ISTANBUL TECHNICAL UNIVERSITY
FACULTY OF COMPUTER AND INFORMATICS

Predicting Traffic Flow with Deep Learning

Graduation Project Final Report

Ece Sancar
150170049

Department: Computer Engineering
Division: Computer Engineering

Advisor: Prof. Dr. Gözde Ünal

June 2022

Statement of Authenticity

I/we hereby declare that in this study

1. all the content influenced from external references are cited clearly and in detail,
2. and all the remaining sections, especially the theoretical studies and implemented software/hardware that constitute the fundamental essence of this study is originated by my/our individual authenticity.

İstanbul, Haziran 2022

Ece Sancar

Predicting Traffic Flow with Deep Learning

(SUMMARY)

Traffic flow and transportation play significant roles in urban life. Traffic congestion situations, accidents, increasing cost of traffic coordination, and environmental pollution caused by vehicles are becoming significant problems in big cities due to the increasing population. In order to overcome these problems and reduce the negative effects of traffic in urban life, traffic has to be managed in a smart way. Intelligent Transportation Systems (ITS) are employed in big cities to manage and coordinate traffic. Traffic forecasting is one of the most challenging functions of these systems and accurate traffic flow predictions are substantial for control and guidance. The purpose of traffic forecasting is to anticipate future traffic states in the traffic network based on a series of previous traffic states and the physical roadway network. With the advancements in computational power and the increasing volume of traffic flow data, recent works in this area are commonly conducted with machine learning methods.

In our work, we aimed to propose a deep learning model for predicting future traffic maps. We tried to model urban traffic flow and implement our model to successfully predict short-term (5-60 minutes) future traffic maps of both seen and unseen cities. We worked on a dataset consisting of dynamic traffic volume and speed data, and static city map data of ten big cities. Dynamic data is formed as traffic volume and speed data at the end of each 5 minutes interval.

Our studies started with visualization of the data and evaluating the baseline methods. Our first objective was to predict traffic flow after the Covid-19 pandemic using the data before the pandemic started. For this task, we used 6 months of traffic data from 2019 to train, and 100 one-hour test slots from 2020 to evaluate our models. For our second objective, predicting the traffic flow of unseen cities, we used training data from 2019 and 2020 but we tested the model on two new cities. For both of the tasks, models output predictions of 5, 10, 15, 30, 45, and 60 minutes after the one-hour test slots.

A literature review is conducted focusing on previous works in this area. Various deep learning approaches and model architectures proposed for the traffic flow prediction problem are investigated. We observed that different methods such as Convolutional Neural Networks (CNN), auto-encoders, and General Adversarial Network (GAN) based models are applied to this problem with the combination of various deep learning methods.

We proposed new deep learning models and compared their results with the baseline methods. We tried to improve the performance of our models by applying different approaches including several deep learning applications such as multi-task learning.

Derin Öğrenme ile Trafik Akışı Tahmini

(ÖZET)

Trafik akışı ve ulaşım, kentsel yaşamda önemli roller oynamaktadır. Artan nüfusa bağlı olarak büyük şehirlerde trafik sıkışıklığı durumları, kazalar, trafik koordinasyon maliyetlerinin artması ve araçların neden olduğu çevre kirliliği önemli sorunlar haline gelmektedir. Bu sorunların aşılabilmesi ve trafiğin kent yaşamındaki olumsuz etkilerinin azaltılabilmesi için trafiğin akıllı bir şekilde yönetilmesi gerekmektedir. Akıllı Ulaşım Sistemleri (AUS), trafiği yönetmek ve koordine etmek için büyük şehirlerde kullanılmaktadır. Trafik tahmini, bu sistemlerin en zorlu işlevlerinden biridir ve doğru trafik akışı tahminleri, trafik kontrolü ve yönlendirilmesi açısından önemlidir. Trafik tahmininin amacı, bir dizi önceki trafik durumuna ve fiziksel karayolu ağına dayalı olarak, trafik ağında gelecekteki trafik durumlarını tahmin etmektir. Hesaplama gücündeki gelişmeler ve artan trafik akışı verisi hacmi ile bu alandaki son çalışmalar yaygın olarak makine öğrenimi yöntemleriyle yürütülmektedir.

Çalışmamızda, gelecekteki trafik haritalarını tahmin etmek için bir derin öğrenme modeli önermeyi amaçladık. Hem görünen hem de görünmeyen şehirlerin yakın gelecekteki (5-60 dakika) trafik haritalarını başarılı bir şekilde tahmin etmek için kentsel trafik akışını modellemeye ve modelimizi hayata geçirmeye çalıştık. 10 büyük şehrin dinamik trafik hacmi ve hız verileri ile statik şehir haritası verilerinden oluşan bir veri seti üzerinde çalıştık. Dinamik veriler, her 5 dakikalık kesit sonundaki trafik yoğunluk ve hız verilerinden oluşmaktadır.

Çalışmalarımız, verilerin görselleştirilmesi ve temel yöntemlerin değerlendirilmesi ile başladı. İlk hedefimiz Covid-19 pandemisi başlamadan önceki verileri kullanarak pandemi sonrası trafik akışını tahmin etmekti. Bu görev için, modellerimizi değerlendirmek amacıyla 2019 yılından 6 aylık trafik verilerini ve 2020 yılından 100 adet bir saatlik test aralıklarını kullandık. İkinci hedefimiz olan görünmeyen şehirlerin trafik akışını tahmin etmek için ise modeli eğitmek için 2019 ve 2020 yıllarından 6 aylık veriler kullandık ancak modeli iki yeni şehir üzerinde test ettik. Her iki görev için de modeller, bir saatlik test aralıklarından sonraki 5, 10, 15, 30, 45 ve 60 dakikalık tahminlerini üretirler.

Bu alanda daha önce yapılmış çalışmalara odaklanarak bir literatür taraması gerçekleştirdik. Trafik akışını tahmin etme problemi için önerilen çeşitli derin öğrenme yaklaşımları ve model mimarilerini araştırdık. Evrişimsel Sinir Ağları (ESA) ve Çekişmeli Üretici Ağlar tabanlı modeller gibi farklı yöntemlerin çeşitli derin öğrenme yöntemleriyle birleştirilerek bu soruna uygulandığını gözlemledik.

Yeni derin öğrenme modelleri tasarladık ve sonuçlarını temel yöntemlerle karşılaştırdık. Çoklu görev öğrenme gibi çeşitli derin öğrenme uygulamaları da dahil olmak üzere farklı yaklaşımlar uygulayarak modellerimizin performanslarını iyileştirmeye çalıştık.

Contents

1 Introduction and Project Summary

In our work, we aim to model urban traffic density and predict future traffic maps. Our main objective is to build deep learning models robust to domain shifts in space and time. Our proposed model will predict future traffic maps of the cities using past traffic flow data and traffic maps of unseen cities using traffic flow data of other cities.

Traffic flow prediction for big cities can be helpful in many respects. As a result of population growth and increasing urbanization, traffic problems are experienced in big cities. Tracking current traffic volumes and forecasting short-term future traffic flow maps allow the sooner detection of anomalies, proper navigation of vehicles, or control of the traffic flow using traffic lights and other existing sources [1].

We got the dataset from the competition Traffic4Cast 2021 and determined the two tasks explained above according to the competition. For evaluating our results, we used the leaderboard submissions on the website of the competition.

The dataset for the explained tasks consists of dynamic and static data from ten big cities (Istanbul, Berlin, Chicago, Melbourne, Antwerp, Bangkok, Barcelona, Moscow, New York, and Vienna). Static data includes a gray-scale city map and eight-channel binary connection map for cardinal and ordinal directions of the city. Dynamic data is comprised of volume and the average speed information for four ordinal directions. Some of the cities will be used for training and testing, and the other ones are used for only testing.

Firstly, we analyzed the traffic flow data from different cities and different time intervals in order to come up with a proper solution. We visualized static and dynamic data. We used traffic data of the different cities to observe changes in traffic volume and speed at different time intervals. We monitored the shift in traffic densities in these cities after the restrictions for the Covid-19 pandemic.

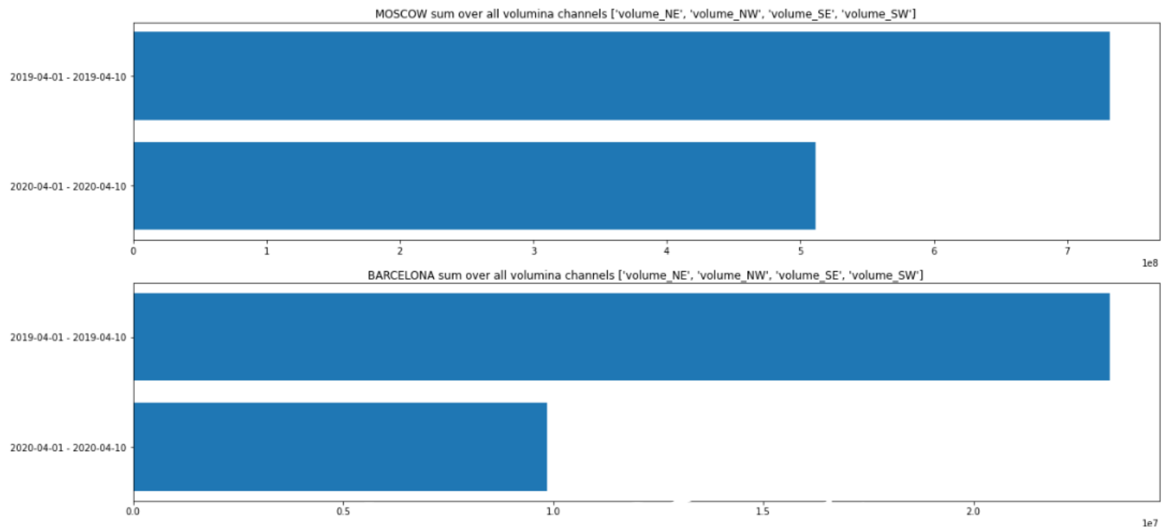

Figure 1.1: Traffic density information of Moscow and Barcelona before and after Covid-19 pandemic

We began our experiments with evaluating and comparing Graph-ResNet [1] and U-Net [2] results. Our purpose was to achieve better results than these baseline methods.

The Graph-ResNet approach begins with graph extraction. Graph extraction is achieved by extracting and masking the pixels on the street network using the static data explained above and transforming the remaining pixels into graph data. This graph data is used by the model to create node-level predictions. Finally, the transformation from graph domain to image domain is performed. The architecture consists of a simple Graph Convolutional Network (GCN) as proposed in [3] with the combination of Residual Blocks inspired by ResNet Architecture [4].

We used the results of these experiments with Graph-ResNet and U-Net methods to compare evaluations of our proposed deep learning models to predict future traffic maps of seen and unseen cities.

1.1 Engineering Standards and Multiple Constraints

Traffic flow prediction applications are used in real-life systems such as navigation and ITS. We tried to maintain the usability of our models in real-life systems while making design choices for our networks. As an example, we decided not to train different models for predicting different cities' future traffic flow since it will affect our model's usability. We used the dataset we got from the competition explained above for training our models. We did not use any data which may cause legal complications. We faced time constraints during our project. We tried to schedule our experiments according to the training time of our models and the time schedule of our project.

2 Literature Survey

Many short-term traffic forecasting deep learning methods are available, yet comparing these methods and applications is not consistent since studies apply these tasks using different spatio-temporal data resolution and data indicators. However, various methods and approaches to this task are examined in order to come up with an appropriate solution and implementation to our problem. In this section, we will discuss these examined previous works.

As stated before, many deep learning approaches have been applied to short-term traffic forecasting tasks. In order to benefit from spatial features, a number of studies use convolutional neural networks (CNN) in their proposed model. However, this approach restrains the model to only perform on data structured as images, videos, etc. The street network could be provided to the model directly by modeling the traffic network as a graph to fully make use of spatial information and to reduce problem complexity. Graph Convolutional Networks (GCN) [3] are commonly used in the studies for the traffic forecasting task to take advantage of using the traffic network as graph data. Some of the proposed studies using GCNs for traffic forecasting are discussed in the following paragraph.

Martin et al. proposed a Graph Residual Network architecture [1] which is inspired by the ResNet presented in [4] and GCN presented in [3] with combination of skip connections. This combined network architecture allows the training of deeper GCNs without the over-smoothing problem. Their experiment results with different GCN architectures show that the GCN-based methods generalize better on seen and unseen cities in the task of traffic flow prediction. Yu et al. [5] introduced a Spatio-Temporal GCN model consisting of spatio-temporal convolutional blocks, which are a combination of graph convolutional layers and convolutional sequence learning layers, to model spatial and temporal dependencies. They focused on the traffic forecasting task as a typical time-series prediction problem. Work introduced by Cui et al. [6] combines GCN architecture with Long Short-Term Memory (LSTM). Li et al. proposed to model the traffic flow as a diffusion process on a directed graph and introduced Diffusion Convolutional Recurrent Neural Network (DCRNN) [7].

Other than GNNs, several studies include different architectures such as U-Net [2] [8] [9]. Martin et al. [10] applied simple U-Net architecture to the problem of predicting the traffic flow of known cities. Wiedemann et al. [11] combined the U-Net architecture with the operations of dividing the input data into randomly cut patches and merging the predictions. Their results showed that these dividing and merging operations improved the results on unseen data.

Generative Adversarial Networks (GANs) [12] are also present in the recent studies conducted in the area of traffic forecasting [13] [14]. In one of the most successful GAN-based studies, Zhang et al. [15] proposed a traffic prediction model called TrafficGAN which focuses on network-level predictions instead of road-level or region-level predictions. Both CNN and LSTM models are combined with the GAN structure in order to capture the spatio-temporal correlations among the road links of a road network.

As stated by Guo et al. [16], traffic flow is a kind of spatio-temporal data that exhibits both spatial and temporal correlation and heterogeneity, thus 3 Dimensional convolutions can be used to capture these correlations to increase prediction performance rather than 2 Dimensional designs. There are several studies available that apply this method on traffic forecasting problem [16] [17] [18] [19]. We adapted the 3D U-Net network [20] to the traffic flow prediction task in a proper manner to apply to our dataset. 3D U-Net network is proposed by Çiçek et al. by replacing all 2D operations in previous U-Net architecture [2] with their 3D counterparts in order to perform volumetric segmentation.

Convolutional LSTM Network [21] is proposed by Shi et al. as a machine learning approach for the task of the future precipitation prediction. Since both precipitation and traffic forecasting problems can be formulated as spatiotemporal sequence forecasting problems, we adapted their network to our dataset and used in our experiments.

Numerous deep learning applications take advantage of multi-task learning which allows the model to learn multiple tasks simultaneously. There are several studies on traffic forecasting using multi-task learning by separating the prediction approach into different tasks [22] [23] [24]. In our work, we try to predict future traffic by volume and speed on the road networks. Thus, we adapted our 3D U-Net with multi-task learning on speed and volume prediction and conducted experiments.

As explained in [25] attention gates are commonly used in many deep learning tasks. By suppressing irrelevant regions and highlighting significant features, models implemented with attention gates can perform better. We used 3D Skip Attention Units proposed in [26] and our 3D U-Net model to build a 3D Attention U-Net model suitable for our task and data.

3 Method

Our two tasks were to predict future traffic flow after Covid-19 using training data from 2019 and to predict the future traffic flow in the unseen cities using training data from other cities. For comparing different models and proposing a better model for these tasks, we conducted different experiments with different networks. These models will be explained in the Experiments section. In this part, experiment setup and constant operations will be stated.

For constructing train and validation sets, we used the parameters limit and fraction given as arguments as the number of input files to be read and the fraction of the train/validation sets. We used randomly selected input files as we could not train our models with the whole available training set which is constructed by the six months of traffic data of each city. We read these randomly selected time slots as 2-hour intervals and used the first 1-hour frame for training, and the second hour data for ground truth. The test set was fixed and constructed as 100 one-hour time slots of traffic data for each city.

As learning setup, we used Adam optimizer with a learning rate of 1e-4. We used a batch size of 2. Mean Squared Error is used for evaluation. (Exceptions of these situations are stated in the Experiments section.)

For each experiment, we trained our models for 100 epochs and saved the model state when the validation loss is smaller than the current minimum. At the end of the training, the model state with the minimum validation loss is loaded and the tests are performed.

4 Dataset

Our dataset for both tasks consists of static and dynamic data. Their only difference is the time interval of the traffic data that the dynamic dataset demonstrates.

4.1 Static Data

Each city has one 9-channel static connectivity data. The first channel is a gray-scale city map with the same resolution as the dynamic data (495, 436). The remaining 8 channels are binary connectivity maps of whether a cell is connected to its neighbor cell in eight directions (N, NE, NW, S, SE, SW, E, W).

These static data are used for masking in our experiments for extracting the road network to construct the results or building the graph from road data.

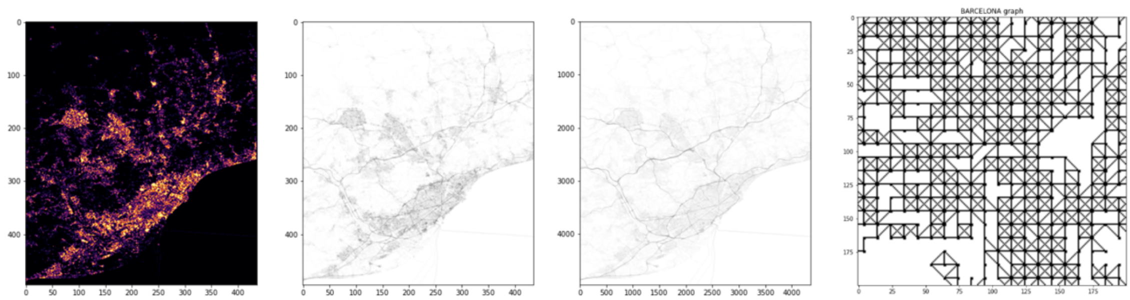

Figure 4.1: Static data of Barcelona

4.2 Dynamic Data

Dynamic data is the input and the output type of our models. It consists of 8 channels. The First 4 channels represent the volume and the second 4 channels represent the average speed on the cells of the road network in four directions (NE, NW, SE, SW).

Figure 4.2: Dynamic data visualization of Istanbul

5 Experiments

After starting our experiments with the baselines we determined in the previous term as U-Net and Graph-ResNet which were described in the Introduction and Literature Survey sections, we constructed different models and used various approaches for achieving better results on the task of traffic flow prediction. We conducted experiments on these models using the methods explained in the Method section. The results of these experiments will be demonstrated in the next part.

5.1 3D U-Net

We began our attempts with 3D U-Net implementation. We made use of the network proposed by Çiçek et al. [20]. We modified our data loaders and transformers to properly read data into 3D format and construct the output in an accurate format to evaluate with Mean Squared Error using the ground truth data.

5.2 Excluding Sparse Input

We tried to increase the performance of our 3D U-Net model with different approaches. After visualizing the input dynamic data, we realized that some of the inputs are representing the traffic flow when the roads are mostly empty such as in the early morning and late-night hours. We tried to use more balanced inputs and exclude these sparse maps while randomly selecting the input data to help the model learn better. These experiments are demonstrated in the results section with the extension of "without the sparse inputs".

5.3 Extracting Random Patches

We conducted some experiments by using random patches taken from the input data to allow the model to generalize better on unseen maps. We used 20x20 sized windows and selected random regions from the dynamic input data. During the test time, we did not do any patching, we made the model predict the whole map. These experiments are demonstrated with the extension of "with patch" in the results section.

5.4 Loss Weighting

We examined the scores of our model separately for volume and speed prediction. We realized that these scores were highly different from each other. The model was predicting better for speed values. We tried to use weighted losses for these two tasks. Before back propagation, we calculated the losses for speed and volume separately by using the relevant channels from the output of the model and weighted these loss values to calculate the overall loss of the model. These experiments are demonstrated with the extension of "with loss weighting" in the results section.

5.5 Multi-Task Learning

We tried to take advantage of Multi-Task Learning since our model was learning to predict the traffic volume and speed at the same time. We partitioned our prediction tasks into these two sub-tasks and applied multi-task learning approach to our model. We separated the channels of speed and volume in the last layer of our network for achieving this. These experiments are demonstrated with the extension of "with multi-task" in the results section.

5.6 3D Attention U-Net

We wanted to take advantage of attention gates to increase the performance of our 3D U-Net model. We used the 3D Skip Attention Units proposed in [26] and replaced the skip connections in our model. With this implementation, we achieved the highest scores in both tasks from the models we proposed. We tried to increase the score of this model by using the approaches described above as excluding sparse input, extracting random patches, loss weighting, and multi-task learning. We also tried to improve the performance using a larger batch size and distributed learning with two GPUs.

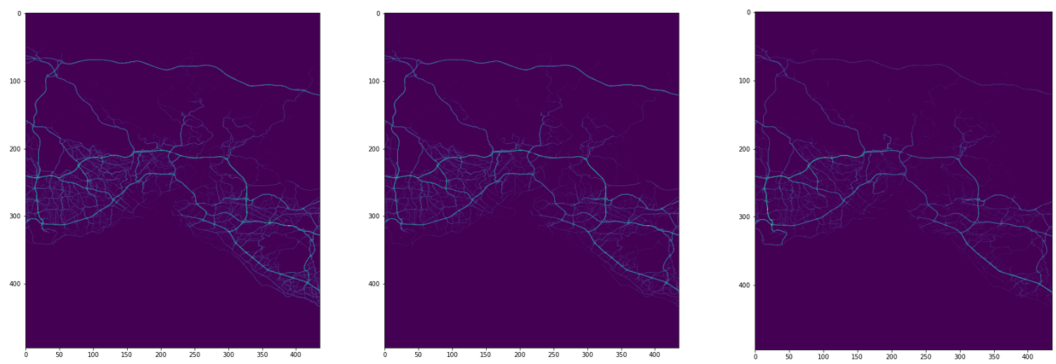

Figure 5.1: Test set examples for Istanbul

6 Results

Table 6.1: Experiment Results on Predicting Traffic Flow After Covid-19 Using Pre-Covid Data

Model	MSE Score
Best Score on the Leaderboard	48,422
Graph-ResNet	54,158
3D Attention U-Net	54,285
3D Attention U-Net with multi-task	54,597
3D Attention U-Net with multi-task and loss weighting	54,612
3D Attention U-Net with loss weighting	54,905
3D U-Net	55,415
U-Net	57,561
3D U-Net with patch	58,563
3D U-Net without the sparse inputs	59,218
3D Attention U-Net with batch size 4	61,552
3D U-Net with multi-task	67,111

Table 6.2: Experiment Results on Predicting Traffic Flow Of Unseen Cities

Model	MSE Score
Best Score on the Leaderboard	59,559
Graph-ResNet	63,36
3D Attention U-Net	63,673
3D Attention U-Net with multi-task	64,023
3D Attention U-Net with multi-task and loss weighting	64,078
3D Attention U-Net with loss weighting	64,183
3D U-Net	64,441
3D U-Net without the sparse inputs	66,324
U-Net	66,528
3D U-Net with patch	66,648
3D Attention U-Net with batch size 4	67,989
3D U-Net with multi-task	69,294

Tables 6.1 and 6.2 demonstrate the results of the experiments we constructed with the models explained in the Experiments section for the two tasks we worked on. Graph-ResNet and U-Net models were the baselines we determined. 3D Attention U-Net has the best scores on both tasks among the models we proposed during our studies.

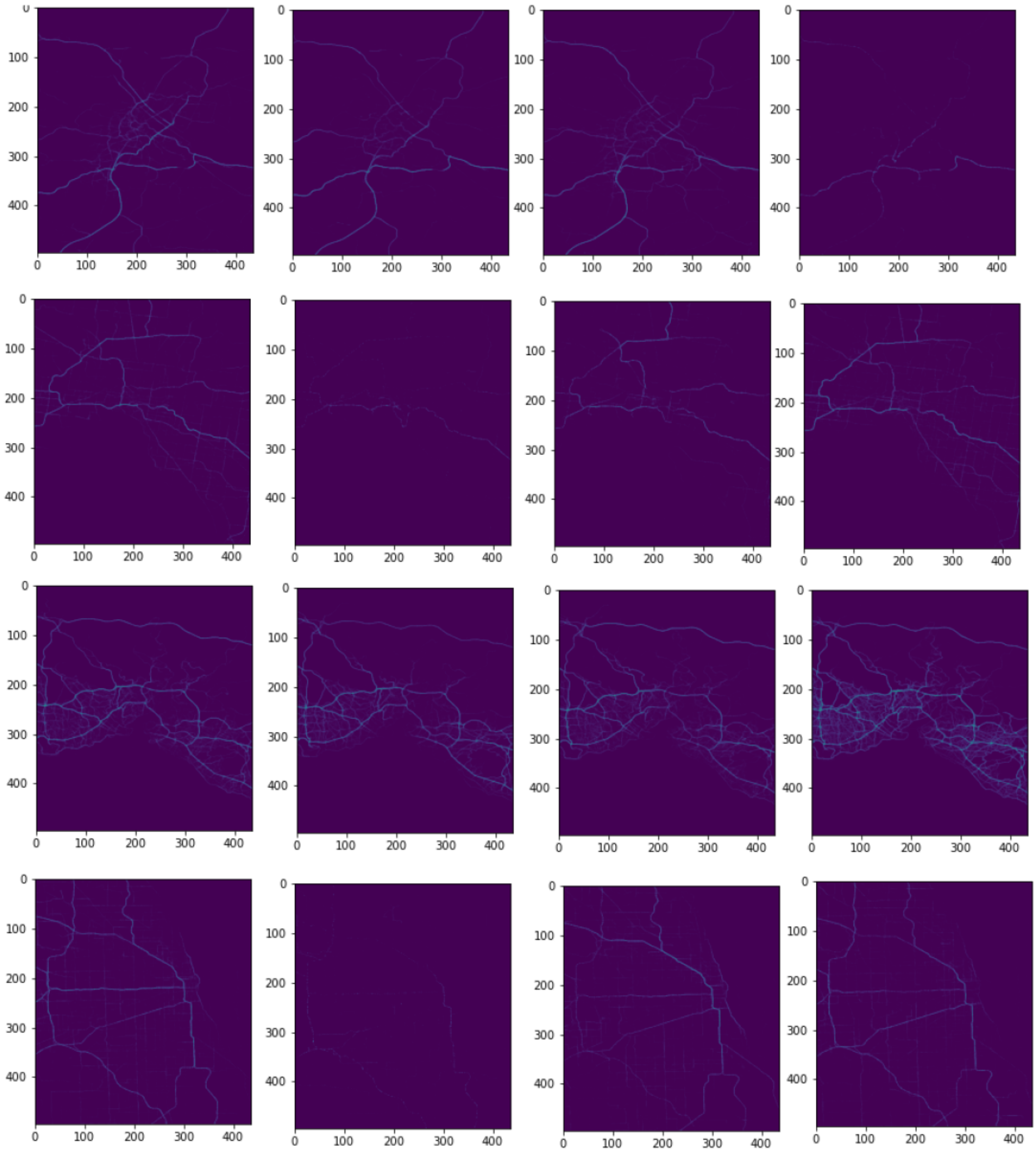

Figure 6.1: Our 3D Attention U-Net model's test results for New York, Melbourne, and Istanbul on randomly selected test slots

7 Conclusion

In conclusion, traffic forecasting is an important task for city planning, traffic control, navigation, and Intelligent Transportation Systems. There are many traffic forecasting with deep learning methods available and this topic is still open for improvement. During our project, we examined previous works conducted in this area and studied different deep learning applications. We implemented and evaluated deep learning models for the task of traffic flow prediction. We tried to improve the performance of our models. We observed that different deep learning applications can give disparate results when applied to different problems or data formats. We achieved higher results than one of the two baseline methods we determined in the previous term. Our work is still open for improvement.

References

[1] H. Martin, D. Bucher, Y. Hong, R. Buffat, C. Rupprecht, and M. Raubal, "Graph-resnets for short-term traffic forecasts in almost unknown cities," in *NeurIPS*, 2019.

[2] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *CoRR*, vol. abs/1505.04597, 2015. [Online]. Available: http://arxiv.org/abs/1505.04597

[3] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," *CoRR*, vol. abs/1609.02907, 2016. [Online]. Available: http://arxiv.org/abs/1609.02907

[4] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *CoRR*, vol. abs/1512.03385, 2015. [Online]. Available: http://arxiv.org/abs/1512.03385

[5] B. Yu, H. Yin, and Z. Zhu, "Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting," *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*, Jul 2018. [Online]. Available: http://dx.doi.org/10.24963/ijcai.2018/505

[6] Z. Cui, K. Henrickson, R. Ke, Z. Pu, and Y. Wang, "Traffic graph convolutional recurrent neural network: A deep learning framework for network-scale traffic learning and forecasting," 2019.

[7] Y. Li, R. Yu, C. Shahabi, and Y. Liu, "Diffusion convolutional recurrent neural network: Data-driven traffic forecasting," 2018.

[8] V. Konyakhin, N. Lukashina, and A. Shpilman, "Solving traffic4cast competition with u-net and temporal domain adaptation," 2021.

[9] J. Santokhi, P. Daga, J. Sarwar, A. Jordan, and E. Hewage, "Temporal autoencoder with u-net style skip-connections for frame prediction," 2020.

[10] H. Martin, Y. Hong, D. Bucher, C. Rupprecht, and R. Buffat, "Traffic map movie forecasting," 2019. [Online]. Available: https://arxiv.org/abs/1910.13824

[11] N. Wiedemann and M. Raubal, "Traffic forecasting on traffic moving snippets," 2021.

[12] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," 2014.

[13] Y. Liang, Z. Cui, Y. Tian, H. Chen, and Y. Wang, "A deep generative adversarial architecture for network-wide spatial-temporal traffic state estimation," 2018.

[14] T. Sun, B. Sun, Z. Jiang, R. Hao, and J. Xie, "Traffic flow online prediction based on a generative adversarial network with multi-source data," *Sustainability*, vol. 13, p. 12188, 11 2021.

[15] Y. Zhang, S. Wang, B. Chen, J. Cao, and Z. Huang, "Trafficgan: Network-scale deep traffic prediction with generative adversarial nets," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 1, pp. 219–230, 2021.

[16] S. Guo, Y. Lin, S. Li, Z. Chen, and H. Wan, "Deep spatial-temporal 3d convolutional neural networks for traffic data forecasting," *IEEE Transactions on Intelligent Transportation Systems*, vol. PP, pp. 1–14, 04 2019.

[17] C. Chen, K. Li, S. G. Teo, G. Chen, X. Zou, X. Yang, R. C. Vijay, J. Feng, and Z. Zeng, "Exploiting spatio-temporal correlations with multiple 3d convolutional neural networks for citywide vehicle flow prediction," *2018 IEEE International Conference on Data Mining (ICDM)*, pp. 893–898, 2018.

[18] H. Li, X. Li, L. Su, D. Jin, J. Huang, and D. Huang, "Deep spatio-temporal adaptive 3d convolutional neural networks for traffic flow prediction," *ACM Trans. Intell. Syst. Technol.*, vol. 13, no. 2, jan 2022. [Online]. Available: https://doi.org/10.1145/3510829

[19] S. Zhang, L. Zhou, X. M. Chen, L. Zhang, L. Li, and M. Li, "Network-wide traffic speed forecasting: 3d convolutional neural network with ensemble empirical mode decomposition," *Computer-Aided Civil and Infrastructure Engineering*, vol. 35, no. 10, pp. 1132–1147, 2020. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/mice.12575

[20] Özgün Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3d u-net: Learning dense volumetric segmentation from sparse annotation," 2016.

[21] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-k. Wong, and W.-c. Woo, "Convolutional lstm network: A machine learning approach for precipitation nowcasting," 2015. [Online]. Available: https://arxiv.org/abs/1506.04214

[22] K. Zhang, L. Wu, Z. Zhu, and J. Deng, "A multitask learning model for traffic flow and speed forecasting," *IEEE Access*, vol. 8, pp. 80 707–80 715, 2020.

[23] Y. Lu, "Learning to transfer for traffic forecasting via multi-task learning," 2021. [Online]. Available: https://arxiv.org/abs/2111.15542

[24] F. Jin and S. Sun, "Neural network multitask learning for traffic flow forecasting," in *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, 2008, pp. 1897–1901.

[25] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker, and D. Rueckert, "Attention u-net: Learning where to look for the pancreas," 2018. [Online]. Available: https://arxiv.org/abs/1804.03999

[26] M. Islam, V. VS, V. J. M. Jose, N. Wijethilake, U. Utkarsh, and H. Ren, "Brain tumor segmentation and survival prediction using 3d attention unet," 2021. [Online]. Available: https://arxiv.org/abs/2104.00985