Transportation Research Part C 141 (2022) 103739

Contents lists available at ScienceDirect

Transportation Research Part C

journal homepage: www.elsevier.com/locate/trc

Traffic volume prediction using aerial imagery and sparse data from road counts

Arman Ganji *, Mingqian Zhang, Marianne Hatzopoulou

Department of Civil and Mineral Engineering, University of Toronto, Toronto, Ontario, Canada

ARTICLE INFO

Keywords:
Google Aerial images
Vehicle detection
AADT
Traffic prediction
Pattern Recognition

ABSTRACT

Around the world, metropolitan areas invest in infrastructure for traffic data collection, albeit focusing on highway networks, thus limiting the amount of data available on inner-city roads. For this purpose, various modelling techniques have been developed to interpolate traffic counts spatially and temporally across an entire road network. However, the predictive accuracy of these models depends on the quality and coverage of traffic count data. In this study, we extend the power of spatio-temporal interpolation models with vehicle detection from aerial images, developing a new approach to estimate Annual Average Daily Traffic (AADT) across all roads in an urban area. Using Google aerial images, we extracted the number of vehicles on a road segment and treated these values as observed traffic counts collected over a short period of time. This information was used as input and merged with traffic count data at stations with longer record lengths to predict traffic on all urban roads. This approach was compared against a hold-out sample of roads with observed traffic count data and images, indicating an R-squared (R^2) = 90% and RMSE = 7675 between predicted and observed daily traffic counts and R^2 = 58% and RMSE = 18918 between observed and predicted AADT. The higher prediction accuracy for daily traffic indicates the power of the proposed method for predicting daily values from images; while the lower accuracy of AADT prediction stresses the need for longer-term data to achieve accurate annual averages based on counts derived from images.

1. Introduction

Many metropolitan areas around the world collect vehicle counts on highways and a select number of major roads, yet with insufficient coverage due to the high costs of permanent ground-based traffic monitoring stations (Gastaldi et al., 2014). In addition to sparse but long-term traffic count stations in urban areas, short-term counts (on the order of a few days) typically have wider coverage, albeit representing a shorter time frame. Short-term traffic counts are usually collected by municipalities in the context of specific projects (traffic impact assessment or safety analysis) and can extend across many locations on the road network. For this reason, different approaches for traffic count prediction have been proposed in the literature, mostly capitalizing on the use of short-term traffic counts, merged with road and land use information to predict Annual Average Daily Traffic (AADT) using a range of interpolation methods (Wang & Kockelman, 2009; Gastaldi et al., 2014; Bagheri et al., 2015; Ganji et al., 2020a). The challenge with such methods is that traffic fluctuations on short time scales as well as time-of-day, day-of-week, and seasonal patterns, often distort the accuracy of the AADT estimates (Ganji et al., 2020a). The US Federal Highway Administration recommends that a traffic density

* Corresponding author.
E-mail address: arman.ganji@utoronto.ca (A. Ganji).

https://doi.org/10.1016/j.trc.2022.103739
Received 11 April 2021; Received in revised form 3 March 2022; Accepted 22 May 2022
Available online 9 June 2022

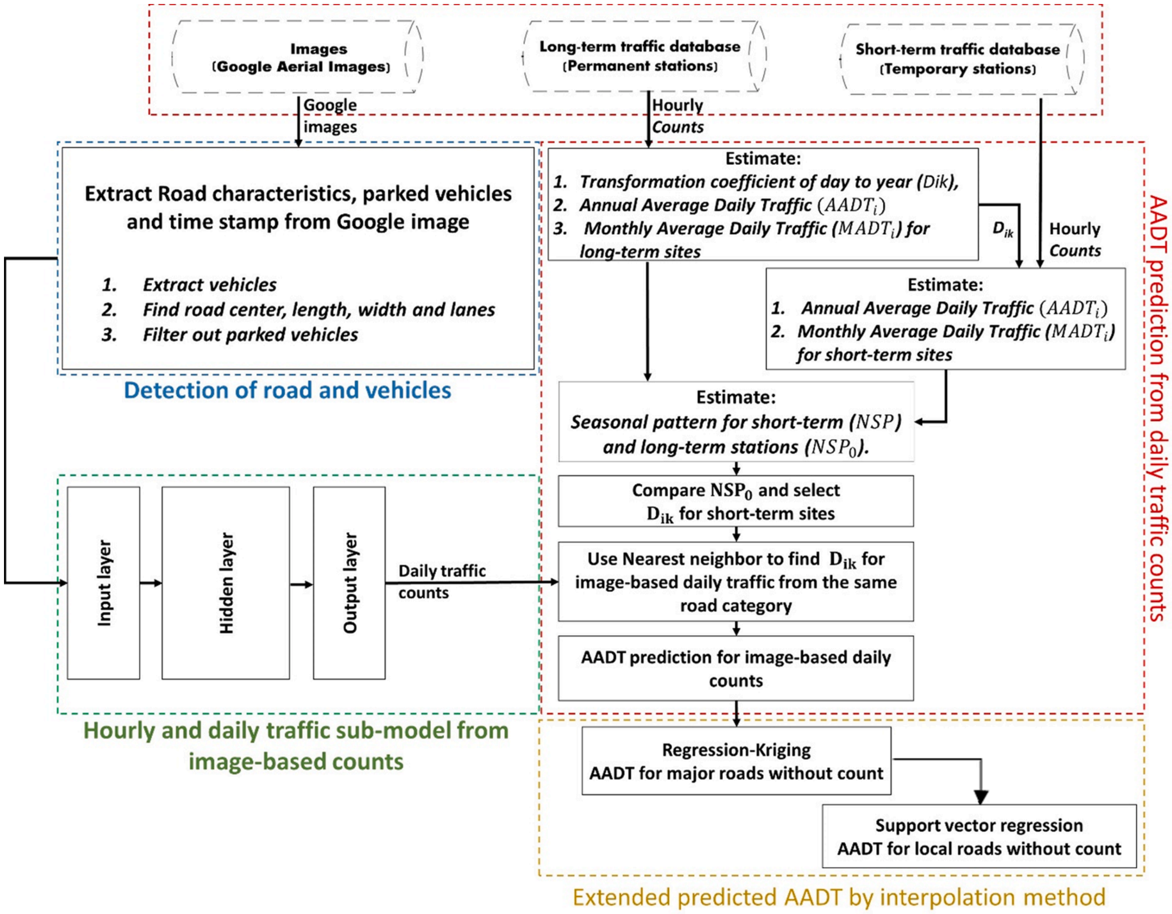

Fig. 1. Overall approach and sequence of data and modules, including image and traffic data (top), road and vehicle detection (upper left), neural networks for daily traffic prediction (upper right), transformation coefficients of daily to monthly traffic (lower left), AADT estimation on roads with no data nor images (lower right).

variation factor be applied to account for time-of-day, day-of-the-week, and monthly variations when using less-than-a-day counts to estimate AADT (US Federal Highway Administration, 2016). To overcome this challenge, in a few studies, the relationship between long-term and short-term traffic counts was captured (Bagheri et al., 2015; Kaack et al., 2019; Ganji et al., 2020a). Kaack et al. (2019) approximated the density variation factor from nearby regions where traffic was monitored. The authors estimated the factor as the conditional average of normalized hourly count values (the hourly count data divided by the mean of all hourly counts in the year). In Ganji et al. (2020a), a Traffic Emission Prediction scheme (TEPs) uses a long record of traffic counts to extend downstream daily traffic counts; a pattern recognition approach further identifies a coefficient to estimate AADT from daily values. This technique provides a unique basis to use short-term traffic counts (e.g., one day) for long-term average traffic prediction (AADT). Advances in such hybrid methods that rely on short and long-term traffic counts provide a promising avenue to improve traffic prediction models by harnessing all data collected in an urban network.

Alternative ways of expanding the database of short-term counts include using satellite imagery (Larsen et al., 2009), GPS data from cell phones, and aerial imagery. These techniques capitalize on the use of various sources of information that are generally more ubiquitous than costly traffic counters and cameras (Larsen et al., 2009). With image processing, many roads can be covered, especially those not equipped with traffic sensors (Larsen et al., 2009). While the application of this technique in traffic prediction has been limited due to the short-term nature of the traffic captured in an image (Cao et al., 2016), the detection of vehicles in images has attracted extensive research focus in the past years (Hinz, 2004; Lenhart et al., 2008; Holt et al., 2009; Kozempel and Reulke, 2009; Larsen et al., 2009; Kembhavi et al., 2010; Li et al., 2021). Among these vehicle detection methods, Larsen et al. (2009) constructed a vehicle shadow mask to provide information on the type of vehicles and traffic counts. The proposed method overcomes the explicit segmentation methods, which may fail to capture vehicles of very low contrast to the local background, especially when the low contrast segments are only slightly brighter than the road. Recent progress in object detection in aerial images benefited largely from region-based convolutional neural networks (R-CNN) (Girshick et al., 2014; Girshick, 2015; Ren et al., 2015; Cheng et al., 2016; Deng et al., 2017; Long et al., 2017). The literature's most successful aerial-based object detection research is based on the region proposal method (Uijlings et al., 2013) and R-CNNs (Girshick et al., 2014) that use bounding boxes as regions of interest (RoIs) and then rely on region features for category identification (Deng et al., 2017). Region proposal methods and R-CNNs have already been successfully applied to detect vehicles in satellite images (Chen et al., 2014; Jiang et al., 2015; Deng et al., 2017).

In this paper, we propose to use image-based vehicle counts and long-term traffic data to generate AADT for all roads in an urban area. This approach overcomes the weakness of using solely satellite or aerial images for traffic count prediction due to short-term traffic fluctuations. Our image processing method is faster than remote sensing methods and uses Google aerial images instead of high-resolution satellite images, which are more computationally intensive (Bridgelall et al., 2016, Bowen et al., 2004). Google aerial images are generally taken at an appropriate resolution over time, making archiving possible (Ganji et al., 2020b). We also take advantage of recent advances in R-CNNs for vehicle detection in images. We propose a new method for traffic volume interpolation, treating vehicles detected in Google aerial images as short-term traffic counts and establishing relationships with long-term traffic count stations to correct the short-term prediction and estimate AADT. We also developed a method to extract road characteristics (width and directionality) from images. Compared to previous studies, this paper proposes innovations along multiple dimensions. We developed a novel technique to extract nearby traffic density factors (i.e., daily to monthly and yearly volumes) for roads that lack any traffic data. This is important since the counts derived from aerial images (which we use to complement other short-term counts) must use traffic density factors aligned with those of nearby roads (of the same category) to generate unbiased AADT values. In previous studies, when no information about traffic variation was available, the density factors were extracted from any available monitoring site. In this study, a monthly/seasonal traffic pattern is extracted for all roads with traffic counts to identify appropriate density factors for roads that only have aerial images. A nearest neighbor (NN) algorithm is used to extract the similarity in diurnal/seasonal traffic patterns. Furthermore, our proposed method can detect vehicles on each lane of road as well as differentiate between vehicles on the road and parked vehicles. This is a departure from previous studies (e.g. Kaack et al., 2019) which estimate the sum of ADDT in two directions. Our approach also includes a method to generate a time stamp for the images based on the shadow length and image coordinates. Finally, we use Kriging as a complementary prediction tool to generate predictions on roads that do not have images nor count data and hence complete the traffic predictions across the entire road network.

This new approach was applied to traffic volume prediction in Toronto, which collects traffic data using different technologies: Short Period Traffic Counts (SPTCs), Turning Movement Counts (TMCs), and permanent traffic counts stations. The predicted traffic volumes and AADT values were validated against traffic counts collected over the entire city using a hold-out sample.

2. Methodology

The proposed modelling framework (Fig. 1) consists of a set of mathematical and statistical models presented in the following sections. The framework is driven by three different data inputs, including aerial images, long-term, and short-term traffic counts. The image processing module extracts the road characteristics from images, including the road width and direction and subsequently detects the number of vehicles on the road. Then, a neural network approach generates hourly and daily traffic volumes for each road based on vehicles extracted from images and road characteristics. To convert the hourly and daily image-based traffic counts to AADT, considering the traffic fluctuations across hours of the day and weeks, a nearest neighbor approach is used to generate coefficients of daily to yearly traffic volumes. These coefficients are then used to estimate AADT for roads with aerial images and/or short-term traffic counts. Finally, Regression Kriging and Support Vector Regression are used to extend the predicted AADT to those roads that do not have any aerial images (poor quality or image showing no traffic) nor traffic counts.

2.1. Detection of vehicle and road characteristics

2.1.1. Aerial images

We downloaded Google aerial images with the associated metadata at 223 locations in Toronto. These locations correspond to the locations of permanent road counters in the city. The image size is 2048 × 2048 patches at zoom level 19 at scale 2, corresponding to 4096 × 4096 pixels, and 72.24 × 72.24 cm. At this zoom level and assuming an average latitude of 43.6532 for Toronto and earth's radius equal to 6,378,137 m, each image captures an area close to 880 × 880 m.

2.1.2. Vehicle detection from aerial images

We adopted a region-based convolutional neural network (R-CNN) that was trained by Ding et al. (2018) using more than 2806 aerial images to detect 15 common object categories, including large and small vehicles (https://github.com/dingjiansw101/AerialDetection). This large dataset for object detection in aerial images, known as DOTA, has lower biases when dealing with Google aerial images since it was collected from multiple sensors and platforms with multiple resolutions. The model was pre-trained with the DOTA database with mean Average Precision (mAP) of 0.808 and 0.798 for the horizontal and oriented object detections, respectively, as cited in the DOTA website (https://captain-whu.github.io/DOTA/results.html). The expanded DOTA is the largest dataset for object detection in Earth vision, to the best of our knowledge. The OBB annotations of DOTA provide a large-scale benchmark for object detection in Earth vision and pose interesting algorithmic questions and challenges to generalized object detection in computer vision.

R-CNNs are a branch of CNNs, developed by LeCun et al. (1998) based on Artificial Neural Networks (ANNs). Compared to Artificial Neural Networks (ANNs), CNNs have a different architecture, which consists of layers including convolutional layers (that learn the convolutions and provide the best performance for each data category), pooling layers (govern overfitting, allow for stable conversion), and the rectified linear unit (enhances the nonlinear properties of the network). Furthermore, CNN input data are images and/or are interpreted as images, significantly reducing the number of parameters and resulting in more rapid processing. R-CNNs follow the

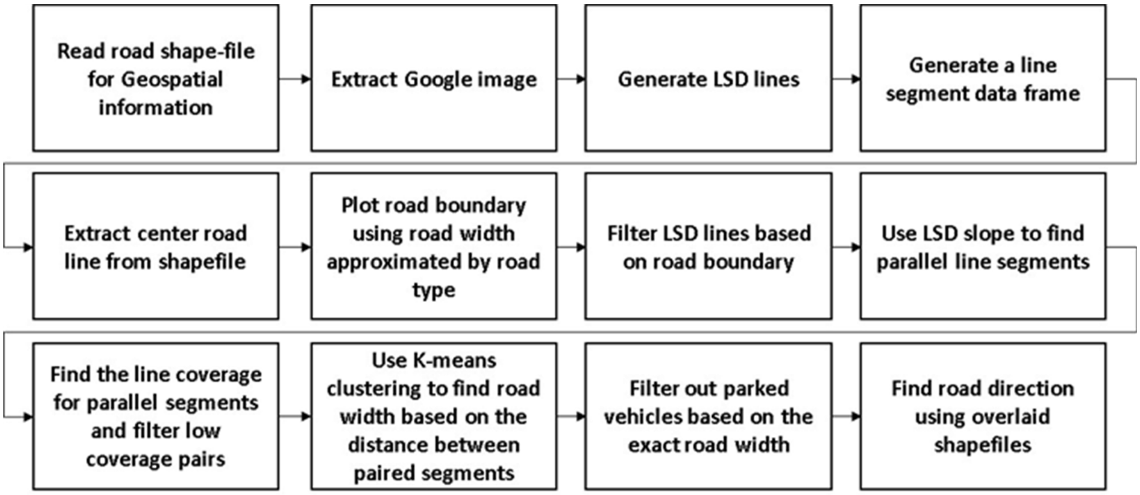

Fig. 2. Steps for detection of road characteristics.

region proposals method. This is a candidate detection set available to the detector. In contrast to CNNs that run the sliding windows over an entire image, R-CNNs select just a few windows or regions of an image.

In this study, the proposed R-CNN takes a Google aerial image and a set of object proposals as inputs. R-CNN initially starts with processing the entire Google aerial image using convolutional and max-pooling layers (calculate the maximum value for each feature patch) to produce a convolutional feature map. Given the feature map, a fixed-length feature vector is extracted from each object proposal using a region of interest (RoI) pooling layer. Assuming **H** and **W** as layer hyper-parameters, the RoI pooling later converts the features inside each RoI into a small feature vector with a fixed spatial extent of $\mathbf{H} \times \mathbf{W}$. These feature vectors are inputs to a sequence of fully connected (**fc**) layers. The model has two output layers that produce four real-valued numbers for each K object class. Another one is the **softmax** probability estimates over **K** object classes plus a catch-all "background" class. Softmax classifies an object with probabilistic values between 0 and 1.

We also computed a mean Average Precision (mAP) index (ranging between 0 and 1) using 100 random images extracted for the city of Toronto. The comparison was made against manual counting conducted by two separate research assistants. This test captures the number of vehicles on the road and therefore reflects the performance of the road detection and vehicle detection modules. Details and results of this method are provided in Appendix II.

2.1.3. Road detection and extraction of road characteristics

Road characteristics (road width, road type, lane numbers, and speed limit) are key factors to transfer the number of detected vehicles in the image to short-term traffic counts (e.g., hourly/daily traffic counts) since road characteristics provide information on road capacity. Our proposed approach for detecting road characteristics from aerial images is presented in Fig. 2. We initially extracted line segments and road centerlines in the Google aerial images. The Line Segment Detector (LSD) (Burns et al., 1984, Xia et al., 2017) was used to detect the centerline (Von Gioi et al., 2012). It is a self-control false detection algorithm (one false alarm is allowed per image during the detection) that can be used for any digital image and does not require tuning. The LSD technique works based on color gradient. For a road segment, the centerline is typically comprised of two parallel gray and dark portions, with the gray level changing from dark to light. The border between the dark and gray portions is called the level line. The level line and gray color gradient are the critical parameters for LSD analysis. LSD initially computes the level-line angle at each pixel of an image and produces a set of unit vector fields tangent to the level line, called the level-line field. Then, all pixels with the same level-line field angle are grouped into an individual region of pixels. This individual region is called the line support region, which can be determined as a line segment. To validate the detected lines, pixels that have the level-line angle corresponding to the angle of the line are counted. Then, a statistical test based on the contrario approach and the Helmholtz principle (Desolneux et al., 2000) was used to identify whether a detected feature had a sufficient number of aligned points to be considered a line. This procedure provides several lines for each road lane that are mostly parallel. Parallel lines detection is necessary for lane area detection, which is a key factor in filtering out parked vehicles from vehicles on a road lane and vehicles on other lanes. As a result, the slope and intercept are used to pair parallel lines.

To identify parallel lines, we also used road shapefiles to filter out the detected parallel lines of a lane. For this purpose, an initial lane width was extracted from road shapefiles. Then, a K-means clustering technique was used to cluster parallel lines. Finally, we assigned the width range of a cluster to each pixel on road centerlines. This width range is viewed as the width of the corresponding lane segment. This method is designed to work on any digital image without parameter tuning.

2.1.4. Timestamp of aerial images

Considering that traffic counts fluctuate over different times of day, the date and time of an image are essential for transferring the image-detected vehicles to short-term traffic counts. Google aerial images lack the exact date and time of images. For this reason, we

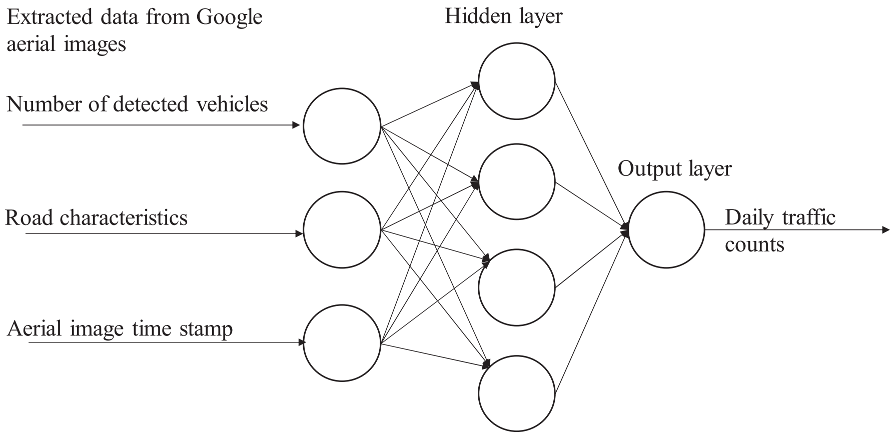

Fig. 3. Schematic diagram of the hourly and daily traffic sub-model.

identified a time and date for each image based on shadows in Google images and information about the sun movement and sunlight phases for specific coordinates and features on aerial images. Sandnes, (2011) used a similar method. Information about the sun's movement and sunlight can be extracted from applications such as Sun Surveyor Lite (https://www.sunsurveyor.com/), LightTrac application (https://www.lighttracapp.com/), or SunCalc (https://suncalc.net/), which use sunlight tracking. These applications use Global Positioning Systems (GPS), the camera's position, and the sun information to provide sunlight direction in a given area. We used SunCalc, which converts the time to the sun's angle above the horizon and azimuth. Using this application to estimate angle, we used Google image shadows to find the closest match in the SunCalc outputs. For this, several object shadows were represented inside a study area using a path from the base of objects to the end of their shadows. The height of objects was also extracted from a Lidar data layer for the City of Toronto and used to estimate the sun's angle.

2.2. Prediction of traffic counts using images

The number of detected vehicles and extracted road characteristics from Google aerial images were used to predict AADT in three different stages. In the first step, municipal traffic count data were prepared as a basis for model development and validation. In the second step, neural network models were trained to predict hourly and daily traffic counts based on counts derived from aerial images. The target values for neural network analysis were the observed traffic counts collected by the City of Toronto. In the third step, a novel approach was developed to generate AADT using neural network model outputs (daily traffic count) as well as traffic counts of the most similar nearby road. This approach generates and uses the nearby transformation coefficient of daily to yearly traffic counts to predict AADT on any road with a high-quality aerial image.

Step 1. Municipal traffic count database.

The City of Toronto manages an extensive traffic monitoring network, with collected traffic counts since 1994. We used data from 223 sites that contain long records of 15-min traffic counts in two different directions. These 223 sites have been collecting data from 2006 to 2016. Traffic counts were extracted at each of the 223 stations for the same time that the only Google aerial image was captured. Small missing periods of traffic counts were imputed using interpolation techniques (less than 1-hr missing interval). Also, for the permanent stations, two trend analysis methods were used to detect and include trends in traffic volume modelling. In addition, the entire database of permanent and temporary traffic count stations was used to estimate a transformation coefficient of daily to yearly counts. Detail on the traffic count program for the City of Toronto can be found in Ganji et al. (2020a). The basic assumption about the automated traffic count equipment is that automatic equipment can collect accurate 48-hour volumes. Equipment error introduces bias which is not affected by sample size. Since we assume that equipment bias is normally distributed with a zero mean, no adjustment was conducted.

Step 2. Hourly and daily traffic sub-model

We developed a neural network model (Fig. 3) to predict hourly and daily traffic counts based on counts extracted from images. The model has one hidden layer with four neurons and an output layer with a single node representing traffic counts. The activation function of the hidden layers, the number of the hidden nodes, and the training epoch numbers were optimized based on the RMSE and the coefficient of determination between the observed (daily and hourly traffic counts) and predicted values. Daily traffic counts (target values) were estimated based on images extracted at 223 locations, coinciding with the 223 permanent traffic count stations. Neural network predictors include the number of detected vehicles, road characteristics, and aerial image timestamp extracted directly from aerial images. The activation function was set to a hyperbolic tangent sigmoid function, leading to the best model performance.

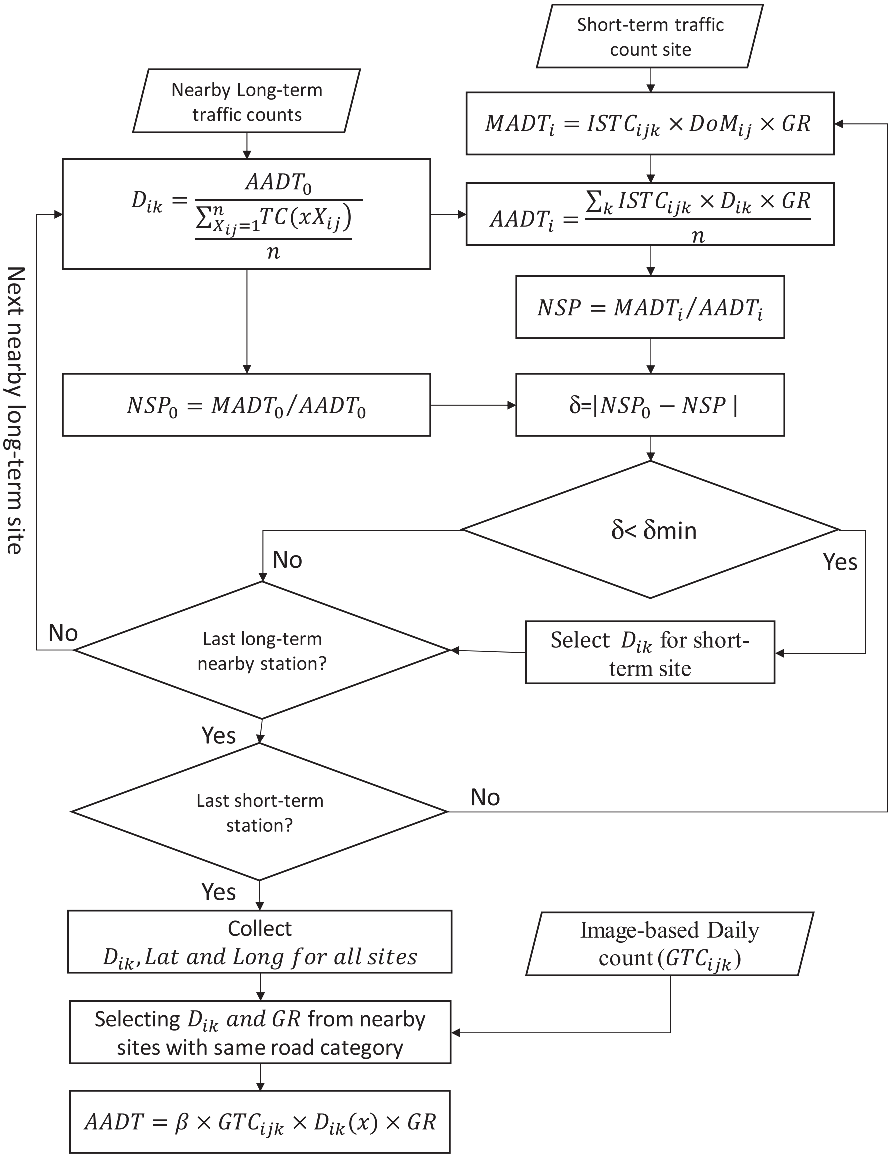

Fig. 4. Schematic diagram to estimate day to year (D_{ik}) transformation coefficient. *MADT*: Mean Annual Daily Traffic, $ISTC_{ijk}$: the available traffic counts of a short-term site at a specific date (i.e., at day ***i***, month ***j*** and year ***k***), *DoM*: the day to month transformation coefficient, *GR*: growth rate, $AADT_i$: Annual Average Daily Traffic, $MADT_i$: Monthly Average Daily Traffic, β: correction factor, GTC_{ijk}: image-based traffic count at day ***i***, month ***j***, and year ***k***, *NSP* normalized seasonal patterns.

This also proved to be the best in depicting the nonlinearity of the modeled natural system, compared with linear and log sigmoid functions.

Step 3. AADT prediction from daily traffic counts

Short-term traffic counts and image-based traffic counts cannot be directly converted to AADT due to traffic fluctuations across different hours of the day and days of the week. Such problems have been addressed in the literature using pattern recognition techniques (Bagheri et al., 2015; Ganji et al., 2020a) that estimate a daily to yearly transformation coefficient. Ganji et al. (2020a) developed state-of-the-art statistical and mathematical models describing the spatial relationship of traffic counts over a network and transferring compiled and extended information from upstream traffic stations to downstream ones. Following Ganji et al. (2020a), we use a Pattern Recognition Traffic Counts (PRTC) approach, which uses information from nearby traffic count sites to estimate AADT using daily traffic counts predicted from images. Using PRTC, AADT is estimated for each road with a daily traffic count based on image-detected vehicles. Pattern recognition approaches typically rely on supervised and unsupervised classifications. In contrast to a supervised classification that uses predefined classes to pool the patterns, unsupervised classification, such as clustering, pools patterns into unknown classes. The nearest neighbor (NN) algorithm is an example of a supervised classification method, which can be applied to find the similarity between a set of points in a multi-dimensional space, such as traffic count patterns. Cai et al., (2016) used the NN algorithm to generate forecasts of short-term traffic. In this study, we use a hierarchical supervised NN method to predict AADT from short-term traffic counts. Fig. 4. presents the algorithm. In a first step, the seasonal pattern of each permanent station is computed for different years. The data from permanent sites for different years are converted to the study year using a growth rate. In a second step, the seasonal pattern of the site with short-term traffic counts is obtained by converting the short-term count collected on a specific day of the week and month to its equivalent monthly volume. The seasonal pattern of the short-term site is then compared with the one from nearby permanent stations, one at a time, until a minimum difference between the short-term seasonal pattern and the permanent station seasonal pattern is reached. At this point, the permanent station day to year transformation coefficient is assigned to the short-term site.

As indicated in Fig. 4, our approach estimates a transformation coefficient of day to year (D_{ik}) using Equation (1) at nearby permanent sites surrounding each short-term count location.

$$D_{ik} = \frac{AADT_0}{\frac{\sum_{X_{ij}=1}^{n} TC(xX_{ij})}{n}} \forall X_{ij} \in \{(i,j)|i = 1, \cdots, 7; j = 1, \cdots, 12\}; n \leq (nyear \times 4) \tag{1}$$

In Eq. (1), $AADT_0$ is the Annual Average Daily traffic at a permanent station, n represents the number of historical traffic counts (X) at day i and month j over a historical data horizon ($nyear$).

In addition to D_{ik} of nearby permanent sites, following the method proposed by Ganji et al. (2020a) and as indicated in Fig. 4, all available historical counts were used to estimate a seasonal pattern for both short-term (NSP) and permanent stations (NSP_0). For this, an initial approximation of Annual Average Daily Traffic ($AADT_i$) and Monthly Average Daily Traffic ($MADT_i$) are estimated using D_{ik} at nearby permanent sites, and the available traffic counts of a short-term site at a specific date (i.e., $ISTC_{ijk}$ at day i, month j and year k), the day-to-month transformation coefficient (DoM), and growth rate (GR). Comparing the normalized seasonal patterns between short-term and permanent stations ($NSP_0 - NSP$), a short-term station is assigned to a permanent station to estimate GR and D_{ik} from the nearby permanent site.

The growth rate or GR is typically estimated using two methods: a least-squares approach and an approximation method proposed by the World Bank (World Bank 2018). Then, given the known D_{ik} values for the permanent and short-term stations and assuming x as a road with Google image-based vehicle counts, the nearest neighbour technique searches the nearby roads of the same classification with known D_{ik} to identify $D_{ik}(x)$ for road x. Finally, the AADT for road x with image-based vehicle count is estimated using Eq. (2).

$$AADT = \beta^* GTC_{ijk}{}^* D_{ij}(x) \times GR \tag{2}$$

In Eq. (2), GTC_{ijk} refers to an image-based traffic count at day i, month j, and year k, GR represents the traffic growth rate for the nearest stations of the same classification based on the nearest neighbor technique. A correction factor (β) is estimated using predicted and observed AADT at all permanent stations (Eq. (3)).

$$\beta = \frac{\sum_{m=1}^{N} \left[GTC_{ijk}{}^* D_{ik}\right]_m}{\sum_{m=1}^{N} [AADT_0]_m} \tag{3}$$

In Eq. (3), N is the number of permanent stations and $AADT_m$ represents the AADT at a permanent station which results from the long-term average of daily traffic counts. This technique is calibrated using the permanent stations to estimate a correction factor (β) and ensure acceptable performance. A long-term hourly average traffic count is used for each permanent site to predict AADT based on the nearby permanent stations, using Eq. (2). Then, Eq. (3) was used to find β. For model validation, the predicted AADTs based on Google aerial images at the permanent stations were also compared with those AADTs that were extracted directly from the permanent traffic count sites. After validation, the D_{ik} and growth rate (GR) were estimated for all count stations with short-term/long-term traffic data and roads with images (last three steps in Fig. 4).

2.3. Capturing the value-added of aerial images

The use of permanent and short-term traffic data collected by the city helps generate daily to yearly factors that were used to convert the daily traffic counts predicted using neural networks based on images into AADT values. While this method generates predictions on all roads where aerial images are available, there is also a need to generate predictions on roads for which high-quality aerial images do not exist. Due to the generally low volumes on these roads, many local roads may also have images showing no vehicles at all. The availability of data from traffic count stations and images treated as short-term counts provides an opportunity to generate predictions on roads without counts nor images, using approaches such as regression kriging (RK) and support vector regression (SVR). For this, aerial images were used to predict daily traffic and AADT on more than 17,800 roads in Toronto that do not have any traffic count value. In turn, RK and SVR use these predictions to estimate AADT on roads where image-based traffic counts nor any observed counts are available. RK was used to predict AADT for major roads and highways, and SVR was used for most local roads. Ganji et al. (2020a) demonstrated that SVR is the most suitable tool for traffic prediction on local roads. SVR and RK modules are discussed in Appendix I and Ganji et al. (2020a).

Since RK and SVR allow for predictions to be made on roads that do not have an image nor count data; we used this opportunity to

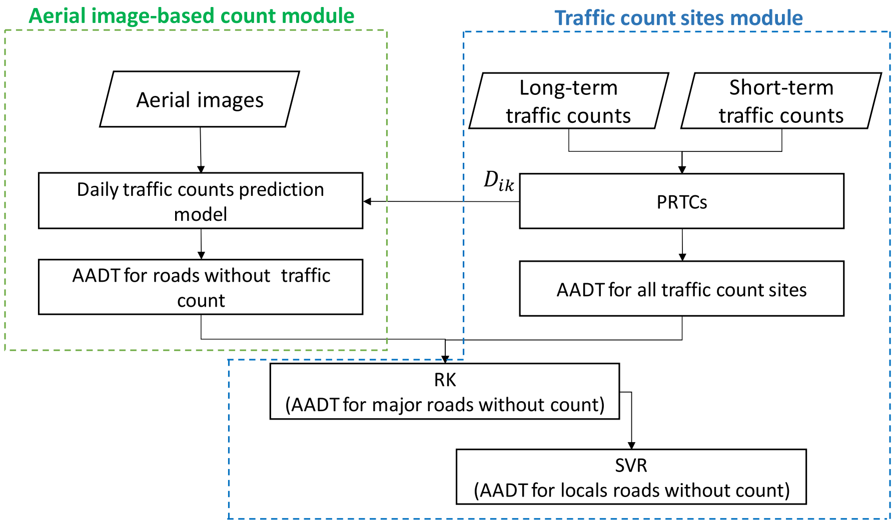

Fig. 5. Schematic diagram to estimate Average Annual Daily Traffic (AADT) with and without aerial images. D_{ik}: transformation coefficient of day-to-year, *PRTC*: Pattern Recognition Traffic Counts, *RK*: regression Kriging, *SVR*: Support Vector Regression.

Fig. 6. Locations of selected traffic monitoring stations (dots on the map).

test the "value-added" of aerial images by using RK and SVR to generate predictions using the City of Toronto traffic count database only and compare the predictions with the approach of using the City of Toronto traffic count database augmented by the traffic counts derived by the images. This test is designed to demonstrate the potential of using aerial images to augment the data collected by cities while treating the counts extracted from images as short-term traffic counts. Fig. 5 summarizes the approach adopted to "augment" the city count database with images. Vehicles were detected in images and treated as short-term traffic counts; daily counts were obtained by associating these short-term counts with actual count data collected by the city. Then using the entire traffic count database, day to year coefficients were assigned to these images to predict AADT. This allows for AADT prediction on any road with an image. Using RK and SVR, the traffic count database collected by the city was used to generate predictions on roads that do not include any counts. This was done using two methods: with only the city collected data or by augmenting the city data with predictions at locations with aerial images. As a result, the city data and the image data become a large database that can be used to generate AADT predictions across the entire road network.

Finally, we conducted two analyses that entail sub-samples to assess the benefit of the images. In these cases, we assume that counts are unavailable on a specific set of roads and test the output of the models that rely on counts from images, using the counts on the roads as ground truth. The first test evaluates the value of image-based counts for the reproduction of daily traffic counts for roads with both short-term and long-term records. In the second test, we evaluated image-based counts in the prediction of AADT on roads with short-term and long-term traffic counts. In this case, note that the AADT that is considered as ground truth may not be the actual count but the output of PRTC model. This is especially the case for roads with short-term records.

3. Results

This study first extracted aerial images at 223 traffic count stations selected from the City of Toronto database for this analysis.

Fig. 7. Detected vehicles using region-based Convolutional Neural Network (R-CNN) model for different types of built environments. The detected vehicles are highlighted using a rectangular border.

These stations monitor traffic in two directions and include more than 20 years of data collection. Fig. 6 presents the distribution of these stations across the city. Vehicles were detected in each aerial image and paired with the hourly and daily traffic volume extracted from each city station for the same day and time as the image.

3.1. Detected vehicles and road characteristics

Fig. 7 presents examples of vehicle detection using R-CNN with RoI transformers under different conditions. The parked vehicles in Fig. 7 were detected by including road characteristics such as road width. The mAP value for R-CNN is 0.71, which is close to what has been proposed by Ding et al and cited by the DOTA website for this methodology (mAPs of 0.808 and 0.798 for the horizontal and oriented object detections, respectively). Fig. 7 shows that the vehicle detection model had an acceptable performance in different situations such as dense residential areas with nearby trees, local roads, parking, and near high-rise buildings with extended shadows and during congestion and fast-moving vehicles on highways.

Fig. 8 presents the steps of identifying the road centerline, edges, and direction of traffic. Fig. 8a and 8b present the Google aerial images as well as level-line and road borders after LSD analysis. Fig. 8c presents a road shapefile that was used to filter out undesirable segments. The width of the road in the shapefile represents the possible search area for parallel border lines, which were estimated using the geospatial information of the shapefile. Fig. 8d presents the parallel lines that remained after filtering out the low coverage parallel lines. These parallel lines help detect the lanes as well as on-road vehicles as indicated in Fig. 8e. Fig. 8f identifies the road direction that has been extracted from the road shapefile. As a result of these steps, all vehicles in the different lanes and directions are extracted and the road widths are also identified.

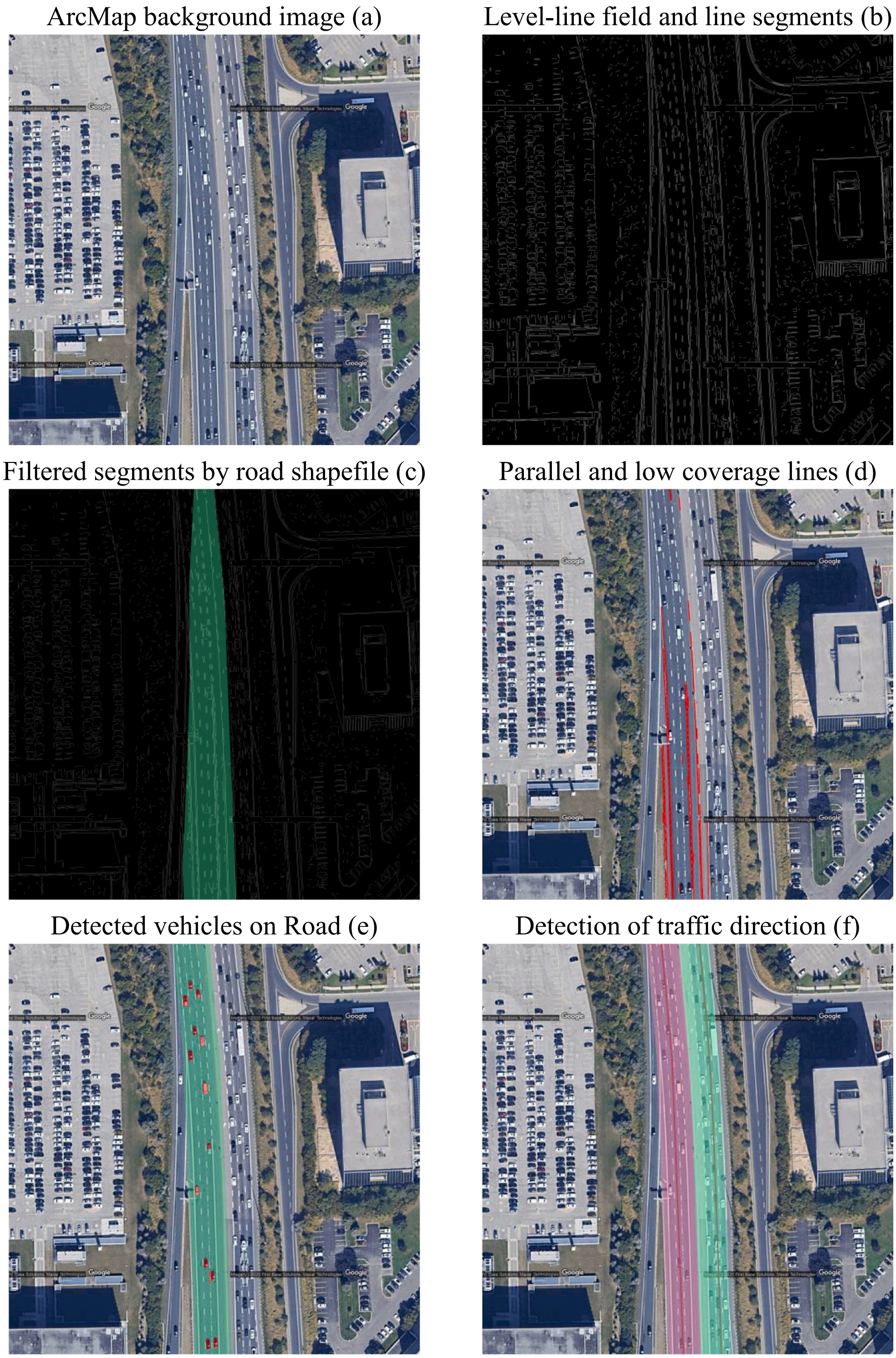

Fig. 8. Road and vehicle detection steps include (a) image extraction, (b) level-line field and line segment extraction, (c) filtering the segments by road shapefile, (d) identifying the parallel and low coverage lines, (e) vehicles detection on roads and (f) traffic direction.

We also calculated the mAP using 100 random images extracted for the city of Toronto. The comparison was made against manual counting conducted by two separate research assistants. The mAP values for all detected vehicles are around 85%. Details and results of this method are provided in Appendix II.

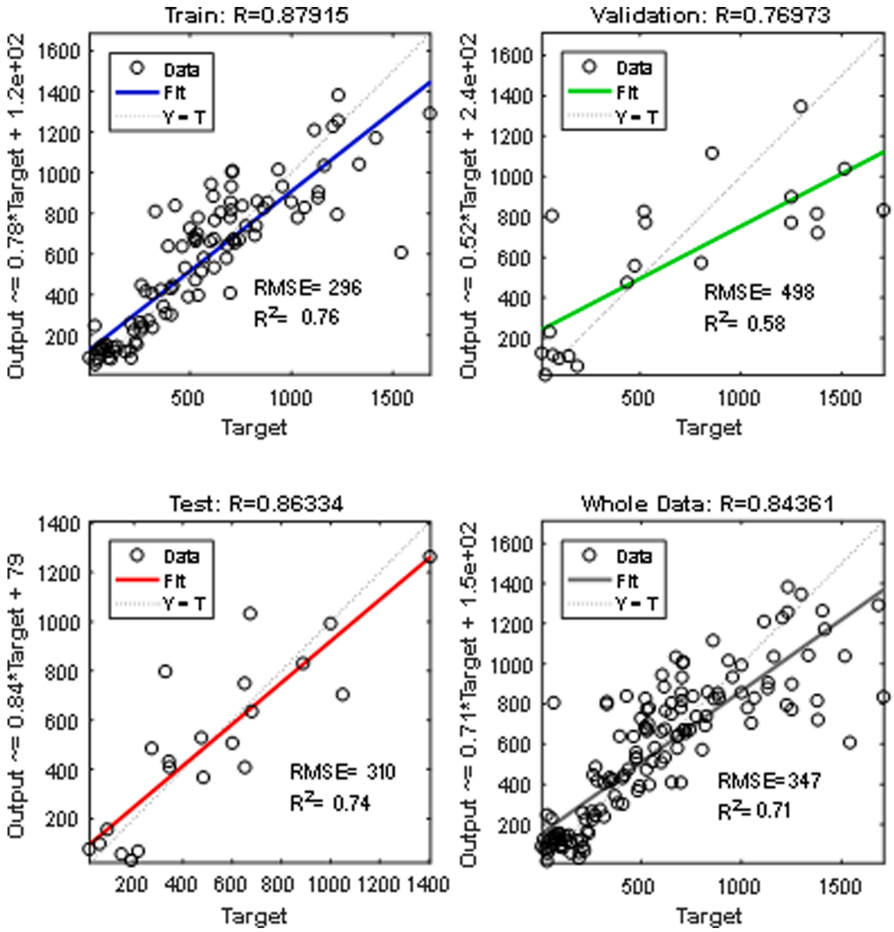

Fig. 9. Artificial Neural Network (ANN) for hourly traffic count prediction: Target represents the desired ANN response; Output represents the neural network response, which guides the learning process involving the changes in weights.

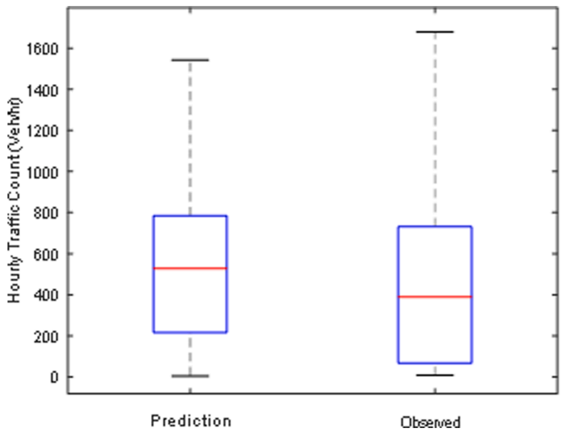

Fig. 10. Comparing the predicted and observed hourly traffic count distributions using Artificial Neural Network (ANN).

3.2. Prediction of traffic counts

3.2.1. Hourly and daily traffic sub-model

Fig. 9 presents the results for neural network training, validation, and testing. Using the number of vehicles detected in each image, the target values are the hourly traffic counts extracted for the 223 different roads (in two directions) from the Toronto traffic count

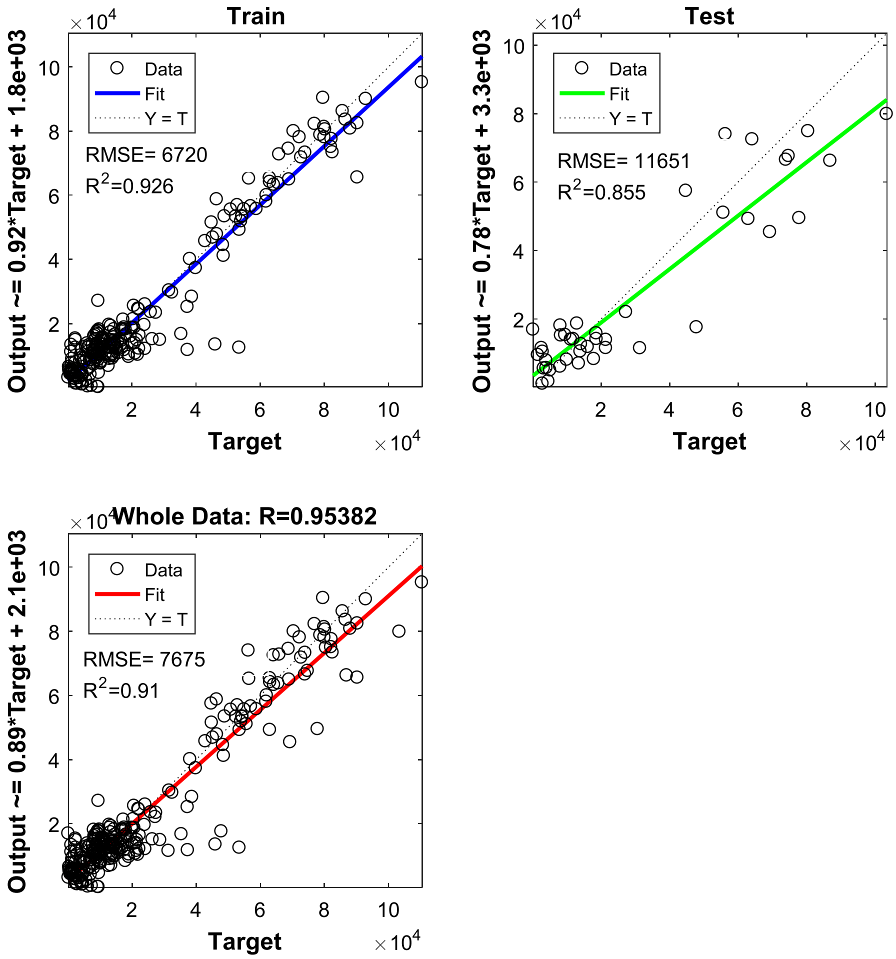

Fig. 11. Bayesian Regularization Artificial Neural Network (BR-ANN) for daily counts prediction: Target represents the desired ANN response; output represents the neural network response, which guides the learning process involving the changes in weights.

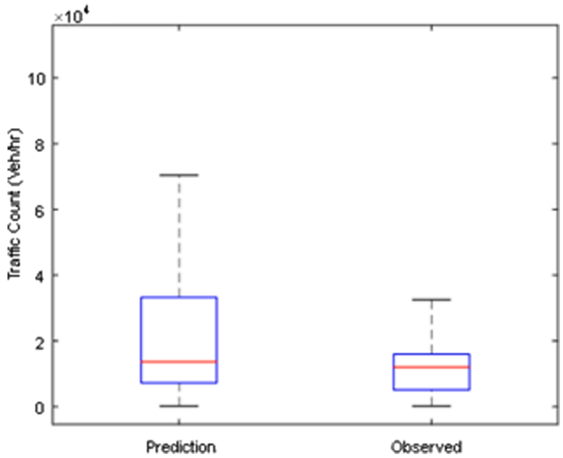

Fig. 12. Comparing the predicted and observed daily traffic count distributions using Bayesian Regularization Artificial Neural Network (BR-ANN).

database. The Levenberg–Marquardt (LM) training approach had the best performance for hourly values. The output represents the estimated hourly traffic counts based on the number of detected vehicles in the image, the road characteristics, and the timestamp of each image. Correlations between observed and predicted traffic counts in the model training and testing phase were around 75%.

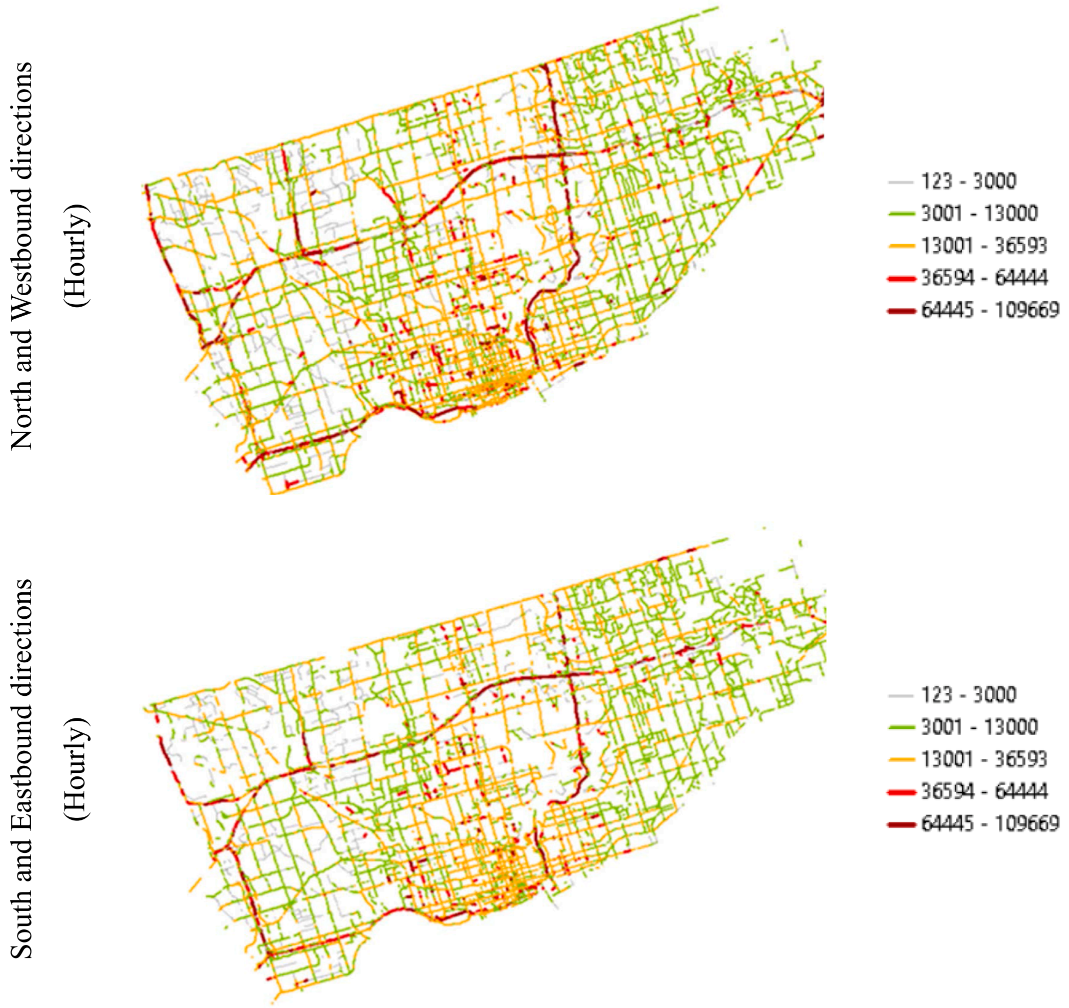

Fig. 13. Estimated daily traffic counts using Google aerial images for 2016 for different traffic directions.

Fig. 10 compares the frequency distributions of predicted and observed hourly traffic counts. This figure indicates a modest but non-significant increase in the 50th percentile of predicted traffic counts compared to the observed values.

In addition to hourly counts, a neural network model was trained using Bayesian regularization (BR) to predict daily traffic counts based on the vehicles detected in the images. For this, the detected vehicle counts from Google aerial images are paired with the daily traffic counts at nearby monitoring sites. The BR method had a better performance for daily traffic count prediction than LM, which worked better for hourly traffic count prediction. Fig. 11 presents the results for training and testing of the ANN model for daily traffic count prediction. The R^2 values for the training, testing and validation stages are higher than 85% and are close to 95% for the whole dataset. The RMSE for the whole dataset is 7675. Comparing the predicted and observed daily counts distribution in Fig. 12 also shows small differences in the frequency distributions.

Based on the training results, the BR neural network model was used to predict daily counts on 17,800 roads that do not have any traffic data but extracted aerial images. The prediction results are presented in Fig. 13 for two different road directions, i.e., North/East and South/West directions. As indicated in these figures, the predicted values showed higher values on highways and major roads and lower values on minor roads.

3.2.2. AADT prediction

Fig. 14 illustrates the results of the PRTC model by comparing predicted AADT values with those extracted from permanent traffic count stations for both east/north and west/south road directions. As indicated in Fig. 14, the model can accurately regenerate the AADT at permanent stations. This means that the PRTC parameters, including estimated D_{ij} are appropriate to predict AADT. Furthermore, to assess the performance of the model for AADT prediction based on the aerial images/short-term traffic counts, we compared predicted AADT from images at the permanent sites with their corresponding AADT values (extracted from long-term traffic data) in Fig. 15. The AADT values for the permanent sites were estimated by averaging long-term daily values over a year. All

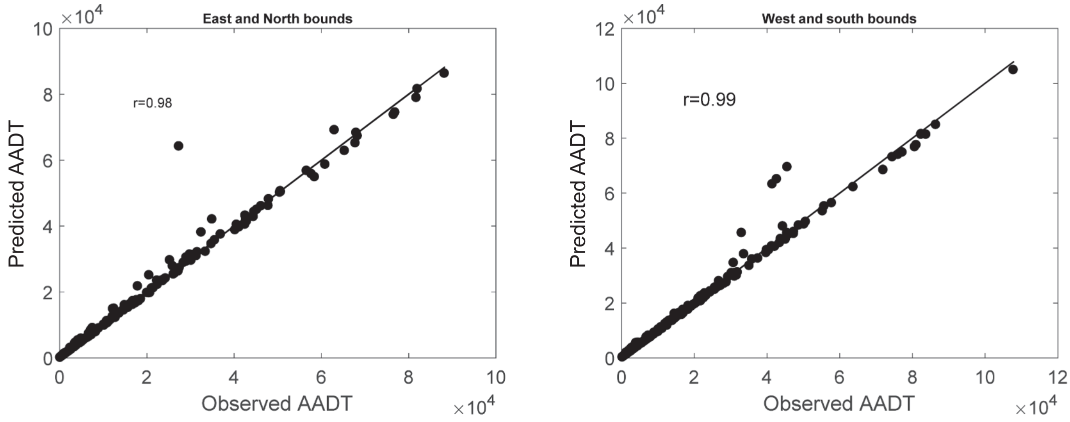

Fig. 14. Comparing observed (x-axis) and predicted AADTs (y-axis) at permanent station locations.

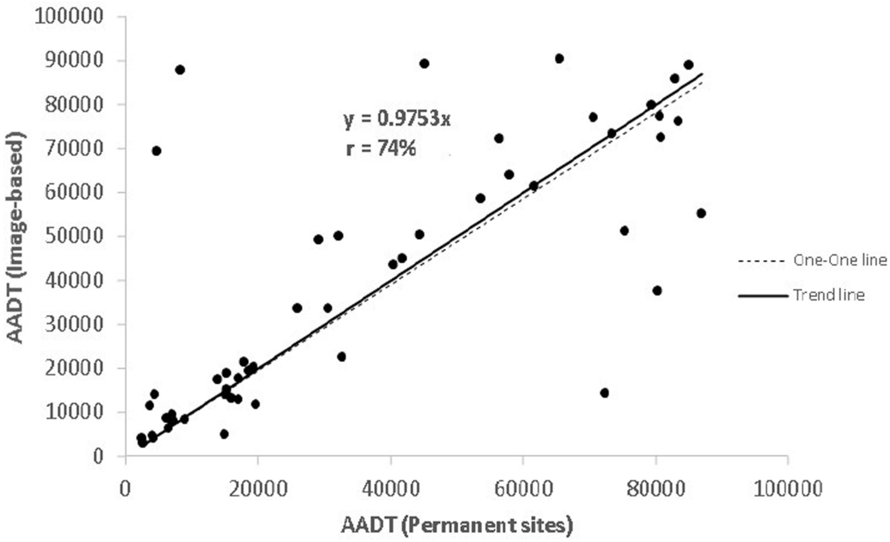

Fig. 15. Comparing observed (x-axis) and image-based predicted AADTs (y-axis) at permanent station locations for 2016.

permanent sites with more than 274 days of continuous daily traffic counts were used for this analysis. Fig. 15 compares the AADT values predicted using image-based vehicle counts with the observed AADT values at permanent stations for 2016. This figure indicates that the predicted AADT values are comparable with observed AADT at permanent sites ($R^2 = 0.58$). While most points follow the ideal one–one line between predicted and observed data, some points occasionally show a departure from the ideal line. These anomalies could be due to image quality, vehicle detection approach, and/or inherent uncertainty of a snapshot itself for traffic prediction. However, the One-One line doesn't show a systematic bias in prediction, which means that the predicted values were not over/underestimated in any domain. This is important since comparing the spatial distribution of traffic counts before and after including Google aerial images, will provide valuable insights for possible improvements in traffic prediction. This also helps us identify locations that may need further data collection to better estimate AADT in the future.

It should also be noted that Fig. 14 is developed based on a long-term mean hourly traffic count, which increases the accuracy of the method. In contrast, Fig. 15 presents AADT estimated based on PRTC and images (instead of long-term mean hourly traffic count) at permanent stations, reducing the model performance for AADT prediction.

3.3. Augmenting city count data with counts derived from images

Fig. 16 shows the predicted AADT for all roads in the City of Toronto in two different directions (North/East and South/West). To better understand the impact of image-based vehicle counts as a new source of information, the predicted AADT values in Fig. 16 are further compared by those estimated without the aid of Google aerial images (abbreviated GAI in the figure); estimated using a

Fig. 16. Estimated Annual Average Daily Traffic (AADT) for points without counts using Google aerial images for 2016 for different traffic directions.

combination of PRTC, RK and SVR. The boxplots of Fig. 17 compare the distributions of predicted traffic counts across the entire city using the City of Toronto database only (without Google aerial images) against the distribution of predictions using both the City of Toronto database and Google aerial images. While it is not possible to compare these predictions against ground truth since they occur across the entire network, including locations without images nor counts, we observe that the AADTs predicted with Google aerial images have a higher variability but with smaller values for different quantiles compared to the predictions made without Google aerial images. The largest effect of images occurs in the prediction of AADT along expressways. Overall, Fig. 17 demonstrates that using image-based vehicle counts as a new source of information offers lower AADT values for the roads without counts than interpolation-based models such as the Kriging model. Fig. 18 further visualizes the percentage change in AADT along all roads in the city.

Finally, the results for independent samples are presented in Figs. 19 and 20. First, an analysis was performed to show the neural network performance for daily traffic count prediction using a hold-out sample (Fig. 19). In this analysis, a set of daily traffic data was excluded from the training and validation process, and they were used as an independent dataset to show how the neural network can reproduce the daily values based on the image counts and road characteristics. The model exhibits a good performance even for those datasets that were never used for model development ($R^2 = 0.72$ and RMSE = 4881). We also compared the image-based AADT for roads that do not have any counts (Fig. 20). Here, we took a sample of roads for which short-term and long-term traffic counts are available and used this data as ground truth. It is important to keep in mind that it is not the short-term counts that are considered as

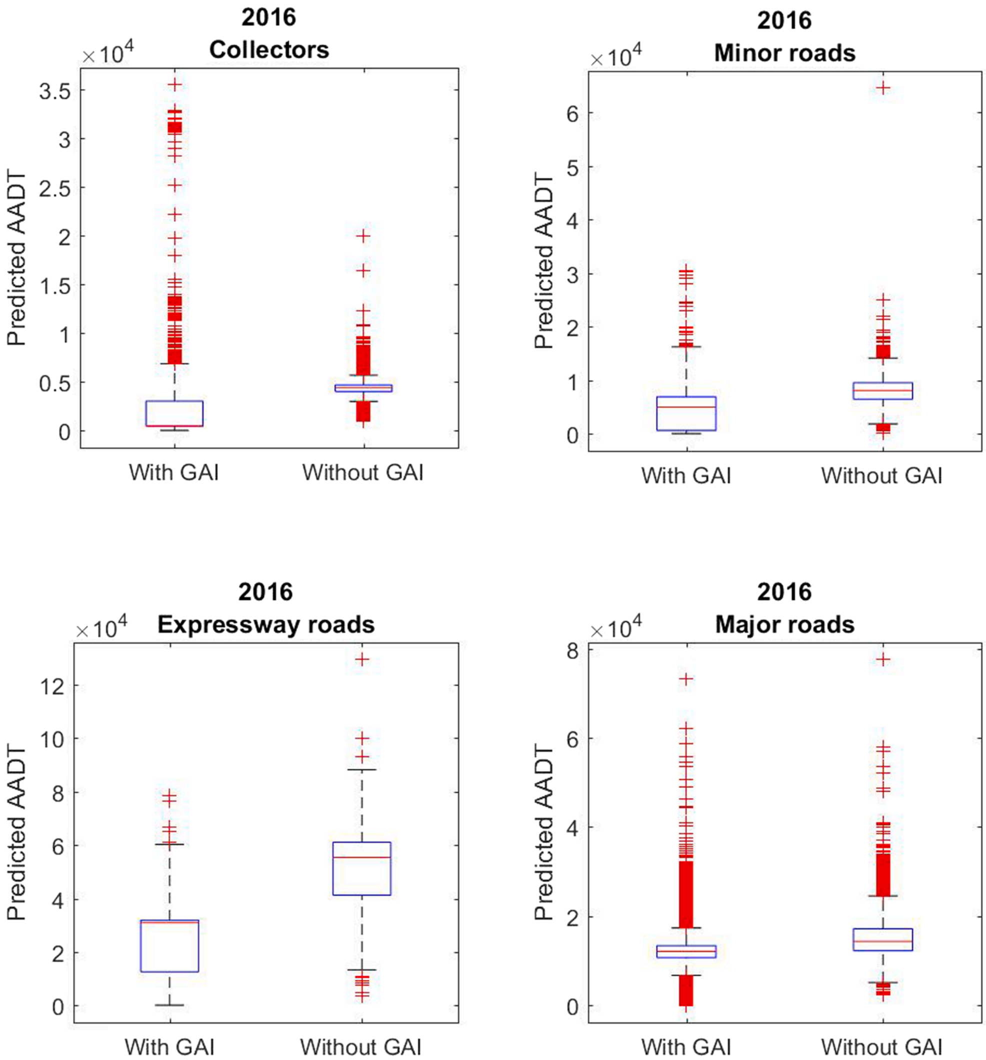

Fig. 17. Comparing Annual Average Daily Traffic (AADT) predictions with and without aerial images for different types of roads (GAI: Google Aerial Image).

Fig. 18. Percent change in predicted AADT after adding the Google aerial images. The negative values show a decrease in predicted values after adding the aerial images in percentage.

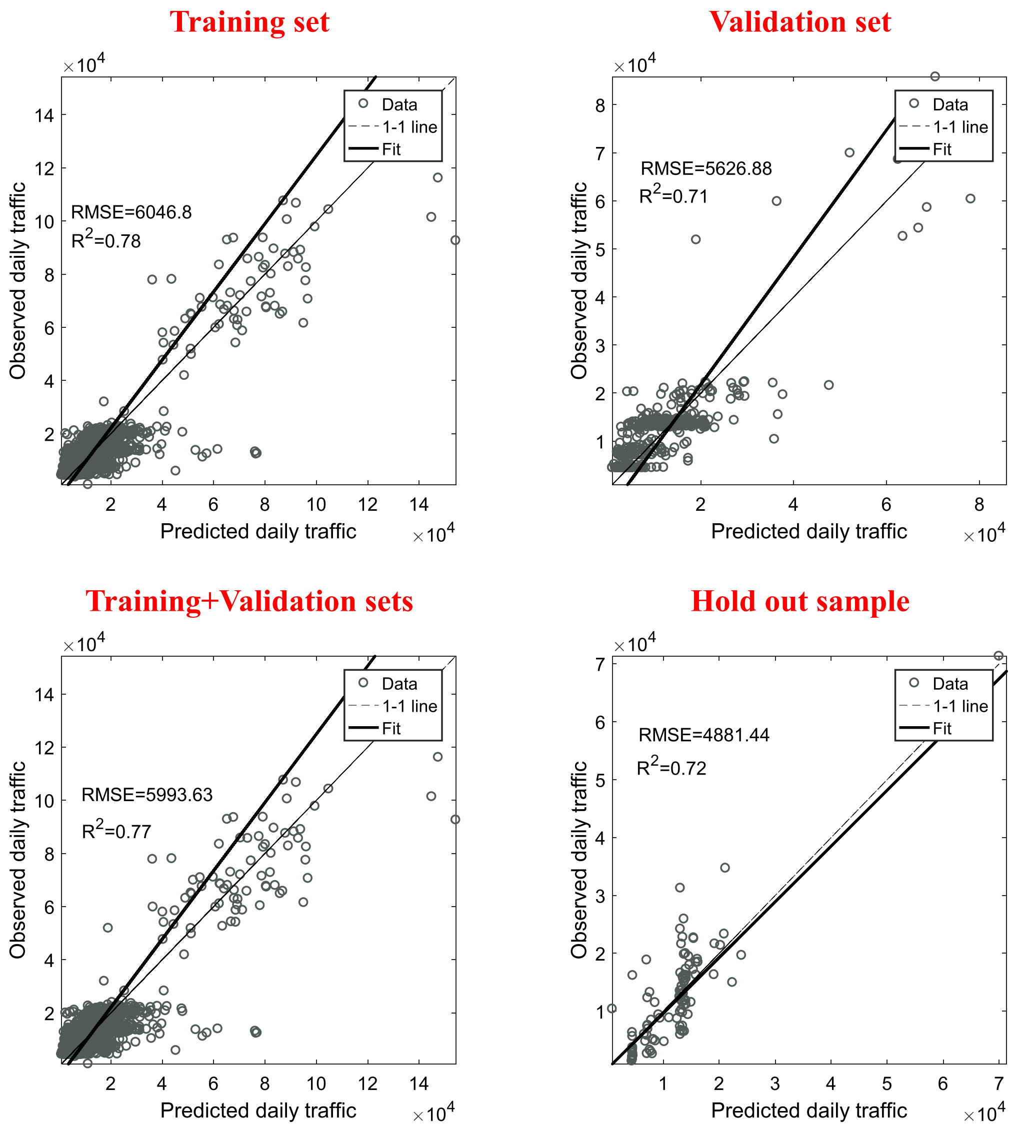

Fig. 19. Performance of neural network approach in predicting daily traffic based on images using a hold-out sample of short-term and long-term traffic count sites.

ground truth, but rather the AADT. So the short-term counts are first converted to AADT using the appropriate coefficients derived from the PRTC model. We compare the output with the AADT derived from image-based counts and PRTC and obtain $R^2 = 0.83$ and RMSE = 4950.

4. Conclusion

In this study, we developed a model of AADT prediction based on vehicle detection and road characteristics extraction from Google aerial images. We used R-CNN for vehicle detection since, among the CNN algorithms for vehicle detection tested in the literature, R-CNN exhibited higher predictive power (Deng et al., 2017; Li and Wang, 2017; Long et al., 2017; Ding et al., 2018). Since the oriented bounding boxes R-CNN has reported even higher detection performance (Ding et al., 2018), we also used the oriented bounding boxes as regions of interests (RoIs) to eliminate the mismatch of objects from Google aerial images.

Vehicle and road characteristics were captured from Google aerial images for more than 17,800 roads using image detection algorithms. We developed methods to extract road area, road width, road direction and captured the time stamp for the aerial image. The detected vehicles were used to estimate hourly and daily traffic counts. The results were validated against observed traffic counts, showing the high accuracy of the proposed method.

Furthermore, because of the availability of Google aerial images, the geographic coverage of vehicle count estimation can be

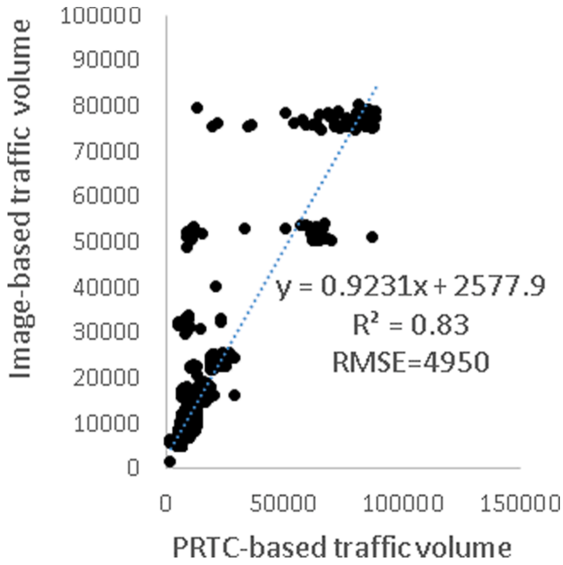

Fig. 20. Performance of modelling approach in predicting AADT traffic based on images using a hold-out sample of short-term and long-term traffic count sites.

expanded in areas where traffic counts are not available. As a benefit to those models that use only observed traffic counts (Ganji et al., 2020a), the integration of images with observed traffic data provides an accurate resource since municipalities do not frequently update their datasets. This is especially the case in metropolitan areas such as Toronto, which has more than 65,000 roads.

We also compared the performance of the ANN model and regression model for daily traffic volume prediction based on the detected vehicles from Google aerial images and road characteristics. Although both ANN and regression models minimize the mean squared error between the observed and predicted (target) values, the form of modeling leads to different performances between these two approaches. The regression approach assumes a pre-defined mathematical functional form with several statistical assumptions (e.g., normality, independent predictors). In contrast, the ANN model can model undefined nonlinear relationships between the input and output variables. However, it should be noted that a major issue for ANN techniques is the potential for overfitting and overtraining, which leads to a fitting of the noise and a loss of generalization of the network. In this research, Bayesian Regularization was used to reduce the potential for overfitting.

Since the proposed methodology works based on the observed traffic counts and images, using relationships extracted directly from the data, it is transferable to other cities. It is important to note that this approach requires a database of short and long-term traffic counts as a foundation and cannot be solely based on images. Although alternative image databases have been used in the literature to predict traffic volume, most previous studies tend to multiply the detected vehicles with time to estimate long-term traffic volumes. This is a significant limitation because long-term traffic volume prediction from short-term counts needs a robust coefficient of transformation that must be extracted from nearby sites using advanced statistical methods. This stresses the importance of city-specific traffic counts. This paper introduces a novel methodology, which consists of a pattern recognition submodule to estimate the coefficient of transformation for image-based traffic counts. Our validation results using the Google Aerial Images indicate that the predicted AADT values are not biased systematically, and they are close to observed values.

CRediT authorship contribution statement

Arman Ganji: Conceptualization, Methodology. **Mingqian Zhang:** Conceptualization, Methodology. **Marianne Hatzopoulou:** Supervision, Conceptualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This study received funding from the Canada Research Chairs program, of which Prof. Marianne Hatzopoulou is the recipient (950-23120). The authors also wish to acknowledge funding from the City of Toronto and The Atmospheric Fund (TAF).

Appendix I

1. Regression kriging (RK)

RK uses nearby AADT values as input and interpolation techniques to estimate AADT at locations without traffic monitoring stations. This method outperforms other interpolation techniques such as ordinary kriging and co-kriging (Ganji et al., 2020a). RK estimator function ($Z = \beta_0 + X_1\beta_1 + \ldots + X_k\beta_k + \varepsilon$) is a linear regression equation plus residuals which are respectively considered as deterministic and stochastic parts of the model. Z is the vector of AADT values in this estimator, and X and b represent the land use predictors and regression parameters, respectively. In this study, predictors are land-use values specific to the road segment where the monitoring station is located, such as the number of lanes, population density within a buffer of 200 m, and speed limit. Except for population, which varies with the year of analysis, all independent variables were extracted from 2011 land use databases. The residual part of the model is a function of the covariance of the residuals, which can be estimated from a semi-variogram of the residual part; however, it also depends on the regression parameters (β). Ganji et al., (2020a) proposed a generalized least squares equation to estimate β *(Eq. (4))*.

$$P(Z;s) = X(s)^T\beta + \sigma(\theta)^T\Sigma(\theta)^{-1}(Z(s) - X(s)\beta) \tag{4}$$

In Eq. (4), P represents the predicted AADT, and a Generalized Least Square (GLS) method was proposed to estimate β as follows:

$$\beta = \left(X(s)^T\Sigma(\theta)^{-1}X(s)\right)^{-1}X(s)^T\Sigma(\theta)^{-1}Z(s) \tag{5}$$

Since β is a function of θ, the steps presented in Fig. I–1 are proposed by Ganji et al., (2020a) to estimate $\beta and \theta$:

In Fig. I–1, semi-variogram represents the relationship between distances and the spatial correlation of residuals. The distance represents the network distances extracted using the shortest path algorithm. In this study, the regression predictors are land-use within a 300 m buffer around the roads, including residential, commercial, industrial, as well as population density, number of lanes, speed limit, and road types (highways, major and minor arterial roads, local road, and collectors). Following up Ganji et al., (2020a), The performance of the RK model is evaluated by comparing the predicted and observed AADT scatter plot (with 1–1 reference line plot), the correlation between observed and predicted values, predicted and observed histograms and a map of relative errors over the entire study area.

2. Support vector regression (SVR)

SVR, which works based on Support Vector Machine (SVM) (Ganji et al., 2020a), estimates annual average daily traffic (AADT) at locations where counts are not available. SVR develops a relationship between land use characteristics and traffic counts on local roads. Assuming a set of predictors ($\Phi(x)$), the generic SVR function can be represented as follow:

$$f(x) = (w.\Phi(x)) + b \tag{6}$$

where $wandb$ are the parameters are estimated by minimizing the expressed in Eq. (7).

$$Lost = \frac{1}{2}||w||^2 + C\sum\Gamma(f(x_i) - y_i) \tag{7}$$

where C is a constant value and $\Gamma(f(x_i) - y_i)$ is a cost function. The appropriate form of cost function has been presented in Ganji et al., (2020a). Equation (7) can be minimized by solving a quadratic optimization problem as follows:

$$\begin{cases} \frac{1}{2}\sum_{i,j=1}^{l}(\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j)k(x_i, xj) - \sum_{i=1}^{l}(\alpha_i^*(y_i - \varepsilon) - \alpha_i\,(y_i - \varepsilon) \\ \qquad s.t. \\ \sum_{i=1}^{l}(\alpha_i^* - \alpha_i) = 0\alpha_i^*, \alpha_i \in [0, C] \end{cases} \tag{8}$$

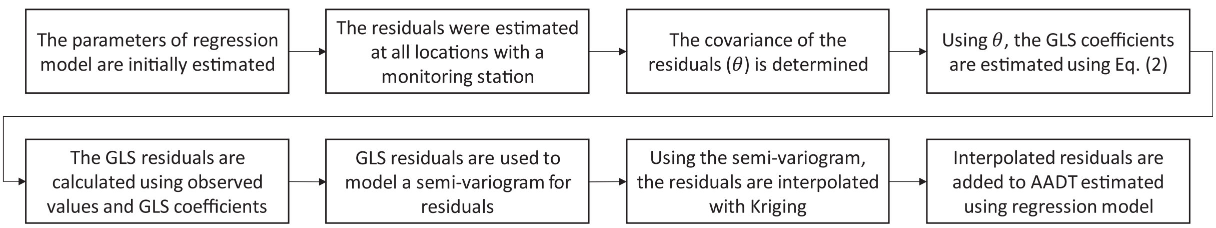

Fig. I–1. Proposed steps to estimate AADT using regression kriging.

α_i^*, α_i in Eq. (8) are the Lagrange multipliers and represent solutions to the above quadratic problem, acting as forces pushing predictions toward a target value. In SVR, the predictors included land-use within a 300 m buffer around the roads, including residential, commercial, industrial, as well as population density, number of lanes, speed limit and road types (highways, major and minor arterial roads, local road, and collectors). Also, all estimated and observed AADT for the major roads were integrated within 300 m buffers around the local roads and used as predictors in the SVR model. SVR performance was evaluated by comparing observed and predicted scatter plots with 1–1 reference line plot and statistical indices for the estimated SVR parameters.

Appendix II

mAP (mean Average Precision) for Object Detection.

Average precision (AP) is a well-known index of object detectors such as R-CNN. AP computes the average precision value, which varies between 0 and 1, and it measures what percentage of predictions are correct (Table A1–II).

$$Precision = TP/(TP + FP)$$

where TP and FP reflect the true and False positives, respectively. A list of 100 random images from Toronto was analyzed to compute

Table A1–II
mAP for 100 random images extracted City of Toronto images database.

Road number	TP	Total positive	Precision	Road number	TP	Total positive	Precision
415	20	26	0.77	913,149	26	31	0.84
534	13	19	0.68	913,354	25	25	1.00
1256	14	15	0.93	913,864	7	7	1.00
1293	10	12	0.83	913,980	27	27	1.00
1732	15	15	1.00	1,137,655	2	2	1.00
1865	13	13	1.00	1,137,725	7	21	0.33
1971	22	26	0.85	1,140,684	0	0	1.00
7777	10	30	0.33	1,140,767	14	15	0.93
8834	6	6	1.00	1,145,316	10	18	0.56
104,239	10	10	1.00	1,145,433	26	33	0.79
105,548	2	2	1.00	2,216,243	16	16	1.00
107,958	43	43	1.00	2,311,915	2	2	1.00
108,256	0	0	1.00	2,945,648	5	5	1.00
109,452	1	1	1.00	3,054,783	40	40	1.00
109,505	4	8	0.50	3,719,873	37	37	1.00
110,060	13	17	0.76	5,343,674	14	14	1.00
110,882	13	13	1.00	6,027,119	22	26	0.85
111,590	70	71	0.99	6,786,837	32	32	1.00
111,718	3	6	0.50	7,195,813	22	25	0.88
111,997	18	18	1.00	7,274,839	19	25	0.76
436,793	6	8	0.75	7,792,433	13	15	0.87
438,506	26	28	0.93	8,005,169	3	31	0.10
439,018	6	6	1.00	8,457,327	4	29	0.14
441,226	11	12	0.92	8,676,766	9	10	0.90
442,124	8	8	1.00	9,066,085	7	8	0.88
443,321	32	59	0.54	9,085,295	18	27	0.67
443,573	43	46	0.93	9,134,281	29	29	1.00
443,654	1	1	1.00	9,655,343	2	2	1.00
443,702	12	12	1.00	9,722,504	15	15	1.00
443,792	20	23	0.87	9,779,365	20	21	0.95
444,113	1	1	1.00	9,879,826	26	27	0.96
444,214	33	33	1.00	10,071,118	8	9	0.89
445,576	10	11	0.91	10,080,008	2	2	1.00
445,617	24	25	0.96	10,133,718	1	1	1.00
445,699	6	15	0.40	10,223,904	16	34	0.47
906,786	13	13	1.00	10,558,821	3	8	0.38
908,032	19	20	0.95	10,635,840	25	35	0.71
908,165	56	56	1.00	11,130,039	12	14	0.86
908,485	21	21	1.00	13,501,859	16	18	0.89
910,356	1	1	1.00	14,017,414	6	11	0.55
912,310	10	10	1.00	14,025,568	15	25	0.60
912,722	35	37	0.95	14,048,690	7	10	0.70
912,906	8	19	0.42	14,229,843	0	0	1.00
14,646,674	8	8	1	30,001,930	2	2	1
14,646,827	2	9	0.22	30,005,230	76	76	1
14,659,946	4	4	1	30,007,983	16	22	0.72
14,660,984	30	30	1	30,008,322	4	7	0.57
14,664,636	10	14	0.71	30,015,981	19	23	0.83
20,043,234	62	62	1	30,029,635	66	68	0.97

the mAP for the vehicle detection module. The comparison was made against manual counting conducted by two separate research assistants.

References

Bagheri, E., Zhong, M., Christie, J., 2015. Improving AADT Estimation Accuracy of Short-Term Traffic Counts Using Pattern Matching and Bayesian Statistics. J. Transp. Eng. 141, A4014001.

Burns, J., Hanson, A., Riseman, E. In Proc. of the Seventh International Conference on Pattern Recognition (July 30-August 2, 1984), Montreal, Canada, 1984.

Cai, P., Wang, Y., Lu, G., Chen, P., Ding, C., Sun, J., 2016. A spatiotemporal correlative k-nearest neighbor model for short-term traffic multistep forecasting. Transport. Res. Part C Emerg. Technol. 62, 21–34.

Cao, L., Wang, C., Li, J., 2016. Vehicle Detection from Highway Satellite Images Via Transfer Learning. Inf. Sci. 366, 177–187.

Chen, X., Xiang, S., Liu, C.-L., Pan, C.-H., 2014. Vehicle Detection in Satellite Images by Hybrid Deep Convolutional Neural Networks. IEEE Geosci. Remote Sens. Lett. 11, 1797–1801.

Cheng, G., Zhou, P., Han, J., 2016. Learning Rotation-Invariant Convolutional Neural Networks for Object Detection in Vhr Optical Remote Sensing Images. IEEE Trans. Geosci. Remote Sens. 54, 7405–7415.

Deng, Z., Sun, H., Zhou, S., Zhao, J., Zou, H., 2017. Toward Fast and Accurate Vehicle Detection in Aerial Images Using Coupled Region-Based Convolutional Neural Networks. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 10, 3652–3664.

Desolneux, A., Moisan, L., Morel, J.-M., 2000. Meaningful Alignments. Int. J. Comput. Vis. 40, 7–23.

Ding, J., Xue, N., Long, Y., Xia, G.-S., Lu, Q., 2018. Learning Roi Transformer for Detecting Oriented Objects in Aerial Images. arXiv preprint: Xiv,1812.00155.

Ganji, A., Shekarrizfard, M., Harpalani, A., Coleman, J., Hatzopoulou, M., 2020a. Methodology for Spatio-Temporal Predictions of Traffic Counts across an Urban Road Network and Generation of an on-Road Greenhouse Gas Emission Inventory. Comput.-Aided Civ. Infrastruct. Eng. 35, 1063–1084.

Ganji, A., Minet, L., Weichenthal, S., Hatzopoulou, M., 2020b. Predicting Traffic-Related Air Pollution Using Feature Extraction from Built Environment Images. Environ. Sci. Technol. 54, 10688–10699.

Gastaldi, M., Gecchele, G., Rossi, R., 2014. Estimation of Annual Average Daily Traffic from One-Week Traffic Counts. A Combined Ann-Fuzzy Approach. Transport. Res. Part C, Emerg. Technol. 47, 86–99.

Girshick, R., 2015. Proceedings of the IEEE international conference on computer vision, pp. 1440-1448.

Girshick, R., Donahue, J., Darrell, T., Malik, J., 2014. Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 580–587.

Hinz, S., 2004. Detection of Vehicles and Vehicle Queues in High Resolution Aerial Images. Photogrammetrie-Fernerkundung-Geoinformation.

Holt, A.C., Seto, E.Y., Rivard, T., Gong, P., 2009. Object-Based Detection and Classification of Vehicles from High-Resolution Aerial Photography. Photogramm. Eng. Remote Sens. 75, 871–880.

Jiang, Q., Cao, L., Cheng, M., Wang, C., Li, J., 2015. International Symposium on Bioelectronics and Bioinformatics (ISBB), 2015. IEEE 184–187.

Kaack, L.H., Chen, G.H., Morgan, M.G., 2019. Proceedings of the 2nd ACM SIGCAS Conference on Computing and Sustainable Societies, pp. 155-164.

Kembhavi, A., Harwood, D., Davis, L.S., 2010. Vehicle Detection Using Partial Least Squares. IEEE Trans. Pattern Anal. Mach. Intell. 33, 1250–1265.

Kozempel, K., Reulke, R., 2009. Fast Vehicle Detection and Tracking in Aerial Image Bursts. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci. 38, 175–180.

Larsen, S.Ø., Koren, H., Solberg, R., 2009. Traffic Monitoring Using Very High Resolution Satellite Imagery. Photogramm. Eng. Remote Sens. 75, 859–869.

LeCun, Y., Bottou, L., Bengio, Y., Haffner, P., 1998. Gradient-Based Learning Applied to Document Recognition. Proc. IEEE 86, 2278–2324.

Lenhart, D., Hinz, S., Leitloff, J., Stilla, U., 2008. Automatic Traffic Monitoring Based on Aerial Image Sequences. Pattern Recogn. Image Anal. 18, 400–405.

Li, X., Wang, S., 2017. Object Detection Using Convolutional Neural Networks in a Coarse-to-Fine Manner. IEEE Geosci. Remote Sens. Lett. 14, 2037–2041.

Li, J., Xu, Z., Fu, L., Zhou, X., Yu, H., 2021. Domain adaptation from daytime to nighttime: A situation-sensitive vehicle detection and traffic flow parameter estimation framework. Transport. Res. Part C Emerg. Technol. 124, 102946.

Long, Y., Gong, Y., Xiao, Z., Liu, Q., 2017. Accurate Object Localization in Remote Sensing Images Based on Convolutional Neural Networks. IEEE Trans. Geosci. Remote Sens. 55, 2486–2498.

Ren, S., He, K., Girshick, R., Sun, J., 2015. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. Adv. Neural Inf. Process. Syst. 91–99.

Sandnes, F.E., 2011. Determining the geographical location of image scenes based on object shadow lengths. J. Sign. Process Syst. 65, 35–47. https://doi.org/10.1007/s11265-010-0538-x.

The World Bank., 2018. World development indicators. Retrieved from https://datacatalog.worldbank.org/dataset/world-development-indicators.

Uijlings, J.R., Van De Sande, K.E., Gevers, T., Smeulders, A.W., 2013. Selective Search for Object Recognition. Int. J. Comput. Vision 104, 154–171.

Von Gioi, R.G., Jakubowicz, J., Morel, J.-M., Randall, G., 2012. Lsd, A Line Segment Detector. Image Processing On Line 2, 35–55.

Wang, X., Kockelman, K.M., 2009. Forecasting Network Data, Spatial Interpolation of Traffic Counts from Texas Data. Transp. Res. Rec. 2105, 100–108.

Xia, Z., Zang, Y., Wang, C., Li, J., 2017. Road width measurement from remote sensing images. IEEE Int. Geosci. Remote Sens. Symposium 902–905.