Automated Pavement Distress Detection and Deterioration Analysis Using Street View Map

XU LEI¹, (Member, IEEE), CHENGLONG LIU², LI LI¹, AND GUIPING WANG¹

¹School of Electronics and Control Engineering, Chang’an University, Xi’an 710064, China
²Key Laboratory of Road and Traffic Engineering of the Ministry of Education, Tongji University, Shanghai 201804, China

Corresponding author: Chenglong Liu (lcl_tongji@126.com)

This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFB1600600, in part by the Key Research and Development Program of Shaanxi Province, China, under Grant 2019ZDLGY15-04-02, Grant 2018ZDCXL-GY-05-07-02, and Grant 2019GY-059, in part by the National Natural Science Foundation of China under Grant 71901040, in part by the Key Science and Technology Program of Shaanxi Province, China, under Grant 2019JQ-442, and in part by the Fundamental Research Funds for the Central Universities under Grant 300102329502.

ABSTRACT Automated pavement distress detection benefits road maintenance and operation by providing the condition and location of various distress rapidly. Existing work generally relies on manual labor or specific algorithms trained by dedicated datasets, which hinders the efficiency and applicable scenarios of methods. Street view map provides interactive panoramas of a large scale of urban roadway network, and is updated in a recurrent manner by the provider. This paper proposed a deep learning method based on a pre-trained neural network architecture to identify and locate different distress in real-time. About 20,000 street view images were collected and labeled as the training dataset using the Baidu e-map. Eight types of distress are notated using Yolov3 deep learning architecture. The scale-invariant feature transform (SIFT) descriptors combined with GPS and bounding boxes were applied to judge the deterioration of the distress. A decision tree was designed to evaluate the change of the distress over some time. A typical road in Shanghai was selected to verify the effectiveness of the proposed model. The images of the road from 2015 to 2017 were collected from the street view map. The results showed that the mean average precision of the proposed algorithm is 88.37%, demonstrating the vast potential of applying this method to detect pavement distress. 43 distress were newly generated, and 49 previous distress were patched in the two years. The proposed method can assist the authorities to schedule the maintenance activities more effectively.

INDEX TERMS Pavement distress detection, street view map, deterioration mode, deep learning, scale-invariant feature transform.

I. INTRODUCTION Pavement distress detection plays a vital role in road maintenance and management. It provides essential information for various pavement performance evaluation, including crack, pothole, corrugation, reveling, patching, etc. Many pavement performance evaluation indexes, such as pavement condition index (PCI), maintenance quality indicator (MQI), are established based on the condition of pavement distress. According to the World Bank report: on an annualized basis each dollar spent on patching and overlays saves at least $1.4 in operating costs and can save as much as $44 depending on traffic volume [1]. The statistic estimated that for every additional $1 a developing country spends road maintenance, road users save $3 [2]. Accurate and efficient pavement distress detection provides reliable data support for maintenance decisions, thereby improving the utilization of maintenance budgets.

Most traditional pavement distress detection techniques were carried out manually or semi-automated by trained technicians [3]. The manual visual survey is labor-intensive and time-consuming, which requires lots of images collected in advance by dedicated vehicles. The survey process is significantly affected by human subjectivity, especially in the case when the working time is extended. Moreover, some distress are quite similar and easy to be confusing. Thus, all the labeling people have to be well-trained and familiar with this area. According to the Canadian infrastructure report, 71% of municipalities collect data on their roads at least once every five years [4], yet the data update frequency of which...
is still relatively low due to the complexity of the detection methods.

![Pavement distress detection using different methods.](image)

Many automated devices have been invented to improve the efficiency and accuracy of pavement distress detection in the recent decade. They can be roughly divided into three types according to the equipment. The first type utilizes the 3D LiDAR or depth camera to automatically identify various distress, such as ARAN profile system [5], ARRB Hawkeye [6], and others. This method makes use of ultraviolet, visible, or near-infrared light to image objects [7], as shown in Fig. 1(a) and (b). Some distress with noticeable height change, such as pothole and rutting, can be easily located using 3D data, the others can be identified using the combination of RGB and point cloud images. LiDAR creates a considerable point cloud image for the subject and, is therefore a relatively expensive tool that limits the number of users [8]. The second type applied ground penetrating radar (GPR) or infrared camera to detect pavement distress [9]. This approach can identify the defects in deep pavement and the distress with the depth information [10]. It’s particularly sensitive to the cracks, as shown in Fig. 1(c). However, the GPR or thermal devices are also expensive and require massive work to be mounted on the vehicles. The third type is the most commonly used method using 2D RGB images, as shown in Fig. 1(d), which relies on a simple fixed camera to collect the pavement images along the road [11]. Then the image recognition algorithms are implemented on these images to classify and locate the distress automatically. Compared with the other two types, the 2D camera-based method is economical and efficient, making it applicable to many scenarios. Nevertheless, the challenges associated with this approach, such as multiple image sources and shooting angle, non-uniformity of various pavement distress, lack of sufficient illumination, and other ambient interferences, continue to keep this area active [12].

Street view map is a technology that provides interactive panoramas from multiple positions along the roads [13]. Many e-map vendors provide street view map services, such as Google, Baidu, and NAVER throughout the world. The images are collected by professional vehicles and have different shooting angles. Besides, they cover the most urban and highway roads in the country, making them very suitable and economical to be the dataset for distress detection. The images are updated periodically to guarantee the effectiveness of the street view map, which enables us to trace the deterioration of different distress. Some studies have involved the street view images into the dataset and calculated image descriptors like SIFT to detect and locate the distress [14]. However, the results are greatly affected by the other elements in street view images, such as humans, vehicles, buildings and trees.

The algorithms of pavement distress recognition using 2D images have come a long way from texture features extraction to deep convolutional neural networks [15]. For model-based methods, rudimentary attempts were undertaken to detect pavement defects using intensity-thresholding methods, such as histogram [16] and background subtraction [17]. Edge detection has then been predominant for crack and pothole detection for a long time, which involves the use of Canny filter, pixel intensity, Sobel operator, etc. [18]–[20]. The multi-scale wavelet transform was another widely recognized algorithm that separates the distress from the background [21]–[23]. Multi-order texture features, such as linear binary pattern (LBP) indicators [24], Multi-order texture features (GLCM) [25], and morphology gradient operator [26] are extensively applied to extract the core characteristics of the distress. The model-based algorithms are stable and rapid, as well as effective in some specific scenarios. However, they may fail to detect diverse distress under other generalized circumstances due to the non-uniformity and variety of distress morphologies.

Deep learning has led to rapid progress in the area of artificial intelligence, such as recognizing images, controlling automated vehicles, and identifying semantics [27]. The potential of applying this technology on pavement distress detection has been verified through many studies in recent years, of which transfer learning involves the knowledge gained from one task to a different but related task [28]. For example, Gopalakrishnan et al. developed a transfer learning structure with a deep convolutional neural network (CNN) to identify the cracks in the imageNeT database [29]. Zhang and Cheng conducted the transfer learning model on the pixel level of images, which achieved an outstanding performance in terms of recall rate [30]. Furthermore, Nie et al. proposed a transfer learning method based on R-CNN to detect pavement cracks considering multiple complex environments [31]. Peng et al. developed deep clustering method, which jointly learns clustering assignment and representation [32]. Many Pre-trained CNNs have been successfully applied in this area. Song and Wang trained 20 faster R-CNNs to select the optimal network architecture for distress inspection [33]. Wu et al. utilized the Single Shot MultiBox Detector (SSD) with mobileNet to classify the damages for infrastructure based on UVA images [34]. Du et al. improved the YOLO
network to make it more suitable for detecting multiple types of pavement distress [35]. However, most of them are trained based on specified datasets, whose image quality and shooting angles are entirely distinct. Hence, the neural network trained by some dataset may not be applicable to the other one. Even for the large and popular dataset like ImageNet, there is only a small proportion of the images depicting pavement conditions.

The texture identification algorithms and end-to-end deep neural networks are difficult to directly apply on street view images because (1) street view images contain enormous information, such as persons, buildings, and trees. Compared with them, pavement distress are unimpressive and tiny, making the texture-based method hard to identify them from complicated scenarios; (2) the morphologies of the same type of distress are quite similar. Matching the distress at different times requires lots of labeled data to train the model, and the matching performance is easy to be affected by the perspective, rotation, and scale of the image. Therefore, we designed a novel hybrid pipeline to estimate the large-scale pavement condition using street view map.

This paper combines the advantages of deep convolutional neural network and image texture features to propose a hybrid model for distress detection and matching based on the street view images. More than 19665 street view images with multiple pavement distress are collected via the Baidu application program interface (API) for training. A two-stage pipeline is developed to identify, locate, and match the distress automatically. We further used the identified data to evaluate the pavement performance and compared the change of the distress over a span of time. The rest of the paper is organized as follows: the dataset and its acquisition method are introduced in Section 2. Section 3 demonstrated the methodology of using Yolov3 to identify different distress. Section 4 represents how to match and trace the change of distress in varied images from different times. Section 5 discusses the results of the proposed method and draws some implications. Section 6 concludes the paper.

II. IMAGE COLLECTION USING STREET VIEW MAP API

A. IMAGE COLLECTION USING STREET VIEW MAP

Compared with the open datasets, such as Google open-accessed street view dataset, the BaiduMap API can provide us the images from a specific view. Besides, the API can trace the images for the same position at different times. Therefore, we utilized the Baidu map to analyze the deterioration of pavement. The images of pavement are extracted from the BaiduMap platform, which allows us to request the street view images in an HTTP URL form using the BaiduMap API. By defining the URL parameters including the GPS (location), the heading angle (heading), the pitch angle (pitch), the size of image (width and height) and the horizontal field view (fov), users can get static street view images in any angle for any point where the street view is available. An example of the HTTP URL is shown as follows:

http://api.map.baidu.com/panorama/v2?ak=XoZbvs2GX1PYUWBp48vuMHg2B2zCbfq9&width=1024&height=512&location=121.2205276,31.26821327&fov=90&heading=90&pitch=30&coordtype=wg84II

In this study, the GPS information of the testing road was firstly collected using a high-resolution vehicle-mounted GPS module. Then, a series of GPS data with a uniform space interval of 10 m was calculated using cubic spline interpolation. By defining the GPS information and generating the corresponding HTTP URL, the images collected for distress identification were returned with a size of 1024 × 512. To better collect the pavement information instead of the other objects such as cars, buildings and the sky, we set the pitch angle and the horizontal field view to 30° and 90°. The street view images of the previous time can also be obtained by the “Timeline” function, as shown in Fig.2. This paper came up with the idea of using the street view map to track the change of large-scale pavement conditions. The proposed method is efficient and economical without the need of dedicated measurement devices. Multiple pavement distress can be detected and located based on the deep neural network. The change of each distress can also be identified using the Timeline function of the street view map. Although the street view images are not real-time collected, they can reflect the pavement condition of the whole road network. In addition, the street view images are updated periodically, which can be used to analyze the deterioration and the change of pavement conditions at different times.

(a) Street view images in 2015.01

(b) Street view images in 2015.12

FIGURE 2. Examples of the images collected by street view map at different time.

B. DATASET DESCRIPTION

Five types of distress and three types of maintained distress are labeled for street view images, including deformation,
pothole, loose, net-crack, cracks, patched-pothole, patched-net, patch-crack. These categories can cover most common pavement defects. Through the detection of patched distress, we can compare the deterioration of various distress at different times.

Statistics on the number of labeled distress are shown in Table 1. Due to the different probabilities of occurrence of various distress, the amount of different distress in the dataset is greatly biased. The image augmentation was performed to improve the generalization performance of the model and alleviate the bias. Random rotation and Gaussian blur were combined to enhance the dataset. The statistics after data augmentation are illustrated in Table 1 as well. 80% of the images are used as the training dataset, and the rest belongs to the testing dataset.

### III. CONVOLUTIONAL NEURAL NETWORK FOR DISTRESS DETECTION

A deep learning-based object detection model YOLOv3 (You Only Look Once version 3) was applied to detect and locate various pavement distress at the same time. YOLO is one of the pre-trained models with the best general performance of object detection tasks in ImageNet datasets [36]. The main advantage of Yolo is that it is a one-stage model, which eliminates the process of traditional region proposal algorithm and unifies the object detection and classification into a regression problem. This one-stage framework takes the original image as input and directly outputs the prediction results, achieving an end-to-end training using multi-scale feature extraction.

Compared with the previous versions of YOLO and other object detection models, YOLOv3 mainly has the following merits. It adjusted the feature extraction network structure to capture the more stable local information and conducted the multi-scale feature fusion for object detection. Also, a multi-level loss function is designed to improve the performance of the model.

#### A. FEATURE EXTRACTION

Compared to model-based algorithms that rely on manually designed features, the most substantial advantage of deep learning is that it can learn features from training data to support subsequent regression and classification. The design of the feature extraction network is the key to identify multiple objects. A stabilized feature extraction network can not only learn accurate and dense features but also speed up the process of local feature learning. The Darknet-53 in Yolov3 is a fully convolutional network that contains 53 convolutional layers, whose network structure is shown in Table 2. It canceled the traditional down-sampling method using max-pooling or average-pooling and replaced it with a convolution kernel with a size of $2 \times 2$ and a stride of 2. Moreover, the practice of the Residual network [37] is borrowed. Shortcut connections are set up between layers in the network, which benefits the network from three perspectives: (1) The structure of residuals can alleviate the disappearance of gradients in the process of deepening the network layer, and ensure that the model is fully trained; (2) The deeper the network layer, the more accurate its feature extraction will be; (3) It also greatly reduces the channel for each convolution, as well as the amount of calculation, and accelerates the model recognition.

#### B. MULTI-SCALE FEATURE FUSION

In YOLOv3, feature maps of three different scales are utilized for distress detection. As shown in Fig. 3, after 79 layers, the convolutional network passes the yellow convolution layer to get the detection result of the first scale. Compared with the input image, the feature map used for detection has been 32 times down-sampling. For example, if the input is $416 \times 416$, the feature map here is $13 \times 13$. It is suitable to detect the large-sized objects in the image because

| Type     | Filters | Size   | Output   |
|----------|---------|--------|----------|
| Convolutional | 32      | 3×3    | 256×256  |
| Convolutional | 64      | 3×3/2  | 128×128  |
| Residual   | 64      | 3×3    | 128×128  |
| Convolutional | 128     | 3×3/2  | 64×64    |
| Convolutional | 64      | 1×1    | 64×64    |
| Residual   | 128     | 3×3    | 64×64    |
| Convolutional | 256     | 3×3/2  | 32×32    |
| Convolutional | 256     | 1×1    | 32×32    |
| Residual   | 512     | 3×3/2  | 16×16    |
| Convolutional | 512     | 3×3    | 16×16    |
| Residual   | 1024    | 3×3/2  | 8×8      |
| Averpool  | 1024    | 3×3    | 8×8      |
| Softmax   |         |        |          |
the down-sampling factor is high and the receptive field of the feature map is relatively large. To achieve fine-grained detection, the feature map after the 79th layer starts to be up-sampled again. A fine-grained feature map of the 91st layer is obtained by concatenating it with the 61st layer in the feature map. It also obtains a feature map of one-sixteenth of the input image size after several convolutional layers. It has a mesoscale receptive field, and it is appropriate for detecting mesoscale objects. Finally, the 91st layer feature map is up-sampled again and fused with the 36th layer feature map. Finally, a feature map with eight times down-sampled relative to the input image is obtained. It has the smallest receptive field and thereby applied for recognizing the small-sized objects.

The size of the anchor also requires to be adjusted according to the change of the number and scale of the feature maps. YOLOv3 uses K-means clustering to obtain the size of the anchor. Each type of down-sampling scale is set with three kinds of anchors, and a total of nine sizes of anchors are clustered. The sizes of the anchors used in this paper are indicated in Table 3.

### TABLE 3. Sizes of each anchor.

| The large receptive field | The mesoscale receptive field | The small receptive field |
|---------------------------|------------------------------|---------------------------|
| 124×76                    | 27×66                        | 17×21                     |
| 253×169                   | 67×52                        | 20×33                     |
| 387×326                   | 67×132                       | 34×27                     |

In practical, a large anchor is applied on the smallest 13 × 13 feature map (with the largest receptive field), which is designed for detecting the large distress, such as net-crack, patched pothole, and deformation. The medium 26 × 26 feature map (medium receptive field) applies a mesoscale anchor to detect medium-sized distress, like potholes. A smaller anchor is used to the larger 52 × 52 feature map (the minor receptive field) to detect small objects such as cracks and patched cracks.

### C. MULTI-LEVEL LOSS FUNCTION

The loss function is a criterion for judging the quality of model training. YOLOv3 has designed a multi-level loss function, as it is a one-stage and completely end-to-end object detection network, as shown in (1). In this unified loss function, the center coordinate error, width and height coordinate error, confidence error and, classification error are included.

\[
\text{Loss} = \hat{\theta} \sum_{i=0}^{N^2} \sum_{j=0}^{B} R^\text{obj}_{i,j} \left[ -\hat{y}_i \log y_i + (1 - \hat{y}_i) \log(1 - y_i) \right] \\
- \hat{x}_i \log x_i + (1 - \hat{x}_i) \log(1 - x_i) \\
+ \hat{\theta} \sum_{i=0}^{N^2} \sum_{j=0}^{B} R^\text{obj}_{i,j} \left[ (\sqrt{w_i} - \sqrt{w}_{i,j})^2 + \left(\sqrt{h_i} - \sqrt{h}_{i,j}\right)^2 \right] \\
+ \sum_{i=0}^{N^2} \sum_{j=0}^{B} R^\text{obj}_{i,j} (-\hat{C}_i \log C_i + (1 - \hat{C}_i) \log(1 - C_i)) \\
+ \sum_{i=0}^{N^2} \sum_{j=0}^{B} R^\text{obj}_{i,j} (-\hat{p}_i(c) \log p_i(c)) \\
+ (1 - \hat{p}_i(c)) \log(1 - \hat{p}_i(c)) \right)
\]

where, \(R^\text{obj}_{i,j}\) represents whether there is an object in the \(j\)th prediction box of the \(i\)th grid. If true then \(R^\text{obj}_{i,j} = 1\), otherwise \(R^\text{obj}_{i,j} = 0\). \(R^\text{nonobj}_{i,j}\) indicates the opposite. \((\hat{x}, \hat{y}, \hat{w}, \hat{h}, \hat{C}, \hat{p})\) are predicted value, while \((x, y, w, h, C, p)\) are the ground truth.

The parameters of the pre-trained deep neural network are calibrated by the training dataset collected from the street view map. The application of the trained Yolov3 network mainly includes three stages: feature extraction, pre-selection region proposal, and regression of category and coordinate. First, images of patch-potholes were utilized to fine-tune the proposed feature extraction network because the patch-potholes and the surrounding pavement have a large gray difference. Using such images for network fine-tune makes the original feature extraction network more applicable for pavement distress. Based on the feature map, we used k-means method to cluster the bounding box sizes for each type of disease as the size of anchors, as shown in Table 3. A pre-selection box of this size can be closer to the size of the disease, and the accuracy of the bounding box coordinates estimation can also be improved.

### IV. DISTRESS DETERIORATION ANALYSIS

#### A. IMAGE RETRIEVAL BASED ON GPS

As mentioned in the introduction, the street view images are updated periodically, which can be applied to evaluate the deterioration of the pavement. The data update frequency can be a year to two years. During this time, there are three
potential situations for distress: (1) the distress is repaired or patched; (2) the distress deteriorates to another status; (3) the distress stays the same condition, as shown in Fig. 4.

The first step of tracing the change of distress is to match the images from different times. As the street view images contain the GPS information of the position pictured, we preliminarily used the Euclidean distance of different images to screen out the distress at the same location. The Haversine equation was adopted to estimate the distance of two images, as shown in (2).

\[
d = 2r \arcsin \left( \sqrt{\sin^2 \left( \frac{\varphi_2 - \varphi_1}{2} \right) + \cos (\varphi_1) \cos (\varphi_2) \sin^2 \left( \frac{\lambda_2 - \lambda_1}{2} \right)} \right)
\]

(2)

where \( d \) is the Euclidean distance between image 1 and 2, \( \varphi_1 \) and \( \varphi_2 \) are the latitude coordinates, \( \lambda_1 \) and \( \lambda_2 \) are the longitude coordinates. However, the measurement error of GPS is about 0.1m to 1m, which may lead to the inaccurate matching of images. In this study, we conservatively collected the images whose distances are less than three meters for the next comparison. Even though the images are collected in the same position, the shooting angles may be slightly different, which potentially causes the error as well. Therefore, we proposed a pixel-level matching based on the results of GPS matching. The framework of tracing the change of distress at different times is illustrated in Fig. 5.

**B. IMAGE PERSPECTIVE TRANSFORMATION**

One of the advantages of street view maps is that it can output the external parameters of the vehicle-mounted camera when retrieving the image from a fixed GPS. This feature facilitates the process of perspective conversion, which is the prerequisite of comparing two images. The relationship between world coordinates \((x_w, y_w, z_w, 1)\) and pixel coordinates \((u, v, 1)\) is expressed as (3) [38]:

\[
Z_c \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = K \cdot \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix}
\]

(3)

where \( K \) indicates the intrinsic camera parameters, which is constant in this case. The external camera parameter matrix includes the rotation matrix \( R \) and translation matrix \( T \). For street view map, the translation matrix \( T \) can be regarded as zero, and the rotation matrix \( R \) is able to be output by API. Therefore, the perspectives of the two images could be transformed into the same one for distress comparison according to (4).

\[
\begin{bmatrix} u_2 \\ v_2 \\ 1 \end{bmatrix} = \begin{bmatrix} u_1 \\ v_1 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} R_1 & T_1 \\ 0 & 1 \end{bmatrix}^{-1} \cdot \begin{bmatrix} R_2 \\ T_2 \\ 0 \end{bmatrix}
\]

(4)

**C. DETERIORATION EVALUATION USING BOUNDING BOX AND SIFT FEATURE POINTS**

The distress type and bounding box generated by the Yolov3 network is utilized to analyze the condition of pavement distress during the data update period \([t_1, t_2]\). If the distress detected at different time belong to different types, then either the distress has deteriorated, or it has been maintained. However, if the distress detected remains the same type, there would be three possibilities, either the distress has not changed, or it has grown but kept the original type, or the new distress occurred in this area. Therefore, the scale-invariant feature transform (SIFT) descriptor is introduced to compare the similarity of two distress. Compared with the deep neural network, the SIFT method can detect and describe local features automatically, which filters the impact of the background. The SIFT descriptors are invariant to image scale and rotation, they are also robust to changes in illumination, noise, and minor changes in viewpoint, which make them easy to be used for image matching. The SIFT describes the local features by applying the Difference of Gaussian (DoG) on different scale-spaces of the image [39]. Fig. 6(a) illustrates the SIFT descriptors in two images. However, we only
reserved the descriptors in the bounding box because the others are from the background, which may affect the matching results. The matching of the local features based on SIFT is illustrated in Fig. 6(b). Random sample consensus (RANSAC) is used to eliminate the outliers of wrongly matched points [40]. The mean Euclidean distance (MEuD) and matching rate (MR) are applied to determine whether the two distress are identical. MEuD represents the matching confidence of the matched pixels, as formulated in (5):

\[
MEuD(S, P) = \frac{1}{m} \sum_{i=1}^{m} \left( \min_{1 \leq j \leq m} \left( \sum_{k=1}^{128} (s_{ik} - p_{jk})^2 \right) \right)^{1/2}
\]

where \(MEuD(S, P)\) denotes the root mean square distance between image \(S\) and \(P\). \(m\) is the amount of the matched points. \(i\) and \(j\) is the sequence number, \(s_{ik}\) and \(p_{jk}\) are the SIFT matrix. MR indicates the ratio of matched pixels to all pixels. The larger the MR, the higher the proportion of similar parts in the two images. If both the MEuD and MR are large, then the distress at \(t_1\) and \(t_2\) stay the same. If the MEuD is relatively small, but the MR is large, then the distress is very likely to grow bigger. If both the MEuD and MR are small, then the distress at \(t_1\) and \(t_2\) are distinct, indicating there is new distress generated during this time. The threshold of the abovementioned indexes can be referred to Du et al.’s research [41].

Therefore, we only observe the distress with high confidence (>0.85) to guarantee the evaluation results are instructive. In turn, the evaluation method of combining the bounding box and SIFT helps to remove the misidentified results of Yolov3.

V. RESULTS AND ANALYSIS
A. TRAINING RESULTS OF YOLOV3
The testing environment is configured as follows: the graphics card GPU is 8 GeForce GTX 1070Ti, CUDA9.0, Ubuntu 16.04, and 32GB of memory. The algorithms are implemented using Tensorflow for training and testing. After approximately thirty-six hours of training, the optimal weights were obtained. Fig. 8 presents the change of the loss function during iteration.

The training results are shown in Fig.9. The precision-recall curve represents how accurate the algorithm is at different recall rates. The area covered between the curve and the coordinate axis is the average accuracy. The average
The accuracy rates of the eight diseases were: deformation: 87.3%, pothole: 88.53%, loose: 88.08%, net-crack: 87.76%, cracks: 88.85%, patch-pothole: 88.65%, patch-net: 88.89%, patch-crack: 88.89%. And the mean average precision (mAP) was 88.37%. Because the patched distress is deeper in gray, and its features are more obvious, making it easy to be detected. With the increase of the dataset, the pre-trained model performs better.

B. RESULTS OF DETERIORATION ANALYSIS
A typical urban road: Yang shupu road in Shanghai, China was selected to conduct the evaluation of distress deterioration. As the street view map in this area was updated every two years, the time interval for deterioration analysis is set as the same one. Fig. 10(a) illustrates the distribution of various pavement distress in 2015, and Fig 10(b) shows the condition in 2017. Note that the evaluation results only roughly reflect the pavement conditions, but cannot accurately record all distress, because some parts of the pavement may be covered by vehicles, shades, pedestrians or other obstacles in images.

As shown in Fig. 10, the predominant distress type on the Yangshupu road is crack in 2015. However, a large proportion of cracks develop into net cracks in 2017, which is denoted by the yellow dots. Most distress happened in the first half of the road, especially the case for intersections. Compared with the situation in 2015, 34 cracks, 6 net-crack, and 9 potholes have been maintained, and distress maintenance rate is about 87.5%. 8 cracks, 29 net-crack, and 6 potholes were newly generated, as shown in Fig. 11, indicating that the deterioration of this road is quite significant in these two years. For the record, the ultra-long longitudinal cracks will be counted as multiple-segmented cracks according to the part captured in the images.

VI. CONCLUSION
This paper presents a deep learning framework for automated pavement distress detection and deterioration evaluation based on street view map. 19665 images with eight types of distress were collected via Baidu API. Given the distribution bias of multiple distress, the original images are augmented by random rotation and Gaussian blur as the training dataset. A pre-trained convolutional neural network named Yolov3 was implemented to identify and locate various distress. A three-level “GPS-Bounding box-SIFT” distress matching framework is constructed to compare the change of the distress at different time. A comprehensive decision tree is designed for evaluating the deterioration condition of the pavement. A typical urban road in Shanghai was used as a real case. The results show that the mAP of the proposed model is 88.37%, and the average precision (AP) for all the distress exceeds 80%, which may attribute to the great shooting angle of the street view images. The distress distribution in 2017 is significantly different from 2015. Up to 87.5% previous distress were patched, but there were still 7 of them deteriorated, and more than 40 distress were newly generated due to the heavy traffic.
Overall, the proposed method takes advantage of the street view map to collect the roadway pavement images, which serves as a cheap yet effective data source while guarantees a good coverage. Moreover, based on the timeline function provided by the street view map, we are able to evaluate the temporal-varied deterioration condition of a large-scale network.

Despite the high accuracy of the deep learning algorithm, the robustness and stability still need further testing in a more comprehensive environment. In addition, street view images are inevitably obscured by obstacles such as trees and vehicles. The pavement distress detection in such condition is going to be studied in the future.

REFERENCES

[1] I. G. Heggie and V. Piers, Commercial Management and Financing of Roads. Washington, DC, USA: World Bank, 1998.
[2] Road Funds and Road Maintenance: An Asian Perspective, Asian Develop. Bank, Mandaluyong, Philippines, 2003.
[3] M. Al-Falahi and A. Kassim, “Automated data collection system of pavement distresses: Development, evaluation & validation of distress types and severities,” in Proc. IOP Conf. Ser. Mater. Sci. Eng., vol. 471, 2019, Art. no. 062015.
[4] Canadian Infrastructure Report Card-Informing the Future. Can. Infras.-tract., Ottawa, ON, Canada, 2016.
[5] G. M. Rowe, J. N. Meegoda, A. A. Jumikis, M. J. Sharrock, N. Bandara, and H. Hettiarachchi, “Detection of segregation in asphalt pavement materials using the ARAN profile system,” Northeast Asphalt User, Producer Group Newport Marriott, Newport, RI, USA, 2002, vol. 30.
[6] P. Cairney, “ARRB group at 50,” J. Australas. College Road Saf., vol. 21, no. 4, p. 58, 2010.
[7] M. Jaboyedoff, T. Oppikofer, A. Abellán, M.-H. Derron, A. Loy, R. Metzger, and A. Pedrazzini, “Use of LIDAR in landslide investigations: A review,” Natural Hazards, vol. 61, no. 1, pp. 5–28, Mar. 2012.
[8] T. B. J. Coenen and A. Golroo, “A review on automated pavement distress detection methods,” Cogent Eng., vol. 4, no. 1, 2017, Art. no. 1374822.
[9] M. Solla, S. Laguila, H. González-Jorge, and P. Arias, “Approach to identify cracking in asphalt pavement using GPR and infrared thermographic methods: Preliminary findings,” NDT & E Int., vol. 62, pp. 55–65, Mar. 2014.
[10] A. Benedetto and S. Pensà, “Indirect diagnosis of pavement structural damages using surface GPR reflection techniques,” J. Appl. Geophys., vol. 62, no. 2, pp. 107–123, Jun. 2007.
[11] M. Eisenbach, R. Stricker, D. Seichter, K. Amende, K. Debes, M. Sesselmann, D. Ebersbach, U. Stoeckert, H.-M. Gross, “How to get pavement distress detection ready for deep learning? A systematic approach,” in Proc. Int. Joint Conf. Neural Netw. (IJCNN), Anchorage, AK, USA, May 2017, pp. 2039–2047.
[12] K. Gopalakrishnan, “Deep learning in data-driven pavement image analysis and automatic distress detection: A review,” Data, vol. 3, no. 3, p. 28, 2018.
[13] J. S. Apte, K. P. Messier, S. Gani, M. Brauer, T. W. Kirchstetter, M. M. Lundin, J. D. Marshall, C. J. Portier, R. C. H. Vermeulen, and S. P. Humphb, “High-resolution air pollution mapping with Google street view cars: Exploiting big data,” Environ. Sci. Technol., vol. 51, no. 12, pp. 6999–7008, Jun. 2017.
[14] D. B. A. Chacra and J. S. Zelek, “Fully automated road defect detection using street view images,” in Proc. 14th Conf. Comput. Robot Vis. (CRV), May 2017, pp. 353–360, doi: 10.1109/CRV.2017.50.
[15] K. C. P. Wang, A. Zhang, J. Q. Li, Y. Fei, C. Chen, and B. Li, “Deep learning for asphalt pavement cracking recognition using convolutional neural network,” in Proc. Airfield Highway Pavements, Aug. 2017, pp. 166–177.
[16] K. R. Kirschke and S. A. Velinsky, “Histogram-based approach for automated pavement-crack sensing,” J. Transp. Eng., vol. 118, no. 5, pp. 700–710, Sep. 1992.
[17] H. N. Koutsopoulos and A. B. Downey, “Primitive-based classification of pavement cracking images,” J. Transp. Eng., vol. 119, no. 3, pp. 402–418, May 1993.
XU LEI (Member, IEEE) received the B.S. degree in automation and the M.S. and Ph.D. degrees in traffic information engineering and control from Chang'an University, Xi'an, Shaanxi, China, in 2002, 2005, and 2016, respectively.

He is currently a Senior Engineer with the School of Electronics and Control Engineering, Chang'an University. He has published more than 20 articles in academic journals and conferences. His current research interests include simultaneous localization and mapping (SLAM), robotic perception, and autonomous driving.

CHENGLONG LIU received the B.S. and Ph.D. degrees in transportation engineering from Tongji University, Shanghai, China, in 2014 and 2019, respectively.

He currently holds a Postdoctoral position with the Key Laboratory of Road and Traffic Engineering of the Ministry of Education, Tongji University. His research interests include pavement performance detection, maintenance optimization, and intelligent transportation infrastructure.

GUiping Wang received the B.S. degree in automation from Xian Jiaotong University, Xi’an, China, in 1983, the M.S. degree in system engineering from the Huazhong University of Science and Technology, Wuhan, China, in 1989, and the Ph.D. degree in carrying tools application project from Chang’an University, China, in 2009. He is currently a Professor with the School of Electronics and Control Engineering, Chang’an University.

He has been engaged in the research of measurement and control technology for electric vehicles and smart cars for a long time. He is also the Deputy Director of the Intelligent Transportation Committee of the China Artificial Intelligence Society, the Vice Chairman of the Provincial Instrumentation Society, the Vice Chairman of the Provincial Power Supply Society, and the Executive Director of the Provincial Automotive Engineering Society. He has hosted and participated in more than 20 scientific research projects. He has published more than 40 academic articles in domestic and foreign journals, and has obtained 16 national patents.

LI Li received the B.S. and M.S. degrees in automation from Chang’an University, Xi’an, China, and the Ph.D. degree in transportation engineering from Tongji University, Shanghai, China.

He is currently an Associate Professor with the School of Electronics and Control Engineering, Chang’an University. His research interests include intelligent transportation systems, driver behavior, and traffic flow.

LI Li received the B.S. and M.S. degrees in automation from Chang’an University, Xi’an, China, and the Ph.D. degree in transportation engineering from Tongji University, Shanghai, China.

He is currently an Associate Professor with the School of Electronics and Control Engineering, Chang’an University. His research interests include intelligent transportation systems, driver behavior, and traffic flow.