Research Proposal

Rui Li

Wuhan University, China

1. Title

Integrating semantic segmentation and linear attention mechanism for remotely sensed image classification.

2. Background

The research goal of segmentation using remote sensing imagery is to automatically provide labels describing the represented physical land use and land cover (LULC). And the information about LULC is essential for a wide range of application scenarios, including urban and regional planning, environmental monitoring and management, and yield estimation [1-3]. LULC information can also provide insights from a panoramic perspective to tackle a multitude of socioeconomic and environmental challenges, such as food insecurity, poverty, climate change and disaster risk [4]. Conventionally, large-scale field surveys are the primary method to obtain the condition of land use and land cover. Despite the outcomes of surveys are of high quality, the investigative procedures are time-consuming and labor-intensive [5, 6]. Meanwhile, the information about the geographical distribution of land cover is often missing. With recent advances in Earth observation technology, a constellation of satellite and airborne platforms have been launched. Profiting from a great many of very fine spatial resolution (VFSR) remotely sensed imagery captured from satellites, scholars have increasingly focused on automatic land use and land cover classification using satellite images.

3. Research Motivation

Semantic segmentation from remote sensing imagery, a process to assign the precise category to each pixel of an image, has a wide range of application scenarios such as land resource management, yield estimation, and economic assessment [11-13].

Vegetation indices are commonly used features extracted from multispectral and hyperspectral images to characterize land surface physical properties. The normalized difference vegetation index (NDVI) [21] and soil-adjusted vegetation index (SAVI) [22] highlight vegetation over other land resources, whereas the normalized difference bareness index (NDBaI) [23] and the normalized difference bare land index (NBLI) [24] emphasize bare land. The normalized difference water index (NDWI) [25] and modified NDWI (MNDWI) [26] indicate water. These indices have been developed and applied widely in the remote sensing community. Meanwhile, different classifiers have been designed from diverse perspectives, from traditional methods such as logistic regression [27], distance measures [28] and clustering [29], to more advanced machine learning methods such as the support vector machine (SVM) [30], random forest (RF) [31] and artificial neural networks (ANN) [32] including the multi-layer perceptron (MLP) [33]. These classifiers depend critically on the quality of features that are extracted for pixel-level land cover classification.

However, the high dependence on manual descriptors restricts the flexibility and adaptability of traditional methods. The emergency of Deep Learning (DL), which is powerful to capture nonlinear and hierarchical features automatically, tackles the above deficiency to a great extent. Meanwhile, Deep Learning (DL) is powerful to automatically capture nonlinear and hierarchical features, which has influenced many domains such as computer vision (CV), natural language processing (NLP), as well as automatic speech recognition (ASR). As a typical classification task, a great many DL methods have been introduced to semantic segmentation from remote sensing imagery. On the other hand, attention mechanism inspired by biological neurons has shown powerful ability in computer vision tasks by exploiting long-range dependencies of feature maps and facilitating neural networks to attain global contextual information [14]. Such attention module has been developed and applied in image classification [15], object detection [16], and semantic segmentation [17] with increased accuracy.

Referring to the method of deep learning and attention mechanism, I have designed a novel framework for hyperspectral image classification, and the research has been published in Remote Sensing [7]. However, as the memory and computational costs of the conventional dot-product attention mechanism increase quadratically with the spatio-temporal size of the input, the application of attention mechanism with large inputs (e.g. large-scale videos, long sequences, or high-resolution images) is extremely challenging. We previously proposed a linear attention mechanism with $O(N)$ complexity to solve the large-scale issues [8] and then designed a Multi-Attention-Network to boost the performance of baseline [20]. The linear attention mechanism has the advantages of decreasing the complexity and enhancing the flexibility of attention. This Ph.D. research will go a step further to integrate semantic segmentation and linear attention mechanism for remote sensing imagery. This is a promising field in computer vision, AI and machine intelligence. Meanwhile, I also take notice of high-resolution remote sensing images and multi-temporal satellite images using the attention mechanism [9, 10].

4. Research Aims

The main objective of this proposal is to combinate semantic segmentation and linear attention mechanism to enhance the performance of deep learning methods on remotely sensed semantic segmentation. This specific question to be addressed:

- How the position of the linear attention mechanism (in the encoder, decoder, low-level layer, high-level layer or multi-level attention) influence the performance of a deep network?
- What is the best option to combine semantic segmentation and linear attention mechanism to achieve the highest accuracy?
- Can linear attention mechanism be applied in a 3D-CNN based network for hyperspectral images or time-series satellite imagery?

5. Methodologies

I will design a novel DilatedFCN-based network (similar to DeepLab V3 [18]) and EncoderDecoder-based networks (e.g. U-Net [19]) with attention, and demonstrate how the position of the attention mechanism could influence the performance of a deep network.

Based on the method development, I will design a paradigm to integrate semantic segmentation and attention mechanism that could achieve the highest accuracy. The multi-level attention will be explored to find the optimal solution.

Thereafter, I will extend the linear attention mechanism to 3D for semantic segmentation from the hyperspectral image and multi-temporal remote sensing imagery.

6. Timetable

Task	Topic	Months
1	The position of the linear attention mechanism	3
	Data analysis and article writing	3
2	The best way to combine semantic segmentation and attention	6
	Data analysis and article writing	3
3	The application of attention in the 3D-CNN based network	6
	Data analysis and article writing	3
Total		24

References

[1] Liu X, He J, Yao Y, et al. Classifying urban land use by integrating remote sensing and social media data[J]. International Journal of Geographical Information Science, 2017, 31(8): 1675-1696.

[2] Zhang C, Sargent I, Pan X, et al. Joint Deep Learning for land cover and land use classification[J]. Remote sensing of environment, 2019, 221: 173-187.

[3] Sulla-Menashe D, Gray J M, Abercrombie S P, et al. Hierarchical mapping of annual global land cover 2001 to present: The MODIS Collection 6 Land Cover product[J]. Remote Sensing of Environment, 2019, 222: 183-194.

[4] Stürck J, Schulp C J E, Verburg P H. Spatio-temporal dynamics of regulating ecosystem services in Europe–the role of past and future land use change[J]. Applied Geography, 2015, 63: 121-135.

[5] B. Basso and L. Liu, "Seasonal crop yield forecast: Methods, applications, and accuracies," in Advances in Agronomy, vol. 154: Elsevier, 2019, pp. 201-255.

[6] C. Zhang et al., "A New CNN-Bayesian Model for Extracting Improved Winter Wheat Spatial Distribution from GF-2 imagery," Remote Sens., vol. 11, no. 6, p. 619, 2019.

[7] Li R, Zheng S, Duan C, et al. Classification of Hyperspectral Image Based on Double-Branch Dual-Attention Mechanism Network[J]. Remote Sensing, 2020, 12(3): 582.

[8] Li R, Jianlin Su, Duan C, Zheng S. Linear Attention Mechanism: An Efficient Attention for Semantic Segmentation[J]. arXiv preprint arXiv:2007.14902, 2020.

[9] Li R, Duan C, Zheng S. MACU-Net Semantic Segmentation from High-Resolution Remote Sensing Images[J]. arXiv preprint arXiv:2007.13083, 2020.

[10] Li R, Zheng S, Duan C. Land Cover Classification from Remote Sensing Images Based on Multi-Scale Fully Convolutional Network[J]. arXiv preprint arXiv:2008.00168, 2020.

[11] Zhang J, Feng L, Yao F. Improved maize cultivated area estimation over a large scale combining MODIS–EVI time series data and crop phenological information[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2014, 94: 102-113.

[12] Zhang C, Sargent I, Pan X, et al. Joint Deep Learning for land cover and land use classification[J]. Remote sensing of environment, 2019, 221: 173-187.
[13] Sulla-Menashe D, Gray J M, Abercrombie S P, et al. Hierarchical mapping of annual global land cover 2001 to present: The MODIS Collection 6 Land Cover product[J]. Remote Sensing of Environment, 2019, 222: 183-194.
[14] Wang X, Girshick R, Gupta A, et al. Non-local neural networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 7794-7803.
[15] Zhang H, Goodfellow I, Metaxas D, et al. Self-attention generative adversarial networks[C]//International Conference on Machine Learning. 2019: 7354-7363.
[16] Bello I, Zoph B, Vaswani A, et al. Attention augmented convolutional networks[C]//Proceedings of the IEEE International Conference on Computer Vision. 2019: 3286-3295.
[17] Fu J, Liu J, Tian H, et al. Dual attention network for scene segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019: 3146-3154.
[18] Chen L C, Papandreou G, Schroff F, et al. Rethinking atrous convolution for semantic image segmentation[J]. arXiv preprint arXiv:1706.05587, 2017.
[19] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation[C]//International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, 2015: 234-241.
[20] Li R, Zheng S, Duan C, et al. Multi-Attention-Network for Semantic Segmentation of High-Resolution Remote Sensing Images[J]. arXiv preprint arXiv:2009.02130, 2020.
[21] C. J. Tucker, "Red and photographic infrared linear combinations for monitoring vegetation," 1978.
[22] A. Huete, "Huete, AR A soil-adjusted vegetation index (SAVI). Remote Sensing of Environment," Remote sensing of environment, vol. 25, pp. 295-309, 1988.
[23] H. Zhao and X. Chen, "Use of normalized difference bareness index in quickly mapping bare areas from TM/ETM+," in International geoscience and remote sensing symposium, 2005, vol. 3, p. 1666.
[24] H. Li et al., "Mapping urban bare land automatically from Landsat imagery with a simple index," Remote Sensing, vol. 9, no. 3, p. 249, 2017.
[25] B.-C. Gao, "NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space," Remote sensing of environment, vol. 58, no. 3, pp. 257-266, 1996.
[26] H. Xu, "Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery," International journal of remote sensing, vol. 27, no. 14, pp. 3025-3033, 2006.
[27] G. Rutherford, A. Guisan, and N. Zimmermann, "Evaluating sampling strategies and logistic regression methods for modelling complex land cover changes," Journal of Applied Ecology, vol. 44, no. 2, pp. 414-424, 2007.
[28] Q. Du and C.-I. Chang, "A linear constrained distance-based discriminant analysis for hyperspectral image classification," Pattern Recognition, vol. 34, no. 2, pp. 361-373, 2001.
[29] U. Maulik and I. Saha, "Automatic fuzzy clustering using modified differential evolution for image classification," IEEE Transactions on Geoscience and Remote Sensing, vol. 48, no. 9, pp. 3503-3510, 2010.
[30] Y. Fu et al., "An improved combination of spectral and spatial features for vegetation

classification in hyperspectral images," Remote Sensing, vol. 9, no. 3, p. 261, 2017.
[31] K. Tatsumi, Y. Yamashiki, M. A. C. Torres, and C. L. R. Taipe, "Crop classification of upland fields using Random forest of time-series Landsat 7 ETM+ data," Computers and Electronics in Agriculture, vol. 115, pp. 171-179, 2015.
[32] C. Adede, R. Oboko, P. W. Wagacha, and C. Atzberger, "A Mixed Model Approach to Vegetation Condition Prediction Using Artificial Neural Networks (ANN): Case of Kenya's Operational Drought Monitoring," Remote Sensing, vol. 11, no. 9, p. 1099, 2019.
[33] C. Zhang et al., "A hybrid MLP-CNN classifier for very fine resolution remotely sensed image classification," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 140, pp. 133-144, 2018.