

METHOD OF PREDICTING ROAD ATTRIBUTES, DATA PROCESSING SYSTEM AND COMPUTER EXECUTABLE CODE

TECHNICAL FIELD

[0001] An aspect of the disclosure relates to a method of predicting one or more road segment attributes corresponding to a road segment in a geographical area. Another aspect of the disclosure relates to a data processing system. Another aspect of the disclosure relates to a non-transitory computer-readable medium storing computer executable code. Another aspect of the disclosure relates to a computer executable code. Another aspect of the disclosure relates to a method for training an automated road attribute predictor.

BACKGROUND

[0002] The service of ride-hailing providers significantly relies on the quality of a digital map. Incomplete satellite image such as a missing road or even a missing road attribute can lead to misleading routing decisions or inaccurate prediction of a driver's arrival time. However, the updating of both commercial and free maps still heavily relies on the manual annotations from human. The high cost results in maps with low completeness and inaccurate outdated data. Taking the OpenStreetMap (OSM) as an example, which provides the community a user-generated map of the world, its data completeness and accuracy vary significantly in different cities. For example, in Singapore, while most of the roads are annotated in the map with the one-way or two-way tags, only about 40% and 9% of the roads are annotated with the number of lanes and the speed limit in the downtown area.

[0003] Therefore, current methods of updating satellite image have drawbacks and it is desired to provide for an improved method of updating satellite image.

SUMMARY

[0004] An aspect of the disclosure relates to a method of predicting one or more road segment attributes corresponding to a road segment in a geographical area. The method may include providing trajectory data and satellite image of the geographical area. The method may include calculating one or more image channels based on the trajectory data. The method may

include, using at least one processor, classifying the road segment based on the one or more image channels and the satellite image using a trained classifier, e.g., into prediction probabilities of the road attributes.

[0005] An aspect of the disclosure relates to a data processing system including one or more processors configured to carry out the method of predicting road attributes.

[0006] An aspect of the disclosure relates to a data processing system including one or more processors configured to carry out a method of predicting road attributes. The system may include a first memory configured to store trajectory data of the geographical area. The system may include a second memory configured to store a satellite image of the geographical area. The system may include a processor configured to calculate one or more image channels based on the trajectory data. The system may include a classifier, the classifier may include a neural network configured to predict road attributes based on the one or more image channels and the satellite image. The classifier may include an output layer configured to provide the prediction probabilities of the road attributes. The classifier may be a trained classifier.

[0007] An aspect of the disclosure relates to a non-transitory computer-readable medium storing computer executable code including instructions for predicting one or more road segment attributes according to the method described herein.

[0008] An aspect of the disclosure relates to a computer executable code including instructions for predicting one or more road segment attributes according to the method described herein.

[0009] An aspect of the disclosure relates to a method of training an automated road attribute predictor. The method of training may include performing forward propagation by inputting training data into the automated predictor to obtain an output result, for a plurality of road segments of a geographical area. The training data may include trajectory data and satellite image having an electronic image format. The method of training may include performing back propagation according to a difference between the output result and an expected result to adjust weights of the automated predictor. The method of training may include repeating the above steps until a pre-determined convergence threshold may be achieved. The automated predictor may include the classifier configured to be trained to provide the prediction probabilities of the road attributes. The classifier may include a neural network configured to predict road attributes based on one or more image channels and satellite image. The one or more image channels may be calculating based on the trajectory data.

[0010] An aspect of the disclosure relates to a classifier trained by the method of training as described herein.

BRIEF DESCRIPTION OF THE DRAWINGS

[0011] The invention will be better understood with reference to the detailed description when considered in conjunction with the non-limiting examples and the accompanying drawings, in which:

- FIG. 1 shows an exemplary map including a geographical area 10 which includes a road segment 12;
- FIG. 2 shows a flow chart of a method 100 of predicting one or more road segment attributes 20 in accordance with various embodiments;
- FIG. 3 shows exemplary details of providing trajectory data 110 in the form of a flow chart, in accordance with various embodiments;
- FIG. 4 shows exemplary details of calculating a bearing image channel 113 in the form of a flow chart, in accordance with various embodiments;
- FIG. 5 shows exemplary details of calculating a speed image channel 116 in the form of a flow chart, in accordance with various embodiments;
- FIG. 6 shows exemplary details of data pre-processing before classifying 130 the road segment 12, in the form of a flow chart, in accordance with various embodiments;
- FIG. 7 shows a flowchart of inputting data into the trained classifier when using early fusion;
- FIG. 8 shows a flowchart of inputting data into the trained classifier when using late fusion;
- FIG. 9 shows one example of a structure of the classifier for early fusion in accordance with some embodiments;
- FIG. 10 shows one example of a structure of the classifier for late fusion in accordance with some embodiments;
- FIG. 11 shows one exemplary classifier structure in accordance with various embodiments;
- FIGS. 12A and 12B shows illustration of the challenges for road attribute detection using only one of satellite images or GPS traces alone. In FIG. 12A, a satellite image is

shown with heavy clouds that interfere with the visibility of the roads. In FIG. 12B, sparse GPS samples makes it more difficult to identify a narrow road;

- FIGS. 13A to 13C show GPS rendering results after applying kernel smoothing. FIG. 13A shows the rendering of the original GPS traces. FIG. 13B shows the rendering of the smoothed GPS traces with kernel=3, and FIG. 13C shows the rendering of the smoothed GPS traces with kernel=5;
- FIG. 14 shows the architecture of an exemplary computer 300, which may be used in accordance with various embodiments
- FIG. 15 shows the number of training and testing samples used in examples, for in each category (i.e., each road attribute).
- FIG. 16 shows a table with results of classification accuracy obtained based on satellite images only, GPS traces only, and their fusion;
- FIGS. 17A to 17B shows exemplary results of per-class F1 measure of road attribute detection based on satellite images, GPS traces, and their fusion on the speed limit and road type detection. FIG. 17A shows the speed limit and FIG. 17B shows the road type for data sample of Singapore;
- FIG. 18 shows exemplary results of per-class F1 measure of road attribute detection based on satellite images, GPS traces, and their fusion on the road type detection for Jakarta;
- FIGS. 19A to 19C show confusion matrices of the method of predicting one or more road segment attributes in accordance with various embodiments on: FIG. 19A, the speed limit for Singapore; FIG. 19B, road type detection for Singapore, FIG. 19C, the road type for Jakarta;
- FIG. 20 shows a table with the results of a study on the impact of the number of bins/channels M_b and M_s that are used to render the sensor data into images;
- FIG. 21 shows a table with the results of a study of the impact of the kernel size adopted for GPS smoothing;
- FIG. 22 shows a table reporting the classification accuracy based on different combinations of input modules and model architectures on the Singapore dataset;
- FIG. 23 shows a table reporting the classification accuracy and the performance gain of the disclosed multimodal feature fusion with and without calibration.

[0012] A neural network stream, or also named as flow, is represented in the drawings from left to right, and may be indicated by arrows.

DETAILED DESCRIPTION

[0013] The following detailed description refers to the accompanying drawings that show, by way of illustration, specific details and embodiments in which the disclosure may be practiced. These embodiments are described in sufficient detail to enable those skilled in the art to practice the disclosure. Other embodiments may be utilized and structural, and logical changes may be made without departing from the scope of the disclosure. The various embodiments are not necessarily mutually exclusive, as some embodiments can be combined with one or more other embodiments to form new embodiments.

[0014] Embodiments described in the context of one of the methods for predicting, systems, computer executable codes, non-transitory computer-readable medium, and methods for training, are analogously valid for the other methods for predicting, systems, computer executable codes, non-transitory computer-readable medium, and methods for training. Similarly, embodiments described in the context of a method for predicting are analogously valid for a system, and vice-versa.

[0015] Features that are described in the context of an embodiment may correspondingly be applicable to the same or similar features in the other embodiments. Features that are described in the context of an embodiment may correspondingly be applicable to the other embodiments, even if not explicitly described in these other embodiments. Furthermore, additions and/or combinations and/or alternatives as described for a feature in the context of an embodiment may correspondingly be applicable to the same or similar feature in the other embodiments.

[0016] In the context of various embodiments, the articles “a”, “an” and “the” as used with regard to a feature or element include a reference to one or more of the features or elements.

[0017] As used herein, the term “and/or” includes any and all combinations of one or more of the associated listed items.

[0018] As used herein, the method of predicting one or more road segment attributes may be simply referred to as the method, while the method for training a classifier is explicitly named as such.

[0019] The term “road” as used herein may mean a way which is envisaged for a vehicle locomotion, and its meaning may include motorway (also known as highway), main road, street, or a combination thereof.

[0020] The term “vehicle” may mean a motor vehicle, for example a car, or a bus.

[0021] As used herein the terms “detection” and “prediction” may be used interchangeably.

[0022] According to various embodiments, road attributes may include, among others, one or more of: one-way/two-way road, number of lanes, speed limit, road type. Examples of road types may include, among others, one or more of: “Residential”, “Service”, and “Footway”, “Primary”, “Secondary”, “Tertiary”.

[0023] As used herein the terms “classifier”, “neural network”, and their variants (e.g. CNN), refer to artificial classifier, artificial neural network, and their variants, respectively, which means that they are computer implemented and are processed by an electronic microprocessor, or simply named herein as “processor”. Examples of processor include, among others, CPU, GPU, neural network co-processor, neural-network processor, neural-network chip.

[0024] The skilled person in the art would understand, based on the present disclosure, that, in embodiments and examples not related to training, the neural network is a trained neural network, the classifier is a trained classifier, and the automated predictor is a trained automated predictor. For example, an automated predictor may have been trained based on a training data record including (i) training trajectory data of at least one geographical area, the trajectory data including location, bearing and speed, and (ii) training satellite image (i.e., training images) of the at least one sub-area of the geographical area, wherein the training trajectory data or the training satellite images include associated one or more pre-determined road attributes.

[0025] According to various embodiments a method of predicting one or more road segment attributes corresponding to a road segment in a geographical area is disclosed herein. The method may include providing trajectory data and satellite image of the geographical area.

[0026] As used herein and in accordance with various embodiments, the term trajectory data may include geographical data, such as geospatial coordinate and may further include time, for example, as provided by the global positioning system GPS. Trajectory data may be obtained from the recording of positions of one or more moving vehicles. For example, latitude, longitude and time, the trajectory data may further include elevation. The GPS coordinates may according to the World Geodetic System, WGS 84, for example, version G1674.

[0027] According to various embodiments, the trajectory data may include a plurality of data points, wherein each data point may include latitude, longitude, bearing, and speed.

[0028] A trajectory trace, e.g., a GPS trace, may be defined to be a sequence of records associated with timestamps. Each record (also named as point) includes location, bearing, and speed returned by sensors. The location of a trajectory record may be represented by the latitude and longitude pair. The bearing is the clock-wise angle of the device's moving direction with respect to the earth's true north direction.

[0029] For the identification of road attributes, the location data may be real world data, for example real world GPS data. Correspondently, the geographical area represents an area on earth's surface.

[0030] As used herein and in accordance with various embodiments, the terms 'geographical' and 'geospatial' may be used interchangeably.

[0031] According to various embodiments, the method may include calculating one or more image channels based on the trajectory data.

[0032] According to various embodiments, the method steps may be processed on a processor. The method may include, using at least one processor, classifying the road segment based on the one or more image channels and the satellite image using a trained classifier into prediction probabilities of the road attributes.

[0033] According to various embodiments, the satellite images may be extracted from larger satellite images, for example, by generating cropped images by cropping images from the larger satellite images. The cropped images may be centered at a corresponding road segment. According to various embodiments, the satellite image data may be, or be obtained from, digital maps, for example existing digital maps. Example of existing digital maps are the maps provided from OpenStreetMap® (www.openstreetmap.org). The digital maps may include rules for the visualization of the geographic objects, for example including one or more of roads, highways, buildings. For example, each of the geographical objects and/or each attribute of the geographical objects, may have a different color, a different perimeter line style (e.g. a different line thickness), a different fill pattern, or a combination thereof. A combination of digital maps of different sources, e.g. having different rules of visualization, may also be used as source for the satellite image data. The satellite image may include channels of different color, for example red (R), green (G), and blue (B).

[0034] According to various embodiments, trajectory data may be extracted from vehicle trace databases, for example crowdsourced GPS traces. The trajectory data may be identified as a group of points or traces of the trajectory data that are associated with a road segment of the road segments, for example, that is within a geographically defined perimeter which is a same perimeter used for determining the satellite image for the same road segment.

[0035] According to various embodiments, the data processing system may include a first memory configured to store trajectory data of the geographical area. The data processing system may include a second memory configured to store satellite image. The satellite image may include image data of the geographical area.

[0036] According to various embodiments, the satellite image may be in the form of image data, for example in electronic form configured to be stored in electronic digital format. An example of an electronic digital format for image data is JPEG. The electronic digital form may be, or provide information, to reproduce the image in the form of an array of pixels.

[0037] According to various embodiments calculating one or more image channels based on the trajectory data may include at least one of:

- i) calculating a trajectory image channel, as part of the one or more image channels, based on the trajectory data;

- ii) calculating a bearing image channel, as part of the one or more image channels, based on the trajectory data;

- iii) calculating a speed image channel, as part of the one or more image channels, based on the trajectory data. According to some embodiments, the method may include i), ii), and iii).

[0038] According to various embodiments, calculating the trajectory image channel may include assigning a count of the number of trajectory points, e.g. GPS points, of the trajectory data that may be projected onto each pixel of a matrix of pixels. For example, having GL^i defined to be a single-channel image, a count may be performed on the number of trajectory points that are projected onto each pixel. Therefore, GL^i is updated by $GL^i(x, y, 0) = GL^i(x, y, 0) + 1$ (Eq. 1).

[0039] According to various embodiments calculating a bearing image channel may include providing a multichannel bearing image including multichannel pixels, wherein the number of channels represents a number of bearing bins, and quantizing bearing values, e.g., by degree, into the bearing bins forming a bearing histogram for each of the multichannel pixels.

[0040] For example, GB^i may be defined to be a M_b -channel image where M_b is the number of bins adopted to quantize the bearing values in degree into a histogram at each pixel. Let Bin_b denote the bin size to generate the bearing histogram, GB^i may be updated, given a trajectory point p_j^i with $bearing_j^i$ at (x, y) in the image, as $GB^i(x, y, int(bearing_j^i/Bin_b)) = GB^i(x, y, int(bearing_j^i/Bin_b)) + 1$ (Eq. 2).

[0041] According to various embodiments calculating a speed image channel may include providing a multichannel speed image including multichannel pixels, wherein the number of channels represents a number of speed bins, and quantizing speed values [in m/s] into the speed bins forming a speed histogram for each of the multichannel pixels.

[0042] For example, GS^i may be defined to be a M_s -channel image where M_s is the number of bins adopted to quantize the speed values, e.g., in m/s, into a histogram at each pixel. Let Bin_s denote the bin size to generate the speed histogram, GS^i may be updated, given a trajectory point p_j^i with $speed_j^i$ at (x, y) in the image, as $GS^i(x, y, int(speed_j^i/Bin_s)) = GS^i(x, y, int(speed_j^i/Bin_s)) + 1$ (Eq. 3).

[0043] According to some embodiments, the method may include concatenating the trajectory image channel, the bearing image channel, and the speed image channel into a concatenated trajectory image before classifying the road segment. For example, the concatenated trajectory image is used as input in the classifier. In another example, the concatenated trajectory image fused with the satellite image is used as input in the classifier.

[0044] The location GL^i , bearing GB^i , speed GS^i images form a $(1 + M_b + M_s)$ -channel image, *i.e.*, the concatenated image, as the image-based feature extracted from the trajectory traces for road segment (r_i) .

[0045] According to various embodiments, the method may further include applying a smoothing filter on the concatenated trajectory image before classifying the road segment. With a high rendering resolution at, the projection of the original trajectory points around a road segment can be noisy and sparse. According to some embodiments, each channel of G^i may be smoothed, e.g., by computing the moving average over a square kernel with size K . Alternative weighting functions such as 2D Gaussian kernel may be adopted. The parameters therein may be tuned based on the characteristics of the trajectory data.

[0046] According to various embodiments, the location GL^i , bearing GB^i , speed GS^i images may be normalized. The distribution of trajectory traces can be unbalanced on different types of roads, e.g., highways and residential roads. To reduce the impact of trajectory traces disparity the location channel may be normalized based on the maximum value over all the pixels, while the bearing and speed channels may be normalized based on the sum over all the respective channels at each pixel. For example, the location GL^i , bearing GB^i , speed GS^i images may be normalized according to Eq. 4 as follows:

$$GL^i(x, y) = GL^i(x, y) / \max_{\{x', y'\}} GL^i(x', y')$$

$$GB^i(x, y) = GB^i(x, y) / \sum_{c'=0}^{M_b-1} GB^i(x', y', c')$$

$$GS^i(x, y) = GS^i(x, y) / \sum_{c'=0}^{M_s-1} GS^i(x', y', c')$$

to obtain the final trajectory data rendering result G^i .

[0047] In one example, the pre-processing algorithm may be summarized as follows:

```

Input: A set of GPS points  $P^i$  in the nearby region of a road segment  $r_i$ 
Output: A multi-channel image  $G^i$  as the feature extracted from  $P^i$  for road segment  $r_i$ 
3D Array  $GL^i, GB^i, GS^i$ ;
/*  $GL^i, GB^i$ , and  $GS^i$  are the image channels generated based on
   location, bearing, and speed, respectively. */
for each point  $p_j^i$  in  $P^i$  do
    /*  $p_j^i = (lat_j^i, lon_j^i, bearing_j^i, speed_j^i)$  is a 4-tuple that contains
       the readings of latitude, longitude, bearing, and speed. */
     $x, y = \text{locate\_pixel}(lat_j^i, lon_j^i, r_i)$ ;
    update_location_channel( $GL^i, x, y$ ) //based on Eq. 1;
    update_bearing_channel( $GB^i, x, y, bearing_j^i$ ) //based on Eq. 2;
    update_speed_channel( $GS^i, x, y, speed_j^i$ ) //based on Eq. 3;
 $G^i = \text{Concat}((GL^i, GB^i, GS^i), \text{axis}=-1)$ ;
kernel_smoothing( $G^i$ ) //based on moving average;
normalization( $G^i$ ) //based on Eq. 4;
return  $G^i$ ;

```

[0048] According to various embodiments, the method, further including applying image rotation until the road segment in the concatenated trajectory image may be aligned with the road segment in satellite image before classifying the road segment.

[0049] According to various embodiments, to reduce the impact of road directions on the extraction of road features, the road features may be calibrated in any or both of the following two aspects. Both the satellite images and the one or more image channels (e.g., the trajectory

based multi-channel images location GL^i , bearing GB^i , speed GS^i) may be rotated to ensure that the road direction is always horizontal in the image, this also has the additional benefit of simplifying the computation of bearing. Instead of using the absolute bearing values in the trajectory traces, the angle distance between the moving direction of the vehicle and the direction of road segment r_i may be computed to calculate GB^i . This is based on the observation that some road attributes such as the one-way/two-way road can be more correlated with the relative angle rather than the absolute bearing values.

[0050] Both calibration methods strengthen the features around the roads while weaken the features of the surrounding environments. This is especially helpful for the detection of road attributes such as number of lanes as the road width can be more easily recognized after calibration. However, the detection of some other road attributes such as speed limit and road type may also rely on the features of the surrounding environments, e.g., residential roads are always within residential areas. In some embodiments, only one of the calibration methods may be used, as better results may be obtained in comparison to using both calibration methods, as, in some cases, calibration may weaken the feature consistency of the surrounding environments too much by applying both calibration methods, resulting in less satisfactory detection rates of certain road attributes.

[0051] According to various embodiments classifying may include fusing the satellite image and the concatenated trajectory image into a fused image and input the fused image in a neural network stream of the trained classifier. By rendering trajectory traces into a multi-channel image, multimodal fusion can be directly conducted at the input layer by concatenating the channels of the satellite images (e.g., the RGB channels) and the location, bearing, and speed channels generated from the trajectory traces. This fusion strategy has the advantage of being able to learn filters from multimodal features as the satellite images and the trajectory traces are spatially aligned at the same rendering resolution. This strategy is also referred as early fusion in this disclosure.

[0052] In examples, providing trajectory data and satellite image of the geographical area may be carried out by extracting the image-based road features from both satellite images and crowdsourced GPS traces, respectively. The detection of each road attribute may then modeled as an image classification problem, in accordance with various embodiments. For example, a network consisting of five convolutional layers, followed by two fully-connected layers and one output layer may be adopted for the classification. The kernel size may be, e.g., 3 and the

number of filters may be set to, e.g., 64, 128, 256, 256, and 256, respectively. A stride of 1 may be adopted in the last three convolutional layers, while a stride of 2 may be adopted in the rest convolutional layers and the max pooling layers. However, the disclosure is not limited thereto.

[0053] According to some embodiments the classifier, which may be a trained classifier, may include a trajectory neural network stream. The classifier may include a satellite image neural network stream. The classifier may include a fully connected layer for receiving input from the trajectory neural network stream and the satellite image neural network stream and outputting a fused stream. The classifying may include inputting the concatenated trajectory image into the trajectory neural network stream, and inputting the satellite image in the satellite image neural network stream.

[0054] According to some embodiments, the trajectory neural network stream may be configured to process multiple trajectory images of the same geographical area including different times, wherein the multiple trajectory images may include the concatenated trajectory image 34. For example, each image of the multiple trajectory images may have a different timestamp, thus the multiple trajectory images may have a time axis. In examples, the trajectory neural network stream S1 may be configured as a Convolutional Recurrent Neural Network CRNN or a 3D Convolutional Neural Network.

[0055] According to various embodiments, the classifier may include a Convolutional Neural Network CNN. For example, the CNN may be selected from: a Dense Convolutional Network including a plurality of layers, wherein each layer of the plurality of layers may be feed-forward connected to every other layer (DenseNet); a CNN including a plurality of convolutional layers followed by fully connected layers, wherein pooling of outputs of neighboring groups of neurons may be performed with overlap (AlexNet); a CNN configured to process a plurality of layers via depthwise convolution and pointwise convolution (MobileNet).

[0056] According to various embodiments, the trained classifier may include: a first group including 2 first convolutional layers, followed by a second group including 3 second convolutional layers, followed by a max pooling layer, followed by a third group including 2 fully connected layers, followed by an output layer. Each convolutional layer of the first group and the second group may be followed by an activation unit, for example a Rectified Linear Unit (ReLU), which may be followed by a batch normalization layer and a max pool layer. Each fully connected layer of the third group may be followed by a respective activation unit,

for example, ReLU. Each convolutional layer of the second group 202 may be followed by an activation unit, for example, ReLU. The output layer may be further processed with softmax pooling layer.

[0057] In various embodiments, a ReLU activation function may be used by way of example, however the present disclosure is not limited thereto, and other activation functions may be used instead.

[0058] Embodiments relate to a data processing system including one or more processors configured to carry out the method of predicting road attributes.

[0059] Embodiments relate to a data processing system including one or more processors configured to carry out a method of predicting road attributes. The method may include a first memory configured to store trajectory data of the geographical area. The method may include a second memory configured to store a satellite image of the geographical area. The method may include a processor configured to calculate one or more image channels based on the trajectory data. The method may include a classifier, which may be a trained classifier, including a neural network configured to predict road attributes based on the one or more image channels and the satellite image; and an output layer configured to provide the prediction probabilities of the road attributes.

[0060] According to various embodiments, the system may be configured to concatenate the one or more image channels into a concatenated trajectory image before classifying the road segment.

[0061] According to some embodiments, the classifier may be configured to fuse the satellite image and the concatenated trajectory image into a fused image and input the fused image in a neural network stream of the trained classifier, for example, into a single neural network stream of the trained classifier. Accordingly, the classifier may, in some embodiments, have a single neural network stream.

[0062] In other embodiments, the classifier, which may be a trained classifier, may include a trajectory neural network stream and a satellite image neural network stream. The classifier may include a fully connected layer for receiving input from the trajectory neural network stream and the satellite image neural network stream and outputting a fused stream. The classifier may be configured to classify when the concatenated trajectory image is input into the trajectory neural network stream, and the satellite image is input in the satellite image neural network stream. The input images may be pre-processed, as disclosed herein.

[0063] According to various embodiments, the trajectory neural network stream may be configured to process multiple trajectory images of the same geographical area including different times, wherein the multiple trajectory images may include the concatenated trajectory image. In examples, the trajectory neural network stream may be configured as a Convolutional Recurrent Neural Network or a 3D Convolutional Neural Network.

[0064] An aspect of the disclosure relates to a non-transitory computer-readable medium storing computer executable code including instructions for predicting one or more road segment attributes according to the method of predicting one or more road segment attributes.

[0065] An aspect of the disclosure relates to a computer executable code including instructions for predicting one or more road segment attributes according to the method of predicting one or more road segment attributes.

[0066] An aspect of the disclosure relates to a method of training a classifier or an automated road attribute predictor including the classifier, the method including:

- performing forward propagation by inputting training data into the automated predictor to obtain an output result, for a plurality of road segments of a geographical area, wherein the training data may include trajectory data and satellite image having an electronic image format.

- performing back propagation according to a difference between the output result and an expected result to adjust weights of the automated predictor; and

- repeating the above steps until a pre-determined convergence threshold may be achieved. The automated predictor may include the classifier configured to provide the prediction probabilities of the road attributes. The classifier may include a neural network configured to predict road attributes based on one or more image channels and satellite image. The one or more image channels may be calculating based on the trajectory data. An aspect of the disclosure relates to a classifier trained by the method of training.

[0067] FIG. 1 shows an exemplary map including a geographical area 10 including city blocks 8 (e.g., residential blocks) and road segment 12. FIG. 2 shows a flow chart of a method 100 of predicting one or more road segment attributes 20 in accordance with various embodiments. According to various embodiments, the method 100 of predicting one or more road segment attributes 20 corresponding to a road segment 12 in a geographical area 10, may include:

- providing trajectory data 110 and satellite image 40 of the geographical area 10;

- calculating one or more image channels 120 based on the trajectory data 30; and
- using at least one processor, classifying 130 the road segment 12 based on the one or more image channels 32 and the satellite image 40 using a trained classifier 61 into prediction probabilities of the road attributes 20.

[0068] FIG. 3 shows exemplary details of providing trajectory data 110 in the form of a flow chart, in accordance with various embodiments. According to various embodiments, calculating one or more image channels 120 based on the trajectory data 30 may include at least one of:

- i) calculating a trajectory image channel 111, as part of the one or more image channels 120, based on the trajectory data 30;
- ii) calculating a bearing image channel 113, as part of the one or more image channels 120, based on the trajectory data 30;
- iii) calculating a speed image channel 116, as part of the one or more image channels 120, based on the trajectory data 30. According to various embodiments, the method 100 may include all of i), ii), and iii).

[0069] According to various embodiments, the method 100 may include concatenating 118 the trajectory image channel, the bearing image channel, and the speed image channel into a concatenated trajectory image 34 before classifying 130 the road segment 12. According to various embodiments, calculating the trajectory image channel 111 may include assigning a count of the number of trajectory points, e.g., GPS points, of the trajectory data 30 that may be projected onto each pixel of a matrix of pixels onto that pixel.

[0070] FIG. 4 shows exemplary details of calculating a bearing image channel 113 in the form of a flow chart, in accordance with various embodiments. According to various embodiments calculating a bearing image channel 113 may include providing a multichannel bearing image including multichannel pixels, wherein the number of channels represents a number of bearing bins, and quantizing bearing values, e.g., by degree, into the bearing bins forming a bearing histogram for each of the multichannel pixels.

[0071] FIG. 5 shows exemplary details of calculating a speed image channel 116 in the form of a flow chart, in accordance with various embodiments. According to various embodiments, calculating a speed image channel 116 may include providing a multichannel speed image including multichannel pixels, wherein the number of channels represents a number of speed

bins, and quantizing speed values, e.g., in m/s, into the speed bins forming a speed histogram for each of the multichannel pixels.

[0072] FIG. 6 shows exemplary details of data pre-processing before classifying 130 the road segment 12, in the form of a flow chart, in accordance with various embodiments. According to various embodiments, the method 100, may further include applying a smoothing filter 131 on the concatenated trajectory image 34 before classifying 130 the road segment 12. According to various embodiments, the method 100, may further include applying image rotation 133 until the road segment 12 in the concatenated trajectory image 34 may be aligned with the road segment 12 in satellite image 40 before classifying 130 the road segment 12.

[0073] FIG. 7 shows a flowchart of inputting data into the trained classifier when using early fusion. According to various embodiments, classifying may include fusing 140 the satellite image 40 and the concatenated trajectory image 34 into a fused image and input the fused image in a neural network stream of the trained classifier 61, for example in a single neural network stream.

[0074] FIG. 8 shows a flowchart of inputting data into the trained classifier when using late fusion. According to various embodiments the trained classifier 61 may include:

- a trajectory neural network stream S1;
- a satellite image neural network stream S2;
- a fully connected layer FC1 for receiving input from the trajectory neural network stream S1 and the satellite image neural network stream S2 and outputting a fused stream FS1.

The classifying may include inputting the concatenated trajectory image 34 into the trajectory neural network stream S1, and inputting the satellite image 40 in the satellite image neural network stream S2. According to various embodiments the trajectory neural network stream S1 may be configured to process a multiple trajectory images of the same geographical area 10 including different times, wherein the multiple trajectory images may include the concatenated trajectory image 34. In examples, the trajectory neural network stream S1 may be configured as a Convolutional Recurrent Neural Network CRNN or a 3D Convolutional Neural Network.

[0075] According to various embodiments, a data processing system 60, including one or more processors, may include a first memory configured to store trajectory data 30 of the geographical area 10, for which examples are illustrated in FIGS. 9 to 10. The data processing system 60 may include a second memory configured to store a satellite image 40 of the geographical area. The data processing system 60 may include a processor configured to

calculate one or more image channels 120 based on the trajectory data 30. The data processing system 60 may include a classifier 61 including a neural network configured to predict road attributes 20 based on the one or more image channels 32 and the satellite image 40. The classifier 61 may be a trained classifier. The classifier 61 may include an output layer configured to provide (e.g. to calculate) the prediction probabilities of the road attributes 20. The prediction probabilities may be stored, for example in a database 62. In another example, the prediction probabilities may be used to update map data stored in a database, e.g., database 62.

[0076] FIG. 9 shows one example of a structure of the classifier for early fusion in accordance with some embodiments.

[0077] According to various embodiments, the method 100 of any of the previous claims wherein the trained classifier 61 may include a Convolutional Neural Network (CNN). For example, the CNN may be selected from:

- a Dense Convolutional Network including a plurality of layers, wherein each layer of the plurality of layers may be feed-forward connected to every other layer, for example, DenseNet;
- a Convolutional Neural Network including a plurality of convolutional layers followed by fully connected layers, wherein pooling of outputs of neighboring groups of neurons may be performed with overlap, for example AlexNet;
- a Convolutional Neural Network, configured to process a plurality of layers via depthwise convolution and pointwise convolution, for example MobileNet.

[0078] According to various embodiments, the one or more image channels 32 may include at least two image channels, and wherein the data processing system 60 may be configured to concatenate 118 the one or more image channels 32 into a concatenated trajectory image 34 before classifying 130 the road segment 12. FIG. 9 shows the optional rotation “rot” of the satellite image 40, alternatively or in addition, the rotation may be applied to the one or more image channels 32 (or the concatenated trajectory image 34). FIG. 9 also shows the optional smoothening (“smooth”) of the concatenated trajectory image 34.

[0079] According to some embodiments, as exemplified by FIG. 9, the classifier 61 may be configured to fuse (“fuser”) the satellite image 40 and the concatenated trajectory image 34 into a fused image and input the fused image in a neural network stream of the trained classifier 61. The neural network may include a neural network stream S0, for example, a single neural

network stream. The classifier 61 may include an output layer “OUT1” configured to provide (e.g. to calculate) the prediction probabilities of the road attributes 20. The prediction probabilities may be stored, for example in a database 62. In another example, the prediction probabilities may be used to update map data stored in a database, e.g., database 62.

[0080] FIG. 10 shows one example of a structure of the classifier for late fusion in accordance with some embodiments.

[0081] According to some embodiments, the trained classifier 61 may include a trajectory neural network stream S1. The trained classifier may include a satellite image neural network stream S2. The trained classifier may include a fully connected layer FC1 for receiving input from the trajectory neural network stream S1 and the satellite image neural network stream S2 and outputting a fused stream FS1. FIG. 10 also shows the optional smoothening (“smooth”) of the concatenated trajectory image 34. Fusion of the trajectory neural network stream S1 with the satellite image neural network stream S2 occurs latter within the neural network at the fully connected layer FC1. The fully connected layer FC1 may be, but is not limited to be, the first fully connected layer of the neural network. The classifier may be configured to classify when the concatenated trajectory image 34 is input into the trajectory neural network stream S1, and the satellite image 40 is input in the satellite image neural network stream S2.

[0082] The classifier 61 may include an output layer “OUT1” configured to provide (e.g. to calculate) the prediction probabilities of the road attributes 20. The prediction probabilities may be stored, for example in a database 62. In another example, the prediction probabilities may be used to update map data stored in a database, e.g., database 62.

[0083] FIG. 11 shows one exemplary classifier structure 200, in accordance with some embodiments, for early fusion. According to some embodiments, the classifier 200, which may be a trained classifier, may include a first group 201, e.g., including 2 first convolutional layers (“Conv, 64, 3x3, 2x2” and “Conv, 128, 3x3, 2x2”), which may be followed by a second group 202. The classifier 200 may include the second group 202, e.g., including 3 second convolutional layers (“Conv, 128, 3x3, 2x2”), which may be followed by a max pooling layer (203, “Max Pool., 3x3, 2x2”). The classifier 200 may include the max pooling layer, which may be followed by a third group 204. The classifier 200 may include the third group 204, which may include 2 fully connected layers (“FC, 256”), which may be followed by an output layer 205. The classifier 200 may include the output layer 205. Each convolutional layer of the first group 201 and the second group 202 may be followed by a respective activation unit, for

example ReLU, which activation unit may be followed by a batch normalization layer and a max pool layer. Each convolutional layer of the second group 202 may be followed by a respective activation unit, for example, ReLU. Each fully connected layer of the third group 204 may be followed by a respective activation unit, for example, ReLU. The output layer 205 may be further processed with softmax pooling layer (“Softmax”). While a detailed layer structure and layer configuration is shown, the disclosure is not limited thereto. For example, other activation units, other number of layers, and other layer’s configuration (such as number of neurons) may be used.

[0084] According to some embodiments the classifier 61, which may be the trained classifier, may include a Convolutional Neural Network (CNN). The CNN may be, e.g., selected from: a Dense Convolutional Network including a plurality of layers, wherein each layer of the plurality of layers may be feed-forward connected to every other layer (DenseNet); [0085] a CNN including a plurality of convolutional layers followed by fully connected layers, wherein pooling of outputs of neighboring groups of neurons may be performed with overlap (AlexNet); a CNN configured to process a plurality of layers via depthwise convolution and pointwise convolution (MobileNet).

[0086] Various embodiments may relate to a computer executable code and/or to a non-transitory computer-readable medium storing the computer executable code including instructions for extracting road attributes according to the method of predicting one or more road attributes in accordance with various embodiments.

[0087] According to various embodiments, a data processing system may include one or more processors configured to carry out the method of predicting road attributes. The data processing system may be implemented in a computer. The data processing system may include a first memory configured to store trajectory data of the geographical area. For example, the trajectory data may be obtained from a server via a JavaScript Object Notation (JSON) request. The processing system may include a second memory configured to store satellite image, wherein the satellite image may include image data of the geographical area. For example, the satellite images may be stored in a server providing local and/or global digital maps, e.g., which may be accessed by a location. The processing system may include a satellite image extractor which may, e.g., crop map images to a pre-determined required size and/or for a pre-determined location (e.g. centered at a road segment). The processing system may include a classifier

configured to predict road attributes based on trajectory data and satellite image, in accordance with various embodiments. The classifier may include a neural network.

[0088] According to some embodiments, the trajectory neural network stream S1 may be configured to process multiple trajectory images of the same geographical area 10 including different times, wherein the multiple trajectory images may include the concatenated trajectory image 34. For example, each image of the multiple trajectory images may have a different timestamp, thus the multiple trajectory images may have a time axis. In some embodiments, the trajectory neural network stream S1 may be configured as a Convolutional Recurrent Neural Network CRNN or a 3D Convolutional Neural Network.

[0089] Various embodiments also relate to a method for training the classifier, or for training an automated road attribute predictor including the classifier. The method for training may include performing forward propagation by inputting training data into the automated predictor to obtain an output result, for a plurality of road segments 12 of a geographical area 10, wherein the training data includes trajectory data 30 and satellite image 40 having an electronic image format. The method for training may include performing back propagation according to a difference between the output result and an expected result to adjust weights of the automated predictor. The method for training may include repeating the above steps until a pre-determined convergence threshold may be achieved. The automated predictor may include the classifier 400 configured to provide the prediction probabilities of the road attributes 20. The classifier may include a neural network configured to predict road attributes 20 based on one or more image channels 32 and satellite image 40, wherein the one or more image channels 32 may be calculating based on the trajectory data 30.

[0090] FIGS. 12A and 12B shows illustration of the challenges for road attribute detection using only one of satellite images or GPS traces alone. In FIG. 12A, a satellite image is shown with heavy clouds that interfere with the visibility of the roads. In FIG. 12B, sparse GPS samples makes it more difficult to identify a narrow road.

[0091] As shown in FIG. 13A, with a high rendering resolution at 0.6 m/pixel, the projection of the original trajectory points around a road segment r_i can be noisy and sparse. This can be improved by smoothing each channel of G_i by computing the moving average over a square kernel with size K . FIG. 13B shoes the rendering of the smoothed GPS traces with $K=3$, and FIG. 13C shoes the rendering of the smoothed GPS traces with $K=5$.

[0092] FIG. 14 shows the architecture of an exemplary computer 300, which may be used in accordance with various embodiments to carry out the method of predicting road attributes, the method for training the classifier, and/or to implement the data processing system. The computer 300 includes a bus 301 through which one or more of the devices may communicate with each other. In the example of FIG. 17, the following devices are shown connected to the bus 301: a CPU 310; a main memory 320, for example a RAM; a storage device 330, for example a hard disk drive, a solid state drive, a flash drive; a communication device 340, for example for wired or wireless communication, e.g. WiFi, USB, Bluetooth; a display interface 332, and other user interfaces 360, for example for user input; however the disclosure is not limited thereto, and more or less devices may be included in the computer and the computer and/or bus may have other architectures than the one illustrated.

[0093] Various embodiments relate to a non-transitory computer-readable medium storing computer executable code including instructions for predicting one or more road segment attributes according to the method of predicting one or more road segment attributes. Various embodiments relate to a computer executable code including instructions for predicting one or more road segment attributes according to the method of predicting one or more road segment attributes. The computer executable code may be executed, e.g., in the above described computer architecture.

[0094] In the following examples, an exemplary experimental setup is explained, and then examples of the method for road attribute detection are evaluated. The effectiveness of the satellite images, trajectory data, and their fusion in the detection of road attributes is determined. Examples of road attributes are one-way/two-way road, number of lanes, speed limit, and road type. Examples of road types are “Residential”, “Service”, and “Footway”, “Primary”, “Secondary”, and “Tertiary”. For below experiments, a same input size and resolution for the images generated from different data sources is used for determining the effectiveness of both early fusion and late fusion strategies. Next, an ablation analysis is performed on the settings of bin number and kernel size for trajectory feature generation and smoothing to verify the design disclosed herein. Finally, the multimodal model is integrated into three state-of-the-art network architectures to demonstrate its effectiveness in road attribute detection.

[0095] Two large-scale real-world datasets of Singapore and Jakarta are used for the examples below. To prepare the datasets, the ground-truth labels are derived of four road

attributes, namely one-way/two-way road, number of lanes, speed limit, and road type, from the OpenStreetMap data. The road segments without ground-truth labels are removed and the remaining dataset is divided into 80%-20% splits for training and testing. The number of training and testing samples in each category (*i.e.*, each road attribute) is illustrated in the table of FIG. 15. There is shown one row for the Singapore dataset and one row for the Jakarta dataset, and four columns, showing, for each dataset, the training/testing dataset size for the attributes of one/two way, number of lanes, speed limit, and road type.

[0096] As only a few roads in Jakarta are annotated with the speed limit label no speed limit detection was performed on the Jakarta dataset. For feature extraction, satellite imagery is used (e.g., DigitalGlobe) and real-world trajectory traces of in-transit drivers in Singapore and Jakarta (e.g., GPS traces of Grab drivers). For comparison, the overall classification accuracy and the per-class F-measure are reported as the evaluation metrics.

[0097] FIG. 16 shows a table with the results for (a) Singapore and for (b) Jakarta, of classification accuracy obtained based on satellite images only, trajectory data only, early fusion of satellite images and trajectory data, late fusion of satellite images and trajectory data at the first fully connected layer FC1 (refer to FIG. 11, the first FC of group 204), and late fusion of satellite images and trajectory data at the second fully connected layer FC2 (refer to FIG. 11, the second FC of group 204), with the best result highlighted (bold) in each column. In the comparison, early fusion refers to concatenating the RGB channels of the satellite images and the location, bearing, and speed channels of the GPS traces at the input layer and a classifier architecture as described in FIG. 11 was used. While late fusion refers to processing satellite images and GPS traces with two separate network streams first and fusing the two networks at a later point by concatenating the features outputted by the FC1 or FC2 layers, a classifier architecture as described in FIG. 11 was used with the modification that the input streams are separated until the respective fusion takes place at FC1 or FC2. In this experiment, the number of bins to generate the bearing and speed channels from GPS traces were set to $M_b = 2$ and $M_s = 3$, respectively. The kernel size for GPS smoothing was set to 9×9 . The tuning of these parameters will be evaluated and discussed further below.

[0098] Comparative classifiers trained on satellite images or GPS traces alone have their own limitations. For example, the visibility of roads may not always be good due to occlusions caused by trees, buildings, or even heavy clouds in a satellite image. The crowdsourced trajectory traces, on the other hand, contain intrinsic noise resulting in incorrectly placed

trajectory points off the road. That classifiers trained in Singapore performed better than those trained in Jakarta is attributed to a quality of the Singapore dataset being better as the number of roads with ground-truth labels in each category is much higher than that in Jakarta, and, it is theorized that Singapore tends to have well-structured road networks that might be easier to be recognized. Due to the alignment of the satellite images and trajectory traces in the geospatial space (rotation to alignment or rotation to horizontal as disclosed herein), early fusion is able to learn the pixel-wise correspondences from them, therefore, on both datasets, early fusion outperformed late fusion in most of the cases. Comparatively, late fusion generally performed slightly worse than early fusion. However, late fusion has the flexibility that the method (and classifier) can be extended from the current trajectory data rendering of 2D to 3D (adding a time axis), where different network architectures may be used to process satellite images and GPS traces separately. The early fusion approach significantly improved the classification accuracy compared to the individual classifiers trained on satellite images and GPS traces separately. On the road type detection, the classification accuracy has been improved by 9.8% to 14.3%, which demonstrates the effectiveness of our proposed approach.

[0099] FIGS. 17A to 17B shows exemplary results of per-class F1 measure of road attribute detection based on satellite images, trajectory traces (in this example, GPS), and their fusion (using early fusion) on the speed limit and road type detection. FIG. 17A shows the speed limit and FIG. 17B shows the road type for data sample of Singapore. FIG. 18 shows exemplary results of per-class F1 measure of road attribute detection based on satellite images, GPS traces, and their fusion on the road type detection for Jakarta.

[00100] The per-class F-measure comparison of the classifiers trained based on satellite images, trajectory traces, and their early fusion is shown in FIGS. 17A, 17B and 18. F-measure is computed as $2 \cdot ((precision \cdot recall) / (precision + recall))$, which considers both precision and recall. For speed limit detection, the trajectory data (in the example, GPS) based classifier outperformed the satellite image based classifier in most of the cases except the class of 40 km/h. This is believed to be because the trajectory traces contain vehicle's bearing and speed information, which is missing from the satellite images. Satellite images provide additional information for road attribute detection. For road type detection, the satellite image based classifier performed better on classes of "Residential", "Service", and "Footway", while the trajectory based classifier performed better on classes of "Primary", "Secondary", and "Tertiary". The same trend shows on both Singapore dataset and Jakarta dataset. This is

because the contextual information in a satellite image shows the surrounding environment around a road. Thus, the satellite image helps with the detection of certain road types such as residential roads that are around residential areas and service roads that are for access to parking, driveways, and alleys. On the other hand, classes of “Primary”, “Secondary”, and “Tertiary” define the most, the second most, and the third most important roads in a country's road network system. Both the number of vehicles travelled on these roads and the quality of the corresponding trajectory traces tend to be high. So the trajectory based classifier tends to recognize such road types more easily.

[00101] FIGS. 19A to 19C show exemplary normalized confusion matrices of the method of predicting one or more road segment attributes in accordance with various embodiments. The vertical axis represents the true label and the horizontal axis represents the predicted label. FIG. 19A shows the confusion matrix for the speed limit for Singapore. FIG. 19B shows the confusion matrix for the road type detection for Singapore. And FIG. 19C shows the confusion matrix for the road type prediction for Jakarta. The six speed classes (also named as speed limit classes) chosen are 40, 50, 60, 70, 80, and 90 km/h. For the speed limit detection, the number of test samples for the six classes are 415, 1075, 299, 102, 15, and 17, respectively. We can see that most confusions are between common and rare classes. Due to the class imbalance, the detection of samples from rare classes is more challenging as the classifier tends to favor the majority class. For road type detection, the classifiers trained in both Singapore and Jakarta can recognize “Residential” and “Service” roads easily. “Primary”, “Secondary”, and “Tertiary” roads can be recognized more easily in Singapore than in Jakarta.

[00102] An ablation analysis on the key parameter settings in the trajectory data rendering. The impact of the number of bins/channels M_b and M_s that are used to render the sensor data into images are studied and the results shown in the table of FIG. 20. The first row reports the classification accuracy obtained by training with trajectory (in the example, GPS) locations only, which serves as a baseline when other sensor data is not available in the GPS traces. As previously explained, the GPS locations are rendered into a single-channel image GL that captures the spatial distribution of the GPS points by counting the number of GPS points that are projected onto each pixel. Comparatively, the M_b -channel image GB and the M_s -channel image GS rendered from bearing and speed tend to be more descriptive as they capture not only the spatial distribution of the GPS points, but also the local distribution of vehicles' moving direction and speed at each pixel. By comparing the results shown in rows 2 to 4, it can be seen

that improved classification accuracy can be obtained by increasing the number of channels M_b and M_s especially for the detection of speed limit and road type. The last row in the table shows the classification accuracy obtained by training based on the fusion of GL , GB , and GS , where the input $G = \text{Concat}((GL, GB, GS), \text{axis} = -1)$. As features extracted from GPS location, bearing, and speed capture complementary information, the classifier trained on 7-channel G outperformed the classifier trained on 7-channel GB or 7-channel GS in the six out of the seven cases.

[00103] The results of a study of the impact of the kernel size adopted for GPS smoothing is shown in the table of FIG. 21. The kernel size is set to 1×1 , 3×3 , 5×5 , 7×7 , 9×9 , and 11×11 and the results are reported in FIG. 21. The advantages of applying kernel smoothing to GPS traces are twofold. First, it helps reduce the negative impact caused by sensor noise. Second, crowdsourced GPS traces can be sparse in some regions where pixels on a road can have no GPS points projected to it, especially when we render at a high resolution. As FIG. 21 shows, classifiers trained with smoothed GPS traces outperformed the classifiers trained with the original GPS traces by a large margin. The kernel size of 9×9 is considered to be a good choice based on the experimental results on the used datasets, and this kernel size is used as exemplary setting in the experiments herein.

[00104] The disclosed multimodal fusion solution of road features from satellite images and GPS traces can be easily integrated with any existing network architectures to train an end-to-end classifier for road attribute detection. To demonstrate the effectiveness, the baseline network (FIG. 11) was replaced by a number of state-of-the-art image classification architectures including AlexNet, MobileNet, and DenseNet. The table in FIG. 22 reports the classification accuracy based on different combinations of input modules and model architectures on the Singapore dataset. The size of the AlexNet is similar to the baseline network, both of which consist of five convolutional layers followed by ReLU activations. AlexNet adopted a kernel size of 11×11 and 5×5 respectively in the first two convolutional layers, which is the biggest among the four models. Comparatively, MobileNet and DenseNet went much deeper. To improve a model's efficiency, MobileNet factorized a standard 2D convolution into a depthwise convolution and a 1×1 pointwise convolution. DenseNet divided the network into multiple densely connected dense blocks composed of narrow layers, which are connected by transition layers that perform 1×1 convolution and 2×2 average pooling. From the results it can be seen that all models obtained significant performance gain

by applying our proposed multimodal fusion solution. For the road attribute detection problem, DenseNet achieved the best classification accuracy based on satellite images, but it performed less robust when dealing with GPS traces. Better results were obtained with multimodal fusion, for the baseline network, MobileNet, and DenseNet, on the detection of the four road attributes. [00105] FIG. 23 shows a table reporting the classification accuracy and the performance gain of the disclosed multimodal feature fusion with calibration (as explained above) compared to the classifiers trained without calibration. By applying both image rotation and bearing adjustment calibrations, the model disclosed herein significantly improved the detection accuracy of the one-way/two-way road and the number of lanes by 6.7% and 6.9%, respectively. For speed limit and road type, the best detection rates were obtained by applying the bearing adjustment calibration only. One possible reason is that the detection of these road attributes partially relies on the features of the surrounding environments. For example, highways in the same downtown area may have similar speed limit and residential roads are always within the residential area. The best classification accuracy obtained by the base model is 90.6%, 74.5%, 83.0%, and 80.7% for one-way/two-way road, number of lanes, speed limit, and road type detection, respectively.

[00106] A raw trajectory trace is noisy and does not contain the information of the true route the vehicle travelled. Therefore, traditional GPS-based road attribute detection methods mostly perform map matching algorithms to find the group of traces that are associated with each road segment in the preprocessing phase. However, the effectiveness of map matching algorithms can be significantly degraded by both low-sampling-rate GPS traces and incomplete road networks, not to mention the huge computational cost when pre-processing a large number of GPS traces. To solve the above issues, it is disclosed herein the method to render trajectory traces as a multi-channel image that can be directly passed to a classifier without applying map matching algorithms. The disclosed trajectory rendering is more efficient than the traditional map matching based road feature extraction methods. Moreover, both the sampling rate of the trajectory traces and the completeness of the map have little impact on the effectiveness of the method in accordance with various embodiments.

[00107] A multimodal fusion framework that learns from both satellite images and crowdsourced GPS traces for robust road attribute detection is disclosed herein. Accordingly, a method of predicting one or more road segment attributes, a data processing system, a non-transitory computer-readable medium storing computer executable code, and a method for

training an automated road attribute predictor are disclosed herein, among others. In order to learn multimodal pixel-wise correspondences, trajectory traces may be rendered into a multi-channel image that align with satellite images in the spatial domain. Moreover, the trajectory rendering method does not require map matching to preprocess the raw trajectory traces. Thus, the herein disclosed method is less sensitive to the sampling rate of trajectory traces compared to traditional trajectory based road feature extraction methods.

[00108] While the disclosure has been particularly shown and described with reference to specific embodiments, it should be understood by those skilled in the art that various changes in form and detail may be made therein without departing from the spirit and scope of the invention as defined by the appended claims. The scope of the invention is thus indicated by the appended claims and all changes which come within the meaning and range of equivalency of the claims are therefore intended to be embraced.

CLAIMS

1. A method (100) of predicting one or more road segment attributes (20) corresponding to a road segment (12) in a geographical area (10), the method (100) comprising:
 - providing trajectory data (110) and satellite image (40) of the geographical area (10);
 - calculating one or more image channels (120) based on the trajectory data (30); and
 - using at least one processor, classifying (130) the road segment (12) based on the one or more image channels (32) and the satellite image (40) using a trained classifier (61) into prediction probabilities of the road attributes (20).
2. The method (100) of claim 1, wherein calculating one or more image channels (120) based on the trajectory data (30) comprises at least one of:
 - i) calculating a trajectory image channel (111), as part of the one or more image channels (32), based on the trajectory data (30);
 - ii) calculating a bearing image channel (113), as part of the one or more image channels (32), based on the trajectory data (30);
 - iii) calculating a speed image channel (116), as part of the one or more image channels (32), based on the trajectory data (30).
3. The method (100) of claim 1, comprising i), ii), and iii).
4. The method (100) of claim 3, wherein calculating the trajectory image channel (111) comprises assigning a count of the number of trajectory points of the trajectory data (30) that are projected onto each pixel of a matrix of pixels onto that pixel.
5. The method (100) of claim 3 or claim 4, wherein calculating a bearing image channel (113) comprises providing a multichannel bearing image comprising multichannel pixels, wherein the number of channels represents a number of bearing bins, and quantizing bearing values into the bearing bins forming a bearing histogram for each of the multichannel pixels.
6. The method (100) of any one of claims 3 to 5, wherein calculating a speed image channel (116) comprises providing a multichannel speed image comprising multichannel pixels,

wherein the number of channels represents a number of speed bins, and quantizing speed values into the speed bins forming a speed histogram for each of the multichannel pixels.

7. The method (100) of any one of claims 3 to 6, comprising concatenating (118) the trajectory image channel, the bearing image channel, and the speed image channel into a concatenated trajectory image (34) before classifying (130) the road segment (12).

8. The method (100) of claim 7, further comprising applying a smoothing filter (131) on the concatenated trajectory image (34) before classifying (130) the road segment (12).

9. The method (100) of claim 7 or claim 8, further comprising applying image rotation (133) until the road segment (12) in the concatenated trajectory image (34) is aligned with the road segment (12) in satellite image (40) before classifying (130) the road segment (12).

10. The method (100) of any one of claims 7 to 9, wherein classifying comprises fusing (140) the satellite image (40) and the concatenated trajectory image (34) into a fused image and input the fused image in a neural network stream of the trained classifier (61).

11. The method (100) of any one of claims 7 to 9, wherein the trained classifier (61) comprises:

a trajectory neural network stream (S1);

a satellite image neural network stream (S2);

a fully connected layer (FC1) for receiving input from the trajectory neural network stream (S1) and the satellite image neural network stream (S2) and outputting a fused stream (FS1),

wherein the classifying comprises inputting the concatenated trajectory image (34) into the trajectory neural network stream (S1), and

inputting the satellite image (40) in the satellite image neural network stream (S2).

12. The method (100) of claim 11, wherein the trajectory neural network stream (S1) is configured to process a multiple trajectory images of the same geographical area (10)

comprising different times, wherein the multiple trajectory images include the concatenated trajectory image (34).

13. The method (100) of any of the previous claims wherein the trained classifier (61) comprises a Convolutional Neural Network.

14. The method (100) of claim 13 wherein the trained classifier (61) comprises:

- a first group comprising 2 first convolutional layers, followed by
- a second group comprising 3 second convolutional layers, followed by
- a max pooling layer, followed by
- a third group comprising 2 fully connected layers, followed by
- an output layer;

wherein each convolutional layer of the first group and the second group is followed by an activation unit, and

wherein each fully connected layer of the third group is followed by a respective activation unit.

15. A data processing system (60) comprising one or more processors configured to carry out a method (100) of predicting road attributes (20), the data processing system (60) comprising:

- a first memory configured to store trajectory data (30) of the geographical area (10);
- a second memory configured to store a satellite image (40) of the geographical area;
- a processor configured to calculate one or more image channels (120) based on the trajectory data (30);
- a trained classifier (61) comprising a neural network configured to predict road attributes (20) based on the one or more image channels (32) and the satellite image (40); and
- an output layer configured to classify an output of the neural network into prediction probabilities of the road attributes (20).

16. The data processing system (60) of claim 15, wherein the one or more image channels (32) comprise at least two image channels, and wherein the system (60) is configured to concatenate (118) the one or more image channels (32) into a concatenated trajectory image (34) before classifying (130) the road segment (12).

17. The data processing system (60) of claim 15 or claim 16, wherein the classifier is configured to fuse (140) the satellite image (40) and the concatenated trajectory image (34) into a fused image and input the fused image in a neural network stream of the trained classifier (61).

18. The data processing system (60) of any one of claims 15 to 16, wherein the trained classifier (61) comprises:

- a trajectory neural network stream (S1);

- a satellite image neural network stream (S2);

- a fully connected layer (FC1) for receiving input from the trajectory neural network stream (S1) and the satellite image neural network stream (S2) and outputting a fused stream (FS1),

wherein the trained classifier (61) is configured to classify when the concatenated trajectory image (34) is input into the trajectory neural network stream (S1), and the satellite image (40) is input in the satellite image neural network stream (S2).

19. The data processing system (60) of claim 18, wherein the trajectory neural network stream (S1) is configured to process multiple trajectory images of the same geographical area (10) comprising different times, wherein the multiple trajectory images include the concatenated trajectory image (34).

20. The data processing system (60) of any one of claims 15 to 19, wherein the trained classifier (61) comprises a Convolutional Neural Network.

21. A non-transitory computer-readable medium storing computer executable code comprising instructions for predicting one or more road segment attributes according to the method of any one of claims 1 to 14.

22. A computer executable code comprising instructions for predicting one or more road segment attributes according to the method of any one of claims 1 to 14.