Chair of Transportation Systems Engineering
Department of Mobility Systems Engineering
Technical University of Munich

# Improving spatio-temporal traffic prediction through transfer learning

## Thesis Defense

**João Rodrigo Olenscki**

Chair of Transportation Systems Engineering
Department of Mobility Systems Engineering
Technical University of Munich

February 22nd, 2024

# Analysis of impact on source domain diversity

A thesis presented in part fulfilment of the requirements of the Degree of Master of Science in Transportation Systems at the Department of Civil, Geo and Environmental Engineering, Technical University of Munich.


**Advisors**:

- Cheng Lyu, M.Sc.
- Prof. Constantinos Antoniou, Ph.D.
- Prof. Larissa Driemeier, Ph.D.[1]

[1] Polytechnical School of the University of São Paulo (POLI-USP), Brazil

# Outline


1 Introduction

2 Methodology

3 Experiments

4 Conclusion

# Motivation

- Recent advancements in neural networks have significantly enhanced traffic prediction, enabling for the development of *smart cities*;
- The evolution on data capturing and processing arranged by Big Data have also prepared cities to make predictions on crowd flow [1, 2], traffic flow [3, 4], public transit flow [5, 6], travel demands [7], and traffic speeds [8];
- These models, however, come with a substantial data appetite, and this poses a challenge for initiating new intelligent networks because meaningful predictions can't be made without access to substantial data history [9].

# Motivation

- This problem, also known as the "cold-start" problem [10], has a higher burden on small and medium cities [9];
- Pan et al. [10] proposed the use of transfer learning techniques to solve this issue;
- This enables training models for cities with limited data by taking advantage of patterns observed in cities with abundant data;
- In our problem, we would apply a Domain Adaptation transfer learning, as we lack target labeled data but have plenty of source labeled data.

# Research gap

- Many studies have delved into the application of transfer learning techniques on traffic prediction problems;
- Wang et al., Jin et al. [11, 12] proposed the use of a parallel architecture, where the source and target branches are connected by a MMD loss on the features extracted;
- For a different field (internet traffic prediction), [13] used a similar architecture with a CTD loss to constraint the features to a common latent space;
- Ouyang et al. [14] introduced an adversarial network leveraging attention to obtain predictions;
- Tang et al. [15] came up with a combination of the previous works, proposing an architecture with parallel branches and an adversarial layer composed of a domain discriminator and a Gradient Reversal Layer (GRL).

# Research gap

- But few studies have specialized on determining how the diversity of the source domain affects the prediction accuracy;
- Some of them [16, 15] allow for the possibility of having more than one city as source;
- Therefore, we identify a research gap on the impact of the source domain diversity on the model's accuracy.

# Research questions

Q.1 Is it possible to encompass more than two cities as sources in a transfer learning process?

Q.2 What's the impact of the number of source cities on the model's accuracy?

Q.3 Is there a limit on the number of source cities?

# Outline

1 Introduction

2 Methodology

3 Experiments

4 Conclusion

# Building the city graph

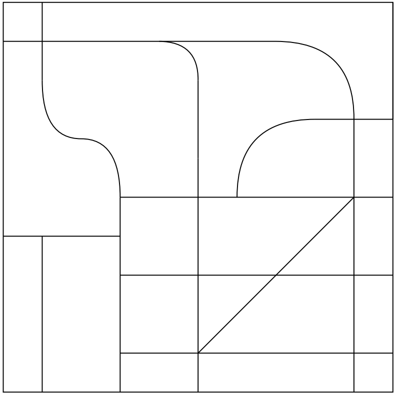

**Figure 1** Example of a road structure from a city.

# Building the city graph

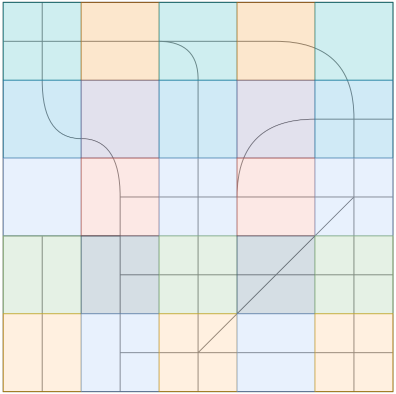

**Figure 2** Example of a road structure from a city divided into a grid.

# Building the city graph

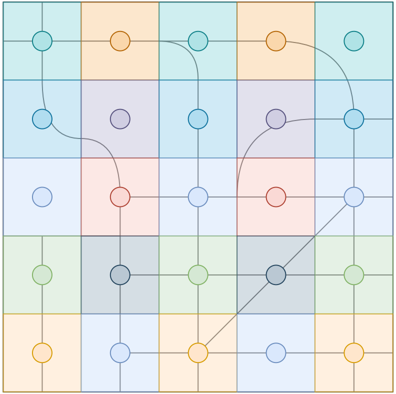

**Figure 3** Example of a road structure from a city divided into a grid, with each region being a node.

# Building the city graph

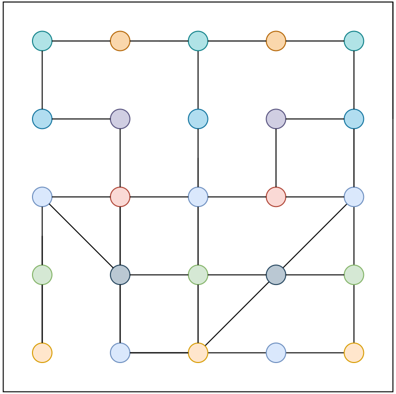

**Figure 4** Example of a road structure from a city represented as a graph.

# Examining the data tensor

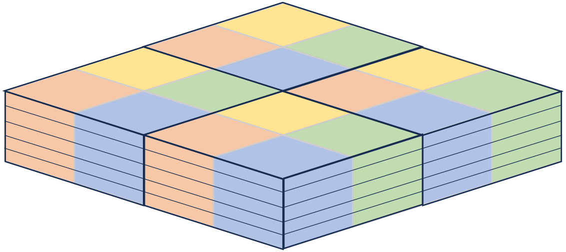

**Figure 5** Visualization of the data (as a tensor).

# Problem Definition

**Problem 1.** *Given a data-scarce target city, $C_T$, and a set of $n$ data-rich source cities $\{C_{S1}, ..., C_{Sn}\}$, the problem proposed is to predict the value of the target city's data at $t_T + 1$ with the historical data of the target city itself to that point and of the source cities:*

$$\min_{\theta} \mathcal{L}(\tilde{\mathcal{X}}_{T,t_T+1}, \mathcal{X}_{T,t_T+1}) \tag{1}$$

*where*

$$\tilde{\mathcal{X}}_{T,t_T+1} = \theta(\mathcal{X}_{T,1:t_T}, \{\mathcal{X}_{S1}, ..., \mathcal{X}_{Sn}\}) \tag{2}$$

Note that $\mathcal{L}$ is the error criterion, which may be tuned depending on our dataset and model specifications. Note also that $t_{Sk} \gg t_T \forall k = 1, ..., n$, indicating the target city's scarcity and the sources' richness.

# Proposed Architecture

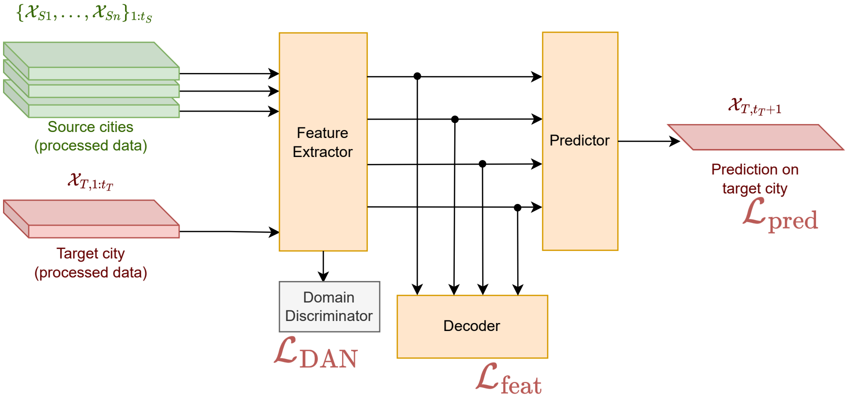


**Figure 6** Simplified version of the proposed model.

# Feature Extractor

- We propose a Feature Extractor based on a GConvLSTM cell, and trained as part of an STGAE (Spatio-Temporal Graph Autoencoder);

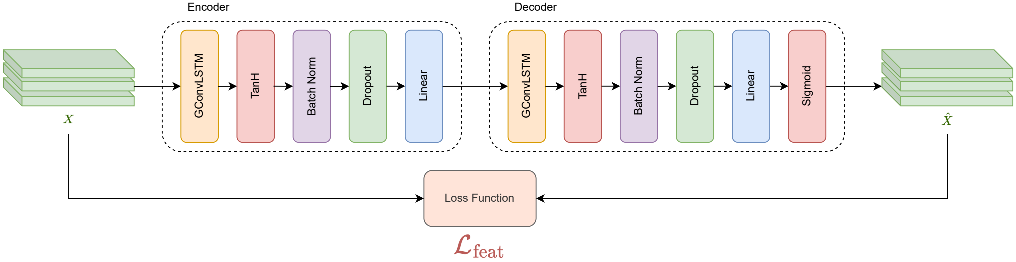


**Figure 7** Diagram representing the autoencoder as a combination of an encoder and a decoder.

# Predictor

- The Predictor is composed of a A3T-GCN cell, an activation function, a Linear layer, and a Sigmoid normalization.

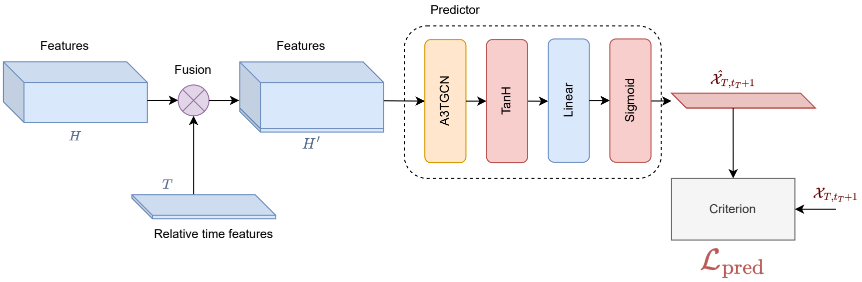


**Figure 8** Diagram representing the architecture for the Predictor network.

# Predictor

- These relative time features are derived from a sine-cosine transformation to exploit the inherent cycles in traffic flow:

$$hour\_sin = \sin\left(2\pi \cdot \frac{curr\_hour}{hours\_in\_day}\right), \quad hour\_cos = \cos\left(2\pi \cdot \frac{curr\_hour}{hours\_in\_day}\right) \tag{3}$$

And for the day-of-the-week, we calculate:

$$day\_sin = \sin\left(2\pi \cdot \frac{curr\_weekday}{days\_in\_week}\right), \quad day\_cos = \cos\left(2\pi \cdot \frac{curr\_weekday}{days\_in\_week}\right) \tag{4}$$

# Domain Adaptation

- Two techniques: Adversarial Training and Parameter Sharing;
- Adversarial Training: composed of a GRL and a Domain Discriminator, tries to force the features to a common latent space;

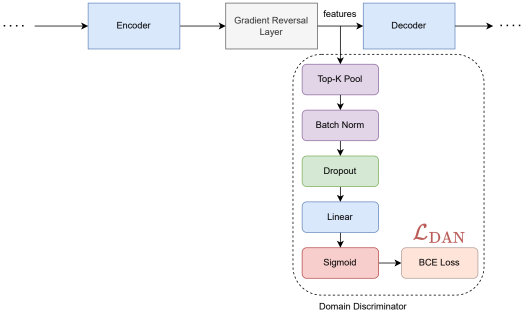


**Figure 9** Diagram representing the architecture for the Domain Discriminator module.

# Parameter Sharing

- Two-Phase Training: Pre-train (with source domain), and Fine-tune (with target domain);
- Losses (from reconstruction and adversarial training) are balanced with the $\lambda$ regularization parameter.

$$\underbrace{\mathcal{L}_{\text{AE}}}_{\text{Total Loss}} = \underbrace{\mathcal{L}_{\text{feat}}}_{\text{Reconstruction Loss}} + \lambda \cdot \underbrace{\mathcal{L}_{\text{DAN}}}_{\text{Adversarial Loss}} \tag{5}$$

# Parameter Sharing

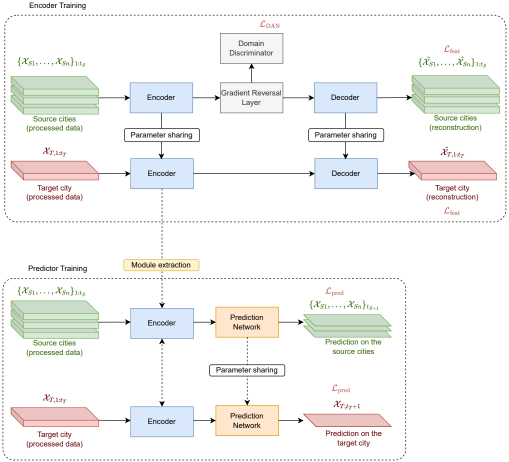


**Figure 10** Diagram representing the training script and domain adaptation process.

# Outline


1 Introduction

2 Methodology

3 Experiments

4 Conclusion

# Data Analysis

- We extracted our dataset from the NeurIPS2021 Traffic4cast competition [17];
- This dataset is composed of 8 cities, each with 360 days of data, each day composed of 240 snapshots;
- Each snapshot is represented as a tensor of size $(495, 436, 8)$, where $495 \times 436$ represents the city grid, and $8$ stands for the channels of each region, $4$ channels for the average speed in each diagonal direction, and $4$ for the volume of cars in each diagonal direction;
- The values of the tensors are in the `uint8` range, meaning values are contained in the $[0, 255]$ interval;

# Dataloaders

It's also crucial to clearly define the dataloaders that will be used in each training phase.

- Pre-training: we propose the use of both source and target cities. We need to have diversity on this phase due to the Domain Discriminator training. In order to make the alignment of the dataloaders possible, the target data was repeated until it had the same size as the source dataset;
- Fine-Tuning: during this phase, only the target data is to be used (meaning we stop to train the autoencoder adversarially).

# Highlights

- Criterion function: during the autoencoder training, we select a function to calculate the “loss” of the model. While the most common choice is MSE, we can also try WMSE, MSLE, WMSLE, LogCosh, or Focal Loss;
- There’s a trade-off between performance and consistency for some of the criterion choices;
- Number of Epochs: we tested the impact of the number of training epochs on the model’s performance;
- As our dataset is large, two epochs seems to already be a plateau on learning.

# Pre-training as Domain Adaptation (Autoencoder)

- Suitability of pre-training as a domain adaptation technique for the autoencoder.

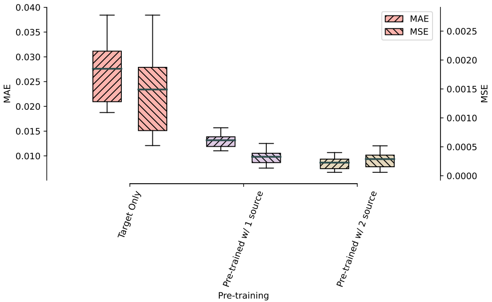


**Figure 11** Results for the pre-training as a domain adaptation technique experiment.

# Domain Discriminator

- For this experiment, we varied the value of $\lambda$, the tunable regularization parameter that balances the weight of the reconstruction and adversarial losses on the autoencoder training.

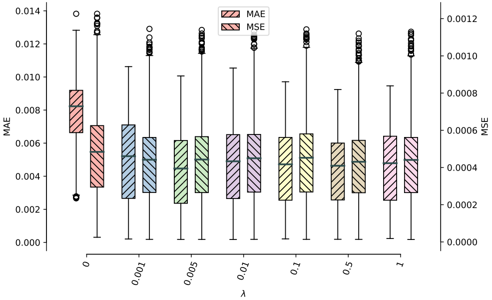


**Figure 12** Results for the domain discriminator experiment.

# Full City Experiments

- Using optimal parameters derived from the previous experiments;
- We reduce the batch size significantly (as the full city demands a huge memory load for each datapoints), and had to implement several changes to adapt the model (replace the Batch Normalization layers for Layer Normalization on all parts of the model [18], implement Gradient Accumulation [19], etc);
- We test the model for 3 values of number of source cities (1, 2, and 3), and 2 values of size of target dataset (1, and 2 weeks).

# Full City Experiments - Results

| | | 1 week | | | | | | 2 weeks | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | **RMSE** | | | **MAE** | | | **RMSE** | | | **MAE** | | |
| Statistical | ARIMA | 0.0821 | | | 0.0523 | | | 0.0785 | | | 0.0499 | | |
| Statistical | HA | 0.1227 | | | 0.0947 | | | 0.1150 | | | 0.0888 | | |
| Data-Driven | ConvLSTM | 0.1373 | | | 0.0853 | | | 0.1227 | | | 0.0847 | | |
| | ours (FT only) | 0.1695 | | | 0.0404 | | | 0.1504 | | | 0.0356 | | |
| Transfer | # Sources | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 |
| Learning | ours | 0.0171 | 0.0171 | 0.0173 | 0.0018 | 0.0017 | 0.0019 | 0.0171 | 0.0166 | 0.0172 | 0.0017 | 0.0020 | 0.0036 |

**Table 1** Results of the model and proposed baselines

# Results

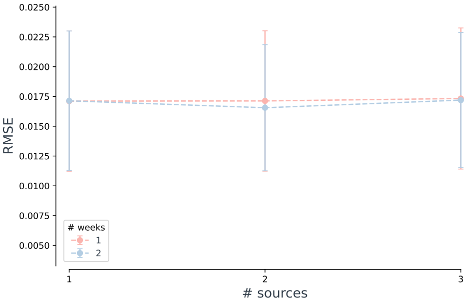


**Figure 13** Prediction errors on full city experiment.

# Autoencoder Reconstruction

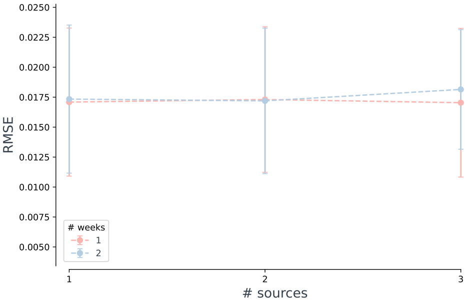


**Figure 14** Reconstruction errors on full city experiment (Autoencoder).

# Outline

1 Introduction

2 Methodology

3 Experiments

4 Conclusion

# Summary

- The growth of *smart cities* increase the demand for predictive traffic models;
- The “cold-start” problem is a key challenge when creating a *smart city*, and it affects specially small and medium sized cities;
- To handle this issue, novel “transfer learning” frameworks were proposed;
- On this work, we developed a transfer learning model using current state-of-the-art components;
- We analyzed the impact that the diversity of the source dataset has on the model’s performance, but we couldn’t show meaningful results on this research question;

# Contributions

- Separating the feature extraction and prediction tasks proved to be an efficient way to systematically tune the model;
- As these tasks are independent, training them separately results in a computational and time gain;
- Using the STGAE to model the feature extractor proved to work, as we could reach reconstruction losses (RMSE) on the order of $\times 10^{-2}$;
- We successfully joined two different domain adaptation techniques (Domain Discriminator with GRL, and Parameter Sharing), and produced experiments to prove their effectiveness.

# Contributions

- The yielded model is capable of taking cities with different sizes (and therefore different number of nodes) as sources and target;
- The final model yields results far ahead from the classical baselines, with errors (RMSE) on the order of $2 \times 10^{-2}$;
- We couldn't observe significant changes on performance with the variation of the main proposed variables (target dataset size, and number of source cities), but other significant results were obtained and can lead to new research paths.

# Limitations

- Computational Resources: as a city is big (by number of nodes), and a data point is dense (by number of snapshots), it requires a great amount of volatile memory to train. Furthermore, hyperparameter search, for a model this complex, is a real challenge with limited computational resources;
- Data Availability: the dataset, despite being very organized and containing a great amount of cities, lacks availability, as it is fairly sparse;

# Future Works

- As we noted, the model can handle cities of different sizes, but future studies can analyze how the size discrepancy affects the model's performance;
- Intra-city learning seems to be possible using the current architecture, even for different data types (meaning learning car traffic patterns from bike sharing, for instance);
- Another possible path to follow is the graph similarity analysis of two cities. As we continue to increase the data lake of possible source cities, it would be very interesting to be able to select source cities based on their graph structure similarity to target cities.

# Thank you!

Questions?

# References I

[1] Junbo Zhang et al. "Predicting citywide crowd flows using deep spatio-temporal residual networks". In: *Artificial Intelligence*. Vol. 259. Elsevier, 2018, pp. 147–166.

[2] Wenwei Jin et al. "Spatio-temporal recurrent convolutional networks for citywide short-term crowd flows prediction". In: *Proceedings of the 2nd International Conference on Compute and Data Analysis*. 2018, pp. 28–35.

[3] Nicholas G Polson and Vadim O Sokolov. "Deep learning for short-term traffic flow prediction". In: *Transportation Research Part C: Emerging Technologies*. Vol. 79. Elsevier, 2017, pp. 1–17.

[4] Yuankai Wu et al. "A hybrid deep learning based traffic flow prediction method and its understanding". In: *Transportation Research Part C: Emerging Technologies*. Vol. 90. Elsevier, 2018, pp. 166–180.

# References II


[5] Yang Liu, Zhiyuan Liu, and Ruo Jia. "DeepPF: A deep learning based architecture for metro passenger flow prediction". In: *Transportation Research Part C: Emerging Technologies*. Vol. 101. Elsevier, 2019, pp. 18–34.

[6] Di Chai, Leye Wang, and Qiang Yang. "Bike flow prediction with multi-graph convolutional networks". In: *Proceedings of the 26th ACM SIGSPATIAL international conference on advances in geographic information systems*. 2018, pp. 397–400.

[7] Xu Geng et al. "Spatiotemporal multi-graph convolution network for ride-hailing demand forecasting". In: *Proceedings of the AAAI conference on artificial intelligence*. Vol. 33. 2019, pp. 3656–3663.

[8] Bing Yu, Haoteng Yin, and Zhanxing Zhu. "Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting". In: *arXiv preprint arXiv:1709.04875*. 2017.

# References III

[9] Yunjie Huang et al. “Traffic Prediction With Transfer Learning: A Mutual Information-Based Approach”. In: *IEEE Transactions on Intelligent Transportation Systems*. Vol. 24. 8. Institute of Electrical and Electronics Engineers (IEEE), Aug. 2023, pp. 8236–8252.

[10] Sinno Jialin Pan and Qiang Yang. “A survey on transfer learning”. In: *IEEE Transactions on knowledge and data engineering*. Vol. 22. 10. IEEE, 2009, pp. 1345–1359.

[11] Senzhang Wang et al. “Spatio-Temporal Knowledge Transfer for Urban Crowd Flow Prediction via Deep Attentive Adaptation Networks”. In: *IEEE Transactions on Intelligent Transportation Systems*. Vol. 23. 5. 2022, pp. 4695–4705.

[12] Yilun Jin, Kai Chen, and Qiang Yang. “Selective Cross-City Transfer Learning for Traffic Prediction via Source City Region Re-Weighting”. In: *Proceedings of the ACM*

# References IV


*SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2022, pp. 731–741.

[13] Yanlong Wang et al. “Application-Based Cooperative Content Caching via Tensor Domain Adaptation Networks”. In: *2022 IEEE 14th International Conference on Wireless Communications and Signal Processing, WCSP 2022*. 2022, pp. 22–27.

[14] Xiaocao Ouyang et al. “CityTrans: Domain-Adversarial Training with Knowledge Transfer for Spatio-Temporal Prediction across Cities”. In: *IEEE Transactions on Knowledge and Data Engineering*. 2023, pp. 1–14.

[15] Yihong Tang et al. “Domain Adversarial Spatial-Temporal Network: A Transferable Framework for Short-term Traffic Forecasting across Cities”. In: *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*. CIKM ’22. Atlanta, GA, USA: Association for Computing Machinery, 2022, pp. 1905–1915.

# References V


[16] Bin Lu et al. "Spatio-Temporal Graph Few-Shot Learning with Cross-City Knowledge Transfer". In: *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. ACM, Aug. 2022.

[17] Christian Eichenberger et al. "Traffic4cast at NeurIPS 2021 - Temporal and Spatial Few-Shot Transfer Learning in Gridded Geo-Spatial Processes". In: *Proceedings of the NeurIPS 2021 Competitions and Demonstrations Track*. Ed. by Douwe Kiela, Marco Ciccone, and Barbara Caputo. Vol. 176. Proceedings of Machine Learning Research. PMLR, June 2022, pp. 97–112.

[18] Jimmy Lei Ba, Jamie Ryan Kiros, and Geoffrey E. Hinton. *Layer Normalization*. 2016.

[19] Yanping Huang et al. *GPipe: Efficient Training of Giant Neural Networks using Pipeline Parallelism*. 2019.

# References VI


[20] A. Palazzi and J. Liu. *ConvLSTM for Pytorch*. `https://github.com/Jimexist/conv_lstm_pytorch`. 2020.

[21] Xingjian Shi et al. "Convolutional LSTM network: A machine learning approach for precipitation nowcasting". In: *Advances in neural information processing systems*. Vol. 28. 2015.

[22] Paulius Micikevicius et al. *Mixed Precision Training*. 2018.

[23] B. Zhou, D. He, and Z. Sun. "Traffic predictability based on ARIMA/GARCH model". In: *2006 2nd Conference on Next Generation Internet Design and Engineering, 2006. NGI '06.* 2006, 8 pp.–207.

[24] Ilya Loshchilov and Frank Hutter. "Fixing Weight Decay Regularization in Adam". In: *CoRR*. Vol. abs/1711.05101. 2017.

# References VII


[25] Hongyang Gao and Shuiwang Ji. "Graph U-Nets". In: *CoRR*. Vol. abs/1905.05178. 2019.

[26] Jiawei Zhu et al. "A3T-GCN: Attention Temporal Graph Convolutional Network for Traffic Forecasting". In: *CoRR*. Vol. abs/2006.11583. 2020.

[27] Michaël Defferrard, Xavier Bresson, and Pierre Vandergheynst. "Convolutional Neural Networks on Graphs with Fast Localized Spectral Filtering". In: *Advances in Neural Information Processing Systems*. 2016.

[28] Youngjoo Seo et al. "Structured Sequence Modeling with Graph Convolutional Recurrent Networks". In: *Lecture Notes in Computer Science*. Springer International Publishing, 2018, pp. 362–373.

[29] Mohammad Sabbaqi et al. "Graph-Time Convolutional Autoencoders". In: *Learning on Graphs Conference*. PMLR. 2022, pp. 24–1.

# References VIII


[30] Yaroslav Ganin and Victor Lempitsky. “Unsupervised Domain Adaptation by Backpropagation”. In: 2015.

[31] Yangxin Fan et al. “Spatio-Temporal Denoising Graph Autoencoders with Data Augmentation for Photovoltaic Timeseries Data Imputation”. In: 2023.

[32] G. E. Hinton and R. R. Salakhutdinov. “Reducing the Dimensionality of Data with Neural Networks”. In: *Science*. Vol. 313. 5786. American Association for the Advancement of Science (AAAS), July 2006, pp. 504–507.

[33] Adam Paszke et al. “PyTorch: An Imperative Style, High-Performance Deep Learning Library”. In: *Advances in Neural Information Processing Systems 32*. Ed. by H. Wallach et al. Curran Associates, Inc., 2019, pp. 8024–8035.

# References IX


[34] Matthias Fey and Jan E. Lenssen. “Fast Graph Representation Learning with PyTorch Geometric”. In: *ICLR Workshop on Representation Learning on Graphs and Manifolds*. 2019.

[35] Benedek Rozemberczki et al. “PyTorch Geometric Temporal: Spatiotemporal Signal Processing with Neural Machine Learning Models”. In: *Proceedings of the 30th ACM International Conference on Information and Knowledge Management*. 2021, pp. 4564–4573.

[36] Arthur Gretton et al. “A Kernel Two-Sample Test”. In: *J. Mach. Learn. Res.* Vol. 13. null. JMLR.org, Mar. 2012, pp. 723–773.

[37] Huaxiu Yao et al. “Deep multi-view spatial-temporal network for taxi demand prediction”. In: *Proceedings of the AAAI conference on artificial intelligence*. Vol. 32. 2018.

# References X


[38] Zheng Zhao et al. “LSTM network: a deep learning approach for short-term traffic forecast”. In: *IET Intelligent Transport Systems*. Vol. 11. 2. Institution of Engineering and Technology (IET), Feb. 2017, pp. 68–75.

[39] Xiaolei Ma et al. “Learning Traffic as Images: A Deep Convolutional Neural Network for Large-Scale Transportation Network Speed Prediction”. In: *Sensors*. Vol. 17. 4. MDPI AG, Apr. 2017, p. 818.

[40] Zafar Ali et al. “Deep learning in citation recommendation models survey”. In: *Expert Systems with Applications*. Vol. 162. Elsevier BV, Dec. 2020, p. 113790.

[41] ChuanTao Yin et al. “A literature survey on smart cities.”. In: *Sci. China Inf. Sci.* Vol. 58. 10. 2015, pp. 1–18.

[42] Leye Wang et al. “Cross-City Transfer Learning for Deep Spatio-Temporal Prediction”. In: *Proceedings of the Twenty-Eighth International Joint Conference on Artificial*

# References XI

*Intelligence*. International Joint Conferences on Artificial Intelligence Organization, Aug. 2019.

[43] Junbo Zhang et al. "DNN-based prediction model for spatio-temporal data". In: *Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. ACM, Oct. 2016.

[44] Huaxiu Yao et al. "Learning from Multiple Cities: A Meta-Learning Approach for Spatial-Temporal Prediction". In: *The World Wide Web Conference*. ACM, May 2019.

[45] Yunjie Huang et al. "Traffic Prediction With Transfer Learning: A Mutual Information-Based Approach". In: *IEEE Transactions on Intelligent Transportation Systems*. 2023, pp. 1–17.

[46] Manh-Phu Nguyen, Minh-Son Dao, and Thanh-Binh Nguyen. "3D-STGPCN: 3D Spatio-Temporal Graph Point-wise Convolutional Network for Traffic Forecasting". In:

# References XII

*Proceedings - 2022 IEEE International Conference on Big Data, Big Data 2022*. 2022, pp. 5998–6007.

[47] Yuan Zhang et al. “Full-scale spatio-temporal traffic flow estimation for city-wide networks: a transfer learning based approach”. In: *Transportmetrica B*. Vol. 11. 1. 2023, pp. 869–895.

[48] Jiping Xing et al. “Transfer learning for robust urban network-wide traffic volume estimation with uncertain detector deployment scheme”. In: *Electronic Research Archive*. Vol. 31. 1. 2023, pp. 207–228.

[49] Junyi Li et al. “Network-scale traffic prediction via knowledge transfer and regional MFD analysis”. In: *Transportation Research Part C: Emerging Technologies*. Vol. 141. 2022.

# References XIII


[50] Jiansheng Lin et al. “A Data-driven Base Station Sleeping Strategy Based on Traffic Prediction”. In: *IEEE Transactions on Network Science and Engineering*. 2021.

[51] Safa Boudabous et al. “Dynamic Graph Convolutional LSTM application for traffic flow estimation from error-prone measurements: Results and transferability analysis”. In: *2021 IEEE 8th International Conference on Data Science and Advanced Analytics, DSAA 2021*. 2021.

[52] Sayda Elmi and Kian-Lee Tan. “Travel Time Prediction in Missing Data Areas: Feature-based Transfer Learning Approach”. In: *Proceedings - 2020 IEEE 22nd International Conference on High Performance Computing and Communications, IEEE 18th International Conference on Smart City and IEEE 6th International Conference on Data Science and Systems, HPCC-SmartCity-DSS 2020*. 2020, pp. 1088–1095.

# References XIV


[53] Peng Xie et al. “Urban flow prediction from spatiotemporal data using machine learning: A survey”. In: *Information Fusion*. Vol. 59. 2020, pp. 1–12.

[54] Leye Wang et al. “Ridesharing car detection by transfer learning”. In: *Artificial Intelligence*. Vol. 273. Elsevier BV, Aug. 2019, pp. 1–18.

[55] Lingyu Zhang et al. “Multi-Modal Graph Interaction for Multi-Graph Convolution Network in Urban Spatiotemporal Forecasting”. In: *Sustainability*. Vol. 14. 19. MDPI AG, Sept. 2022, p. 12397.

[56] Xu Geng et al. “Spatiotemporal Multi-Graph Convolution Network for Ride-Hailing Demand Forecasting”. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 33. 01. Association for the Advancement of Artificial Intelligence (AAAI), July 2019, pp. 3656–3663.

# References XV


[57] Leye Wang et al. "Exploring the Generalizability of Spatio-Temporal Traffic Prediction: Meta-Modeling and an Analytic Framework". In: *IEEE Transactions on Knowledge and Data Engineering*. Vol. 35. 4. Institute of Electrical and Electronics Engineers (IEEE), Apr. 2023, pp. 3870–3884.

[58] Chujie Tian et al. "A transfer approach with attention reptile method and long-term generation mechanism for few-shot traffic prediction". In: *Neurocomputing*. Vol. 452. Elsevier BV, Sept. 2021, pp. 15–27.

[59] Eric L. Manibardo, Ibai Lana, and Javier Del Ser. "Transfer Learning and Online Learning for Traffic Forecasting under Different Data Availability Conditions: Alternatives and Pitfalls". In: *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, Sept. 2020.

# References XVI


[60] Leye Wang, Bin Guo, and Qiang Yang. "Smart City Development with Urban Transfer Learning". In: arXiv, 2018.

[61] Ying Wei, Yu Zheng, and Qiang Yang. "Transfer Knowledge between Cities". In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, Aug. 2016.

[62] Howard W. Bevis. "A Model for Predicting Urban Travel Patterns". In: *Journal of the American Institute of Planners*. Vol. 25. 2. Informa UK Limited, May 1959, pp. 87–89.

[63] Martin J. Beckmann, C. B. McGuire, and C. B. Winsten. *Studies in the Economics of Transportation*. Santa Monica, CA: RAND Corporation, 1955.