



Article

Impact of Traffic Flow Rate on the Accuracy of Short-Term Prediction of Origin-Destination Matrix in Urban Transportation Networks

Renata Żochowska and Teresa Pamuła *

Faculty of Transport and Aviation Engineering, Silesian University of Technology, Krasińskiego 8 Street, 40-019 Katowice, Poland; renata.zochowska@polsl.pl

* Correspondence: teresa.pamula@polsl.pl

Abstract: Information about spatial distribution (OD flows) is a key element in traffic management systems in urban transport networks that enables efficient traffic control and decisions to redirect traffic to less congested sections of the network in emergencies. With the development of modern techniques of remote sensing, more and more advanced methods are used to measure traffic and determine OD flows. However, they may produce results with different levels of errors caused by various factors. The article examines the impact of traffic volume and its variability on the error values of short-term prediction of the OD matrix in the urban network. The OD flows were determined using a deep learning network based on data obtained from video remote sensing devices. These data were recorded at earlier intervals concerning the forecasting time. The extent to which there is a correlation between the size of OD flows and the prediction error was examined. The most frequently used measure of prediction accuracy, i.e., MAPE (mean absolute percentage error), was considered. The analysis carried out made it possible to determine the ranges of traffic flow rate for which the MAPE stabilizes at the level of approximately 6%. A set of video remote sensing devices was used to collect spatiotemporal data. They were located at the entrances and exits from the study area on important roads of a medium-sized city in Poland. The conclusions obtained may be helpful in further research on improving methods to determine OD matrices and estimate their reliability. This, in turn, involves the development of more precise methods that allow for reliable traffic forecasting and improve the efficiency of traffic management in urban areas.



Citation: Żochowska, R.; Pamuła, T. Impact of Traffic Flow Rate on the Accuracy of Short-Term Prediction of Origin-Destination Matrix in Urban Transportation Networks. *Remote Sens.* **2024**, *16*, 1202. <https://doi.org/10.3390/rs16071202>

Academic Editor: Chen Chen

Received: 27 February 2024

Revised: 19 March 2024

Accepted: 27 March 2024

Published: 29 March 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: traffic flow; OD matrix prediction; deep learning; prediction error analysis; short-term prediction; urban transportation network; video remote sensing data

1. Introduction

Effective traffic management in urban transportation networks requires an extensive database. One of the most important pieces of information is the spatial distribution of traffic presented. This knowledge is particularly useful in atypical circumstances, such as various types of road accidents, roadworks, vehicle or infrastructure failures, or other situations that require redirecting the whole or only a part of the traffic to other routes. In such cases, the spatial distribution of traffic, built and updated dynamically based on traffic data obtained using video remote sensing devices at subsequent time intervals, is an important element of the traffic management system in urban road networks and enables the effective determination of optimal routes at a given moment.

One form of presenting the temporal and spatial distribution of traffic is through the OD (origin-destination) matrix, the individual cells of which represent the number of trips between a pair of TAZ (traffic analysis zones) or other locations marked as the origin and destination of the trip in a specific unit of time. These connections are expressed as OD pairs. In the analysis of transportation networks often used in planning and modeling transport systems at the strategic level, each zone is represented by a centroid constituting

the place of accumulation of traffic flows generating and absorbing in this zone. In practical applications, these centroids are often moved using the so-called connectors to the nearest nodes of the technical network (e.g., road or railway networks). In this way, the estimation of the OD matrix between TAZ centroids can be reduced to the estimation of the OD matrix between the nodes of the technical transport network. The same procedure is also followed in traffic management, which requires operational activities and the preparation of short-term traffic predictions.

OD matrices can be constructed in many ways depending on both future application requirements and the available data. An important factor is also the specificity of a given area related to land use, the structure of the transportation network, social characteristics, and available transport subsystems. OD matrices can be built for various means of transport, time intervals, travel destinations, or groups of participants. In this way, individual demand strata are created, expressed in the form of OD matrices, specific to a given area. These matrices can be determined for historical data, the existing state, and future demand. Individual elements of the OD matrix can be presented in absolute units with the interpretation of traffic volume, or in relative units expressing the share of traffic flow for an individual OD pair in the flow on a road section.

For traffic planning in urban transport networks, travel demand models are most often used, in which OD matrices are estimated using classical methods. They require large-scale research and information on the socio-demographic situation, traffic conditions, transport behavior, and spatial development of the area. Therefore, such methods are expensive and time-consuming, and the results obtained from their use may quickly become outdated due to the rapidly changing transport system in urban areas. Moreover, transport demand models built based on classical methods, due to the high level of aggregation of traffic data, do not consider temporary disturbances in the traffic flows resulting from the time of day, seasonality, organization of mass events, road works, or weather conditions. Therefore, for traffic management, which requires frequent updating of input data depending on changes in actual traffic conditions, increasingly more precise information sources and modern measurement techniques are used to build the OD matrix.

Currently, more and more cities are equipped with traffic monitoring devices that continuously record data on traffic intensity in sections of the transport network. Most devices use remote sensing, i.e., the process of remotely obtaining information about objects or phenomena. Spatial data usually takes the form of a digital image. We are then dealing with the so-called imaging remote sensing. A special type of remote sensing is video detection, which uses the principle of processing images provided by cameras installed on the road or at an intersection by video detector modules. Thanks to this, it is possible to, among other things, detect the presence and direction of movement of vehicles, as well as the detection of cyclists, vehicle classification, traffic intensity measurement, and queue measurement. The values of traffic intensities obtained from video remote sensing devices also enable advanced analyses to detect various types of anomalies in traffic volume distributions [1,2].

Acquiring traffic data using video remote sensing devices is also increasingly used to build OD matrices. Research on the optimization of the sensor locations for the estimation of origin-destination demands is also being carried out [3,4]. Therefore, such a method of obtaining data has great potential for development and may be considered a more attractive alternative to classical methods [5,6]. In our study, we also analyzed traffic data obtained using video detectors.

Traffic prediction plays a fundamental role in intelligent transportation systems [7]. Effective operation of such systems requires the use of modern forecasting methods that ensure acceptable results. Reliable short-term forecasts of traffic volume can support the system in route planning, driving vehicles, and minimizing congestion. This directly causes a reduction in the costs incurred both by individual users of the transport system and by traffic organizers and units managing road infrastructure [8,9]. As a result, it has a positive impact on the perception of the urban area as more livable.

However, determining reliable traffic forecasts is difficult due to the complex and dynamic spatial–temporal relationships between traffic flows in different parts of the road network in urban areas [1]. In recent years, much research has been carried out in this field [10,11]. One of the most promising research directions is the use of deep learning methods, which provide more accurate results and greater resistance to missing data and errors than previously used approaches based on traditional machine learning methods, such as Kalman filters, Bayesian networks, or SVMs (support vector machines) [12]. This improves the reliability of traffic prediction.

The level of errors in traffic predictions is influenced not only by the choice of the method, which covers the data acquisition technique, its quality and level of aggregation, the observation range, and the prediction horizon, but also by many different factors of temporal, spatial, and movement nature. The predictability limit depends, among others, on the time and type of day, season, and location in the urban network—factors that also influence changes in traffic volume in road sections. The method of obtaining data is also important. In our research, we assumed that the data would be recorded by video sensing devices. Therefore, the main goal of our research was to check whether and to what extent traffic flow affects the reliability of OD matrix estimation.

This article is based on the research presented in the publication [11], which describes a method for estimating and predicting the OD matrix using a recursive network with LSTM (long short-term memory) and DLNA (deep learning network with autoencoders). This network enables capturing time series characteristics over short and long periods and is often used for traffic forecasting. Traffic predictions using LSTM networks may achieve higher reliability [13,14]. During the research, it was observed that the OD flow prediction error MAPE (mean absolute percentage error) depends on the traffic volume and its variability. Therefore, a research question arose as to whether this trend regarding short-term traffic predictions also holds when estimating the value of the OD flows. The MAPE value is a good measure for assessing error because it provides an average of the prediction errors for the test period, reflecting the degree of dispersion between the predicted values and the actual data. Moreover, this measure is expressed as a percentage, which allows the comparison of the accuracy of predictions for different models.

According to our current knowledge, there are no publications on the dependence of the OD flow prediction error on traffic intensity and its variability. Therefore, a research hypothesis was formulated in the form of a question: is there a significant relationship between the MAPE expressing the prediction error of OD matrix estimation and the size of the traffic flow rate? A positive answer may contribute to further research aimed at determining the limit value of traffic volume (expressed in absolute or relative terms), above which the obtained predictions can be treated as reliable. This issue is important from the point of view of road traffic management, in which short-term traffic predictions are useful tools.

The remainder of this article is structured as follows. Section 2 provides a general review of the literature on short-term OD matrix prediction methods, and mainly includes articles from 2015 to 2022. Section 3 describes the method for estimating the prediction errors of the OD matrix. It also includes a description of the prediction model using the LSTM network. Section 4 presents a case study of the road network of the city of Gliwice (Poland) along with the results. Section 5 analyses the results obtained. The analysis is extended in Section 6 in the form of multiple linear regression functions, which determine the dependence of the average MAPE error for a given range on traffic intensity based on two factors: average traffic intensity and the coefficient of variability of traffic intensity. Finally, Section 7 provides conclusions and plans for further research.

2. Related Work

Methods for determining the OD matrix can be classified in many ways depending on the adopted criteria. A broad overview of this area was presented in publications [15,16], among others. The choice of method mainly depends on the purpose of using the final

OD matrix, the degree of aggregation of available data, and the size of the study area. For traffic planning purposes, the OD matrix can be obtained from a macroscopic demand model for a city or larger area. However, if the study area is too small and is only in one or several TAZs, or if a macroscopic model is missing, estimating the OD matrix becomes a problem for practitioners. In such cases, information on traffic counts may be useful to estimate the OD matrix. This data can also be used for traffic management purposes due to the possibility of continuously updating what is possible by using the video detection systems installed in various parts of the examined area.

Methods for estimating the OD matrix using traffic counts differ primarily in the data acquisition process and the techniques for building the traffic assignment. The traffic data used to estimate the OD matrices can be obtained in various ways. The most widely used data acquisition methods include vehicle license plate registration [17–20], mobile phone data analysis [21–24], FCD (floating car data) [25–27], traffic intensity and speed measurements [28], vehicle GPS position recording [29], and others [30,31]. Data obtained from video sensing devices are also an important source of information on traffic flow in a section of the road [2], especially when it is necessary to estimate the OD matrix in real time.

Among the techniques for estimating OD matrices, approaches based on traffic modeling are often used, including approaches based on gravity models, gravity-opportunity-based models, information minimization and entropy maximization approaches, or linear programming techniques. Mathematical methods also constitute an important group, including gradient-based solution techniques, Kalman filtering techniques, PCA—Principal Component Analysis, the bilevel programming approach, PFE—Path Flow Estimator, and statistics such as the maximum-likelihood method, the generalized least squares method, Bayesian inference, and the Gaussian elimination method.

There are also more and more publications in the literature in which a neural network was used to estimate the OD matrix [32,33]. In [28], the authors proposed a data-driven method for estimating the OD flows in cases where a supply pattern including the speeds and traffic volume is available. They used a simple multilayer perceptron neural network to predict the production and attraction of regions of origins and destinations. The authors showed that no iterative dynamic procedure that leads to equilibrium assignment is needed in the case of these input data.

An important direction in the development of methods to predict traffic volume or estimate the OD matrix is the use of deep learning [7,13–15,24,34–38]. The research results indicate that these methods significantly increase the possibilities and reliability of traffic prediction. The most frequently used networks include:

- LSTM—Long Short-Term Memory network
- TCN—Temporal Convolutional Network
- ConvLSTM—Convolutional LSTM Network
- ST-GCN—Spatial–Temporal Graph Convolutional Network
- MLP—Multi-Layer Perceptron network
- DLNA—Deep Learning Network with Autoencoders
- MGC—Multi-Graph Convolutional network
- ED-MGC network—Encoder–Decoder Multi-Graph Convolutional network
- DySAT—Dynamic Self-Attention Network
- DNEAT—Dynamic Node-Edge Attention Network
- RMGC—Residual Multi-Graph Convolutional network
- ST-ED-RMGC—Spatio–Temporal Encoder–Decoder Residual Multi-Graph Convolutional network
- MF-ResNet—Multi-Fused Residual Network
- CAS-CNN—Channel-wise Attentive Split–Convolutional Neural Network,
- GTFNN—Graph–Temporal Fused Neural Network,
- EEMD-LSTM—Ensemble Empirical Mode Decomposition LSTM Network.

The paper [15] presents a comprehensive review of deep learning approaches used in traffic forecasting from multiple perspectives and summarizes existing traffic forecasting methods. Deep learning methods provide good results and are currently most used in traffic forecasts. The level of forecast error depends on various factors. For example, the authors of [15] found that the precision of the prediction depends mainly on the dataset used and the forecast horizon.

Traffic prediction errors can be assessed using various measures. Widely used methods in the estimation of OD matrices include:

- RMSE (Root Mean Squared Error)
- MAE (Mean Absolute Error)
- MAPE (Mean Absolute Percentage Error)

The RMSE determines how much, on average, the forecast variable's realizations deviate from the calculated forecasts. The values of this error are expressed in units of forecast values and depend on the traffic volume. In turn, the MAE determines by how much, on average, during the prediction period, the actual realizations of the forecast variable will deviate in absolute value from the forecasts. The MAPE was chosen for the analysis because it provides information on the average size of the prediction errors determined as a percentage of the actual value. The MAPE values allow us to compare the accuracy of forecasts obtained for different models. This error is considered one of the best metrics for assessing the prediction accuracy of the model.

To summarize the literature review, it should be noted that deep learning is an important and promising direction of research in the field of OD matrix estimation, but when choosing a type of network to be used, special attention should be paid to the level of prediction estimation error. In the presented publications, the authors estimated OD matrices using various methods, including multiparameter hybrid models that required very high computational power. However, we did not find any studies on the assessment of the error of the predicted OD matrix estimated using deep learning as a function of the traffic volume or its variability. In our opinion, this is an important research problem because OD matrices can only be predicted with an acceptable level of reliability within a certain range of traffic volume values. The purpose of this article is, among other things, to define this range.

The authors' contributions are presented below.

- We examine the impact of the traffic flows and their variability on the accuracy of the estimated OD flows,
- we used the model of direct OD matrix prediction based on traffic volume using deep learning,
- we pre-process several hundred thousand data on traffic intensity from video-sensing detectors located at key points in the city,
- we have shown that the higher the traffic intensity value, the smaller the average MAPE error in estimating the OD matrix,
- we determined regression models of the dependence of the average error of the prognostic OD matrix on the intensity values and their variability recorded using video remote sensing devices in 15 min intervals.

3. Materials and Methods

Analysis of prediction error was performed for the method of estimating and forecasting the OD matrix in the urban road network presented in the publication [8]. The main data were obtained from video remote sensing devices recording information on traffic intensity, which are in key locations of the studied area. A 15 min analysis period was adopted for the prediction of the OD matrix. It was also assumed that there were no traffic disruptions during the measurements. In this case, it is not necessary to consider the path change for the OD pair depending on the level of traffic flow intensity. This assumption allows for the adoption of constant routes in each period.

In the proposed method, the OD pair contains two vertices where a single vehicle trip begins and ends. These points are located at the vertices of the transport network in places where video remote sensing devices have been installed. They recorded traffic intensity in the directions of entry and exit from the study area. Therefore, the structure of the OD matrix corresponds to the connections between these points.

The general scheme of the method for determining the error of OD matrix estimation is shown in Figure 1. The most important input data obtained from video detectors include the registered number of vehicles passing a specific road cross-section in a unit of time, which to prepare the predicted OD matrices was set at 15 min. Moreover, to prepare a road and street network model, it is important to define the boundaries of the analysis area and identify the road infrastructure that can be used to distribute traffic flows. On this basis, minimal paths enabling traffic assignment are determined, both at stage 2 of determining the prior OD matrix and after determining the predicted OD matrix.

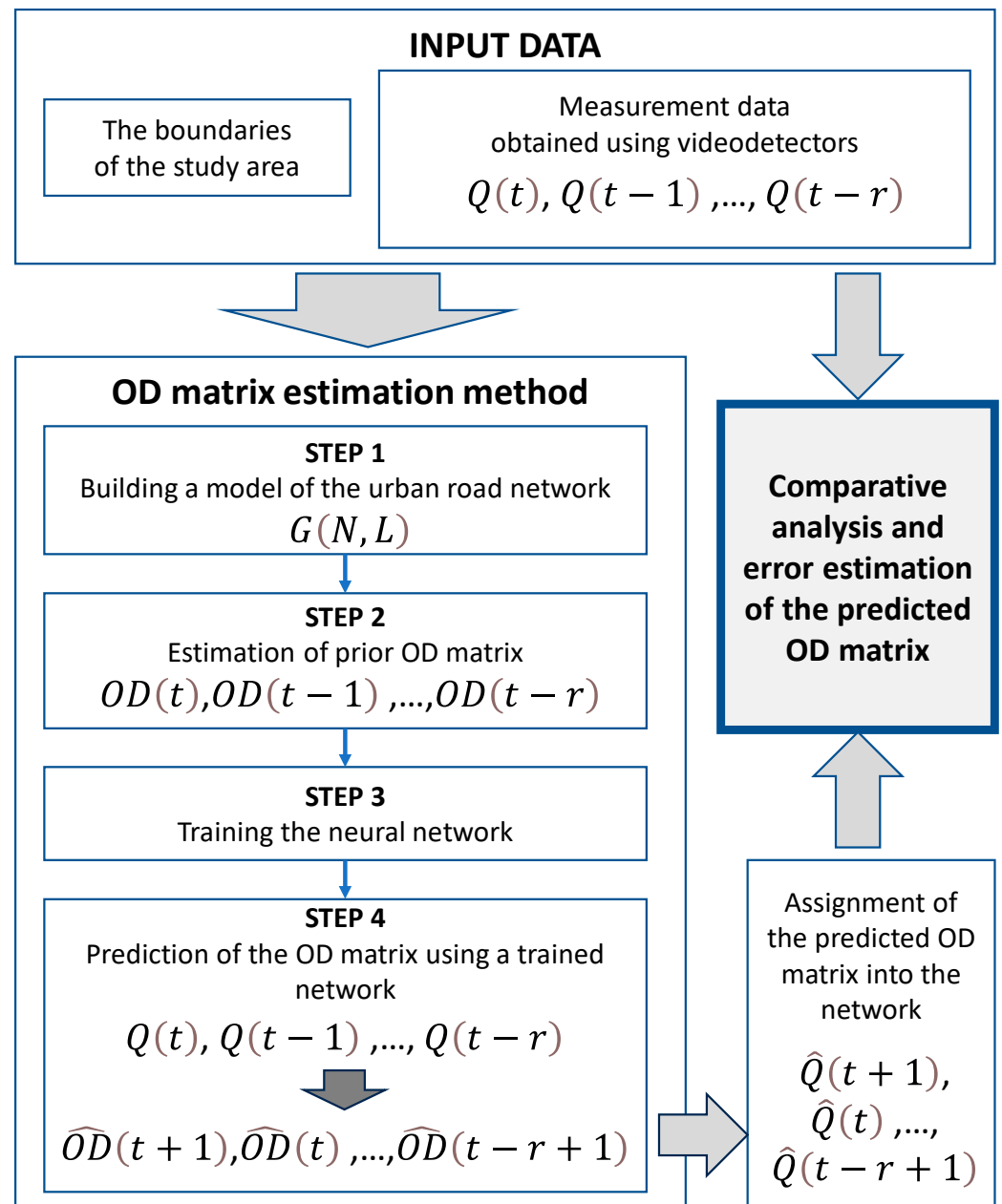


Figure 1. Methodological framework for determining the OD matrix estimation error.

The method of estimating the OD matrix based on traffic counts consists of four steps [11]. First, a road network model should be built to enable further analysis. For this purpose, information about the network structure and the location of video detectors recording traffic data is necessary. The road network model is built based on elements of graph theory.

The road network has been mapped in the form of a directed graph $G(N,L)$, where N corresponds to a set of nodes and L —to a set of links of this graph. The vertices are divided into two subsets:

- NB —a set of border nodes, located on the boundaries of the analysis area; these are places with video detectors that enable data acquisition, and record traffic entering and exiting the analysis area.
- NA —set of nodes inside the study area; these may be places where the main traffic flows divide and merge, as well as those where additional devices enabling vehicle registration are located, i.e., providing information that can be used when building the predicted OD matrix.

These subsets are disjunctive and complementary. Thus, the set of all nodes used to build the road network model can be described as follows:

$$N = \{\dots, i, \dots, j, \dots : i, j \in NB \cup NA, i \neq j, NB \cap NA = \emptyset\} \tag{1}$$

Each section of the road network has been mapped in the form of a pair of nodes between which there is a direct connection by the road infrastructure, i.e., (i,j) , with $i \in N, j \in N, i \neq j$. Therefore, the set of links can be described as follows:

$$L = \{\dots, (i,j), \dots : i \in N, j \in N, i \neq j\} \tag{2}$$

The nodes included in the NB set were used to build the structure of the OD matrix. By connecting the nodes between which there are connections, the OD pairs with the individual elements described as (o,d) were obtained, where $o \in NB, d \in NB, o \neq d$.

The assumptions that the road network between the nodes $o \in NB, d \in NB$ is coherent—which means that at least one path can be provided between each pair of nodes—and that at each point $i \in NB$ traffic data for two directions can be recorded, lead to the conclusion that the OD matrix is square. Furthermore, the presented research covered only OD flows with different origins and destinations, i.e., $o \neq d$.

In the second step, a prior OD matrix is constructed based on data obtained from traffic monitoring devices, which can be formally presented in the form:

$$OD(t) \equiv [x^{od}(t)] = \begin{bmatrix} - & x^{12}(t) & \dots & \dots & x^{1n}(t) \\ x^{21}(t) & - & \dots & \dots & \vdots \\ \vdots & \dots & - & \dots & \vdots \\ \vdots & \dots & \dots & - & \vdots \\ x^{n1}(t) & \dots & \dots & x^{nn-1}(t) & - \end{bmatrix} \tag{3}$$

where n is the number of rows or columns in the square OD matrix.

The iterative procedure supporting the process of building this matrix assumes that after assigning the estimated OD matrix to a road network, the deviations between the registered and estimated traffic flows on individual sections should be as small as possible [11]. Therefore, we are looking for such values of estimated OD flows $\widehat{OD}(t)$ which, when assigned to the network, give minimal deviations between the estimated volumes of traffic flows $\widehat{Q}(t)$ and the real ones $Q(t)$, i.e., registered by the video-sensing detectors during the period t . This corresponds to the formulation of the optimization task for each interval t in the following form:

$$\widehat{OD}(t)^* = \underset{L}{\operatorname{argmin}} F(Q(t), \widehat{Q}(t)) \quad \wedge \quad \widehat{Q}(t) = \operatorname{assign}(\widehat{OD}(t)) \tag{4}$$

where the notation $\text{assign}(\widehat{OD}(t))$ denotes the assignment of the matrix $\widehat{OD}(t)$ to the road network. Furthermore:

$\widehat{OD}(t) \equiv [\hat{x}^{od}(t)]$ —estimated (predicted) OD matrix,

$Q(t) \equiv [q_{ij}(t)]$ —matrix containing intensity values on the sections $(i, j) \in L$, recorded by video-sensing devices,

$\hat{Q}(t) \equiv [\hat{q}_{ij}(t)]$ —matrix containing the intensity values on the sections $(i, j) \in L$, estimated based on the assignment of the estimated OD matrix, i.e., $\widehat{OD}(t)$ to the road network.

The minimization of the objective function is performed separately for each period t . This means that the effect of this procedure is to obtain such a set of OD matrices (for each interval), the assignment of which on the road network leads to the best convergence with the observed results. These matrices constitute reference matrices for the process of training neural networks in step 3 [11].

The spatial scope of the analysis was assumed to cover the area in which the path of every OD pair is traveled by traffic flows in a time shorter than the analysis period t . It was also assumed that the network is not overloaded, which means that for individual OD pairs the shortest minimum paths are constant and independent of the analyzed period. Hence, for each OD pair, one path is defined in all intervals. Therefore, the overall relationship between traffic volumes on the link and OD flows can be determined as:

$$q_{ij}(t) = \sum_{(o,d)} \alpha_{ij}^{od}(t) \cdot x^{od}(t) \tag{5}$$

where:

$q_{ij}(t)$ —registered traffic volumes on the link $(i, j) \in L$ in period t ,

$\alpha_{ij}^{od}(t)$ —fraction (share) of OD flows for the pair (o, d) that uses link $(i, j) \in L$ in period t ,

$x^{od}(t)$ —O-D flows for the pair (o, d) in period t ; the element of the matrix $OD(t)$.

To analyze the impact of traffic flows on the prediction error, the OD matrix prediction error values were used, which were calculated using a deep learning neural network with autoencoders (DLNA). This prediction method turned out to be better than the method using the LSTM network or the historical method [8]. The structure of the neural network with autoencoders used to predict the OD matrix flows is shown in Figure 2. After reaching an acceptable error level, the neural network was used to predict the OD matrix in the last stage of the method.

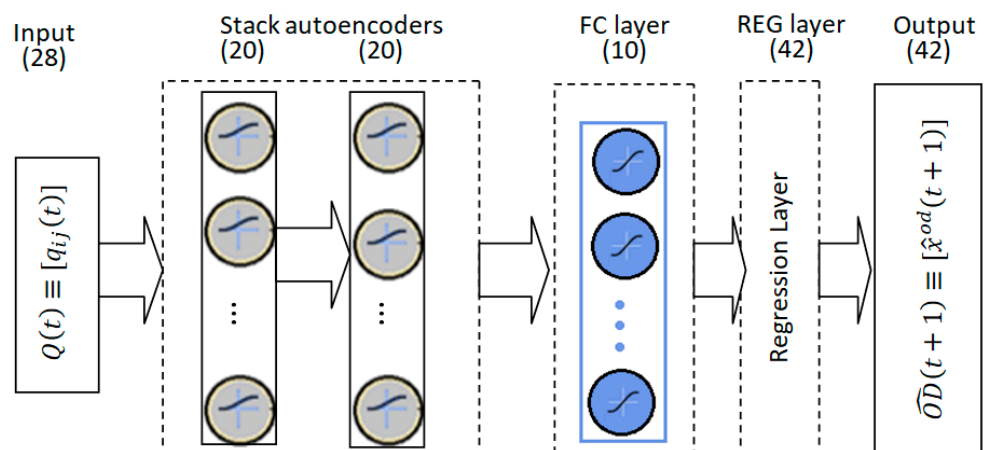


Figure 2. Diagram of the deep learning with autoencoders neural network used in the OD matrix prediction model.

The network maps the function:

$$\widehat{OD}(t+1) = F(Q(t)) \tag{6}$$

During network training, traffic volume matrix sequences are provided for inputs: $Q(t), Q(t-1), \dots, Q(t-r)$, where $Q(t) \equiv [q_{ij}(t)]$, with $q_{ij}(t)$ interpreted as traffic intensity in the road section $(i, j) \in L$ in the interval t (15 min), $(r+1)$ —number of elements of the training set.

The output data are the predicted OD matrices in the next interval, i.e., $\widehat{OD}(t+1), \widehat{OD}(t), \dots, \widehat{OD}(t-r+1)$, where $\widehat{OD}(t+1) \equiv [\hat{x}^{od}(t+1)]$, with $\hat{x}^{od}(t+1)$ interpreted as an element of the OD matrix in the interval $(t+1)$. Traffic flow sequences on the input of the network are time series.

A deep learning network for the prediction of the OD flow matrix consists of an input DLNA layer, an FC layer, and a REG layer as the output layer. The traffic flow corresponding to successive 15 min intervals was given at the network input. The number of entries was 28 values of traffic flows on links registered in the same interval.

Various network configurations were checked by changing the number of neurons in the autoencoder layers and the number of neurons in the FC layer. The best network that contained autoencoder layers consisted of a stack of two layers of autoencoders with 20 neurons in each layer. The number of inputs were 28 and outputs and the remaining network layers were 42 neurons for the OD matrix with dimensions of 7×7 , disregarding the values on the main diagonal corresponding to intrazonal flows. The maximum number of epochs was 100.

The MATLAB 2020a (Academic License) software and a laptop with an Intel (R) Core (TM) i5, 1.19 GHz processor, 8 GB RAM, and 1 TB SDD disk were used for network training. The training time depended mainly on the number of epochs, the number of layers, the number of neurons in autoencoder layers, as well as the mini-batch value, and ranged from 5 min to 4 h.

When training the network, the input was the traffic flow in each interval, and the output was the predicted OD matrix in the next interval. Then, the predicted OD matrix was assigned to the road network, which made it possible to estimate values of traffic intensity at the border nodes (i.e., $i \in NB$), which were compared with the values recorded by video remote sensing devices at these points.

The data obtained from the video detectors included both the values of traffic intensities entering the study area in the period t , marked as $q_i^{\text{in}}(t)$, and exiting this area, marked as $q_i^{\text{out}}(t)$. Therefore, at each location of the video detector (i.e., $i \in NB$), data were obtained in the form of two vectors:

- $Q_i^{\text{in}} \equiv [q_i^{\text{in}}(t)]$ —a vector containing the values of traffic intensity $q_i^{\text{in}}(t)$ entering the study area in subsequent intervals t ,
- $Q_i^{\text{out}} \equiv [q_i^{\text{out}}(t)]$ —a vector containing the values of traffic intensity $q_i^{\text{out}}(t)$ exiting the study area in subsequent intervals t .

As a result of the comparison, the error values MAPE^{in} and MAPE^{out} were calculated for the estimation of the OD matrix for each interval for a given working day of the week and for each border node according to the formulas:

$$\text{MAPE}_i^{\text{in}, wd}(t) = \left| \frac{q_i^{\text{in}, wd}(t) - \hat{q}_i^{\text{in}, wd}(t)}{q_i^{\text{in}, wd}(t)} \right| * 100\% \quad (7)$$

$$\text{MAPE}_i^{\text{out}, wd}(t) = \left| \frac{q_i^{\text{out}, wd}(t) - \hat{q}_i^{\text{out}, wd}(t)}{q_i^{\text{out}, wd}(t)} \right| * 100\% \quad (8)$$

where:

- wd —the type of working day,
- $q_i^{\text{in}, wd}(t)$ —the values of traffic intensity entering the study area on working day wd , registered in interval t by the video remote sensing device located at the node $i \in NB$,
- $q_i^{\text{out}, wd}(t)$ —the values of traffic intensity exiting the study area on working day wd , registered in interval t by the video remote sensing device located at the node $i \in NB$,

$\hat{q}_i^{\text{in},wd}(t)$ —the values of traffic intensity entering the study area on working day wd at the node $i \in NB$ in interval t estimated based on the assignment predicted OD matrix $\widehat{OD}(t)$ to the road network,

$\hat{q}_i^{\text{out},wd}(t)$ —the values of traffic intensity exiting the study area on working day wd at the node $i \in NB$ in interval t estimated based on the assignment predicted OD matrix $\widehat{OD}(t)$ to the road network.

4. Datasets

An important step was to build a road network model for the tested area, which is presented in Section 4.1. Examples of input and output data for training the neural network are presented in Section 4.2, and the method of analyzing traffic intensity recorded by video sensing devices is discussed in Section 4.3.

4.1. Description of the Study Area

The research was carried out based on data obtained from the Traffic Control Center in a medium-sized city in Poland, Gliwice. The analysis area and the locations of the video detectors from which the measurement data were obtained are shown in Figure 3. The border nodes (i.e., included in the NB set) are marked in red. For each of them, the direction corresponding to the vectors Q_i^{in} and Q_i^{out} is marked.

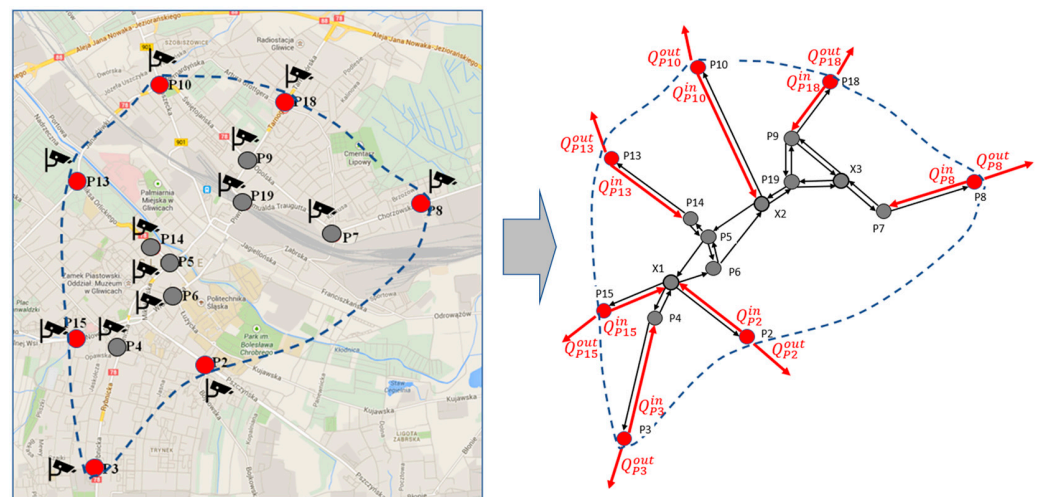


Figure 3. Analysis area with location of detectors and road network model marked.

The lack of traffic congestion allows the assumption of fixed shortest paths between the selected points. Paths for individual OD pairs were determined using the Dijkstra algorithm. On this basis, a schematic structure of the aggregated model of the road network shown in Figure 3 has been constructed.

Based on the symbols in Figure 3, sets of nodes and links for the road network have been presented in Table 1.

At points $P4$, $P5$, $P6$, $P7$, $P9$, $P14$, and $P19$, there are additional devices recording traffic intensity, which were used in the estimation of the OD matrix. The points designated as $X1$, $X2$, and $X3$ are places where traffic flows are divided or merged, and they were introduced to enable the construction of a road network model. In the set of links, the traffic intensities for the following ones were considered for the analysis of errors $MAPE_i^{\text{in},wd}(t)$: $(P2, X1)$, $(P3, P4)$, $(P8, P7)$, $(P10, X2)$, $(P13, P14)$, $(P15, X1)$, and $(P18, P9)$. They are marked in red in Figure 3.

Table 1. Sets of nodes and links for the road network.

Elements of Network	Set Designation	Elements of the Set	Number of Elements of the Set
nodes	NB	{P2, P3, P8, P10, P13, P15, P18}	7
	NA	{P4, P5, P6, P7, P9, P14, P19, X1, X2, X3}	10
links	L	{(P3, P4), (P4, P3), (P4, X1), (P2, X1), (P15, X1), (X1, P6), (P13, P14), (P14, P13), (P14, P5), (X2, P5), (P10, X2), (P19, P9), (P19, X3), (P9, X3), (P9, P18), (P18, P9), (P7, X3), (P7, P8), (P8, P7), (X1, P4), (X1, P15), (X1, P2), (P5, X1), (P5, P14), (P5, P6), (P6, P5), (P6, X2), (X2, P19), (P19, X2), (X2, P10), (P9, P19), (X3, P19), (X3, P9), (X3, P7)}	34

4.2. Input and Output Data for Neural Network

The research was carried out using data from video remote sensing detectors for three months (May, June, and July). Data were collected at 5 min intervals, but for the study, they were processed in 15 min. The camera viewing range and detection field for recording traffic data using video remote sensing are shown in Figure 4.

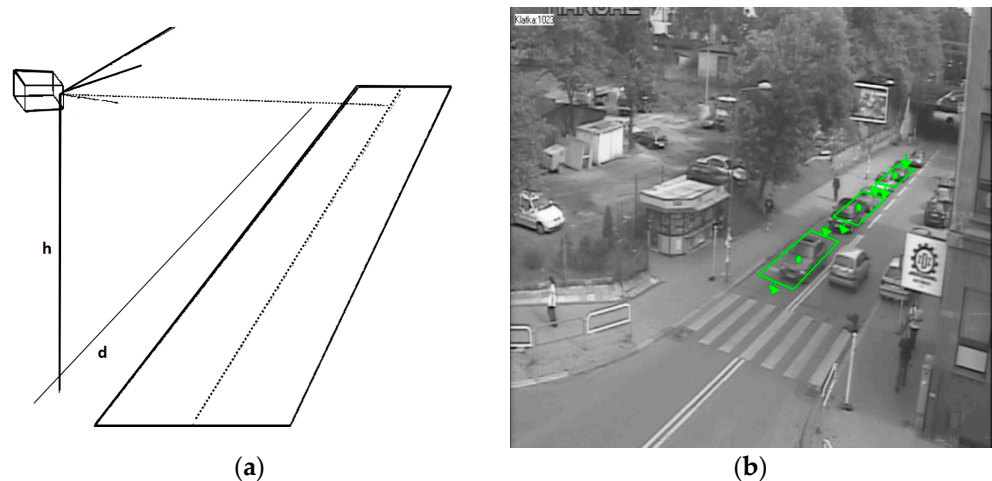


Figure 4. The method of recording traffic data using video remote sensing: (a) the camera viewing range; (b) the detection field.

Based on the selected locations of the measurement points and traffic data recorded by the video detectors, the OD flows for 42 OD pairs were estimated. An interval of 15 min was considered when generating the OD matrix for each of the five working days separately. A total of 2400 prior OD matrices were estimated for the training sequence and 480 (i.e., 5×96) for the test sequence.

Example input and output data (i.e., training data) for the neural network for the selected interval are presented in Tables 2 and 3. The values of the reference OD matrix are expressed as relative units.

The results obtained for test data for five working days, from Monday to Friday and from 6:00 a.m. to 10:00 p.m., were analyzed.

Table 2. Traffic intensities [veh/15 min] on 16.05.2016 (Monday) in the period from 07:45 a.m. to 08:00 a.m.

P2 in	P2 out	P3 in	P3 out	P4 in	P4 out	P5 in
257	233	199	115	148	108	142
P5 out	P6 in	P7 in	P7 out	P8 in	P8 out	P9 in
379	350	164	191	157	138	132
P9 out	P10 in	P10 out	P13 in	P13 out	P14 in	P15 in
100	101	126	57	128	141	211
P15 out	P18 in	P18 out	P19 in	P19 out	P13 in	P13 out
125	154	104	99	77	182	135

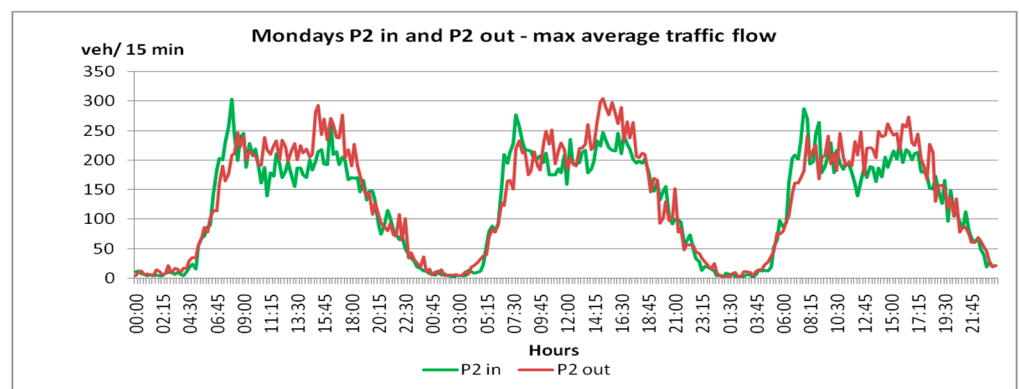
Table 3. OD matrix [-] for 16.05.2016 (Monday) in the period from 07:45 a.m. to 08:00 a.m.

Nodes	P2	P3	P15	P13	P10	P18	P8
P2	-	0.183	0.275	0.194	0.217	0.181	0.140
P3	0.201	-	0.167	0.159	0.167	0.150	0.125
P15	0.257	0.150	-	0.170	0.187	0.157	0.127
P13	0.200	0.144	0.163	-	0.166	0.147	0.125
P10	0.214	0.132	0.156	0.148	-	0.083	0.083
P18	0.153	0.121	0.123	0.125	0.097	-	0.438
P8	0.208	0.160	0.166	0.169	0.147	0.127	-

4.3. Traffic Volume Analysis

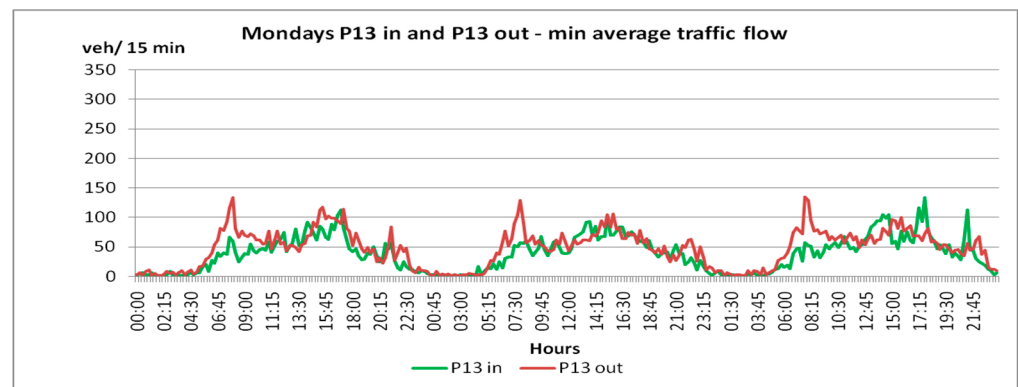
Before we started analyzing the dependence of the OD matrix prediction error on traffic flows, we carried out a preliminary analysis of this intensity. Figure 4a,b show daily traffic patterns at point P2, where the highest average traffic volumes were recorded considering the entire network, and at point P13, with the lowest average traffic volumes.

In Figure 5, for example, traffic intensities recorded by video detectors show significant differences in specific periods of the day. Hence, averaging the flows, e.g., for OD pairs, for the whole road network would not be a good solution. Similarly, the average of peak hours differed depending on the OD pairs. Therefore, we decided to divide the traffic intensity values at the border points for flows entering and exiting the study area into ranges, regardless of the time of occurrence. Seven ranges of traffic intensity were adopted as follows: 0–50, 51–100, 101–150, 151–200, 201–250, 251–300, and 301–351 [veh/15 min].



(a)

Figure 5. Cont.



(b)

Figure 5. Daily traffic flow patterns for selected measurement points: (a) P2 out and P2 in—the highest intensity; (b) P13 out and P13 in—the lowest total intensity.

We averaged the traffic intensity and prediction errors of the OD matrix for the applied traffic intensity ranges in the period from 6:00 a.m. to 10:00 p.m., for each test day of the week and each direction (entry and exit) of OD flows. Sample data for Friday for the vectors Q_i^{out} are presented in Tables 4 and 5. Empty cells mean that at a given measurement point for a specific direction, no traffic intensity within a given range of values was recorded.

Table 4. Averaged values of traffic intensity [veh/15 min] for the ranges.

Ranges of Traffic Intensity	Measurement Points						
	P2 out	P3 out	P15 out	P13 out	P10 out	P18 out	P8 out
0–50	-	48	36	40		32	43
51–100	97	77	79	71	84	77	78
101–150	123	129	114	112	130	120	119
151–200	185	171	-	167	171	154	160
201–250	228	-	-	226	217	-	-
251–300	273	-	-	-	-	-	-
301–350	314	-	-	-	-	-	-

Table 5. Averaged values of MAPE^{out} [%] for the ranges of traffic intensity.

Ranges of Traffic Intensity	Measurement Points						
	P2 out	P3 out	P15 out	P13 out	P10 out	P18 out	P8 out
0–50	-	12.969	9.465	7.922	-	12.652	8.452
51–100	5.315	9.377	7.032	6.261	7.996	8.540	8.235
101–150	7.570	6.704	8.001	6.677	7.453	6.993	8.026
151–200	6.627	6.232	-	5.479	5.802	4.583	5.849
201–250	6.587	-	-	3.201	6.051	-	-
251–300	5.133	-	-	-	-	-	-
301–350	6.584						

During the analysis, we noticed that the variability of traffic intensity was different in particular ranges. Moreover, the calculated values of MAPE for the estimation of the OD matrix depended not only on traffic intensity but also on its variability. Therefore, we also included this parameter in our analyses. As a measure of the spread of traffic intensity,

the coefficient of variation was determined for the given ranges, calculated separately for flows entering and exiting the studied area according to the formulas:

$$CV_{i,r}^{\text{in},wd} = \frac{SD_r^{wd}(Q_i^{\text{in}})}{AVG_r^{wd}(Q_i^{\text{in}})} * 100\% \quad (9)$$

$$CV_{i,r}^{\text{out},wd} = \frac{SD_r^{wd}(Q_i^{\text{out}})}{AVG_r^{wd}(Q_i^{\text{out}})} * 100\% \quad (10)$$

where:

$SD_r^{wd}(Q_i^{\text{in}})$ —standard deviation of the traffic intensities for vehicles entering the study area, recorded at the measurement point i (presented in the form of a vector Q_i^{in}) on the working day wd and falling within the range r ,

$SD_r^{wd}(Q_i^{\text{out}})$ —standard deviation of the traffic intensities for vehicles exiting the study area, recorded at the measurement point i (presented in the form of a vector Q_i^{out}) on the working day wd and falling within the range r ,

$AVG_r^{wd}(Q_i^{\text{in}})$ —average value of the traffic intensities for vehicles entering the study area, recorded at the measurement point i (presented in the form of a vector Q_i^{in}) on the working day wd and falling within the range r ,

$AVG_r^{wd}(Q_i^{\text{out}})$ —average value of the traffic intensities for vehicles exiting the study area, recorded at the measurement point i (presented in the form of a vector Q_i^{out}) on the working day wd and falling within the range r .

Table 6 presents the values of the coefficient of variation of traffic intensity for the given ranges, for the dataset presented in Table 4 (Friday for the vectors Q_i^{out}).

Table 6. Averaged values of coefficient of variation [%] for the ranges of traffic intensity.

Ranges of Traffic Intensity	Measurement Points						
	P2 out	P3 out	P15 out	P13 out	P10 out	P18 out	P8 out
0–50	-	-	26.219	19.592	-	28.684	16.462
51–100	-	22.457	15.554	15.854	18.970	18.759	20.239
101–150	12.015	10.213	10.190	12.434	10.132	10.674	9.727
151–200	5.738	7.813	-	-	7.893	-	-
201–250	7.178	-	-	-	-	-	-
251–300	5.381	-	-	-	-	-	-
301–350	2.429	-	-	-	-	-	-

Empty places in the table indicate no traffic intensity in this range was recorded, or a single value for which the standard deviation cannot be calculated.

5. Results

The prediction errors of the OD matrix were calculated using data recorded over half a year, from January to June, in parallel at all measurement points to take into account spatial-temporal relationships. As previously mentioned, our analysis included data registered for five test working days from Monday to Friday from 6 a.m. to 10 p.m.

To verify our thesis, we analyzed the dependence of the average MAPE prediction error on the average traffic intensity values for a given intensity range for five working days and all measurement points. Charts of these dependencies for, for example, points P2 and P3 for one working day (Thursday) are presented in Section 5.1, along with a discussion.

Next, Section 5.2 presents the results of the correlation analysis for the average traffic intensity values at the analyzed measurement points and the MAPE error value of the predicted OD flows both for the flows entering (i.e., inputs) and exiting (i.e., outputs) the

study area. In turn, Section 5.3 presents the results of correlation analysis for prediction errors and the traffic volume coefficient of variation.

5.1. Analysis of the Dependence of the Average MAPE Prediction Error on the Average Intensity for the Ranges

For the intensity ranges included in Tables 4–6, the average MAPE errors and the average traffic intensity values were calculated. The results were compared for five test working days and all border points where the video remote sensing devices were located. The results for selected points P2 and P3 are shown in Figure 6.

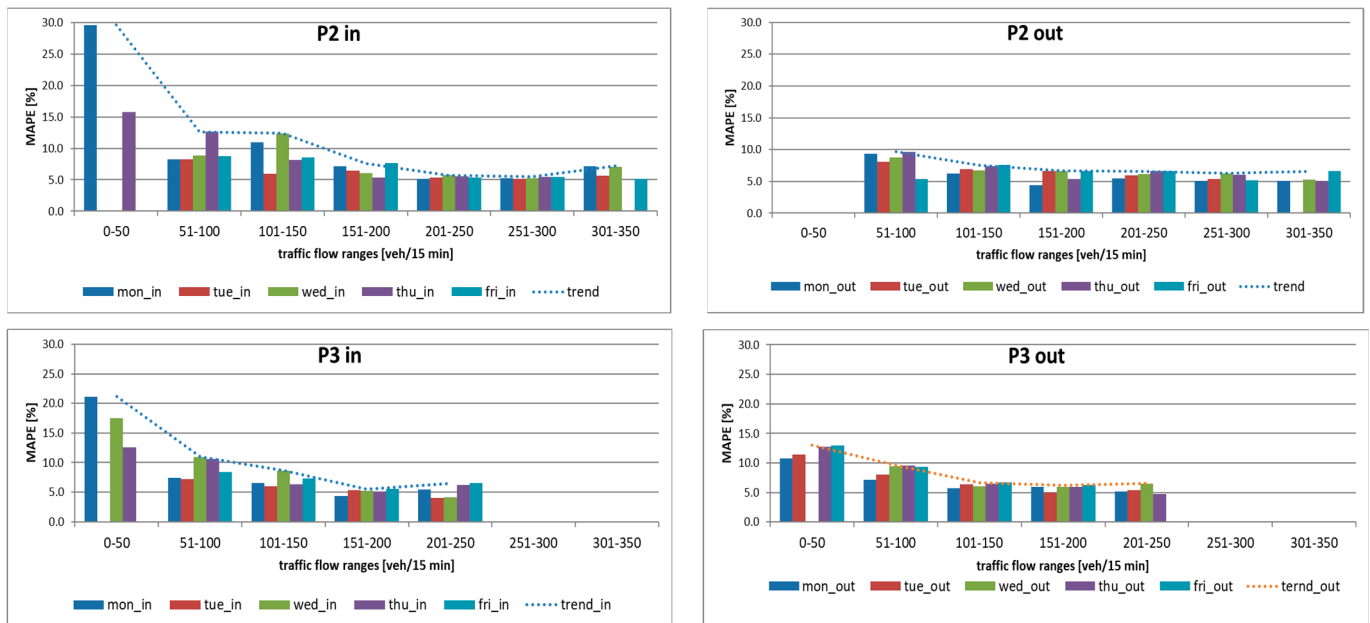


Figure 6. Dependence of the MAPE on the traffic intensity values for the entry (in) and exit (out) directions for measurement points P2 and P3.

For point P2 in, traffic intensity values less than 50 [veh/15 min] occurred only on Monday and Thursday, while for point P2 out, the intensity values exceeded 50 [veh/15 min] on all working days from Monday to Friday. In turn, for the points P3 in and P3 out, the intensity values did not exceed 250 [veh/15 min].

In Figure 6 it can be seen that the MAPE error for both points P2 and P3 stabilizes at a value of approximately 6–7% for the traffic intensity range above 150 [veh/15 min]. A similar situation occurs for the remaining measurement points.

Traffic intensities for individual working days were also analyzed separately. For each day, the intensities at all measurement points were studied simultaneously. The results of these analyses are presented in Figure 7.

In Figure 7, as before, a decreasing trend of the MAPE error with increasing traffic intensity for all measurement points (both inputs and outputs) is observed, except for single deviations.

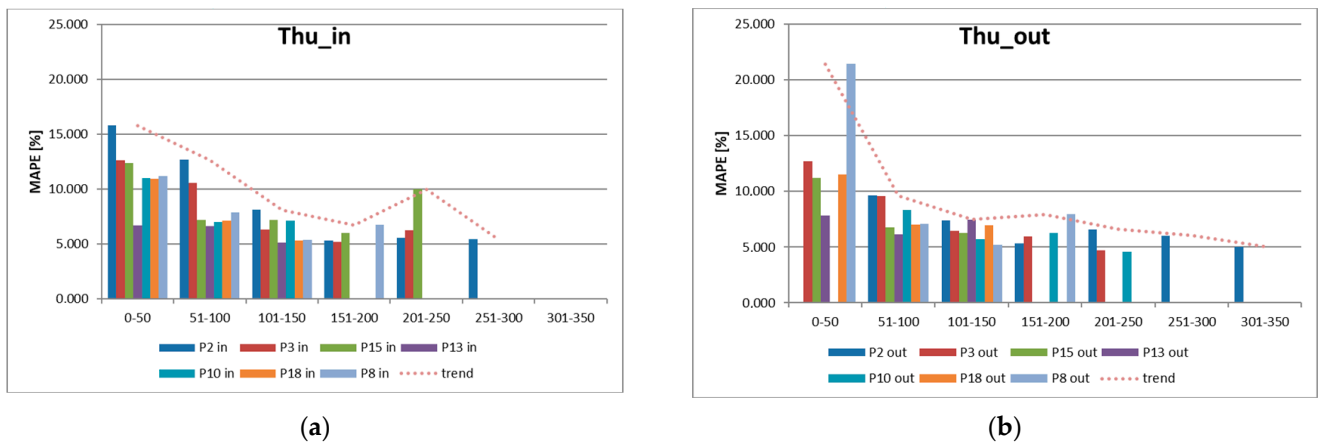


Figure 7. The values of MAPE error for all measurement points for Thursday; (a) for traffic flows entering the studied area (in); (b) for traffic flows exiting the study area (out).

5.2. Correlation between MAPE Error and Traffic Volume

The next step of the analysis was to examine the correlation between the MAPE value of the prediction error of the OD matrix and the average intensity of the traffic. The analysis was performed for five working days for all measurement points for both the inputs and outputs of the OD matrix. Figure 8 shows the results in the form of correlation plots for Thursday.

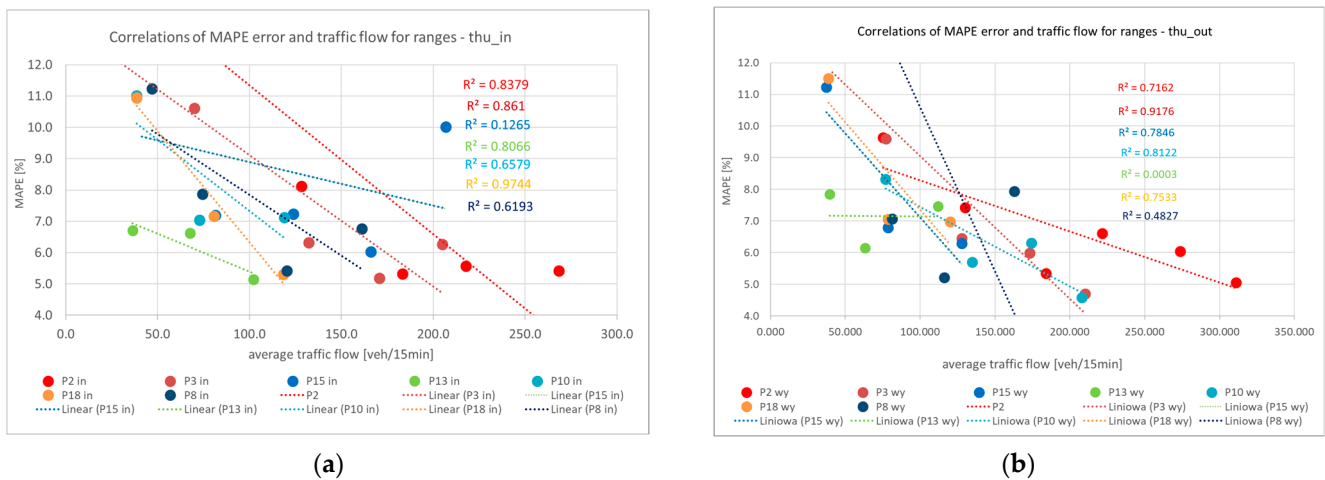


Figure 8. Dependence of the average MAPE error and the average traffic flows for intensity ranges; (a) for traffic flows entering the studied area (in); (b) for traffic flows exiting the study area (out).

The R^2 coefficient for all P in points (Figure 8a) is greater than 0.6, which indicates a high correlation between the prediction error and the intensity value. As the average traffic intensity increases, the prediction error value of the OD matrix decreases. On average, the decrease in the error value is 0.045 per unit of intensity [veh/15 min] for all inputs from the points for Thursday, calculated as the average of the regression line parameters for all measurement points.

For the point outputs of the OD matrix, the coefficients of correlation R^2 are also mostly high. Only in the case of point P13 is the correlation very weak, which is due to the fact that the intensity values at this point do not exceed 150 [veh/15 min]. For the outputs, the average decrease in the error value is 0.042 per unit of intensity [capacity/15 min]—Figure 7b.

5.3. Correlation between MAPE Error and Coefficient of Variation

During our analyses, we also noticed that not only the range of traffic intensity values but also the traffic variability index has a significant impact on the prediction error. Figure 9 shows the correlation plots of the average MAPE error and the calculated coefficient of variation of the traffic intensity for all inputs and outputs of the OD matrix for Thursday.

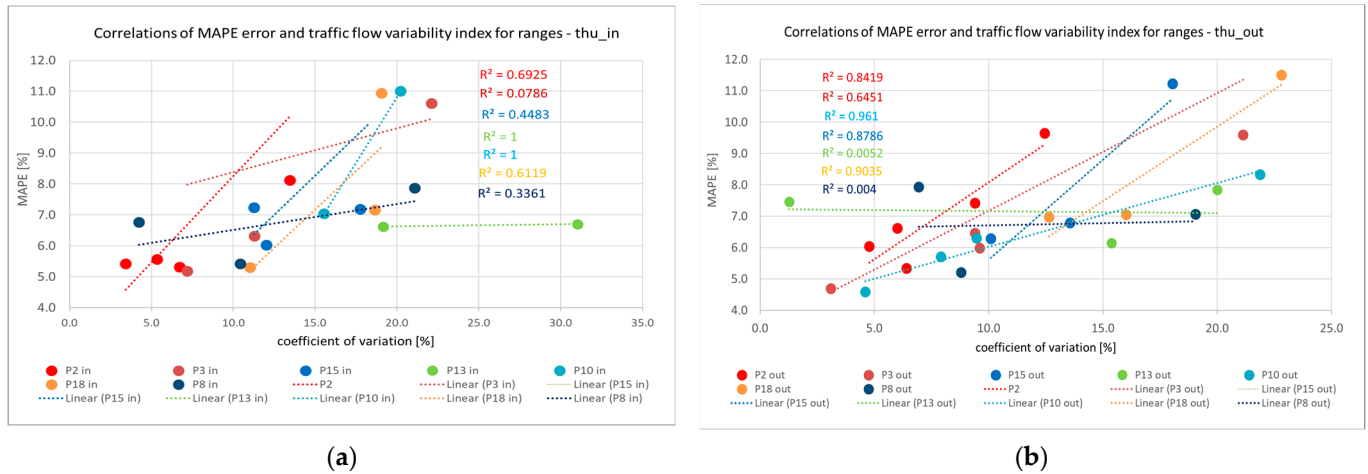


Figure 9. Dependence of the average MAPE error and the coefficient of variation for all measurement points for Thursday; (a) for traffic flows entering the studied area (in); (b) for traffic flows exiting the study area (out).

The graphs presented in Figure 9 clearly show that the correlations between the MAPE error value and the coefficient of intensity variation are high, and in most cases, the value of the coefficient R^2 for all models is greater than 0.6. This confirms our thesis about the significant impact of traffic intensity variability on the value of the MAPE prediction error. The greater the intensity variability, the greater the error. However, in several cases, the impact of the coefficient of variation on the prediction error may not satisfy this thesis. This is mainly due to the small amount of data on intensity in this range of traffic intensity.

6. Discussion

It was shown in Sections 5.1–5.3 that the prediction error MAPE depends on the average value of traffic intensity within the range and the coefficient of variation of this intensity in the range. The relationship models for inputs and outputs were determined in the form of multiple linear regression using the least squares method as follows:

$$MAPE_{avg,r}^{in,wd} = m1^{in} \cdot Q_{avg,r}^{in,wd} + m2^{in} \cdot CV_{avg,r}^{in,wd} + m3^{in} \quad (11)$$

$$MAPE_{avg,r}^{out,wd} = m1^{out} \cdot Q_{avg,r}^{out,wd} + m2^{out} \cdot CV_{avg,r}^{out,wd} + m3^{out} \quad (12)$$

where:

$MAPE_{avg,r}^{in,wd}$ —average value of MAPE error for the traffic flows entering the study area in the working day wd with the recorded value of intensity within the range r ,

$MAPE_{avg,r}^{out,wd}$ —average value of MAPE error for the traffic flows exiting the study area in the working day wd with the recorded value of intensity within the range r ,

$Q_{avg,r}^{in,wd}$ —average values of traffic intensities for the flows entering the study area in the working day wd with the recorded value of intensity within the range r ,

$Q_{avg,r}^{out,wd}$ —average values of traffic intensities for the flows exiting the study area in the working day wd with the recorded value of intensity within the range r ,

$CV_{avg,r}^{in,wd}$ —average values of coefficient of variation of traffic intensities for the flows entering the study area in the working day wd with the recorded value of intensity within the range r ,

$CV_{avg,r}^{out,wd}$ —average values of coefficient of variation of traffic intensities for the flows exiting the study area in the working day wd with the recorded value of intensity within the range r ,

$m1^{in}$, $m2^{in}$, and $m3^{in}$ —values of the parameters of the regression model for the flows entering the study area,

$m1^{out}$, $m2^{out}$, and $m3^{out}$ —values of the parameters of the regression model for the flows exiting the study area.

The multiple linear regression models are presented in Figure 10.

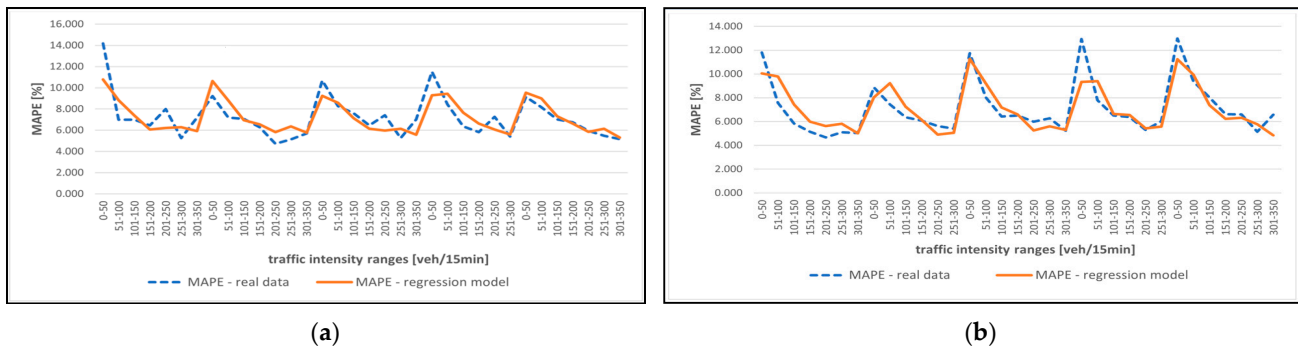


Figure 10. Dependence of the average MAPE error on the traffic flow ranges and the coefficient of variation; (a) for average traffic flows entering the studied area (in); (b) for average traffic flows exiting the study area (out).

The parameters of the multiple regression function (i.e., $m1^{in}$, $m2^{in}$, and $m3^{in}$ for flows entering the study area, and $m1^{out}$, $m2^{out}$, and $m3^{out}$ for the flows exiting the study area) were calculated. Their values are appropriately 0.00089, 0.2696, and 4.4454 for the inputs, and 0.001154, 0.3307, and 3.6735 for the outputs. To calculate the parameters, the values of vectors containing MAPE errors, the corresponding traffic intensities, and coefficients of variation for five test working days for all measurement border points were used. The coefficient of intensity variation has a greater impact on the error value than the average traffic intensity.

The dependence of the prediction error on the average intensity and the coefficient of the variation calculated for the intensity ranges is well shown by the bubble chart in Figure 11. The size of the bubbles determines the value of the coefficient of traffic intensity variation for specific ranges.

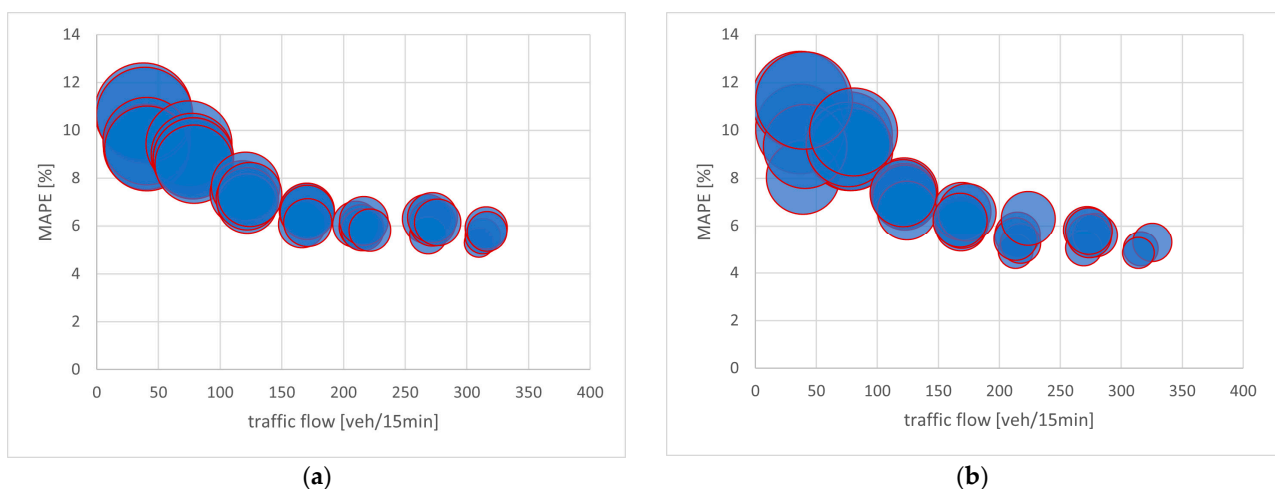


Figure 11. Dependence of the average MAPE error on the traffic flow ranges and the coefficient of variation for all measurement points of the OD matrix; (a) for average traffic flows entering the studied area; (b) for average traffic flows exiting the study area.

The graph clearly shows that as the intensity increases, the prediction error decreases. At the same time, the error is larger when the bubble size is larger. A larger bubble means a larger coefficient of the variation of traffic intensity.

7. Conclusions

Errors in the short-term forecasting of the OD matrix can significantly affect the effectiveness of traffic management and, therefore, also the traffic conditions in the transport network. Thus, efforts should be made to ensure that the methods used are subject to the lowest possible level of error and the obtained predictions are characterized by the highest possible level of reliability.

The article presents the results of the analysis of MAPE prediction errors, calculated based on the prediction of OD flows using the deep learning method [8] as dependent on the traffic intensity and its variability. To facilitate the analysis, the intensity values registered by video remote sensing devices in the measurement points both for the traffic flows entering and exiting the study area were divided into ranges. For the intensity ranges, the average values of errors, intensities, and coefficients of variation were calculated. The influence of traffic intensity, the influence of the coefficient of variation, and the influence of both parameters on the error values were examined separately. The results of the analysis showed that as the average traffic intensity increases, the error decreases, while an increase in the value of the coefficient of variation causes an increase in the prediction error. To check the influence of both parameters simultaneously, the multiple linear regression function was used. The parameters of this function, calculated for all measurement points, showed that the coefficient of variation has a much greater impact on the error value than the intensity itself.

The subject discussed in the article requires further in-depth analysis to detect the dependence of the OD matrix prediction error on factors other than traffic characteristics. An interesting direction for further research is to find a relationship between the location of video remote sensing devices, the structure of the road network, or transport subsystems operating in the studied area, and the degree of their use in transport demand. Additionally, an interesting area of research is to check whether there are also relationships between traffic intensity and its variability for other metrics of error in the estimation of OD matrices (e.g., RMSE, MAE) and whether they are shaped in a similar way as in the case of the MAPE error. It is also worth comparing prediction errors in this respect for different deep learning network models (including MF-ResNet, CAS-CNN, STGCN, STSGCN, and STFGNN).

The results obtained can be used to assess the prediction of the OD flows. Correction of the predicted values of the OD matrix considering the parameters mentioned above can significantly improve the accuracy of the prediction and therefore have a positive impact on traffic management and traffic conditions on the road network.

Author Contributions: Conceptualization, R.Ž. and T.P.; methodology, R.Ž. and T.P.; software, T.P.; validation, T.P.; formal analysis, R.Ž.; investigation, T.P. and R.Ž.; resources, T.P.; data curation, T.P.; writing—original draft preparation, R.Ž. and T.P.; writing—review and editing, R.Ž. and T.P.; visualization, T.P. and R.Ž.; supervision, R.Ž. and T.P.; project administration, R.Ž. and T.P.; funding acquisition, T.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Dataset available on request from the authors.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Riaz, W.; Gao, C.; Azeem, A.; Saifullah; Bux, J.A.; Ullah, A. Traffic Anomaly Prediction System Using Predictive Network. *Remote Sens.* **2022**, *14*, 447. [[CrossRef](#)]
2. Macioszek, E.; Kurek, A. Extracting Road Traffic Volume in the City before and during COVID-19 through Video Remote Sensing. *Remote Sens.* **2021**, *13*, 2329. [[CrossRef](#)]

3. Fu, H.; Lam, W.H.K.; Shao, H.; Ma, W.; Chen, B.Y.; Ho, H.W. Optimization of multi-type sensor locations for simultaneous estimation of origin-destination demands and link travel times with covariance effects. *Transp. Res. Part B Methodol.* **2022**, *166*, 19–47. [[CrossRef](#)]
4. Fu, H.; Lam, W.H.K.; Shao Xu, X.P.; Lo, H.P.; Chen, B.Y.; Sze, N.N.; Sumalee, A. Optimization of traffic count locations for estimation of travel demands with covariance between origin-destination flows. *Transp. Res. Part C Emerg. Technol.* **2019**, *108*, 49–73. [[CrossRef](#)]
5. Żochowska, R. Dynamic Approach to the Origin-Destination Matrix Estimation in Dense Street Networks. *Arch. Transp.* **2012**, *24*, 389–413. [[CrossRef](#)]
6. Żochowska, R. Selected issues in modelling of traffic flows in congested urban networks. *Arch. Transp.* **2014**, *29*, 77–89. [[CrossRef](#)]
7. Pamuła, T. Impact of Data Loss for Prediction of Traffic Flow on an Urban Road Using Neural Networks. *IEEE Trans. Intell. Transp. Syst.* **2019**, *20*, 1000–1009. [[CrossRef](#)]
8. Jacyna, M.; Wasiak, M.; Lewczuk, K.; Karoń, G. Noise and environmental pollution from transport: Decisive problems in developing ecologically efficient transport systems. *J. Vibroeng.* **2017**, *19*, 5639–5655. [[CrossRef](#)]
9. Jacyna, M.; Żochowska, R.; Sobota, A.; Wasiak, M. Scenario Analyses of Exhaust Emissions Reduction through the Introduction of Electric Vehicles into the City. *Energies* **2021**, *14*, 2030. [[CrossRef](#)]
10. Pamula, T. Road Traffic Conditions Classification Based on Multilevel Filtering of Image Content Using Convolutional Neural Networks. *IEEE Intell. Transp. Syst. Mag.* **2018**, *10*, 11–21. [[CrossRef](#)]
11. Pamuła, T.; Żochowska, R. Estimation and prediction of the OD matrix in uncongested urban road network based on traffic flows using deep learning. *Eng. Appl. Artif. Intell.* **2023**, *117 Pt A*, 105550. [[CrossRef](#)]
12. Wang, Y.; Zhang, D.; Liu, Y.; Dai, B.; Lee, L.H. Enhancing transportation systems via deep learning: A survey. *Transp. Res. Part C* **2019**, *99*, 144–163. [[CrossRef](#)]
13. Li, J.; Guo, F.; Sivakumar, A.; Dong, Y.; Krishnan, R. Transferability improvement in short-term traffic prediction using stacked LSTM network. *Transp. Res. Part C* **2021**, *124*, 102977. [[CrossRef](#)]
14. Wang, T.; Hussain, A.; Sun, Q.; Li, S.E.; Cao, J. Congestion Prediction of Urban Road Traffic by Using Deep Stacked LSTM Network. *IEEE Intell. Transp. Syst. Mag.* **2021**, *14*, 102–120. [[CrossRef](#)]
15. Yin, X.; Wu, G.; Wei, J.; Shen, Y.; Qi, H.; Yin, B. Deep Learning on Traffic Prediction: Methods, Analysis and Future Directions. *IEEE Trans. Intell. Transp. Syst.* **2021**, *23*, 4927–4943. [[CrossRef](#)]
16. Wang, Q.; Yuan, Y.; Zhang, Q.; Yang, X.T. Signalized arterial origin-destination flow estimation using flawed vehicle trajectories: A self-supervised learning approach without ground truth. *Transp. Res. Part C Emerg. Technol.* **2022**, *145*, 103917. [[CrossRef](#)]
17. Antoniou, C.; Ben-Akiva, M.; Koutsopoulos, H.N. Dynamic traffic demand prediction using conventional and emerging data sources. *IEEE Proc.—Intell. Transp. Syst.* **2006**, *153*, 97–104. [[CrossRef](#)]
18. Zhou, X.; Mahmassani, H.S. Dynamic origin-destination demand estimation using automatic vehicle identification data. *IEEE Trans. Intell. Transp. Syst.* **2006**, *7*, 105–114. [[CrossRef](#)]
19. Liu, J.; Zheng, F.; Henk; van Zuylen, J.; Li, J. A dynamic OD prediction approach for urban networks based on automatic number plate recognition data. *Transp. Res. Procedia* **2020**, *47*, 601–608. [[CrossRef](#)]
20. Sánchez-Cambronero, S.; Álvarez-Bazo, F.; Rivas, A.; Gallego, I. Dynamic Route Flow Estimation in Road Networks Using Data from Automatic Number of Plate Recognition Sensors. *Sustainability* **2021**, *13*, 4430. [[CrossRef](#)]
21. Zin, T.A.; Lwin, K.K.; Sekimoto, Y. Estimation of originating-destination trips in Yangon by using big data source. *J. Disast. Res.* **2018**, *13*, 6–13. [[CrossRef](#)]
22. Tolouei, R.; Psarras, S.; Prince, R. Origin-destination trip matrix development: Conventional methods versus mobile phone data. *Transport. Res. Proc.* **2017**, *26*, 39–52. [[CrossRef](#)]
23. Alexander, L.; Jiang, S.; Murga, M.; González, M.C. Origin-destination trips by purpose and time of day inferred from mobile phone data. *Transp. Res. Part C Emerg. Technol.* **2015**, *58 Pt B*, 240–250. [[CrossRef](#)]
24. Liu, Z.; Liu, Z.; Fu, X. Dynamic Origin-Destination Flow Prediction Using Spatial-Temporal Graph Convolution Network with Mobile Phone Data. *IEEE Intell. Transp. Syst. Mag.* **2021**, *14*, 147–161. [[CrossRef](#)]
25. Nigro, M.; Cipriani, E.; del Giudice, A. Exploiting floating car data for time-dependent origin-destination matrices estimation. *J. Intell. Transport. Syst. Technol. Plan. Oper.* **2018**, *22*, 159–174. [[CrossRef](#)]
26. Ceccato, R.; Gecchele, G.; Rossi, R.; Gastaldi, M. Cost-effectiveness analysis of Origin-Destination matrices estimation using Floating Car Data. *Experimental results from two real cases. Transp. Res. Procedia* **2022**, *62*, 541–548. [[CrossRef](#)]
27. Vogt, S.; Fourati, W.; Schendzielorz, T.; Friedrich, B. Estimation of origin-destination matrices by fusing detector data and Floating Car Data. *Transp. Res. Procedia* **2019**, *37*, 473–480. [[CrossRef](#)]
28. Krishnakumari, P.; van Lint, H.; Djukic, T.; Cats, O. A data driven method for OD matrix estimation. *Transp. Res. Part C* **2020**, *113*, 38–56. [[CrossRef](#)]
29. Yang, X.; Lu, Y.; Hao, W. Origin-destination estimation using probe vehicle trajectory and link counts. *J. Adv. Transport.* **2017**, *2017*, 4341532. [[CrossRef](#)]
30. Hussain, E.; Bhaskar, A.; Chung, E. Transit OD matrix estimation using smartcard data: Recent developments and future research challenges. *Transp. Res. Part C Emerg. Technol.* **2021**, *125*, 103044. [[CrossRef](#)]
31. Brzeziński, A.; Dybicz, T. Possibility of Big Data application for OD-matrix calibration in transport demand models. *Arch. Civ. Eng.* **2021**, *67*, 215–232. [[CrossRef](#)]

32. Savrasovs, M.; Pticina, I. Methodology of OD Matrix Estimation Based on Video Recordings and Traffic Counts. *Procedia Eng.* **2017**, *178*, 289–297. [[CrossRef](#)]
33. He, Y.; Zhao, Y.; Tsui, K.-L. Short-term forecasting of origin-destination matrix in transit system via a deep learning approach. *Transp. A Transp. Sci.* **2023**, *19*, 2033348. [[CrossRef](#)]
34. Zhang, J.; Che, H.; Chen, F.; Ma, W.; He, Z. Short-term origin-destination demand prediction in urban rail transit systems: A channel-wise attentive split-convolutional neural network method. *Transp. Res. Part C* **2021**, *124*, 102928. [[CrossRef](#)]
35. Ke, J.; Qin, X.; Yang, H.; Zheng, Z.; Zhu, Z.; Ye, J. Predicting origin-destination ride-sourcing demand with a spatio-temporal encoder-decoder residual multi-graph convolutional network. *Transp. Res. Part C* **2021**, *122*, 102858. [[CrossRef](#)]
36. Zheng, H.; Chen, J.; Huang, Z.; Yang, K.; Zhu, J. Short-Term Online Forecasting for Passenger Origin–Destination (OD) Flows of Urban Rail Transit: A Graph–Temporal Fused Deep Learning Method. *Mathematics* **2022**, *10*, 3664. [[CrossRef](#)]
37. Cao, Y.; Hou, X.; Chen, N. Short-Term Forecast of OD Passenger Flow Based on Ensemble Empirical Mode Decomposition. *Sustainability* **2022**, *14*, 8562. [[CrossRef](#)]
38. Khalesian, M.; Furno, A.; Leclercq, L. Improving deep-learning methods for area-based traffic demand prediction via hierarchical reconciliation. *Transp. Res. Part C Emerg. Technol.* **2024**, *159*, 104410. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.