. Superpixels and polygons using simple non-iterative clustering. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 4651–4660.
[2] Muhammad Alam, Jian-Feng Wang, Cong Guangpei, LV Yunrong, and Yuanfang Chen. 2021. Convolutional Neural Network for the Semantic Segmentation of Remote Sensing Images. *Mobile Networks and Applications* 26, 1 (2021), 200–215.
[3] Martin Baatz and Arno Schäpe. 2000. Multiresolution segmentation: an optimization approach for high quality multi-scale image segmentation. *Heidelberg: Angewandte Geographische Information sverarbeitung XII. Wichmann-Verlag* (2000), 12–23.
[4] Aydın Buluç, Henning Meyerhenke, Ilya Safro, Peter Sanders, and Christian Schulz. 2016. Recent advances in graph partitioning. *Algorithm engineering* (2016), 117–158.
[5] Nameirakpam Dhanachandra and Yambem Jina Chanu. 2017. A survey on image segmentation methods using clustering techniques. *European Journal of Engineering and Technology Research* 2, 1 (2017), 15–20.
[6] Yao Ding, Xiaofeng Zhao, Zhili Zhang, Wei Cai, and Nengjun Yang. 2021. Graph sample and aggregate-attention network for hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters* (2021).
[7] Matthias Drusch, Umberto Del Bello, Sébastien Carlier, Olivier Colin, Veronica Fernandez, Ferran Gascon, Bianca Hoersch, Claudia Isola, Paolo Laberinti, Philippe Martimort, et al. 2012. Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote sensing of Environment* 120 (2012), 25–36.
[8] Angel Garcia-Pedrero, Mario Lillo-Saavedra, Dionisio Rodriguez-Esparragon, and Consuelo Gonzalo-Martin. 2019. Deep learning for automatic outlining agricultural parcels: Exploiting the land parcel identification system. *IEEE Access* 7 (2019), 158223–158236.
[9] Markus Immitzer, Francesco Vuolo, and Clement Atzberger. 2016. First experience with Sentinel-2 data for crop and tree species classifications in central Europe. *Remote sensing* 8, 3 (2016), 166.
[10] Asako Kanezaki. 2018. Unsupervised image segmentation by backpropagation. In *2018 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. IEEE, 1543–1547.
[11] Thomas N. Kipf and Max Welling. 2017. Semi-Supervised Classification with Graph Convolutional Networks. In *International Conference on Learning Representations (ICLR)*.
[12] Nataliia Kussul, Guido Lemoine, Francisco Javier Gallego, Sergii V Skakun, Mykola Lavreniuk, and Andrii Yu Shelestov. 2016. Parcel-based crop classification in Ukraine using Landsat-8 data and Sentinel-1A data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 9, 6 (2016), 2500–2508.
[13] Pierre Lassalle, Jordi Inglada, Julien Michel, Manuel Grizonnet, and Julien Malik. 2015. A scalable tile-based framework for region-merging segmentation. *IEEE Transactions on Geoscience and Remote Sensing* 53, 10 (2015), 5473–5485.

[14] Guohao Li, Matthias Muller, Ali Thabet, and Bernard Ghanem. 2019. Deepgcns: Can gcns go as deep as cnns?. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 9267–9276.

[15] Zhenguo Li, Xiao-Ming Wu, and Shih-Fu Chang. 2012. Segmentation using superpixels: A bipartite graph partitioning approach. In *2012 IEEE conference on computer vision and pattern recognition*. IEEE, 789–796.

[16] Han Liu, Jun Li, Lin He, and Yu Wang. 2019. Superpixel-guided layer-wise embedding CNN for remote sensing image classification. *Remote Sensing* 11, 2 (2019), 174.

[17] Li Mi and Zhenzhong Chen. 2020. Superpixel-enhanced deep neural forest for remote sensing image semantic segmentation. *ISPRS Journal of Photogrammetry and Remote Sensing* 159 (2020), 140–152.

[18] Azade Nazi, Will Hang, Anna Goldie, Sujith Ravi, and Azalia Mirhoseini. 2019. Gap: Generalizable approximate graph partitioning framework. *arXiv preprint arXiv:1903.00614* (2019).

[19] Song Ouyang and Yansheng Li. 2021. Combining deep semantic segmentation network and graph convolutional neural network for semantic segmentation of remote sensing imagery. *Remote Sensing* 13, 1 (2021), 119.

[20] Nan Qiao, Yi Zhao, Ruei-Sung Lin, Bo Gong, Zhongxiang Wu, Mei Han, and Jiashu Liu. 2019. Generative-discriminative crop type identification using satellite images. In *2019 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. IEEE, 1–5.

[21] Jianbo Shi and Jitendra Malik. 2000. Normalized cuts and image segmentation. *IEEE Transactions on pattern analysis and machine intelligence* 22, 8 (2000), 888–905.

[22] David Stutz, Alexander Hermans, and Bastian Leibe. 2018. Superpixels: An evaluation of the state-of-the-art. *Computer Vision and Image Understanding* 166 (2018), 1–27.

[23] Wei-Chih Tu, Ming-Yu Liu, Varun Jampani, Deqing Sun, Shao-Yi Chien, Ming-Hsuan Yang, and Jan Kautz. 2018. Learning superpixels with segmentation-aware affinity loss. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 568–576.

[24] Ranjith Unnikrishnan, Caroline Pantofaru, and Martial Hebert. 2007. Toward objective evaluation of image segmentation algorithms. *IEEE transactions on pattern analysis and machine intelligence* 29, 6 (2007), 929–944.

[25] Xiaofang Wang, Yuxing Tang, Simon Masnou, and Liming Chen. 2015. A global/local affinity graph for image segmentation. *IEEE Transactions on Image Processing* 24, 4 (2015), 1399–1411.

[26] Jiadi Yin, Jinwei Dong, Nicholas AS Hamm, Zhichao Li, Jianghao Wang, Hanfa Xing, and Ping Fu. 2021. Integrating remote sensing and geospatial big data for urban land use mapping: A review. *International Journal of Applied Earth Observation and Geoinformation* 103 (2021), 102514.

[27] Chinatsu Yonezawa, Masahiro Negishi, Kenta Azuma, Manabu Watanabe, Naoki Ishitsuka, Shigeo Ogawa, and Genya Saito. 2012. Growth monitoring and classification of rice fields using multitemporal RADARSAT-2 full-polarimetric data. *International journal of remote sensing* 33, 18 (2012), 5696–5711.

[28] Xiao Zhang, Liangyun Liu, Xidong Chen, Yuan Gao, Shuai Xie, and Jun Mi. 2021. GLC_FCS30: Global land-cover product with fine classification system at 30 m using time-series Landsat imagery. *Earth System Science Data* 13, 6 (2021), 2753–2776.

[29] Yang Zhang, Moyun Liu, Jingwu He, Fei Pan, and Yanwen Guo. 2021. Affinity fusion graph-based framework for natural image segmentation. *IEEE Transactions on Multimedia* (2021).

[30] Yanfei Zhong, Rongrong Gao, and Liangpei Zhang. 2016. Multiscale and multifeature normalized cut segmentation for high spatial resolution remote sensing imagery. *IEEE transactions on geoscience and remote sensing* 54, 10 (2016), 6061–6075.

[31] Jie Zhou, Ganqu Cui, Shengding Hu, Zhengyan Zhang, Cheng Yang, Zhiyuan Liu, Lifeng Wang, Changcheng Li, and Maosong Sun. 2020. Graph neural networks: A review of methods and applications. *AI Open* 1 (2020), 57–81.

[32] Lei Zhou and Weiyufeng Wei. 2020. DIC: deep image clustering for unsupervised image segmentation. *IEEE Access* 8 (2020), 34481–34491.