A Research Agenda for Spatial Analysis Chapter

Progress on Machine Learning Applications in Geography

Stephen Law, Shen Yao, Zhong Chen

Abstract

The field of artificial intelligence has expanded rapidly in recent years permeating to many application domains including medical science, climate science, finance, and geography. In Geography, these advances have culminated in the new subdomain of GeoAI which was driven by advances in deep learning, optimised computational tools and the availability of large scale spatially embedded data. In this chapter, we will describe a couple of techniques in deep learning for analysing image, point, graph and text data. We will then provide some projections on the near future for the topic, including increasing application of deep learning on traditional geographical problems and incorporation of geographical thinking in machine learning, open data practices as well as cross disciplinary engagements and teaching. We envisage the use of deep learning in geography will continue to grow rapidly leading to hopefully new spatial insight, knowledge and methods to be discovered in the future.

1.0 Background

Geography is an inclusive discipline without strict boundaries–it pervades multiple topics and is there all around us (Holt-Jensen 2018). There are several debates, divisions, and revolutions in the history of geography. One is the quantitative revolution, a period from the 1950s – 1960s, where geographers increasingly used statistical and mathematical techniques, theorems, and proofs in understanding geographical systems (Burton 1963). This brought 'scientific thinking' to geography, leading to an increased use of quantitative practices. Along this line, the development of GIS in the 1980s has further improved the technical bases and computational power for geographical research. In the late 1990s, Geocomputation was proposed as a new paradigm and defined by Openshaw and Alvanides (1999 pp.270) as the “adoption of a large-scale computationally intensive approach to the problems of physical and human geography in particular, and the geosciences in general”. Artificial intelligence tools were mentioned alongside GIS data as a family of computational approaches (Openshaw 1993; Openshaw and Openshaw 1997).

Since 2012, the field of artificial intelligence have expanded rapidly due to the possibilities offered from deep learning, for example in computer vision (Krizhevsky et al 2012) and language understanding tasks (Sutskever et al 2014). Such progresses have driven innovations in many domains and applications including autonomous driving (Geiger et al 2012), and medical science research (Jumper et al 2021) but also Geospatial artificial intelligence GeoAI in geography (Janowicz et al 2020). In this chapter, we focus on the recent progress in quantitative geography, namely machine learning and deep learning, and provide some projections on this topic for future geographical research and education. We will begin by giving a brief background to the topic.

2.0 From Machine Learning to Deep Learning in Geography

Most definitions of Artificial intelligence/Machine Learning begin with the famous question asking whether machines can think? (Turing 1950) These enquiries motivated early research in Machine Learning (Samuel 1959) which is a subfield or method of AI that signifies the ability for a computer to learn without being explicitly programmed. One of these research strands, inspired by biological neural networks, resulted in the development of the perceptron machine by Rosenblatt (1957). This idea of connecting neurons together to make a synthetic nervous system has led to the development of many different kinds of “Artificial” Neural Networks. Despite these early advances in learning machines, it wasn’t until (1) critical methodological developments in the 1990s, such as the backpropagation algorithm (LeCun, 1990), (2) advances in computational tools, such as the use of graphic processor (Raina et al 2009) and (3) the ubiquity of large scale data (Krizhevsky 2012) that more complex neural computing methods (known as “deep learning”) have become operational. These deep neural networks are composed of multiple processing layers and are able to automatically learn representations from data (LeCun, Bengio, Hinton 2015).

In Geography, the application of machine learning and artificial intelligence started prior to the deep learning revolution (Openshaw 1993; Openshaw and Openshaw 1997). This includes the use of ensemble learning methods in remote sensing (Benediktsson et al 1990), the application of artificial neural networks on spatial interaction models (Openshaw 1993) and the use of unsupervised methods on geodemographic profiling (Harris et al 2005; Singleton and Longley 2009). It wasn't until recently that deep learning has become accepted in quantitative geography, leading to a new subdomain known as Geospatial Artificial Intelligence or GeoAI (Janowicz et al 2020). Here, GeoAI refers to spatially explicit artificial intelligence techniques in geography. The driver of growth is fundamentally led by advances in deep learning methods, development of highly optimised and scalable computational tools, and importantly the increasing availability of spatially embedded big data (Kitchin 2014), like vector-based points data from location based social media (Shen et al 2019), mobility data from travel cards (Zhong et al 2016), street network data from OpenStreetMaps (Boeing 2017), spatially embedded textual data such as surnames (Longley et al 2011) and Earth-Observation data, for example from the Sentinel satellites (Drusch et al 2012).

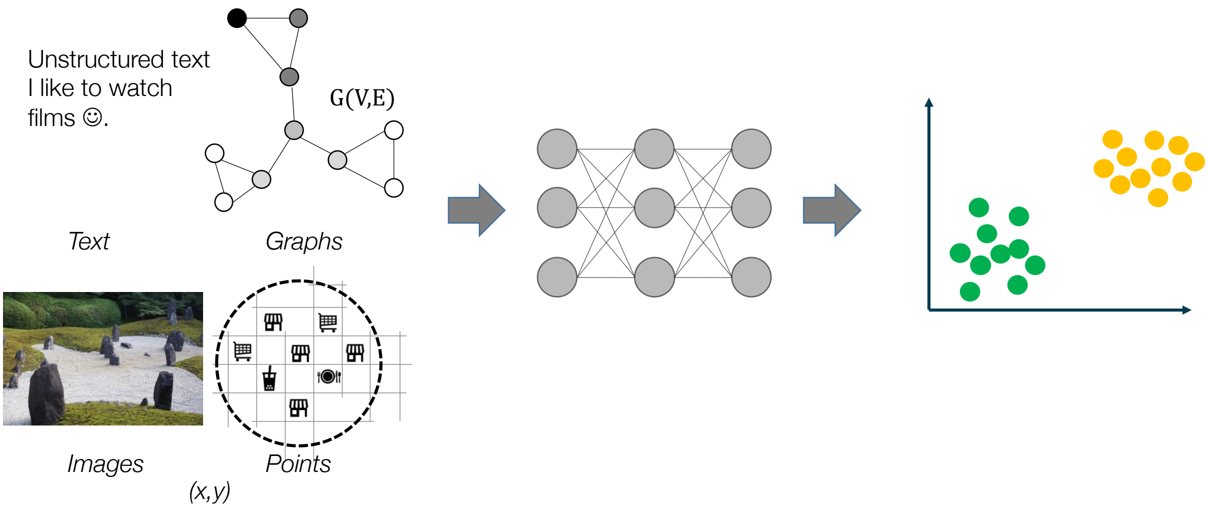

Fig1. Deep learning research focuses on learning a data representation (eg. text, graphs, points, images) that can capture the underlying structure of the data.

Most of deep learning research entails the use of deep neural networks to learn a mapping between the input data and a data-specific representation that can be useful for a down-stream task such as prediction, generation and discovery. We will now summarise a couple of example applications of deep learning methods applied to four types of geographical data, namely image, point, graph and text.

2.1 Image Data

Machine learning methods have been applied to remotely sense data in geography since the 1990s (Benediktsson et al 1990). One early example is the use of ensemble learning methods (such as random forests) for land use and land cover classification (Gisalson et al 2006). Due the increasing volume and resolution of open remotely sensed data coupled with the increasing capability of deep learning as a feature extractor, these techniques have been used extensively on many remote sensing tasks such as image restoration (Zhang et al 2014) and pixel level classification (Lagrange et al 2015).[1] The flexibility of deep learning techniques allows easy adaptation for different spatial, spectral and temporal resolutions, as well as integration with different data types. Once a deep learning model has been trained, forecasting can be done at scale. Recent applications include monthly sea-ice prediction (Andersson et al 2021), daily rainfall nowcasting (Ravuri et al 2021), tree canopy estimation (Francis and Law 2022; Weinstein et al 2019) and citywide crowd prediction (Zhang et al 2017).

Machine learning were also applied to imagery captured at the street level prior to deep learning. Early examples include the retrieval of classical computer vision features such as Histogram of Oriented Gradients (Dalal et al 2005) and Scale Invariant Features Transform (Lowe 2004) to discriminate urban architectural features in Paris (Doersch 2012), and the prediction of safety perception in the United States (Naik et al 2014). Similar to other domains, most of the recent research on street imagery

[1] A separate chapter in the book focuses on this topic.

(specially from 2015 onwards) focused predominantly on using deep learning methods such as convolutional neural networks and more recently vision transformers.[2] Examples include estimating demographic profiles in the States (Gebru et al 2017), nowcasting gentrification (Glaeser et al 2018), estimation of real estate values (Law et al 2020) and on predicting urban design quality (Law et al 2017; 2019). These methods, when used innovatively, can help the monitoring of urban environments for urban and transport planning.

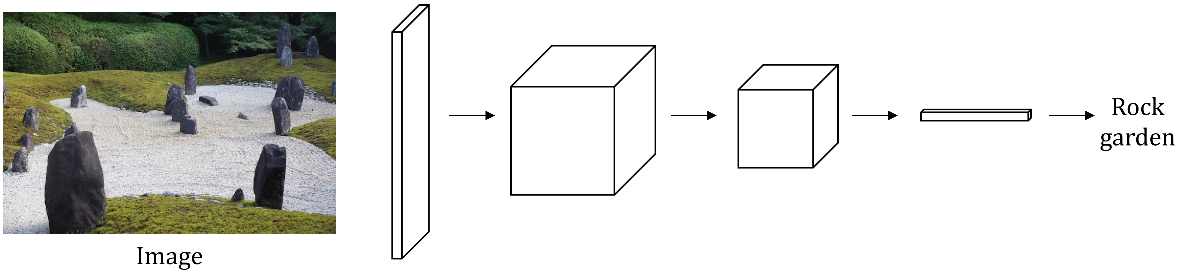

Fig 2. A standard Convolutional Neural Network classifier

Method

In terms of methods for image data, the most popular deep learning computer vision methods used in geography is a type of artificial neural network called Deep Convolutional Neural Networks (LeCun et al 1990; LeCun et al. 2015; Krizhevsky 2012). Convolutional Neural Networks (ConvNets) can learn a function that maps inputs, images in this case, to a representation that can then be used to classify a target class/value, for example in recognising whether the image is a rock garden or not (see Figure 1). This architecture consists of stacks of convolutional layers which learns a set of local spatial features hierarchically from image data.[3] These representative features can then be used in task such as scene recognition (Simonyan and Zisserman 2014), object detection (Girshick 2016) and semantic segmentation (Chen et al. 2017). ConvNets can also be adapted for unsupervised/generative learning. One variant is the variational autoencoder which embeds the data into a latent distribution through a set of encoding and decoding layers (Kingma et al 2013; Pu et al. 2016). Another popular variant is Generative Adversarial Network (GAN), which is composed of a generator that synthesizes an image from a random vector and a discriminator that distinguishes whether the generated image is a real or fake image (Goodfellow et al 2014). These generative models are able to synthesize photo-realistic images and are beginning to be used in geography (Zhu et al 2020).

2.2 Points Data

Much research in deep learning concerns mapping the input data into a data specific representation. This is not only true for image data but also for point data which is one of the most common data type in spatial analysis. A common example of this data type includes event data where each point corresponds to a particular spatial temporal event such as crime (Huang et al 2018). Deep learning with point data have been used in a wide array of studies including point of interest prediction (Mai et al 2020), next location prediction (De Brebisson et al 2015), point cloud classification (Qi et al 2017) and geo-aware image-classification (Chu et al 2019). Recent research encodes location information (Chu et al 2019) and contextual information for spatially explicit modelling (Mai et al 2020). Another common example of this data type is trajectory data where points in space and time are linked sequentially such as those captured from high frequencies GPS. Data mining on trajectory data to study human mobility pattern is widely studied in spatial analysis (Gonzalez et al 2008; Barbosa et al 2018). Deep learning methods have also been applied extensively on mobility data, for example in studying human mobility flow prediction (Alahi et al 2016; Xu et al 2018).

---

[2] For a more comprehensive review on the recent use of computer vision methods in urban analytics, please see Ibrahim et al (2020).

[3] More recently we have seen the rising popularity of Vision Transformers, inspired from natural language processing, as an alternative architecture for visual representation learning (Dosovitskiy et al 2020).

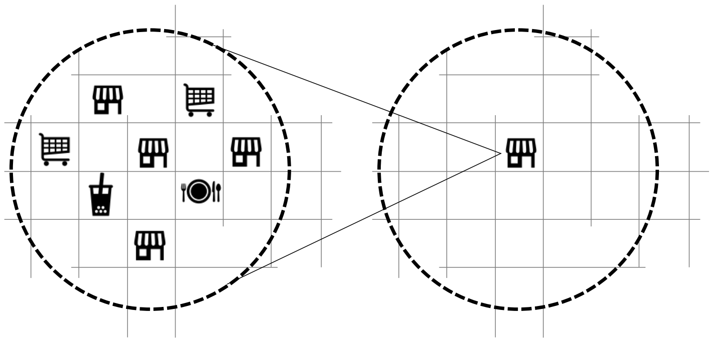

Fig 3. A key task in spatial data encoding is to capture both positional and contextual information for spatially explicit machine learning

Method
Early research on spatial point representation embeds space into some discrete value like a grid or postcode not dissimilar to the Bag of Words representation (Tang et al 2015) or embedding coordinates directly within a machine learning based image classifier (Chu et al 2019). Both of these results show improvement over geo-unaware classification models. In more recent research, we are seeing spatial representation learning that not only encode the points explicitly but also accounting for spatial relationship between points (see Figure 2). One example is Tile2Vec (Jean et al 2018), which is an unsupervised learning method that uses a triplet loss to learn a contextual continuous representation similar to Word2Vec for tasks such as poverty prediction with remotely sensed data. A more recent example is Space2Vec, inspired from research in neuroscience (Banino et al 2018), which uses an encoder-decoder architecture that encodes positional information and a space-aware graph attention network to capture multi-scale spatial relationships for tasks such as POI prediction or image classification (Mai et al 2020).

2.3 Graph Data
From streets to transportation networks, geospatial network (graph) data is another popular forms of data to study in transport geography (Haggett and Chorley 1970), urban planning and architecture (Hillier and Hansen 1989). This type of geospatial data model relationships between entities as complex networks that are embedded in space (Batty 2013). An example is the street network, where each junction is a node and each street is an edge, or a commuter flow network, where each city is a node and the flows between cities are edges (Barthelemy 2018). Geospatial networks are often modelled as graphs that are composed of a set of nodes and edges embedded in space (see Figure 3). Until recently, there have been limited applications of graph-based machine learning methods in geography. A recent example is Zhu et al (2020) which applied Graph Convolutional Neural Networks (GCN) on a nearest neighbours graph to predict place characteristics. Another popular use is traffic forecasting (Zhao et al 2019) where graph neural network are combined with Recurrent Neural Network (Fu et al. 2016) to capture spatial temporal dependencies on graphs.

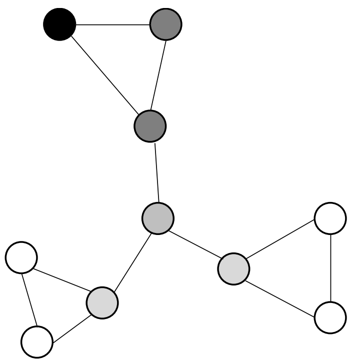

Fig 4. Geospatial networks are graphs that are made up of a set of nodes (vertices) connected by a set of edges embedded in space.

Method
Graph-based deep learning methods were popularised by Duvenaud et al (2015) and Kipf and Welling (2016), who developed a specific type of Graph Neural Network called a Graph Convolutional Neural Network (GCN). GCN, which adapts from ConvNet is similarly composed of multiple layers, where each layer aggregates the attributes of neighbouring nodes. The resulting node embeddings can be used for various downstream tasks such as node or graph classification. These seminal works led to many subsequent and related graph neural network methods such as GraphSage (Hamilton et al 2017) and Graph Attention Network (Velickovic et al. 2017).

2.4 Textual Data
Natural language processing (NLP) is a subfield within computational linguistics and artificial intelligence that tries to achieve a better understanding of natural language through computational methods (Manning and Schutze 1999). For example it can be used to understand the topic of a tweet (Lansley and Longley 2016) or to understand the sentiment trajectories of novels or films (Del Vecchio et al 2018). There are many subtasks within text analysis such as information retrieval and extraction as well as text classification, translation, generation and summarisation. In geography, text analysis is often used to recover textual information in a specific geographic location. This includes spatially embedded textual data such as geo-tagged tweets or spatially implicit data where the document describes a place that needs to be geocoded such as Wikipedia or news articles. Examples include what opinions or emotions are associated with a place (Ballatore and Adams 2015), how sentiment relates to specific environmental features (Wartmann and Purves 2018), social media topic modelling (Lansley and Longley 2016), the extraction of perceptual neighbourhood boundaries from social media data (Mckenzie and Adams 2017). More recently we have seen the increasing popularity of geographic knowledge graph that extracts information from geographic semantic entities that can be used for information retrieval and Geographical Question and Answering (Mai et al 2020b; Janowiscz et al 2022). To provide a general framework for understanding NLP tasks in geography, we find Hu et al. (2018)'s summary of geo-text data analysis useful. Hu et al. (2018) divide geo-text analysis tasks into the following groups:

- *Geoparsing*, which extracts and geolocate places,
- *Place relations,* which measures the co-occurrence of places within text,
- *Place sequencing,* which tracks individuals trajectories through text,
- *Place opinions* which models the sentiment of text about places such as districts or restaurants
- *Place zones*, which seeks to detect neighbourhoods from textual data about cities
- *Place impacts.* which identifies the real time attitudes of people in places during high impact events, as in disaster management.

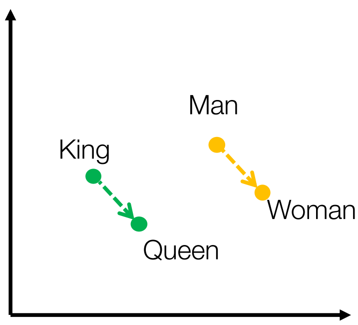

Fig 5. The idea of Neural Word Embedding adapted from Word2Vec concept (Mikolov et al 2013)

Method
A key concept in NLP is embedding text either at the character level, word level, sentence level or document level into a numerical representation. Prior to advances in neural computing, words were embedded by counting the term frequency, either in an unweighted (BagOfWords) or weighted form (Term-Frequency-Inverse-Document-Frequency). This type of representation produces a sparse matrix[4] that does not consider the ordering and context of words. These limitations lead to more efficient

[4]space inefficient with many zeroes in the matrix.

methods that analyze text using a continuous representation such as neural word embedding methods (Bengio et al 2001). An early variant of this type of embedding, such as Word2Vec (Mikolov et al 2013), uses shallow neural networks that embed terms and immediate context into a continuous representation where similar concepts would be nearer to each other than dissimilar concepts in the embedded vector space. A famous example is that the distance between “king” and “man” in the Word2Vec embedding is similar to the distance between “queen” and “woman” (see Figure 4). Other methods for text embedding include GloVe (Pennington et al 2014), fastText (Joulin et al 2016) and Bert embeddings. The latter makes use of self-attention layers in considering contextual relations between words (Vaswani et al. 2017; Devlin et al 2018). These neural word embedding methods can be used in downstream language modelling tasks such as text classification, document summarisation, text translation and text generation such as GPT-3 (Brown et al 2020). To note, these NLP architecture have since inspired Vision Transformer models

rising popularity of Vision Transformers, inspired from natural language processing, as an alternative architecture for visual representation learning (Dosovitskiy et al 2020).

To summarise, we have described a couple of examples on how deep learning methods are being applied on different types of spatial data commonly used in Geography including, image, point, network and text data. This list is not exhaustive and it is constantly being extended but it demonstrates the possibilities and potential of applying these methods in quantitative geography.

3.0 Towards Greater Integration

To continue, we envisage greater integration between geography and machine learning on three topics:

1. *Tackling geographical problems with deep learning methods while incorporating geographical thinking and representation in machine learning (ML <-> GEO)*
2. *Embracing cross disciplinary research*
3. *New norm in geographical research - Open research practice*

Tackling geographical problems with AI while incorporating geographical thinking in machine learning
Despite the continuous progress of technical advances in quantitative geography, we should focus on geographical problems, not solutions. Research on the creative use of deep learning techniques for traditional geographical problems are ascendant. Simultaneously, we are also seeing the incorporation of geographical thinking and representation in machine learning. We will describe a couple of recent examples here demonstrating applications of modern machine learning methods such as deep learning on traditional geographical problems such as spatial interaction, spatial interpolation and hedonic price models.

The first recent example is the estimation of spatial interaction models using a deep learning architecture entitled Deep Gravity Model (Simini et al 2021). In this research, the author framed an origin constrained spatial interaction model (Wilson 1971) as a two stage problem; in the first stage, the model estimates a set of destination probabilities using a fully connected feedforward neural network classifier, and in the second stage, the method multiplies the destination probabilities by the number of residents in the origin location to give the corresponding commuting flows to each destination. The authors found the deep gravity model achieves better results than a linear gravity model. These results are expected but more importantly it shows how deep learning can provide novel ways to study a classical geographical problem in estimating commuting flows between origins and destinations (Fotheringham and O’Kelly 1989).

The second is the use of a spatial auxiliary task for spatial interpolation (Klemmer et al 2021). The authors proposed a method for embedding the autoregressive nature of spatial data by utilising multi-resolution local Moran’s I as a model agnostic auxiliary task learner. The auxiliary task is a model agnostic method that can be plugged into standard deep learning models (eg. GAN and ConvNet) coupling a spatial loss into a multi-objective optimisation problem. The model agnostic method was tested on a spatial interpolation task using a standard Digital Elevation Modelling dataset. The Moran’s I embedded ConvNet outperformed benchmarks from geography including Inverse Distance Weighting (IDW), Ordinary Kriging and Universal Kriging. This research demonstrates how traditional geography theory in spatial analysis (Miller 2004; Longley et al. 2005) can be incorporated into modern deep learning approaches showing a bi-directional relationship between the two disciplines.

The third is the use of a semi-interpretable framework on house price prediction (Law et al 2019). A key criticism with machine learning method is its lack of interpretability as these models (often non-linear) focus on prediction accuracy rather than explainability (Molnar et al. 2020). In responding to this concern, Law et al (2019), developed a semi-interpretable model in predicting house price through a two stage process. In the first stage, street and aerial images were compressed into a visual desirability feature through two fine-tuned pretrained convolutional neural network models (Simonyan and Zisserman 2014). In the second stage, the visual feature can then be used as part of an hedonic price regression model in predicting house price. The research shows that a semi-interpretable framework that integrates non-interpretable models (like Artificial Neural Networks) with interpretable models (like linear models) allows for greater interpretability of the visual features geographically while achieving similar accuracy to a fully non-interpretable model. Importantly, this research shows how geography and mapping can help with interpreting the implicit representations of deep learning models, and also how non-interpretable models can be used in geographic regression problems such as hedonic price models.

Embracing Interdisciplinary Research

Another ascendant trend we are observing is increasing cross institute and departmental collaborative research and education in quantitative geography (Turing Institute as an example)[5]. A scan of recent papers published in the seminal GeoAI workshop in 2019 (Gao et al 2019) and 2021 (Lunga et al 2021) showed that 16 of the 27 papers presented in the workshop have included authors from multiple institute and/or across multiple departments within the same institute. These research efforts show cross discipline and institute collaborative research is becoming more common in the domain, which can help facilitate novel integration between geography and machine learning.

To further encourage geographic data science and spatially-explicit machine learning research in the long term, we need to integrate the teaching of quantitative data science methods in geography. An encouraging sign is that an online search shows at least twelve (at the time of writing) Postgraduate level masters courses in the UK that are teaching topics related to geographic data science or urban analytics (University College London, London School of Economics, Birkbeck University of London, University of Bristol, University of Liverpool, University of Leeds, University of Glasgow, Kings College London, University of Exeter, University of Manchester, Newcastle University). We believe this trend will continue in the near future, when there will be greater levels of geography-specific data science teaching resources.

New norm in geographical research - Open research practice

Machine learning is fundamentally about learning from data. Therefore, the quality of data sets and the reproducibility of research are essential qualities to building useful and transparent machine learning models. A welcoming trend that is happening concurrently is the acceptance of open data practices and products in geography (Arribas-Bel et al 2021) to ensure research in geographic data science is open, transparent and reproduceable. Examples include the production of well documented open teaching resources (Arribas-Bel 2019), the proliferation of computational notebooks (Boeing and Arribas-Bel 2021), the use of open data application interface such as osmNX (Boeing 2017) and Google Earth Engine (Gorelick et al 2017), the use of open source analytical tools such as PySAL (Rey et al 2021) and GeoPandas (Jordahl 2014) and the launch of journal sections for data and novel software or data products, such as urban data/code: a new section in *Environment and Planning B*[6] and *Scientific Data*[7]. These trends are also expected to continue in the near future.

4.0 Discussion and Conclusion

In summary, the utilisation of deep learning in geography will continue to grow due to rapid advances in AI and the growing abundance of spatial data. We have described some examples on how deep learning is currently applied on four spatial data types namely; images, points, graphs and texts. Additionally, we

[5] The Alan Turing Institute https://www.turing.ac.uk/
[6] Environment and Planning B: Urban Analytics and City Science. https://journals.sagepub.com/home/epb
[7] Nature Scientific Data. https://www.nature.com/sdata/

have described three ascendant topics that we see in the near future of GeoAI. With the aid of aforementioned advancements and increase cross-disciplinary engagements and open data practices, we envisaged new spatial insights, knowledge and methods to be discovered in the future.

Due to the topic's popularity, this summary is not exhaustive. We did not cover several importance topics, such as research on data ethics, bias and privacy in geography (Proctor 1998; Krumm 2009; Finn et al 2012), the application of reinforcement learning on multi-agent simulations (Heppenstall et al 2021; Crooks et al 2018; Martinez-Gil et al 2014) and the application of machine learning frameworks to optimise the analysis of large scale geospatial data (Richardson et al 2020). These topics are vital for the domain and are discussed elsewhere in this book or are planned to be addressed in future summaries.

Acknowledgement
This study has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 949670), and from ESRC under JPI Urban Europe/NSFC (grant No. ES/T000287/1). . The authors would also like to thank the referees and editors for their valuable comments.

References
Alahi, A., Goel, K., Ramanathan, V., Robicquet, A., Fei-Fei, L., & Savarese, S. (2016). Social lstm: Human trajectory prediction in crowded spaces. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 961-971).

Andersson, T.R., Hosking, J.S., Pérez-Ortiz, M., Paige, B., Elliott, A., Russell, C., Law, S., Jones, D.C., Wilkinson, J., Phillips, T. and Byrne, J., (2021). Seasonal Arctic sea ice forecasting with probabilistic deep learning. Nature communications, 12(1), pp.1-12.

Arribas-Bel, D., Green, M., Rowe, F., & Singleton, A. (2021). Open data products-A framework for creating valuable analysis ready data. Journal of Geographical Systems, 23(4), 497-514.

Arribas-Bel, D (2019). A course on geographic data science. The journal of open source education. 2(14).

Ballatore, A. and Adams, B. (2015). Extracting place emotions from travel blogs. In Proceedings of AGILE, volume 2015, pages 1–5.

Banino, A., Barry, C., Uria, B., Blundell, C., Lillicrap, T., Mirowski, P., Pritzel, A., Chadwick, M.J., Degris, T., Modayil, J. and Wayne, G., (2018). Vector-based navigation using grid-like representations in artificial agents. Nature, 557(7705), pp.429-433.

Barthelemy, M. (2018). Morphogenesis of spatial networks. Cham, Switzerland: Springer International Publishing.

Barbosa, H., Barthelemy, M., Ghoshal, G., James, C.R., Lenormand, M., Louail, T., Menezes, R., Ramasco, J.J., Simini, F. and Tomasini, M., (2018). Human mobility: Models and applications. Physics Reports, 734, pp.1-74.

Batty, M. (2013). The new science of cities. MIT press.

Bay, H., Tuytelaars, T., & Gool, L. V. (2006). Surf: Speeded up robust features. In European conference on computer vision (pp. 404-417). Springer, Berlin, Heidelberg.

Benediktsson, J. A., Swain, P. H., & Ersoy, O. K. (1990). Neural network approaches versus statistical methods in classification of multisource remote sensing data. Vancouver, Canada, July 10-14, 1989) IEEE Transactions on Geoscience and Remote Sensing.

Bengio, Y., Ducharme, R. & Vincent, P. (2001). A neural probabilistic language model. In Proc. Advances in Neural Information Processing Systems 13 932–938.

Boeing, G. (2017). OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. Computers, Environment and Urban Systems, 65, 126-139.

Boeing, G., & Arribas-Bel, D. (2021). GIS and computational notebooks. arXiv preprint arXiv:2101.00351.

Boulemtafes, A., Derhab, A., Challal, Y., (2020). A review of privacy-preserving techniques for deep learning. Neurocomputing Vol.384.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J.D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A. and Agarwal, S., (2020). Language models are few-shot learners. Advances in neural information processing systems, 33, pp.1877-1901.

Burton, I. (1963). The quantitative revolution and theoretical geography. Canadian Geographer. 7 (4): 151–62. doi:10.1111/j.1541-0064.1963.tb00796.x.--as cited in Johnston, Ron and Sideway James (2016). Geography and Geographers: Anglo-American Human Geography since 1945 (7th ed). New York: Routledge.

Chehata, N., Guo, L., & Mallet, C. (2009). Airborne lidar feature selection for urban classification using random forests. In Laserscanning.

Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2017). Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE transactions on pattern analysis and machine intelligence, 40(4), 834-848.

Chen, G., Li, S., Knibbs, L.., Hamm, N. A. S., Cao, W., Li, T., Guo, J., Ren, H., Abramson, M., Guo, Y. (2018). A machine learning method to estimate PM2.5 concentrations across China with remote sensing, meteorological and land use information. Science of the total environment. 636. Pp.52-60.

Crooks, A., Malleson, N., Manley, E., & Heppenstall, A. (2018). Agent-based modelling and geographical information systems: a practical primer. Sage.

Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. In 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05) (Vol. 1, pp. 886-893). Ieee.

De Brébisson, A., Simon, É., Auvolat, A., Vincent, P., & Bengio, Y. (2015). Artificial neural networks applied to taxi destination prediction. In Proceedings of the 2015th International Conference on ECML PKDD Discovery Challenge - Volume 1526 (ECMLPKDDDC'15). CEUR-WS.org, Aachen, DEU, 40–51.

Del Vecchio, M., Kharlamov, A., Parry, G., & Pogrebna, G. (2018). The Data science of Hollywood: Using emotional arcs of movies to drive business model innovation in entertainment industries. Available at SSRN 3198315.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Vol.1. pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Doersch, C., Singh, S., Gupta, A., Sivic, J., & Efros, A. (2012). What makes paris look like paris?. ACM Transactions on Graphics, 31(4).

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S. and Uszkoreit, J., (2020). An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929.

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P. and Meygret, A., 2012. Sentinel-2: ESA's optical high-resolution mission for GMES operational services. Remote sensing of Environment, 120, pp.25-36.

Duvenaud, D. K., Maclaurin, D., Iparraguirre, J., Bombarell, R., Hirzel, T., Aspuru-Guzik, A., & Adams, R. P. (2015). Convolutional networks on graphs for learning molecular fingerprints. Advances in neural information processing systems, 28.

Fischer, MM and Gopal, S. (1994). Artificial Neural Network: a new approach to modelling inter-regional communication flows. Journal of Regional science 34(4): 503-27.

Finn, R. Wright, D. (2012) Unmanned aircraft systems: Surveillance, ethics and privacy in civil applications. Computer Law & Security Review 28. pp184-194.

Fotheringham, A. S., & O'Kelly, M. E. (1989). Spatial interaction models: formulations and applications (Vol. 1, p. 989). Dordrecht: Kluwer Academic Publishers.

Fu, Rui, Zuo Zhang, and Li Li. (2016). Using LSTM and GRU neural network methods for traffic flow prediction. In 2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC), pp. 324-328. IEEE, 2016.

Gao, S., Newsam, S., Zhao, L., Lunga, D., Hu, Y., Martins, B., Zhou, X. and Chen, F. eds., (2019). Proceedings of the 3rd ACM SIGSPATIAL International Workshop on AI for Geographic Knowledge Discovery. Associations for Computing Machinery (ACM).

Gebru, T., Krause, J., Wang, Y., Chen, D., Deng, J., Aiden, E. L., & Fei-Fei, L. (2017). Using deep learning and Google Street View to estimate the demographic makeup of neighborhoods across the United States. Proceedings of the National Academy of Sciences, 114(50), 13108-13113.

Geiger, A., Lenz, P., & Urtasun, R. (2012). Are we ready for autonomous driving? the kitti vision benchmark suite. In 2012 IEEE conference on computer vision and pattern recognition (pp. 3354-3361). IEEE.

Glaeser, E. L., Kim, H., & Luca, M. (2018). Nowcasting gentrification: using yelp data to quantify neighborhood change. In AEA Papers and Proceedings, 108, 77-82.

Gislason, P. O., Benediktsson, J. A., & Sveinsson, J. R. (2006). Random forests for land cover classification. Pattern recognition letters, 27(4), 294-300.

Gonzalez, M. C., Hidalgo, C. A., & Barabasi, A. L. (2008). Understanding individual human mobility patterns. nature, 453(7196), 779-782.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. and Bengio, Y., (2014). Generative adversarial nets. Advances in neural information processing systems, 27.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote sensing of Environment, 202, 18-27.

Girshick, R. (2015). Fast r-cnn. In Proceedings of the IEEE international conference on computer vision (pp. 1440-1448).

Haggett, P., & Chorley, R. J. (1970). Network analysis in geography (Vol. 1). St. Martin's Press.

Harris, R., Sleight, P., & Webber, R. (2005). Geodemographics, GIS and neighbourhood targeting (Vol. 7). John Wiley and Sons.

Handy, S. L., & Niemeier, D. A. (1997). Measuring accessibility: an exploration of issues and alternatives. Environment and planning A, 29(7), 1175-1194.

Heppenstall, A., Crooks, A., Malleson, N., Manley, E., Ge, J., & Batty, M. (2021). Future developments in geographical agent-based models: Challenges and opportunities. Geographical Analysis, 53(1), 76-91.

Hillier, B., & Hanson, J. (1989). The social logic of space. Cambridge university press.

Holt-Jensen, A., (2018). Geography: history and concepts. Sage.

Hu, Y. (2018). Geo-text data and data-driven geospatial semantics. Geography Compass, 12(11), e12404.

Hu, Y., Gao, S., Lunga, D., Li, W., Newsam, S., & Bhaduri, B. (2019). GeoAI at ACM SIGSPATIAL: progress, challenges, and future directions. Sigspatial Special, 11(2), 5-15.

Huang, C., Zhang, J., Zheng, Y., & Chawla, N. V. (2018). DeepCrime: Attentive hierarchical recurrent networks for crime prediction. In Proceedings of the 27th ACM international conference on information and knowledge management (pp. 1423-1432).

Ibrahim, M. R., Haworth, J., & Cheng, T. (2020). Understanding cities with machine eyes: A review of deep computer vision in urban analytics. Cities, 96, 102481.

Jordahl, K. (2014). GeoPandas: Python tools for geographic data. URL: https://github. com/geopandas/geopandas, 3.

Klemmer, K., & Neill, D. B. (2021). Auxiliary-task learning for geographic data with autoregressive embeddings. In Proceedings of the 29th International Conference on Advances in Geographic Information Systems (pp. 141-144).

Janowicz, K., Gao, S., McKenzie, G., Hu, Y., & Bhaduri, B. (2020). GeoAI: spatially explicit artificial intelligence techniques for geographic knowledge discovery and beyond. International Journal of Geographical Information Science, 34(4), 625-636.

Janowicz, K., Hitzler, P., Li, W., Rehberger, D., Schildhauer, M., Zhu, R., Shimizu, C., Fisher, C.K., Cai, L., Mai, G. and Zalewski, J., (2022). Know, Know Where, KnowWhereGraph: A densely connected, cross-domain knowledge graph and geo-enrichment service stack for applications in environmental intelligence. AI Magazine, 43(1), pp.30-39.

Joulin, A., Grave, E., Bojanowski, P., & Mikolov, T. (2016). Bag of tricks for efficient text classification. arXiv preprint arXiv:1607.01759.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A. and Bridgland, A., (2021). Highly accurate protein structure prediction with AlphaFold. Nature, 596(7873), pp.583-589.

Jurafsky, D. (2000). Speech & language processing. Pearson Education India.

Kingma, D. P., & Welling, M. (2013). Auto-encoding variational bayes. The 2nd International Conference on Learning Representations (ICLR2014). arXiv preprint arXiv:1312.6114.

Kipf, T. N., & Welling, M. (2016). Semi-supervised classification with graph convolutional networks. The 5th International Conference on Learning Representations (ICLR2017) arXiv preprint arXiv:1609.02907.

Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. Big data & society, 1(1), 2053951714528481.

Krizhevsky, A. Ilya Sutskever, Geoffrey E Hinton (2012). ImageNet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems.

Krumm, J. (2009). A survey of computational location privacy. Personal and Ubiquitous Computing, 13(6), 391-399.

Lagrange, A., Le Saux, B., Beaupere, A., Boulch, A., Chan-Hon-Tong, A., Herbin, S., Randrianarivo, H. and Ferecatu, M., (2015) Benchmarking classification of earth-observation data: From learning explicit features to convolutional networks. In 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS) (pp. 4173-4176). IEEE.

Lansley, G., & Longley, P. A. (2016). The geography of Twitter topics in London. Computers, Environment and Urban Systems, 58, 85-96.

Law, S., Seresinhe, C.I., Shen, Y., Gutierrez-Roig, M.,(2020). Street-frontage-net: urban image classification using deep convolutional neural networks.International Journal of Geographical Information Science,1–27

Law, S., Paige, B., & Russell, C. (2019). Take a look around: using street view and satellite images to estimate house prices. ACM Transactions on Intelligent Systems and Technology (TIST), 10(5), 1-19.

Le, Q., & Mikolov, T. (2014). Distributed representations of sentences and documents. In International conference on machine learning (pp. 1188-1196). PMLR.

LeCun, Y.; Boser, B.; Denker, J. S.; Henderson, D.; Howard, R. E.; Hubbard, W.; Jacker, L. D. (1990). Handwritten digit recognition with a back-propagation network. Advances in Neural Information Processing Systems.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. nature, 521(7553), 436-444.

Li, X., Zhang, C.,Li, W.,Ricard, R.,Meng, Q.,Zhang, W. (2015). Assessing street-level urban greenery using Google Street View and a modified green view index. Urban Forestry and Urban Greening. Vol.14-3. pp.675-685.

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. International journal of computer vision, 60(2), 91-110.

Longley, P. A., Goodchild, M. F., Maguire, D. J., & Rhind, D. W. (2005). Geographic information systems and science. John Wiley & Sons.
Longley, P. A., Cheshire, J. A., & Mateos, P. (2011). Creating a regional geography of Britain through the spatial analysis of surnames. Geoforum, 42(4), 506–516.

Lunga, D., Yang, L., Gao, S., Martins, B., Hu, Y., Deng, X., & Newsam, S. (Eds.). (2021). Proceedings of the 4th ACM SIGSPATIAL International Workshop on AI for Geographic Knowledge Discovery. Associations for Computing Machinery (ACM).

Manning, C., & Schutze, H. (1999). Foundations of statistical natural language processing. MIT press.

Mai, G.C., Janowicz, K., Yan, B., Zhu, R., Cai, L., Lao, N. (2020). "Multi-scale representation learning for spatial feature distributions using grid cells." In: Proceedings of ICLR 2020, Apr. 26 - 30, 2020, Addis Ababa, ETHIOPIA .

Mai, G., Janowicz, K., Cai, L., Zhu, R., Regalia, B., Yan, B., Shi, M. and Lao, N., (2020b). SE-KGE: A location-aware knowledge graph embedding model for geographic question answering and spatial semantic lifting. Transactions in GIS, 24(3), 623-655.

Martinez-Gil, F., Lozano, M., & Fernández, F. (2014). MARL-Ped: A multi-agent reinforcement learning based framework to simulate pedestrian groups. Simulation Modelling Practice and Theory, 47, 259-275.

Miller, D., Kaminsky, E., Rana, S. (1995). Neural Network classification of Remote sensing data. Computers and Geosciences 21(3): 377-386.

Miller, H. J. (2004). Tobler's first law and spatial analysis. Annals of the association of American geographers, 94(2), 284-289.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G., Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In Advances in neural information processing systems, pp. 3111–3119, 2013.

Michell, T. (1997). Machine Learning. McGraw Hill.

Molnar, C., Casalicchio, G., & Bischl, B. (2020). Interpretable machine learning–a brief history, state-of-the-art and challenges. In Joint European Conference on Machine Learning and Knowledge Discovery in Databases (pp. 417-431). Springer, Cham.

Nadai, M., Xu, Y., Letouze, E., Gonzalez, M., Lepri, B. (2020). Socio-economic, built environment, and mobility conditions associated with crime: a study of multiple cities. Scientific Reports. 10. 13871.

Naik, N., Philipoom, J., Raskar, R., & Hidalgo, C. (2014). Streetscore-predicting the perceived safety of one million streetscapes. In Proceedings of the IEEE conference on computer vision and pattern recognition workshops (pp. 779-785).

Newell, A. & Simon, H. (1956). The logic theory machine: A complex information processing system. IRE Transactions on Information Theory, 2, 61-79.

Openshaw, S. (1993). Modelling spatial interaction using a neural net. In Geographic information systems, spatial modelling and policy evaluation (pp. 147-164). Springer, Berlin, Heidelberg.

Openshaw, S. and Openshaw, C., (1997). Artificial intelligence in geography. 1st ed. New York, NY: John Wiley & Sons, Inc. ISBN 0471969915.

Openshaw, S. and Alvanides, S., (1999). Applying geocomputation to the analysis of spatial distributions. Geographical Information Systems: Principles and Technical issues, 1, pp.267-282.

Proctor, J. D. (1998). Ethics in geography: giving moral form to the geographical imagination. Area, 30(1), 8-18.

Pu, Y., Gan, Z., Henao, R., Yuan, X., Li, C., Stevens, A., & Carin, L. (2016). Variational autoencoder for deep learning of images, labels and captions. Advances in neural information processing systems, 29.

Qi, C. R., Su, H., Mo, K., & Guibas, L. J. (2017). Pointnet: Deep learning on point sets for 3d classification and segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 652-660).

Raina, R., Madhavan, A. & Ng, A. Y. (2009). Large-scale deep unsupervised learning using graphics processors. In Proc. 26th Annual International Conference on Machine Learning 873–880.

Ravuri, S., Lenc, K., Willson, M., Kangin, D., Lam, R., Mirowski, P., Fitzsimons, M., Athanassiadou, M., Kashem, S., Madge, S. and Prudden, R., (2021). Skilful precipitation nowcasting using deep generative models of radar. Nature, 597(7878), pp.672-677.

Rey, S.J., Anselin, L., Amaral, P., Arribas-Bel, D., Cortes, R.X., Gaboardi, J.D., Kang, W., Knaap, E., Li, Z., Lumnitz, S. and Oshan, T.M., (2021). The pysal ecosystem: Philosophy and implementation. Geographical Analysis.

Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. Psychological Review, 65(6), 386–408.

Richardson, B., Rees, B., Drabas, T., Oldridge, E., Bader, D. A., & Allen, R. (2020). Accelerating and Expanding End-to-End Data Science Workflows with DL/ML Interoperability Using RAPIDS. In Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (pp. 3503-3504).

Samuel, A. L. (1959). Some studies in machine learning using the game of checkers. IBM Journal of research and development, 3(3), 210-229.

Shen, Y., Karimi, K., Law, S., & Zhong, C. (2019). Physical co-presence intensity: Measuring dynamic face-to-face interaction potential in public space using social media check-in records. PloS one, 14(2), e0212004.

Simonyan, K. and Zisserman, A. (2015) Very Deep Convolutional Networks for Large-Scale Image Recognition. The 3rd International Conference on Learning Representations (ICLR2015). arXiv preprint arXiv:1409.1556.

Simini, F., Barlacchi, G., Luca, M., & Pappalardo, L. (2021). A Deep Gravity model for mobility flows generation. Nature communications, 12(1), 1-13.

Singleton, A. D., & Longley, P. A. (2009). Geodemographics, visualisation, and social networks in applied geography. Applied Geography, 29(3), 289-298.

Stewart, A. J., Robinson, C., Corley, I. A., Ortiz, A., Ferres, J. M. L., & Banerjee, A. (2021). TorchGeo: deep learning with geospatial data. arXiv preprint arXiv:2111.08872.

Sutskever, I. Vinyals, O. & Le. Q. V. (2014). Sequence to sequence learning with neural networks. In Proc. Advances in Neural Information Processing Systems 27 3104–3112.

Turing, A. M. (1950). I. - Computing machinery and intelligence. Mind, 59(236), 433-460.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł. and Polosukhin, I., (2017). Attention is all you need. Advances in neural information processing systems, 30.

Wartmann, F. M., Acheson, E., and Purves, R. S. (2018). Describing and comparing landscapes using tags, texts, and free lists: an interdisciplinary approach. International Journal of Geographical Information Science, pages 1–21.

Wilson AG (1971) A family of spatial interaction models, and associated developments. Environment and Planning A 3: 1–32. CrossRef.

Xu, Y., Piao, Z., & Gao, S. (2018). Encoding crowd interaction with deep neural network for pedestrian trajectory prediction. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 5275-5284).

Yan, B., Janowicz, K., Mai, G., & Zhu, R. (2019). A spatially explicit reinforcement learning model for geographic knowledge graph summarization. Transactions in GIS, 23(3), 620-640.

Zhang, J. , Zhong, P., Chen, Y., S. Li, (2014). “L(1/2)-regularized deconvolution network for the representation and restoration of optical remote sensing images,” IEEE Trans. Geosci. Remote Sensing.,vol. 52, no. 5, pp. 2617–2627, 2014.

Zhang, J., Zheng, Y., & Qi, D. (2017). Deep spatio-temporal residual networks for citywide crowd flows prediction. In Thirty-first AAAI conference on artificial intelligence.

Zhao, L., Song, Y., Zhang, C., Liu, Y., Wang, P., Lin, T., Deng, M. and Li, H., (2019). T-gcn: A temporal graph convolutional network for traffic prediction. IEEE Transactions on Intelligent Transportation Systems, 21(9), pp.3848-3858.

Zhong, C., Batty, M., Manley, E., Wang, J., Wang, Z., Chen, F., & Schmitt, G. (2016). Variability in regularity: Mining temporal mobility patterns in London, Singapore and Beijing using smart-card data. PloS one, 11(2), e0149222.

Zhu, D., Cheng, X., Zhang, F., Yao, X., Gao, Y., & Liu, Y. (2020). Spatial interpolation using conditional generative adversarial neural networks. International Journal of Geographical Information Science, 34(4), 735-758.