Integrating XAI and GeoAI

Jin Xing
School of Engineering, Newcastle University, Newcastle Upon Tyne, United Knigdom
jin.xing@newcastle.ac.uk

Renee E. Sieber
Department of Geography, McGill University, Montreal, Canada
renee.sieber@mcgill.ca

Abstract

While eXplainable Artificial Intelligence (XAI) has significant potential to glassbox Deep Learning, there are challenges in applying it in the domain of Geospatial Artificial Intelligence (GeoAI). A land use case study highlights these challenges, which include the difficulty of selecting reference data/models, the shortcomings of gradients to serve as explanation, the limited semantics and knowledge scope in the explanation process of GeoAI, and underlying GeoAI processes that are not amenable to XAI. We conclude with possibilities to achieve Geographical XAI (GeoXAI).

1 Introduction

The acronym eXplainable Artificial Intelligence (XAI) is, simply put, AI whose functioning can be understood by humans, although XAI more commonly describes a suite of computational algorithms that are applied to AI algorithms to render their output and corresponding training processes more interpretable for given users [1][15]. XAI has the potential to 'glassbox' the blackbox of AI, specifically in Deep Learning (DL). In DL we lack control over how the model detects and classifies features, which means that the features can be misclassified even as the model optimizes performance or features may be classified in unexpected ways. To date XAI has largely not been actively applied to the domain of Geospatial Artificial Intelligence (GeoAI) (cf., [3]). Our concern is that GeoAI is not well-suited to XAI and therefore may generate misleading interpretations.

We briefly describe some challenges of integrating XAI and GeoAI. We illustrate these challenges with a land use classification case study using an XAI called SHapley Additive exPlanations (SHAP). We conclude with possibilities to realize a GeoXAI.

2 Challenges Integrating XAI and GeoAI

We envision four potential issues in integrating GeoAI and XAI. These include the difficulty of selecting reference data/models, the shortcomings of gradients as explanation, the limited semantics and knowledge scope in the explanation process of GeoAI, and underlying GeoAI processes that are not amenable to XAI. To a certain extent this latter issue is the most important because difficulties in integrating 'geo' into AI complicates the application of any explainability approach.

First, most XAI algorithms require reference data points to serve as a baseline of feature and model explanation [22]. Reference data points or datasets are features where the

XAI results are selected to measure a neutral contribution of neurons to the output at a particular layer [15]. Usually, a good reference neither classifies nor misclassifies elements in a convolutional layer. These non-reactive reference points can be challenging to find and any cartographic attributes (e.g, locations, distances, coordinates, and projections) can be neglected. GeoAI models are so spatially explicit that even neutral data will likely activate in some layers [8]. A popular XAI technique, Taylor Decomposition, deconstructs neurons in the layers' choices in terms of the contributions of input variables. In a Taylor Decomposition, such reference points are treated as hyperparameters that require onerous tuning [13]. Hyperparameter tuning is useful as it often occurs in the input layer but the process emphasizes model performance and not domain-specific attributes like geography. The explicit integration of geographic attributes (e.g., adhering to Tobler's Law) should increase progress in both GeoAI and GeoXAI [11].

Second, gradients are one of the founding optimization algorithms in DL and play a pivotal role in a large number of XAI techniques [1]. They offer a kind of sensitivity test of the impacts on the output of tweaking the input data. Balduzzi et al. [2] formally described what is called the shattered gradient problem, in which differentials among gradients decay as the number of layers increase. Algorithms like SmoothGrad [17] flatten differentials between layers but can blur layer boundaries and, more important for GeoAI, ignore geographic boundaries (e.g., between land uses). Such XAI approaches can distort the importance of activation of the boundaries in the original geographic datasets and thus reduce interpretability.

A third challenge of realizing GeoXAI lies in gaps in geographic semantics in its output interpretation [9]. Research on geospatial semantics and ontologies (e.g., [10]) are largely absent in many GeoAI applications and are challenging to insert into XAI. Without an 'explanation of the explanation', XAI might fail to inform us if the model structure is adequate, if the input data is sufficient, or if the training process is implemented correctly. Semantics could reconcile colloquial labels to model results of terms like mount to describe large mounds and tall mountains. Knowledge representation and approaches like qualitative spatial reasoning could contribute to GeoXAI as well as GeoAI [7].

As part of this challenge, the knowledge scope required for any AI is usually larger and more complicated than prosaic AI tasks, which suggests additional knowledge for interpretation, even with adequate training/input datasets. DL has been largely applied for highly specific tasks, such as cat/dog recognition from images. However, GeoAI tasks are complicated due to their close connections with geographic context [6]. In remote sensing-based land use change detection, decisions are not only associated with slight pixel value differences among images acquired at different times, but also the semantics of land use changes [21]. Autonomous driving systems do not only depend on current traffic conditions, but also are subject to local transportation regulations [4]. Such additional knowledge should be analyzed by XAI along with the input geographic datasets.

Lastly, most training processes in GeoAI are not geographic because they can fail to preserve scale, geometry, and topology. Several mature neural networks (e.g., VGG16 and Resnet-101) have been deployed for GeoAI [24]. These networks usually enforce geospatial datasets to be split into small chunks (i.e., reduction in spatial extents), which introduces problems when decomposing boundaries [20]. Hierarchical feature extraction of DL alters the resolution and may distort topological and geometric relationships in the original datasets, such as the maxpooling [14]). No current XAI framework informs us of the degree and impact of such geographic distortion in the training and testing of GeoAI. We also should pay additional attention to ontological differences in how scale is defined in XAI and GeoAI. Most review works in XAI treat scale as an issue of quality (i.e., of the explanations for a given audience) or scope, but XAI algorithms usually interpret scale as global explanations (i.e., XAI for the whole model) or local explanations (i.e., XAI for some portion of the input data).

3 A Case Study

To illustrate the challenges and opportunities of integrating XAI with GeoAI, we look at land use classification. Land use classification represents a typical GIScience application and has had a number of applications with GeoAI (e.g., [18]). The reason that it is a typical application is that it is full of scale (resolution/extent), geometry/topology, and boundary issues. Additionally, land use classification often requires place-based context. Janowicz et al. [8] mention that spatially explicit GeoAI models should not be invariant under relocation of the studied phenomena. Any DL classification modelling requires considerable training data; we use a standard training dataset called the University of California Merced Land Use datasets (UCMLC) developed by Yang and Newsam [23]. The UCMLC contains 100 labelled images for each of 21 land use classes (e.g., from agricultural to storage tanks to airplanes and runways – http://weegee.vision.ucmerced.edu/datasets/landuse.html).

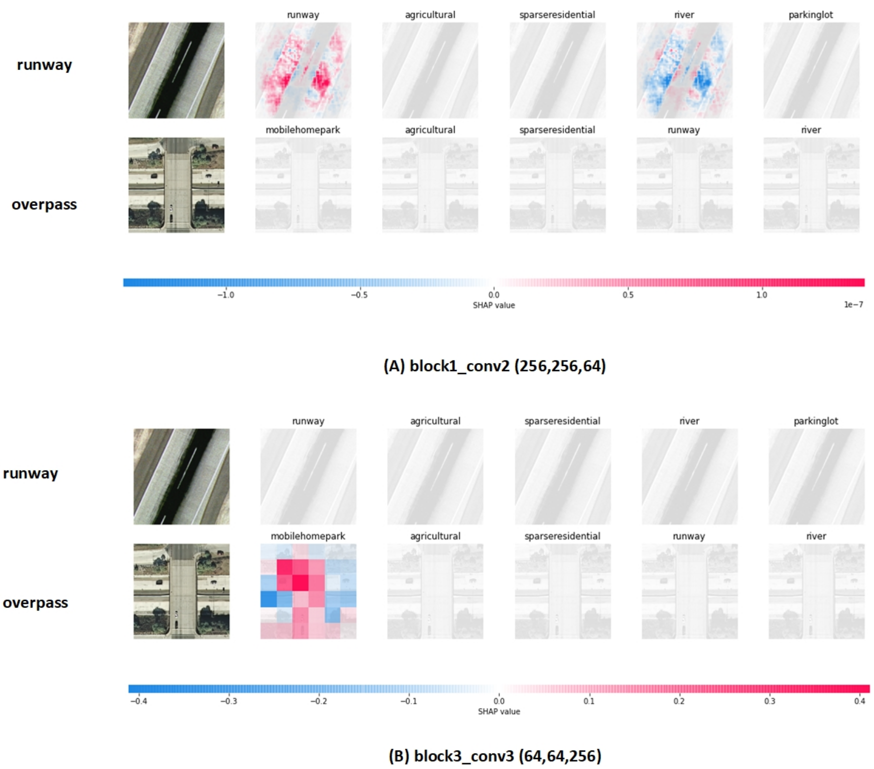

Figure 1 (A) SHAP values depicted with top 5 labels for a runway and overpass example at the 2nd layer of VGG16 model and (B) SHAP explanation at the 9th layer. Red colour ramp depicts impact on the output of positive classification. Blue ramp indicates the negative influence.

Our XAI case study uses land use classification with the UCMLC dataset on the 16-layer University of Oxford Visual Geometry Group (VGG16). VGG16 is a Convolutional Neural Network (CNN) that is widely used for computer vision image classification [16]. Without fine-tuning the VGG16 model to optimize classification results for UCMLC, we still achieve an accuracy of 89.1 percent. We used an XAI called SHAP. Albeit a simple approach, it focuses on feature importance by identifying which patches of features (e.g., from images) from the training or input data contributed to the model's output [12]. This glassboxing algorithm allows the user to determine what is important or what should be done in terms of "feature engineering". Figure 1 shows preliminary XAI results using SHAP method of the classification results to investigate the performance of two layers of VGG16.

We randomly chose one correctly classified example (i.e., runway) and a misclassified one (i.e., an overpass is labelled a "mobilehomepark"). In Figure 1(A), the XAI identifies whichever the convolutional layer is identifying for the runway seems to confirm, according to the SHAP values, that it is contributing to the predicted output. Likewise, the SHAP rejects the "river" classification (or rather which feature patches are being used in the convolutional layer) as a negative contribution to the output. XAI often also can provide insight into multiple layers of the neural network. Figure 1(A) shows that, for the second convolutional layer, there is no predictive value identified by SHAP for overpass due to the selected reference points.

Figure 1(B) shows that the ninth convolutional layer does not register for runway in terms of SHAP, as a good selection of reference. Conversely, the feature patches used in Layer 9 confuse the classification of overpasses. Examination of the underlying feature patches suggests the difficulty of AI to infer complex land use patterns, and the way in which VGG16 is likely misled by the coincidence of green space and cement. The overpass in Figure 1(B) shows the necessity of finding good reference points against which to explain the misclassification. As we mentioned earlier SHAP can be used for feature engineering, it can be used to eliminate 'outlier' regions in the input data. We can potentially create synthetic data to emphasize the most widely recognized SHAP areas or enhance edges. SHAP also employs gradients as the explanation for neuron/layer contributions, in which the shattered gradient problem could lessen the explanation, while smoothing techniques could generate better explanation values but distort the object boundaries. This again highlights the importance of considering geographic attributes in the reference selection.

Figure 1 implies that it is easy to observe raw image regions attributed to the right/wrong classification results but these patches do not contain meaningful geographic semantics. We cannot attach labels such as "overpass" or "runway" to the patches because intermediate image features extracted by VGG16 are computationally but not geographically meaningful. Moreover, the SHAP values stand for the contribution of these patches to the classification results and not the likelihood of geographic features of interest. We cannot be sure if the SHAP values generated are semantically consistent with specific locations, especially if we are interested in invariant spatially-explicit models. It is not only necessary for the final results of the classification to be semantically understandable, but also the XAI outputs to be geographically interpretable, otherwise knowledge generated by XAI will be inaccessible to non-GeoAI expert users.

Lastly there may only be 21 class labels in the UCMLC datasets but non-experts can infer, with additional knowledge, concepts such as "grassland" and "cement pavement". Moreover, a group of trees can be classified under "green space", "park", or even "forest" labels; a small water area can be labelled as "pond", "lake", or "meander". Appropriate labels should originate from disciplines related to geography, not from computer science, which chooses labels based on feature similarity to other labelled data. Current XAI techniques can provide a fitness score for each individual label but cannot suggest if the labels are optimal for the given task or whether additional labels are needed in the training and testing of given GeoAI methods. Therefore, we might have to adopt an over-provision strategy. We could, for example, develop additional rules in which the XAI constrains layers (e.g., deep Taylor Decomposition [15]) or we could develop ensemble models, not to improve the performance but to achieve better GeoXAI.

4 Future Research

XAI is essential in glassbox DL and ensure GeoAI is more understandable and trustworthy. This comprehension cannot be achieved by simply applying XAI techniques to GeoAI. In this paper we argue that geographic interpretation should be integrated with XAI to develop specific explanation frameworks for GeoXAI. (1) Current XAI techniques only offer low-level abstractions, which are difficult to utilize without considerable expertise in AI, a GeoXAI should be designed by and for geospatial information scientists. (2) Current XAI can be incompatible with geospatial data because most explanation techniques are feature-based and not location-based (e.g., retaining boundaries and wholeness of features). For instance, XAI can treat geospatial data as plain tensors that can be arbitrarily split. Among other remedies, we recommend a recomposition approach (cf., [20]) that superimposed the original geographic coordinates to recover geographic context in GeoXAI. (3) XAI visualization tools like https://github.com/yosinski/deep-visualization-toolbox could be modified to provide insight into impacts of geographic scales on explanatory power. (4) Geographic knowledge graphs [19] could supply background information to enhance GeoXAI's explanatory power. We could add spatiality to the neural network layers to create an explanation 'Space time atoms' similar to Xing and Sieber [21]. (5) Social sciences, especially methods to address explainability to different users [5], offer an important path to achieving GeoXAI. (6) Most XAI outputs lack comparability (e.g., SHAP to Taylor Decomposition), although initial work with SHAP might offer such an unified explanation framework [12]. Overall, the geospatial information domain knowledge needs to be integrated into the design of future XAI techniques, in addition to being considered for specific groups of users like GIScientists and cartographers, as well as individuals impacted by GeoAI.

References

1 Alejandro Barredo Arrieta, Natalia Díaz-Rodríguez, Javier Del Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador García, Sergio Gil-López, Daniel Molina, Richard Benjamins, et al. Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai. *Information Fusion*, 58:82–115, 2020.

2 David Balduzzi, Marcus Frean, Lennox Leary, JP Lewis, Kurt Wan-Duo Ma, and Brian McWilliams. The shattered gradients problem: If resnets are the answer, then what is the question? In *International Conference on Machine Learning*, pages 342–350. PMLR, 2017.

3 Ximeng Cheng, Jianying Wang, Haifeng Li, Yi Zhang, Lun Wu, and Yu Liu. A method to evaluate task-specific importance of spatio-temporal units based on explainable artificial intelligence. *International Journal of Geographical Information Science*, pages 1–24, 2020.

4 Laura Garcia Cuenca, Javier Sanchez-Soriano, Enrique Puertas, Javier Fernandez Andres, and Nourdine Aliane. Machine learning techniques for undertaking roundabouts in autonomous driving. *Sensors*, 19(10):2386, 2019.

5 Julie Gerlings, Millie Søndergaard Jensen, and Arisa Shollo. Explainable ai, but explainable to whom? *arXiv preprint arXiv:2106.05568*, 2021.

6 Yingjie Hu, Song Gao, Dalton Lunga, Wenwen Li, Shawn Newsam, and Budhendra Bhaduri. Geoai at acm sigspatial: progress, challenges, and future directions. *Sigspatial Special*, 11(2):5–15, 2019.

7 Yi Huang, May Yuan, Yehua Sheng, Xiangqiang Min, and Yuwei Cao. Using geographic ontologies and geo-characterization to represent geographic scenarios. *ISPRS International Journal of Geo-Information*, 8(12):566, 2019.

8 Krzysztof Janowicz, Song Gao, Grant McKenzie, Yingjie Hu, and Budhendra Bhaduri. Geoai: spatially explicit artificial intelligence techniques for geographic knowledge discovery and beyond, 2020.

9 Robert Laurini, Sylvie Servigne, and Franck Favetta. An introduction to geographic rule semantics. In *DMS2016-22nd International Conference on Distributed Multimedia Systems*, pages 91–97, 2016.

10 Wenwen Li. Geoai: Where machine learning and big data converge in giscience. *Journal of Spatial Information Science*, 2020(20):71–77, 2020.

11 Wenwen Li, Chia-Yu Hsu, and Maosheng Hu. Tobler's first law in geoai: A spatially explicit deep learning model for terrain feature detection under weak supervision. *Annals of the American Association of Geographers*, pages 1–19, 2021.

12 Scott M Lundberg and Su-In Lee. A unified approach to interpreting model predictions. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, pages 4768–4777, 2017.

13 Grégoire Montavon, Sebastian Lapuschkin, Alexander Binder, Wojciech Samek, and Klaus-Robert Müller. Explaining nonlinear classification decisions with deep taylor decomposition. *Pattern Recognition*, 65:211–222, 2017.

14 Sara Sabour, Nicholas Frosst, and Geoffrey E Hinton. Dynamic routing between capsules. *arXiv preprint arXiv:1710.09829*, 2017.

15 Wojciech Samek, Grégoire Montavon, Sebastian Lapuschkin, Christopher J Anders, and Klaus-Robert Müller. Explaining deep neural networks and beyond: A review of methods and applications. *Proceedings of the IEEE*, 109(3):247–278, 2021.

16 Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

17 Daniel Smilkov, Nikhil Thorat, Been Kim, Fernanda Viégas, and Martin Wattenberg. Smoothgrad: removing noise by adding noise. *arXiv preprint arXiv:1706.03825*, 2017.

18 Ava Vali, Sara Comai, and Matteo Matteucci. Deep learning for land use and land cover classification based on hyperspectral and multispectral earth observation data: A review. *Remote Sensing*, 12(15):2495, 2020.

19 Shu Wang, Xueying Zhang, Peng Ye, Mi Du, Yanxu Lu, and Haonan Xue. Geographic knowledge graph (geokg): A formalized geographic knowledge representation. *ISPRS International Journal of Geo-Information*, 8(4):184, 2019.

20 Jin Xing, Renée Sieber, and Margaret Kalacska. The challenges of image segmentation in big remotely sensed imagery data. *Annals of GIS*, 20(4):233–244, 2014.

21 Jin Xing and Renee E Sieber. A land use/land cover change geospatial cyberinfrastructure to integrate big data and temporal topology. *International Journal of Geographical Information Science*, 30(3):573–593, 2016.

22 Feiyu Xu, Hans Uszkoreit, Yangzhou Du, Wei Fan, Dongyan Zhao, and Jun Zhu. Explainable ai: A brief survey on history, research areas, approaches and challenges. In *CCF international conference on natural language processing and Chinese computing*, pages 563–574. Springer, 2019.

23 Yi Yang and Shawn Newsam. Bag-of-visual-words and spatial extensions for land-use classification. In *Proceedings of the 18th SIGSPATIAL international conference on advances in geographic information systems*, pages 270–279, 2010.

24 Liangpei Zhang, Lefei Zhang, and Bo Du. Deep learning for remote sensing data: A technical tutorial on the state of the art. *IEEE Geoscience and Remote Sensing Magazine*, 4(2):22–40, 2016.