Development of Deep Satellite Imagery Analytics for Vegetation and Urban Growth Monitoring

Xin Cai

1 Introduction

Remotely sensed data, including those obtained from space-borne sensors, constitutes an integral part of Earth Observation (EO) data which is a valuable source for measuring spatiotemporal dynamics of the surface of our planet. With the advancement of remote sensing technology, the volume of remotely sensed data has been growing drastically, thereby necessitating the automatic interpretation by employing big data analytics techniques.

Deep neural networks, as the most representative data-driven approach so far, have been applied to various remote sensing-related applications, including but not limited to land cover mapping [1], crop type classification [2], crop yield prediction [3], change detection [4], earth surface forecasting [5], and multi-modal data fusion [6]. Additionally, deep neural networks have been applied to different data modalities in remote sensing, such as hyper-/multi-spectral images [1, 2, 5], synthetic aperture radar (SAR) [6], LiDAR [7], and optical images [4]. Despite the promising results achieved, applying deep learning techniques to process remotely sensed data still faces many challenges, which has been outlined in the survey paper [8], for example, multi-modal information fusion among data obtained through different types of sensors, fusing with a broad scope of geographical information, the temporal variable, and the integration with established physical models. Apart from the challenges specific to remotely sensed data, there are some common challenges needing to be addressed such as data efficiency and the generalization capability of deep learning algorithms.

When it comes to my PhD research project, I plan to confine my attention to the following challenges:

- developing novel neural network architectures that can better exploit spatiotemporal structural information in satellite image time series (SITS);
- employing unsupervised learning methods to distil transferable feature representations that can benefit downstream applications such as vegetation and urban growth monitoring;
- developing deep learning models with better generalization capability that can achieve satisfactory results on unseen scenarios, i.e., deep domain adaptation approaches [9];

2 Overview of Key Research Areas

In the remote sensing community, the availability of large-scale, high-quality datasets for benchmarking deep learning models is limited. These datasets are of crucial importance for developing deep learning models specifically tailored for remotely sensed data. Due to some obstacles in the remote sensing community, e.g., proprietary licenses required to access satellite data, substantial domain expertise for annotation, the use of self-created datasets to evaluate model performance has been commonly seen in many publications. In the paper [10], the authors have summarized a proportion of annotated datasets of remote sensing imagery,

highlighting the fact that most of existing datasets are unsuitable for training deep learning models. Some researchers have made pioneering efforts towards the end of open-sourcing high-quality benchmark datasets [1]. These recently released large-scale datasets, especially the DENETHOR (DynamicEarthNet) [11], EarthNet2021 [5], and MUDS [12], will significantly facilitate my PhD research.

It can be seen from recently released datasets that dealing with spatiotemporal signals has become a key area of interest in the remote sensing community with unresolved challenges for current deep learning models. For example, TempCNN [13] and LSTM [14] combined with spatial encoders such as mobilenetv3 [15] or resnet18 [16] performed even worse than a random forest classifier with hand-crafted features [11], demonstrating the importance of devising bespoke neural network architectures for effectively using SITS represented in consistent spatiotemporal grids. Inroads have been made [2] by combining pixel-set encoders [17] with the self-attention mechanism [18], which has shown superior performance to the random forest classifier and other deep learning models. Therefore, how to effectively and efficiently explore the spatiotemporal structural information in raw SITS has become a particular research area of interest.

Unsupervised representation learning has also been an active research area in the machine or deep learning community. Mainstream solutions include deep generative models, deep clustering methods, and self-supervised learning. There has been rapid progress in deep generative models, such as GANs [19], VAEs [20], deep autoregressive models [21], normalizing flow-based models [22] and their hybrid variants [23], which have proven to be adept at capturing complex distributions of modelled data. The core motivation of using generative models to perform unsupervised representation learning is based on a simple principle "You cannot understand what you cannot generate". As clustering algorithms such as k-means, gaussian mixture models (GMM), and spectral clustering are primary unsupervised learning approaches in the machine learning community, it is natural to extend these methods by integrating them into modern deep learning frameworks. Additionally, self-supervised learning [24] as an emerging research field has been actively studied in recent years, where discriminative approaches have been adopted rather than generative methods based on the assumption that pixel-level generation is computationally costly and may not be necessary for representation learning. One of the most common strategies for self-supervised learning is to predict future, missing or contextual information. Therefore, obtaining high-quality transferable features using unsupervised learning methods is another research problem that I plan to address in my future research.

Unsupervised domain adaptation (UDA) [9] is an active research area in machine or deep learning community. Supervised learning commonly assumes that training and testing datasets are drawn from the same distribution, but this assumption rarely holds in real-world settings. Domain Adaptation refers to developing algorithms that can generalize well to the target domain by training models on a semantic related but distribution different source domain. There are many approaches proposed to tackle this challenge, such as instance re-weighting adaptation, feature adaptation, classifier adaptation, and adversarial adaptation [25]. Adversarial domain adaptation models [26] have achieved great success by enforcing alignment either in raw data space or high-level feature space using adversarial losses to regularize the feature learning process. The domain adaptation problem has also been noticed in recently published papers for crop type mapping [11, 27], where trained models need to be tested in data collected in a different year, which is referred to as out-of-year evaluation. It has been reported that models designed when completely disregarding the domain shift to a different year would suffer a dramatic decrease in performance, which amounts to a 12 percentage points drop in accuracy

[1] Details of those recently released large-scale and high-quality datasets are listed in Appendix A.

on DENETHOR [11]. Image-to-image (I2I) translation techniques [28, 29] have been used to tackle the domain shift problem. For example, in papers [4, 30], researchers proposed to use I2I translation techniques to suppress differences in unchanged regions caused by significant seasonal variations when performing bi-temporal change detection. Therefore, devising deep learning models that can handle the domain shift problem to a satisfactory degree using UDA or I2I translation techniques is the third research problem that I plan to address in my future research.

3 Conclusion & Research Plan

The overall aim of my PhD research is to develop deep learning algorithms that can exploit SITS with a particular focus on vegetation and urban growth monitoring. To tackle these challenges I will:

- investigate methods proposed for video understanding/video action recognition and other research fields which involve spatiotemporal learning to study how the spatiotemporal structural information can be exploited;
- investigate methods related to Graph Convolutional Neural Networks (GCNs) which aim at generalizing operations from regular grids to irregular ones represented as graphs to study how irregular data can be efficiently processed by deep learning models [2];
- investigate deep generative models, deep clustering models, and self-supervised learning to study how to employ these methods to pre-train models on remotely sensed data without labels to extract transferable feature representations;
- investigate UDA and I2I translation techniques to study how to exploit complementary information in multi-modality remotely sensed data and how to improve generalization capability of deep learning models to unseen scenarios.

Besides, before the confirmation assessment, I will submit at least one paper to international conferences or journals.

References

[1] R. Interdonato, D. Ienco, R. Gaetano, and K. Ose, "Duplo: A dual view point deep learning architecture for time series classification," *ISPRS journal of photogrammetry and remote sensing*, vol. 149, pp. 91–104, 2019.

[2] V. S. F. Garnot, L. Landrieu, S. Giordano, and N. Chehata, "Satellite image time series classification with pixel-set encoders and temporal self-attention," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 12 325–12 334.

[3] J. Shook, T. Gangopadhyay, L. Wu, B. Ganapathysubramanian, S. Sarkar, and A. K. Singh, "Crop yield prediction integrating genotype and weather variables using deep learning," *Plos one*, vol. 16, no. 6, p. e0252402, 2021.

[4] B. Fang, L. Pan, and R. Kou, "Dual learning-based siamese framework for change detection using bi-temporal vhr optical remote sensing images," *Remote Sensing*, vol. 11, no. 11, p. 1292, 2019.

[5] C. Requena-Mesa, V. Benson, J. Runge, J. Denzler, and M. Reichstein, "Earthnet2021: A novel large-scale dataset and challenge for forecasting localized climate impacts," 2020. [Online]. Available: https://www.climatechange.ai/papers/neurips2020/48

[6] L. T. Luppino, M. Kampffmeyer, F. M. Bianchi, G. Moser, S. B. Serpico, R. Jenssen, and S. N. Anfinsen, "Deep image translation with an affinity-based change prior for unsupervised multimodal change detection," *IEEE Transactions on Geoscience and Remote Sensing*, 2021.

[2] Crop fields are characterized by irregular boundaries, and therefore using rectangular images as basic processing units are inefficient. Besides, removing the constraints imposed by regular spatiotemporal grids and exploring more flexible representations such as graphs may lead to novel algorithms.

[7] L. Zhang, Z. Shao, J. Liu, and Q. Cheng, "Deep learning based retrieval of forest aboveground biomass from combined lidar and landsat 8 data," *Remote Sensing*, vol. 11, no. 12, p. 1459, 2019.

[8] X. X. Zhu, D. Tuia, L. Mou, G.-S. Xia, L. Zhang, F. Xu, and F. Fraundorfer, "Deep learning in remote sensing: A comprehensive review and list of resources," *IEEE Geoscience and Remote Sensing Magazine*, vol. 5, no. 4, pp. 8–36, 2017.

[9] G. Wilson and D. J. Cook, "A survey of unsupervised deep domain adaptation," *ACM Trans. Intell. Syst. Technol.*, vol. 11, no. 5, pp. 51:1–51:46, 2020.

[10] M. Schmitt, L. H. Hughes, C. Qiu, and X. X. Zhu, "Sen12ms – a curated dataset of georeferenced multi-spectral sentinel-1/2 imagery for deep learning and data fusion," in *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. IV-2/W7, 2019, pp. 153–160.

[11] L. Kondmann, A. Toker, M. Rußwurm, A. Camero, D. Peressuti, G. Milcinski, P.-P. Mathieu, N. Longépé, T. Davis, G. Marchisio *et al.*, "Denethor: The dynamicearthnet dataset for harmonized, inter-operable, analysis-ready, daily crop monitoring from space," *NeurIPS Track on Datasets and Benchmarks*, 2021.

[12] A. Van Etten, D. Hogan, J. M. Manso, J. Shermeyer, N. Weir, and R. Lewis, "The multi-temporal urban development spacenet dataset," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 6398–6407.

[13] C. Pelletier, G. I. Webb, and F. Petitjean, "Temporal convolutional neural network for the classification of satellite image time series," *Remote Sensing*, vol. 11, no. 5, p. 523, 2019.

[14] M. Rußwurm and M. Körner, "Multi-temporal land cover classification with sequential recurrent encoders," *ISPRS International Journal of Geo-Information*, vol. 7, no. 4, p. 129, 2018.

[15] A. Howard, M. Sandler, G. Chu, L.-C. Chen, B. Chen, M. Tan, W. Wang, Y. Zhu, R. Pang, V. Vasudevan *et al.*, "Searching for mobilenetv3," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 1314–1324.

[16] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *European conference on computer vision*. Springer, 2016, pp. 630–645.

[17] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "Pointnet: Deep learning on point sets for 3d classification and segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 652–660.

[18] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in neural information processing systems*, 2017, pp. 5998–6008.

[19] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. C. Courville, and Y. Bengio, "Generative adversarial networks," *Commun. ACM*, vol. 63, no. 11, pp. 139–144, 2020.

[20] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," in *2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, Conference Track Proceedings*, 2014.

[21] A. van den Oord, N. Kalchbrenner, and K. Kavukcuoglu, "Pixel recurrent neural networks," in *Proceedings of the 33nd International Conference on Machine Learning, ICML 2016, New York City, NY, USA, June 19-24, 2016*, ser. JMLR Workshop and Conference Proceedings, vol. 48. JMLR.org, 2016, pp. 1747–1756.

[22] I. Kobyzev, S. J. D. Prince, and M. A. Brubaker, "Normalizing flows: An introduction and review of current methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 11, pp. 3964–3979, 2021.

[23] A. Makhzani, J. Shlens, N. Jaitly, and I. Goodfellow, "Adversarial autoencoders," in *International Conference on Learning Representations*, 2016.

[24] A. v. d. Oord, Y. Li, and O. Vinyals, "Representation learning with contrastive predictive coding," *arXiv preprint arXiv:1807.03748*, 2018.

[25] L. Zhang and X. Gao, "Transfer adaptation learning: A decade survey," *arXiv preprint arXiv:1903.04687*, 2019.

[26] J. Hoffman, E. Tzeng, T. Park, J.-Y. Zhu, P. Isola, K. Saenko, A. Efros, and T. Darrell, "Cycada: Cycle-consistent adversarial domain adaptation," in *International conference on machine learning*. PMLR, 2018, pp. 1989–1998.

[27] G. Weikmann, C. Paris, and L. Bruzzone, "Timesen2crop: A million labeled samples dataset of sentinel 2 image time series for crop-type classification," *IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens.*, vol. 14, pp. 4699–4708, 2021.

[28] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1125–1134.

[29] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2223–2232.

[30] R. Kou, B. Fang, G. Chen, and L. Wang, "Progressive domain adaptation for change detection using season-varying remote sensing images," *Remote Sensing*, vol. 12, no. 22, p. 3815, 2020.

[31] M. Rußwurm, C. Pelletier, M. Zollner, S. Lefèvre, and M. Körner, "Breizhcrops: A time series dataset for crop type mapping," *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences ISPRS (2020)*, 2020.

[32] Q. Wang, L. Zhang, L. Bertinetto, W. Hu, and P. H. S. Torr, "Fast online object tracking and segmentation: A unifying approach," in *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019*. Computer Vision Foundation / IEEE, 2019, pp. 1328–1338.

Appendices

A Large-scale & High-quality Datasets in Remote Sensing

Weikmann et al. [27] released a large-scale time-series dataset called TimeSen2Crop that is comprised of about 1 million of labelled samples belonging to 16 crop types extracted from Sentinel-2 multispectral images. Rußwurm et al. [31] proposed a novel large-scale satellite image time-series dataset for crop type mapping termed BreizhCrops from the region of Brittany, France, which contains more than 600k multivariate time-series extracted from Sentinel-2 satellite imagery. These datasets are characterized by focusing on exploiting high temporal resolution of Sentinel-2 SITS with a revisit time of 5 days for vegetation monitoring. But the spatial extent is neglected in these two datasets by taking mean values of each spectral band in each parcel field, simplifying spatiotemporal signals to only temporal signals. The recently released DENETHOR dataset [11] has filled this gap by providing a high-quality dataset with harmonized, declouded and daily revisit times SITS. The high density in temporal dimension combined with 3m spatial resolution poses great challenges for the existing deep learning methods developed for crop type classification. More importantly, the DENETHOR dataset is made publicly available under a larger project called DynamicEarthNet initiated by Technical University of Munich (TUM) and German Aerospace Center (DLR), aiming at making more multi-temporal EO data accessible. Another recently released dataset which also contains high-density spatiotemporal signals combined with weather variables at the mesoscale ($> 1km$) is EarthNet2021 [5]. The objective of EarthNet2021 is to enable climate impact projection at a more fine-grained scale, e.g., $< 100m$, which will be beneficial for a wide array of downstream tasks, such as crop yield prediction, forest health assessment, and biodiversity monitoring. The above-mentioned datasets are focused on vegetation monitoring or weather forecasting because these tasks can make the most of the rich features in the temporal dimension. Additionally, a recently published dataset called Multi-Temporal Urban Development SpaceNet Dataset (MUDS) [12] is focused on urban growth monitoring, which consists of SITS with 4m spatial resolution and on a monthly basis, covering > 100 unique geographies and totalling $> 11M$ annotations. Compared to traditional bi-temporal change detection which is essentially using two static images acquired at different times, MUDS reframes change detection as change and object tracking tasks because the time span of SITS is from $18 \sim 26$ months. As pointed out in the paper [12], visual object tracking (VOT) algorithms [32] developed in the computer vision community are not suitable for being directly used for this dataset because of substantial image-to-image variations and overly-dense annotated target objects in SITS.

B Summary of Progress To Date & Training Planned

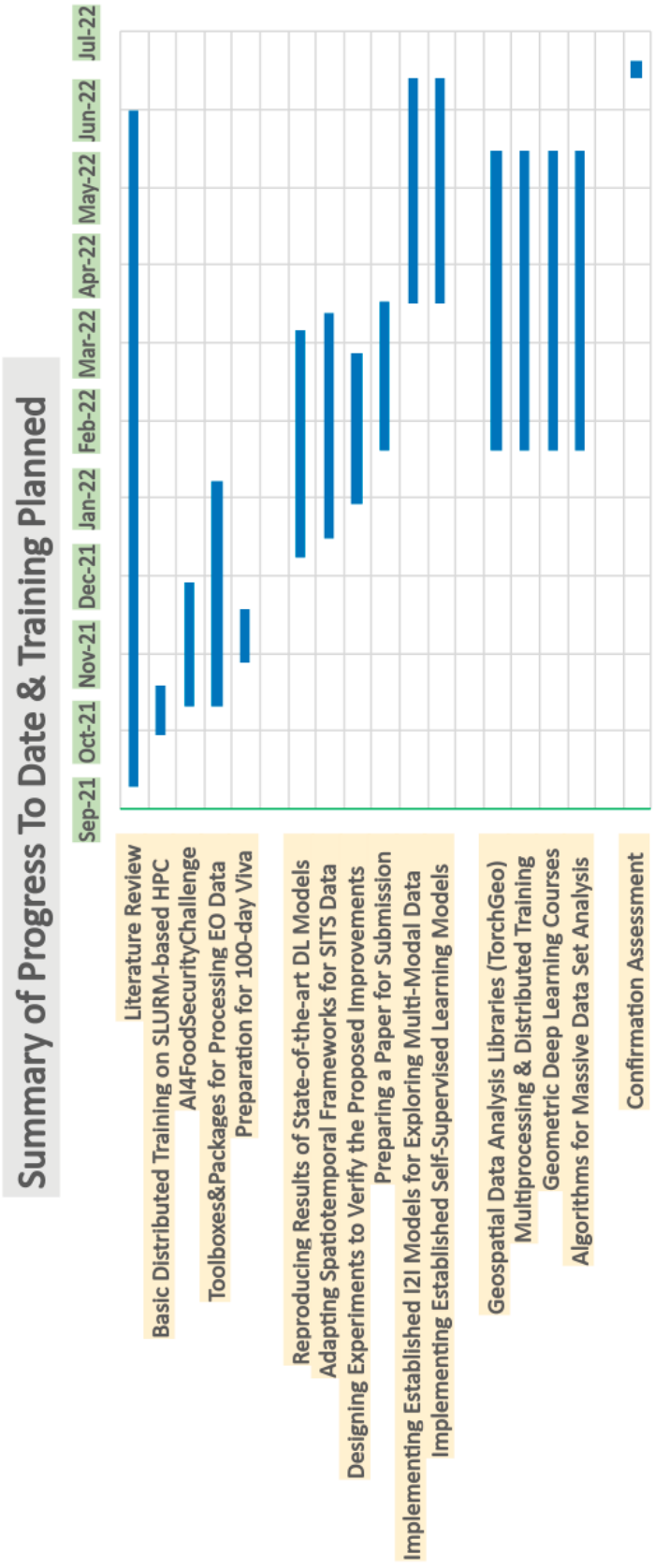

Figure 1: Gantt Chart