Machine Learning Meets Big Spatial Data

Ibrahim Sabek
Department of Computer Science and Engineering
University of Minnesota, Minnesota, USA
sabek001@umn.edu

Mohamed F. Mokbel*
Qatar Computing Research Institute
Hamad bin Khalifa University, Doha, Qatar
mmokbel@hbku.edu.qa

***Abstract*—The proliferation in amounts of generated data has propelled the rise of scalable machine learning solutions to efficiently analyze and extract useful insights from such data. Meanwhile, spatial data has become ubiquitous, e.g., GPS data, with increasingly sheer sizes in recent years. The applications of big spatial data span a wide spectrum of interests including tracking infectious disease, climate change simulation, drug addiction, among others. Consequently, major research efforts are exerted to support efficient analysis and intelligence inside these applications by either providing spatial extensions to existing machine learning solutions or building new solutions from scratch. In this 90-minutes tutorial, we comprehensively review the state-of-the-art work in the intersection of machine learning and big spatial data. We cover existing research efforts and challenges in three major areas of machine learning, namely, data analysis, deep learning and statistical inference. We also discuss the existing end-to-end systems, and highlight open problems and challenges for future research in this area.**

I. Introduction

There has been a recent wide deployment of machine learning (ML) solutions, with their different areas (e.g., data analysis, deep learning), in various big data applications, including public health [20], information extraction [51], data cleaning [40], among others. Meanwhile, spatial applications have witnessed unprecedented explosion in the amounts of generated and collected data. For example, space telescopes generate up to 150 GB weekly spatial data, medical devices produce spatial images (X-rays) at a rate of 50 PB per year, while a NASA archive of satellite earth images has more than 500 TB. To efficiently process such tremendous amounts of spatial data, researchers and developers worldwide have proposed either spatial extensions to existing machine learning systems (e.g., Azure Geo AI [2]) or new end-to-end solutions (e.g., ESRI ArcGIS [11]). Such extensions and new solutions have motivated a wide variety of applications in biology [55], environmental science [56], climatology [14], among others.

In this tutorial, we aim to provide a comprehensive review of existing machine learning systems and approaches that efficiently support big spatial data. Figure 1 depicts the landscape of the intersection between machine learning and big spatial data worlds that will be covered in this tutorial. The horizontal axis in Figure 1 represents the type of each machine learning solution, whether it takes the distinguishing spatial data properties into account or not, while the vertical axis represents the type of application employing such machine learning solution, whether the application is spatial or not. We mainly focus on the three quarters Q_1, Q_2, and Q_3 in Figure 1 because they cover the spatial dimension in the machine learning solutions and/or the big data applications. We skip the quarter Q_0 as it is already covered by previous SIGMOD tutorials about the techniques and challenges in machine learning for big data in general [6], [28]. Also note that another previous VLDB tutorial focused on big spatial data management [10]. Unlike this tutorial, our tutorial aims to combine the two worlds of *scalable machine learning* and *big spatial data* together, which is beyond just applying techniques from one area to another.

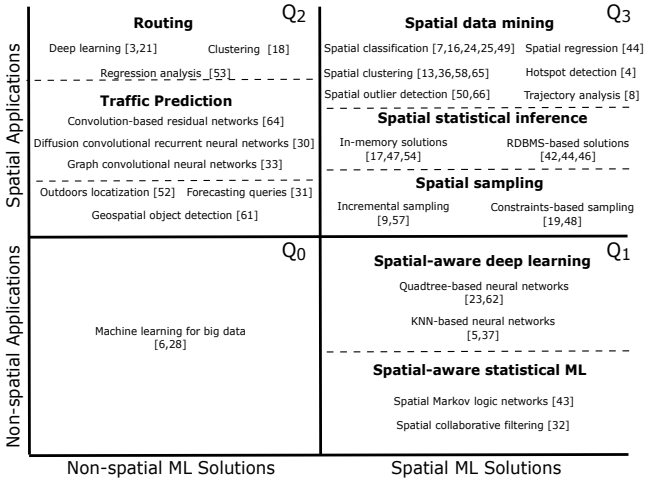

Fig. 1. Landscape of Machine Learning for Big Spatial Data.

In each of the three quarters Q_1, Q_2, and Q_3, we explain the main ideas, architectures, strengths and weaknesses of existing machine learning solutions. We also highlight the strong bond between spatial data management and spatial machine learning workflows, discuss the related technical challenges, and outline the open research opportunities.

II. Tutorial Outline

Figure 2 gives the **90-minutes** tutorial outline, composed of five parts. The first part motivates the need for machine learning systems to support big spatial data, and provides the basic background on these two worlds (Section II-A). The second, third, and fourth parts delve into the ongoing machine

*Also affiliated with University of Minnesota, MN, USA.

[1]This work is partially supported by the National Science Foundation, USA, under Grants IIS-1907855, IIS-1525953 and CNS-1512877.

2375-026X/20/$31.00 ©2020 IEEE
DOI 10.1109/ICDE48307.2020.00169

- **Part 1: Spatial Data and ML Synergy (10 mins)**
  - Importance of ML with big spatial data
  - Quick review of spatial ML landscape
- **Part 2: Spatial ML Solutions for Non-spatial Apps (20 mins)**
  - Spatial-aware deep learning solutions
  - Spatial-aware statistical ML models
- **Part 3: Non-spatial ML Solutions for Spatial Apps (25 mins)**
  - Deep learning and data analysis for routing apps
  - Deep and reinforcement learning for traffic prediction apps
  - Deep learning for localization and spatial object detection
- **Part 4: Spatial ML Solutions for Spatial Apps (20 mins)**
  - Scalable spatial data mining techniques
  - Scalable spatial statistical inference techniques
  - Scalable spatial sampling techniques
- **Part 5: End-to-end Spatial Data Analysis Systems (15 mins)**
  - Spatial support in existing big data analysis systems
  - Full-fledged big spatial data analysis systems

Fig. 2. Tutorial Outline (90 minutes)

learning efforts and challenges in the quarters Q_1, Q_2, and Q_3 from Figure 1, respectively (Sections II-B to II-D). The fifth part reviews the existing end-to-end systems for big spatial data analysis (Section II-E).

A. Part 1: Spatial Data and ML Synergy

This part advocates for the need to develop machine learning systems and techniques for big spatial data that go beyond simple extensions of existing work for general data. We start by describing some motivating applications, introducing the world of big spatial data, and discussing its machine learning related concepts. We then quickly review the landscape of spatial machine learning systems, algorithms, applications, and needs, which will be heavily discussed in the next parts.

B. Part 2: Spatial ML Solutions for Non-spatial Apps

This part covers the role of injecting the spatial awareness inside the underlying machine learning algorithms used in non-spatial applications (e.g., knowledge base construction [43], recommendation systems [32], computer vision [23]) to improve the performance of these applications. We start by highlighting how the spatial data management techniques improve the performance of various deep learning tasks when applying on big spatial data. For example, Quad-tree partitioning [15] is used for: (a) balancing the convolution computation in Convolutional Neural Networks (CNN) for object detection applications [23] and (b) efficient automatic features extraction and matrix factorization operations inside deep learning models [62]. Meanwhile, k-nearest neighbor operations are used to efficiently build specific neural network architectures from big spatial datasets [5], [37]. Then, we discuss the improved spatial variations of other statistical machine learning techniques (i.e., not deep learning) used inside knowledge base construction [43], [45] and recommendation [32] models, while assuring their impact in obtaining more accurate outputs.

C. Part 3: Non-spatial ML Solutions for Spatial Apps

This part covers the usage of existing machine learning techniques, without spatial variations, as "black boxes" in improving the performance of spatial applications. We start by discussing the recent machine learning techniques used inside two specific core applications; routing and traffic prediction. For routing, we show the deep learning [21] and regression analysis [53] techniques used to prepare the routing meta-data (e.g., finding weights of routes). We also present the incremental learning [3] and clustering [18] approaches that are used to make routing maps and perform the routing itself, respectively. For traffic prediction, we discuss its existing deep learning (e.g., convolution-based residual networks [64], diffusion convolutional recurrent neural networks [30], graph convolutional neural networks [33]), as well as reinforcement learning [59] approaches in details. Finally, we give a brief about the machine learning approaches used in other spatial applications including outdoors localization [52], forecasting queries [31], and geospatial object detection [61].

D. Part 4: Spatial ML Solutions for Spatial Apps

This part covers the research efforts for scaling up the performance of three main categories of spatial machine learning and analysis techniques: (1) *Spatial data mining*: common operations in this category include spatial outlier detection [50], [66], spatial classification [7], [16], [24], [25], [49], spatial regression [44], spatial clustering [13], [36], [58], [65], hotspot detection [4], and trajectory analysis [8]. (2) *Spatial statistical inference*: existing spatial inference approaches are categorized into: (a) *in-memory* solutions, where the input dataset of the inference model is first spatially partitioned into a grid. Then, each partition is analyzed using a Bayesian spatial process model (e.g., [17]). Finally, an approximate posterior inference for the entire dataset is obtained by optimally combining the individual posterior distributions from each partition [17], [47], [54]. (b) *RDBMS-based* solutions, where the assumption of fitting the whole model data in memory is no longer valid. Hence, RDBMSs are exploited to support scalable spatial inference computation (e.g., TurboReg [44] and Flash [42], [46]). (3) *Spatial sampling*: due to the massive amounts of spatial data that are available for training any spatial machine learning algorithm, spatial sampling becomes a critical task to efficiently select a set of representative data objects while taking the spatial distribution into account. Existing sampling techniques over big spatial data can be either incremental (i.e., samples are refined over many iterations) [9], [57] or satisfying certain locality constraints (e.g., zooming level) [19], [48].

E. Part 5: End-to-end Spatial Data Analysis Systems

This part covers the big spatial data analysis systems from two aspects: (1) The research efforts of adding spatial support in existing big data analysis systems, which are either: (a) in the form of add-ons libraries and tools that enable processing spatial data with classical operations (e.g., clustering, classification). Examples include spatial extensions to Spark

core (e.g., Simba [60], Magellan [34], GeoSpark [63], GeoMesa [22], UlTraMan [8]) to enable using Spark MLib [35] with spatial data, ESRI spatial data analysis extensions for Hive [12], and PostGIS [38] that can be used along with MADLib [20] to support spatial analytics for PostgreSQL [39], or (b) in the form of built-in native support of spatial analysis operations (e.g., hot spot detection, spatial co-location) inside existing data analysis engines. (2) The research efforts of providing full-fledged big spatial data analysis systems and tools. In such systems, all execution steps in any data analysis operation are optimized for efficient and scalable processing of spatial data. We will classify existing work based on the underlying architecture, which could be either (a) *in-memory systems* (e.g., CrimeStat [29], GeoDa [1], PySAL [41]), (b) *RDBMS-based systems* (e.g., ESRI ArcGIS [11], Flash [46]), or (c) *cloud-based services* (e.g., IBM PAIRS [26]). For all these systems and services, we will give motivational case studies, and a brief on their supported spatial analysis operations and running time efficiency.

III. Target Audience

This tutorial targets researchers, developers, and practitioners, who are interested in large-scale machine learning and big spatial data. No prior knowledge is required to understand the systems and approaches in the tutorial. The tutorial will also be very beneficial for graduate students as it will help in identifying various topics and challenges for PhD topics. Practitioners will get to know the state-of-the-art systems for enriching their machine learning systems and tools with spatial data support. This tutorial will act as an invitation to the database community to join arms for satisfying the emerging needs of big spatial data analysis and machine learning applications.

IV. Relevance to ICDE

Research in the areas of spatial data and scalable machine learning has been always active in the database community in general, and in the ICDE community in particular. With the proliferation of proposed systems and approaches in these areas, it becomes inevitable to present a tutorial that surveys the current state-of-the-art techniques and suggests future research directions for the community. Many of the research efforts covered in this tutorial were recently published in major database conferences including ICDE, VLDB, and SIGMOD [4], [8], [18]–[21], [27], [36], [42], [43], [45], [46], [48], [51], [57], [58], [60], [65].

V. Prior Offerings

Mohamed Mokbel and Ibrahim Sabek have recently presented a 90-minutes tutorial about the same topic in the Very Large Data Bases (VLDB) conference 2019[2], which focused more on the individual machine learning algorithms that are used to extract useful insights and patterns from big spatial data. In contrast, this tutorial probes the whole landscape of the machine learning and big spatial data while equally focusing on both algorithmic and application sides. In addition, the tutorial delves into the internals of the existing end-to-end systems of big spatial data analysis.

[2]http://www.cs.umn.edu/∼sabek/vldb-2019-tutorial/

VI. Biographical Sketches

Ibrahim Sabek is a PhD candidate at the department of Computer Science and Engineering, University of Minnesota. He received his M.Sc. degree at the same department in 2017. His research interests lie in the intersection area between big spatial data management, spatial computing, and scalable machine learning systems. Ibrahim has been awarded the University of Minnesota Doctoral Dissertation Fellowship in 2019 for this dissertation focus on scalable machine learning for big spatial data and applications. His research work has won the first place of ACM SIGSPATIAL Student Research Competition (SRC) 2019, and has been nominated for the Best Paper Award of ACM SIGSPATIAL 2018. During his PhD, he has collaborated with NEC Labs America, and Microsoft Research (MSR) in Redmond. For more information, please visit: http://www.cs.umn.edu/∼sabek.

Mohamed F. Mokbel is the Chief Scientist of Qatar Computing Research Institute and a Professor at University of Minnesota. His current research interests focus on systems and machine learning techniques for big spatial data and applications. His research work has been recognized by the VLDB 10-years Best Paper Award, four conference Best Paper Awards, and the NSF CAREER Award. Mohamed has delivered seven tutorials in VLDB/SIGMOD/ICDE/EDBT conferences, in addition to tutorials in other communities' first-tier venues, including IEEE ICDM and ACM CCS. Mohamed is the past elected Chair of ACM SIGPATIAL, current Editor-in-Chief for Distributed and Parallel Databases Journal, and on the editorial board of ACM Books, ACM TODS, VLDB Journal, ACM TSAS, and GoeInformatica journals. He has also served as PC Vice Chair of ACM SIGMOD and PC Co-Chair for ACM SIGSPATIAL and IEEE MDM. For more information, please visit: www.cs.umn.edu/∼mokbel.

References

[1] L. Anselin et al. GeoDa: An Introduction to Spatial Data Analysis. *Journal of Geographical Analysis*, 38(1):5–22, 2006.

[2] Azure Geo AI. https://azure.microsoft.com/en-us/blog/microsoft-and-esri-launch-geospatial-ai-on-azure/.

[3] F. Bastani, S. He, S. Abbar, M. Alizadeh, H. Balakrishnan, S. Chawla, S. Madden, and D. DeWitt. RoadTracer: Automatic Extraction of Road Networks from Aerial Images. In *CVPR*, pages 4720–4728, 2018.

[4] S. Bhadange, A. Arora, and A. Bhattacharya. GARUDA: A System for Large-scale Mining of Statistically Significant Connected Subgraphs. *PVLDB*, 9(13):1449–1452, 2016.

[5] C.-R. Chen and U. T. Kartini. K-Nearest Neighbor Neural Network Models for Very Short-Term Global Solar Irradiance Forecasting Based on Meteorological Data. *Journal of Energies*, 10(2):186–203, 2017.

[6] T. Condie, P. Mineiro, N. Polyzoti, and M. Weimer. Machine Learning for Big Data (Tutorial). In *SIGMOD*, pages 939–942, 2013.

[7] E. Diday. Spatial Classification. *Journal of Discrete Applied Mathematics*, 156(8):1271–1294, 2008.

[8] X. Ding, L. Chen, Y. Gao, C. S. Jensen, and H. Bao. UlTraMan: A Unified Platform for Big Trajectory Data Management and Analytics. *PVLDB*, 11(7):787–799, 2018.

[9] O. Dovrat, I. Lang, and S. Avidan. Learning to Sample. In *CVPR*, 2019.

[10] A. Eldawy and M. F. Mokbel. The Era of Big Spatial Data (Tutorial). *PVLDB*, 10(12):1992–1995, 2017.

[11] ESRI ArcGIS. https://www.esri.com/en-us/arcgis/about-arcgis/overview.

[12] ESRI Tools for Hive. github.com/Esri/spatial-framework-for-hadoop.

[13] M. Ester, H. Kriegel, J. Sander, and X. Xu. A Density-based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In *SIGKDD*, pages 226–231, 1996.

[14] J. H. Faghmous and V. Kumar. *Spatio-temporal Data Mining for Climate Data: Advances, Challenges, and Opportunities*, pages 83–116. Springer, 2014.

[15] R. Finkel and J. Bentley. Quad Trees a Data Structure for Retrieval on Composite Keys. *Acta Informatica*, 1974.

[16] R. Frank, M. Ester, and A. Knobbe. A Multi-relational Approach to Spatial Classification. In *SIGKDD*, pages 309–318, 2009.

[17] R. Guhaniyogi and S. Banerjee. Meta-Kriging: Scalable Bayesian Modeling and Inference for Massive Spatial Datasets. *Journal of Technometrics*, 60(4):430–444, 2018.

[18] C. Guo, B. Yang, J. Hu, and C. Jensen. Learning to Route with Sparse Trajectory Sets. In *ICDE*, pages 1073–1084, 2018.

[19] T. Guo, K. Feng, G. Cong, and Z. Bao. Efficient Selection of Geospatial Data on Maps for Interactive and Visualized Exploration. In *SIGMOD*, pages 567–582, 2018.

[20] J. M. Hellerstein, C. R'e, F. Schoppmann, D. Z. Wang, E. Fratkin, A. Gorajek, K. S. Ng, C. Welton, X. Feng, K. Li, and A. Kumar. The MADlib Analytics Library: or MAD Skills, the SQL. *PVLDB*, 5(12):1700–1711, 2012.

[21] J. Hu, C. Guo, B. Yang, and C. S. Jensen. Stochastic Weight Completion for Road Networks Using Graph Convolutional Networks. In *ICDE*, pages 1274–1285, 2019.

[22] J. Hughes et al. GeoMesa: A Distributed Architecture for Spatio-temporal Fusion. In *SPIE Defense+Security*, 2015.

[23] P. K. Jayaraman et al. Quadtree Convolutional Neural Networks. In *ECCV*, pages 546–561, 2018.

[24] Z. Jiang, Y. Li, S. Shekhar, L. Rampi, and J. Knight. Spatial Ensemble Learning for Heterogeneous Geographic Data with Class Ambiguity: A Summary of Results. In *SIGSPATIAL*, pages 23:1–23:10, 2017.

[25] Z. Jiang and S. Shekhar. *Spatial Big Data Science: Classification Techniques for Earth Observation Imagery*. Springer Publishing Company, 1st edition, 2017.

[26] L. J. Klein et al. PAIRS: A Scalable Geo-spatial Data Analytics Platform. In *IEEE Big Data*, pages 1290–1298, 2015.

[27] M. Koubarakis, M. Datcu, C. Kontoes, U. Giammatteo, S. Manegold, and E. Klien. TELEIOS: A Database-powered Virtual Earth Observatory. *PVLDB*, 5(12):2010–2013, 2012.

[28] A. Kumar, M. Boehm, and J. Yang. Data Management in Machine Learning: Challenges, Techniques, and Systems (Tutorial). In *SIGMOD*, pages 1717–1722, 2017.

[29] N. Levine. *CrimeStat: A Spatial Statistical Program for the Analysis of Crime Incidents*, pages 381–388. Springer, 2017.

[30] Y. Li, R. Yu, C. Shahabi, and Y. Liu. Diffusion Convolutional Recurrent Neural Network: Data-Driven Traffic Forecasting. In *ICLR*, 2018.

[31] Y. Lin et al. Exploiting Spatiotemporal Patterns for Accurate Air Quality Forecasting Using Deep Learning. In *SIGSPATIAL*, pages 359–368, 2018.

[32] Z. Lu, D. Agarwal, and I. S. Dhillon. A Spatio-temporal Approach to Collaborative Filtering. In *RecSys*, pages 13–20, 2009.

[33] Z. Lv, J. Xu, K. Zheng, H. Yin, P. Zhao, and X. Zhou. LC-RNN: A Deep Learning Model for Traffic Speed Prediction. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 3470–3476, 2018.

[34] Magellan: Geospatial analytics using spark. https://github.com/harsha2010/magellan.

[35] X. Meng et al. MLlib: Machine Learning in Apache Spark. *Journal of Machine Learning Research*, 17(1), 2016.

[36] R. T. Ng and J. Han. Efficient and Effective Clustering Methods for Spatial Data Mining. In *VLDB*, pages 144–155, 1994.

[37] T. Plötz and S. Roth. Neural Nearest Neighbors Networks. In *NIPS*, pages 1087–1098, 2018.

[38] PostGIS. http://postgis.net/.

[39] PostgreSQL. https://www.postgresql.org/, 2019.

[40] T. Rekatsinas, X. Chu, I. F. Ilyas, and C. Ré. HoloClean: Holistic Data Repairs with Probabilistic Inference. *PVLDB*, 10(11):1190–1201, 2017.

[41] S. Rey et al. *PySAL: A Python Library of Spatial Analytical Methods*, pages 175–193. Springer, 2010.

[42] I. Sabek. Adopting Markov Logic Networks for Big Spatial Data and Applications. In *VLDB PhD Workshop*, 2019.

[43] I. Sabek and M. F. Mokbel. Sya: Enabling Spatial Awareness inside Probabilistic Knowledge Base Construction. In *ICDE*, 2020.

[44] I. Sabek, M. Musleh, and M. Mokbel. TurboReg: A Framework for Scaling Up Spatial Logistic Regression Models. In *SIGSPATIAL*, pages 129–138, 2018.

[45] I. Sabek, M. Musleh, and M. F. Mokbel. A Demonstration of Sya: A Spatial Probabilistic Knowledge Base Construction System. In *SIGMOD*, pages 1689–1692, 2018.

[46] I. Sabek, M. Musleh, and M. F. Mokbel. Flash in Action: Scalable Spatial Data Analysis Using Markov Logic Networks. *PVLDB*, 12(12):1834–1837, 2019.

[47] Y.-L. K. Samo and S. Roberts. Scalable Nonparametric Bayesian Inference on Point Processes with Gaussian Processes. In *ICML*, pages 2227–2236, 2015.

[48] A. D. Sarma, H. Lee, H. Gonzalez, J. Madhavan, and A. Halevy. Efficient Spatial Sampling of Large Geographical Tables. In *SIGMOD*, pages 193–204, 2012.

[49] M. Sethi, Y. Yan, A. Rangarajan, R. R. Vatsavai, and S. Ranka. Scalable Machine Learning Approaches for Neighborhood Classification Using Very High Resolution Remote Sensing Imagery. In *SIGKDD*, pages 2069–2078, 2015.

[50] S. Shekhar, C.-T. Lu, and P. Zhang. Detecting Graph-based Spatial Outliers: Algorithms and Applications (a Summary of Results). In *SIGKDD*, pages 371–376, 2001.

[51] J. Shin, S. Wu, F. Wang, C. D. Sa, C. Zhang, and C. Ré. Incremental Knowledge Base Construction Using DeepDive. *PVLDB*, 8(11):1310–1321, 2015.

[52] A. Shokry, M. Torki, and M. Youssef. DeepLoc: A Ubiquitous Accurate and Low-overhead Outdoor Cellular Localization System. In *SIGSPATIAL*, pages 339–348, 2018.

[53] R. Stanojevic, S. Abbar, and M. Mokbel. W-edge: Weighing the Edges of the Road Network. In *SIGSPATIAL*, pages 424–427, 2018.

[54] C. R. Stephens, V. Sánchez-Cordero, and C. G. Salazar. Bayesian Inference of Ecological Interactions from Spatial Data. *Journal of Entropy*, 19(12), 2017.

[55] R. Tibshirani and P. Wang. Spatial Smoothing and Hot Spot Detection for CGH Data Using the Fused Lasso. *Biostatistics*, 9(1):18–29, Jan. 2008.

[56] T. VoPham et al. Emerging Trends in Geospatial Artificial Intelligence (geoAI): Potential Applications for Environmental Epidemiology. *Environmental Health*, 2018.

[57] L. Wang, R. Christensen, F. Li, and K. Yi. Spatial Online Sampling and Aggregation. *PVLDB*, 9(3):84–95, 2015.

[58] W. Wang, J. Yang, and R. R. Muntz. STING: A Statistical Information Grid Approach to Spatial Data Mining. In *VLDB*, pages 186–195, 1997.

[59] H. Wei, G. Zheng, H. Yao, and Z. Li. Intellilight: A reinforcement learning approach for intelligent traffic light control. In *SIGKDD*, pages 2496–2505, 2018.

[60] D. Xie, F. Li, B. Yao, G. Li, L. Zhou, and M. Guo. Simba: Efficient In-Memory Spatial Analytics. In *SIGMOD*, pages 1071–1085, 2016.

[61] Y. Xie et al. An Unsupervised Augmentation Framework for Deep Learning Based Geospatial Object Detection: A Summary of Results. In *SIGSPATIAL*, pages 349–358, 2018.

[62] H. Yin, W. Wang, H. Wang, L. Chen, and X. Zhou. Spatial-Aware Hierarchical Collaborative Deep Learning for POI Recommendation. *TKDE*, 29(11):2537–2551, 2017.

[63] J. Yu, Z. Zhang, and M. Sarwat. Spatial Data Management in Apache Spark: The GeoSpark Perspective and Beyond. *Journal of GeoInformatica*, pages 1–44, 2018.

[64] J. Zhang, Y. Zheng, and D. Qi. Deep Spatio-temporal Residual Networks for Citywide Crowd Flows Prediction. In *AAAI*, pages 1655–1661, 2017.

[65] T. Zhang, R. Ramakrishnan, and M. Livny. BIRCH: An Efficient Data Clustering Method for Very Large Databases. In *SIGMOD*, pages 103–114, 1996.

[66] G. Zheng, S. L. Brantley, T. Lauvaux, and Z. Li. Contextual Spatial Outlier Detection with Metric Learning. In *SIGKDD*, pages 2161–2170, 2017.