Junxian LI, Ph.D. Candidate[1]
(Corresponding author)
E-mail: 1911549@tongji.edu.cn
Zhizhou WU, Ph.D.[1]
E-mail: wuzhizhou@tongji.edu.com
Zhoubiao SHEN, M.E.[2]
E-mail: shenzhoubiao@sucdri.com
[1] Key Laboratory of Road and Traffic Engineering
of the Ministry of Education, Tongji University
No.4800 Cao'an Road, Jiading District,
Shanghai 201804, China
[2] Shanghai Urban Construction Design & Research
Institute (Group) Co., Ltd.
No.3447 Dongfang Road, Pudong New District,
Shanghai 200125, China

Information and Communication Technology
Original Scientific Paper
Submitted: 3 Nov. 2021
Accepted: 28 Jan. 2022

OPEN THE BLACK BOX – VISUALISING CNN TO UNDERSTAND ITS DECISIONS ON ROAD NETWORK PERFORMANCE LEVEL

ABSTRACT

Visualisation helps explain the operating mechanisms of deep learning models, but its applications are rarely seen in traffic analysis. This paper employs a convolutional neural network (CNN) to evaluate road network performance level (NPL) and visualises the model to enlighten how it works. A dataset of an urban road network covering a whole year is used to produce performance maps to train a CNN. In this process, a pretrained network is introduced to overcome the common issue of inadequacy of data in transportation research. Gradient weighted class activation mapping (Grad-CAM) is applied to visualise the CNN, and four visualisation experiments are conducted. The results illustrate that the CNN focuses on different areas when it identifies the road network as different NPLs, implying which region contributes the most to the deteriorating performance. There are particular visual patterns when the road network transits from one NPL to another, which may help performance prediction. Misclassified samples are analysed to determine how the CNN fails to make the right decisions, exposing the model's deficiencies. The results indicate visualisation's potential to contribute to comprehensive management strategies and effective model improvement.

KEYWORDS

visualisation; convolutional neural network (CNN); gradient weighted class activation mapping (Grad-CAM); pretrained network; road network performance.

1. INTRODUCTION

The rapid development of intelligent transportation systems (ITS) provides favourable conditions for city managers to improve traffic, of which the broad deployment of traffic data detectors is the most appreciated. Along with significantly enhanced data storage and computing capacity, ITS affords a large amount of data with broader coverage and higher granularity [1]. Benefiting from this, artificial intelligence (AI) that heavily depends on data is introduced to traffic analysis. Among AI algorithms, deep learning (DL) performs well in identifying the underlying non-linear correlations of data. Therefore, it is especially suitable for analysing large-scale road networks involving complex spatiotemporal correlations between links [2]. Various DL models have been applied in traffic data completion [3], traffic forecast [4], traffic performance evaluation [5], and a vast body of results have been achieved [6]. Despite this, DL is constantly criticised for its un-interpretability. The end-to-end training and complex structure render most DL models 'black boxes' [7], making it challenging to explain how they produce ideal results in traffic analysis. This deficiency limits its role in providing an in-depth understanding of traffic problems. Given that this drawback is pervasive for all disciplines using DL models, there has been a noticeable increase in research on explaining DL models' operating mechanisms, producing a new technology called Explainable AI (XAI).

Among XAI methods, visualisation is the most celebrated for it presents the abstract operating process of DL models as images, which are vivid and straightforward even for non-professionals. Taking the model for image classification as an example, its visualisation usually outputs a thermal map with the same size as the original image. On the map, the pixels with higher thermal values correspond to the model's concern when classifying the image. In comparison, the pixels with lower thermal values are what the model considers not significant to explain the image category. Overlaying this thermal map with the original image, one can understand which part of the image renders the model of the current classification decision. As Selvaraju et al. [8] put it, in the fields where AI performance is not as good as that of human, visualisation will help humans to identify the failure modes of AI; in the fields where AI performance exceeds human, visualisation can, in turn, tell humans how to make better decisions. In traffic analysis, visualisation can benefit city managers from both aspects. It can expose the drawbacks of DL models so that managers can adjust the models to better adapt to a particular road network. It can also illustrate the trends DL models capture in an inefficient road network, containing insights that managers have neglected. These uses are of considerable significance in traffic improvement.

Due to the excellent capability of extracting data and image features for convenient displaying, the convolutional neural network (CNN) is taken by many traffic studies as the bottom module to extract features and compress data in the first round [9, 10]. In these studies, CNN determines the follow-up analysis reliability as a critical part of models. Therefore, it will be significant to visualise CNN to know how it functions and improve it accordingly. Fortunately, CNN is especially suitable for being visualised because of its nature of image processing, and various methods have been developed to accomplish this goal. Among these methods, the pixel space gradient visualisation [11, 12] and feature inversion [13, 14] are typical. However, problems have been found with them in that neither is suitable for explaining the classification results generally. Hence, more practical methods are needed.

Localisation approaches are essential to address the issue. Among them, class activation mapping (CAM) [15] is representative. CAM refers to the idea of 'network in network' [16] to obtain a satisfying capability to interpret CNN. However, the CNN's original structure is modified when CAM is conducted, which means the model needs to be retrained overall. This requirement cannot be met once the model has been launched or the training cost is very high, limiting the application scope of CAM. Gradient weighted class activation mapping (Grad-CAM) [8] provides a solution. It shares the same idea with CAM but puts forward two measures to improve [15]. As a result, Grad-CAM inherits the CAM's ability to explain classification results. Furthermore, it prevents the CNN model from being retrained to be visualised, outperforming CAM in operation speed and applicability. More details on CAM and Grad-CAM are beyond this paper's scope, and interested readers may refer to [8] and [15] for more information.

Despite the merits of CNN visualisation in traffic analysis and the abundance of methods to fulfil the task, there has been little research effort to put it into practice to the best of our knowledge. Based on the gaps, this paper makes an exploratory attempt to apply visualisation in dealing with traffic problems. Firstly, a CNN model is trained to evaluate road network performance. In the process, a pretrained network is employed to prevent overfitting. Secondly, the model is visualised further by Grad-CAM to investigate its internal mechanism 'reversely'. Thirdly, four experiments are conducted to illustrate the knowledge the visualisation brings to traffic analysis. The results indicate that the visualisation can show the managers the essential area of the road network for performance deterioration and recovery, which may help evaluate and predict road network performance. Moreover, visualisation allows the managers to understand the model's deficiencies and improve it accordingly.

The rest of this paper begins with describing techniques to use CNN in evaluating network performance and the method to visualise the model in Section 2. Section 3 elaborates on data processing, modelling and model visualising. Section 4 describes the results of the visualisation. The conclusion and future research direction are presented in the last section.

2. METHODOLOGY

It is challenging for managers to online evaluate road network performance level (NPL) in an interval based on the comprehensive conditions of all links, let alone determine which regions contribute the most to the inefficiency. By taking each NPL as a category, the issue can be transformed into a multiclassification task. A CNN will be trained to fulfil the task and visualised to present how it makes decisions. *Figure 1* shows the process schematically.

To realise the process, three essential concerns are found in the way:

- What to input into the CNN model to present all link performances in the road network;
- How to train a satisfactory CNN model with limited data in traffic analysis;
- Which CNN visualisation method is ideal for balancing explanatory capacity and computational cost, and which layers of the CNN model are the most worth visualising to illustrate the model's focus.

The answers are the performance map, the pretrained network and Grad-CAM to visualise the highest layers of the model, respectively.

2.1 Performance map for each interval

Speed, delay and travel time are commonly used as the link performance index (PI), whichever is straightforward to get online with intelligent connected vehicles, high-definition network cameras or loop detectors. The challenge lies in describing all link PIs of some interval in an appropriate format that CNN is best at processing, i.e. a 2D image termed as the performance map (PMap) in this paper.

Intuitively, a GIS map is an ideal material to create PMaps for its image properties. Dividing the GIS map into m×n square grids that share the same size, the map with grids can be regarded as an image with m×n pixels. Referring to coordinates of the GIS map, the location of each grid can be easily obtained. Denoting the *j*th ($1 \leq j \leq n$) grid in row *i* ($1 \leq i \leq m$) as g_{ij}, the pixelized map will be used as the PMap template.

Each link in the road network is broken into multiple short segments with similar lengths. The short lengths make it reasonable to represent each segment with its midpoint. Collection of Q segments whose midpoints are located in g_{ij}, PI of g_{ij} in interval k, denoted as $PI_{i,j,k}$, is governed by *Equation 1*.

$$PI_{i,j,k} = \frac{\sum_{1}^{Q} l_q \cdot PI_{q,k}}{\sum_{1}^{Q} l_q} \quad (1 \leq q \leq Q) \tag{1}$$

where l_q stands for the length of segment q, and $PI_{q,k}$ the PI of it in interval k, which is inherited from its parent link.

For interval k, a copy of the pixelized map is made and each normalised $PI_{i,j,k}$ is used as the gray value to fill g_{ij}. In this way, the PMap of each interval can be yielded. With these PMaps as input, a CNN was trained to classify them into appropriate NPLs automatically.

2.2 Pretrained network of CNN

As an essential DL framework, CNN is expected to function well with sufficient training data. However, typical data sets in traffic analysis are small to limit CNN capacity in practice, since the training is prone to overfitting. Therefore, besides dropout and regularisation to prevent overfitting, other measures should be employed to improve the model. Among these, a pretrained network is recommended for its comprehensibility and low operation cost. Motivated by these advantages, this paper employed a pretrained network to address the insufficient training issue. The pretrained network selected was VGG16 [17], a CNN fully trained on the ImageNet data set.

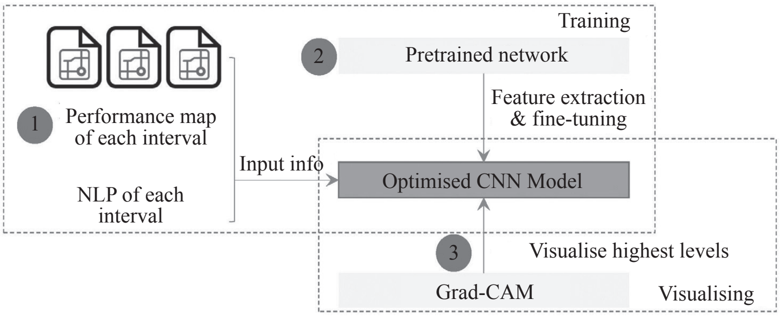
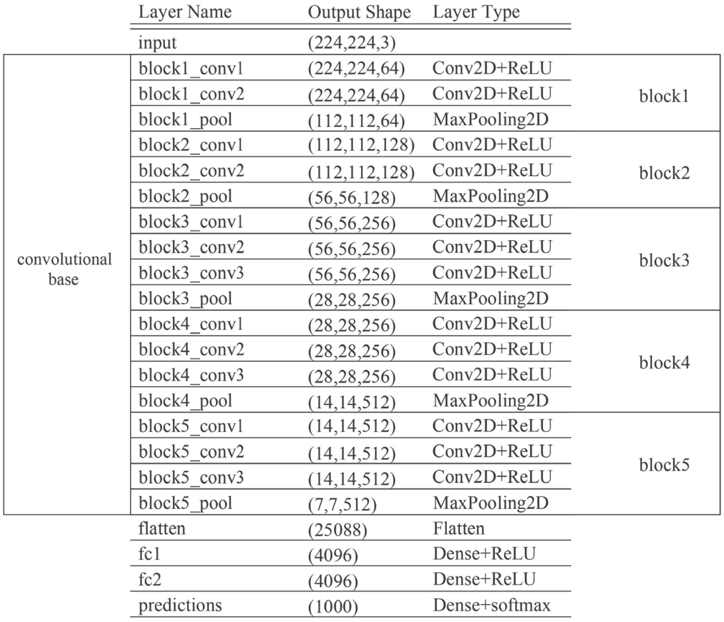


Figure 1 – Process of training and visualising the CNN in this paper



	Layer Name	Output Shape	Layer Type	
	input	(224,224,3)		
convolutional base	block1_conv1	(224,224,64)	Conv2D+ReLU	block1
	block1_conv2	(224,224,64)	Conv2D+ReLU	
	block1_pool	(112,112,64)	MaxPooling2D	
	block2_conv1	(112,112,128)	Conv2D+ReLU	block2
	block2_conv2	(112,112,128)	Conv2D+ReLU	
	block2_pool	(56,56,128)	MaxPooling2D	
	block3_conv1	(56,56,256)	Conv2D+ReLU	block3
	block3_conv2	(56,56,256)	Conv2D+ReLU	
	block3_conv3	(56,56,256)	Conv2D+ReLU	
	block3_pool	(28,28,256)	MaxPooling2D	
	block4_conv1	(28,28,256)	Conv2D+ReLU	block4
	block4_conv2	(28,28,256)	Conv2D+ReLU	
	block4_conv3	(28,28,256)	Conv2D+ReLU	
	block4_pool	(14,14,512)	MaxPooling2D	
	block5_conv1	(14,14,512)	Conv2D+ReLU	block5
	block5_conv2	(14,14,512)	Conv2D+ReLU	
	block5_conv3	(14,14,512)	Conv2D+ReLU	
	block5_pool	(7,7,512)	MaxPooling2D	
	flatten	(25088)	Flatten	
	fc1	(4096)	Dense+ReLU	
	fc2	(4096)	Dense+ReLU	
	predictions	(1000)	Dense+softmax	

Figure 2 – VGG16 structure with original settings

With the image size 224×224×3 and other parameters unchanged in the original paper, *Figure 2* shows the structure of VGG16 with each layer name, which will be used to refer to the specific layer in the rest of the paper.

To reuse VGG16, feature extraction [18] and fine-tuning [19], two of the most conventional methods of dealing with the pretrained network are applied. Feature extraction borrows several general layers from the pretrained network to reuse their structures and weights to extract features and further input these features into a new classifier specially designed for a new problem. By those means, the general layers with a good capacity to extract image features are made full use of, and only the new classifier needs to be trained, which is more effective than training the whole model. Fine-tuning keeps model structure unchanged but adjusts the weights of higher levels to make them more adaptable to the new classification problem at hand.

A CNN model supported by the pretrained network is expected to achieve satisfying performance, and it is instructive to visualise its analysis process.

2.3 Grad-CAM visualisation and layers to be visualised

Grad-CAM is a breakthrough method for visualising CNN classification principles. Although more methods have been derived from it to solve some specific problems, the ability of Grad-CAM is considered appropriate for the task at hand. Its operating speed can especially meet the requirements of online analysis, which is of great significance in traffic management decision-making. Hence, it is applied as the visualisation tool in this paper.

When it comes to the layers to be visualised, the information extracted from the full connection and softmax layers of CNN is challenging to display; however, each convolution layer in the convolution base can be visualised by Grad-CAM. Among these layers, the convolution layers at the highest level contain the most valuable information for illustrating which part of the image contributes to the class identification [20]. Therefore, to better visualise CNN with a low computational burden, it is encouraged to make full use of these layers and focus on the feature map obtained from them to give reasonable explanations of CNN classifications. In this paper, visualisation is carried out on layers of block5 (see *Figure 2*).

3. DATA DESCRIPTION AND PROCESSING

3.1 Data source and the performance index

The data set used came from GAIA Open Data provided by Didi Chuxing. Didi Chuxing collected travel trajectories from a floating vehicle with an installed Didi app in Shenzhen, China, from 1 January 2018 at 00:00 to 30 December 2018 at 23:50. After map matching and data aggregation, the data were provided as the travel time index (TTI) table with a 10 min interval for all 1,172 links. Because of the occasional failure to transmit, only the data of 52,007 intervals were provided, rather than 52,416 with no missing data during the collection period. These missing data cannot be recovered, and the missing intervals were marked to indicate that the records before and after them were not consecutive in time.

For link p, with u_p as its free-flow speed and $v_{p,k}$ its speed during interval k, its TTI in interval k is $TTI_{p,k}=u_p/v_{p,k}, TTI \in [1,\infty)$. For comparison convenience, all $TTI_{p,k}$ are standardised as

$$\overline{TTI}_{p,k} = \frac{TTI_{p,k}}{\max(TTI_{*,k})} \qquad (2)$$

in which max $(TTI_{*,k})$ stands for the largest $TTI_{p,k}$ in interval k. It maps $TTI_{p,k}$ to (0,1], and a higher $\overline{TTI}_{p,k}$ suggests worse link performance. With all link TTIs in an interval, the NPL of this interval was determined and labelled from O to IV, O for the best, and IV for the worst.

3.2 Image preparation for CNN input

A GIS map of the road network was also provided (*Figure 3a*). Its longitude ranges from 113.772855° to 114.179985° and latitude from 22.459465° to 22.856285°, respectively. With 0.005° as the unit in both directions, the map was divided into 82×82 grids, an appreciated resolution for online computing. In this way, each grid holds an approximate side length of 550 m, similar to the average roadway length in the urban road network.

Thus, the pixelized map (*Figure 3b*) is formed.

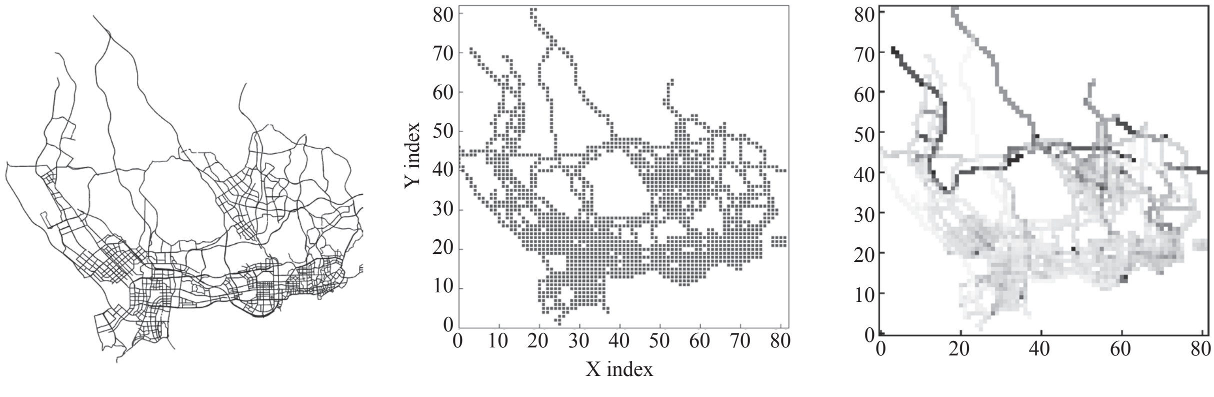
1,172 links were further broken into 80,509 segments with lengths less than 40 m and similar to each other. Taking TTI as the PI, with all segment TTIs in each interval, *Equation 1* was applied to calculate the gray values of the grids to fill the pixelized map of each interval (*Figure 3c*). In this way, a total of 52,007 interval PMaps were prepared to train a CNN to classify PMaps into five categories, e.g. NPL O to IV.

3.3 CNN modelling and optimising

The 52,007 data were numbered from 0 to 52,006. The NO.0~39,999 samples were selected as the training set, and NO.40,000~ 52,006 as the test set. Considering that the subsequent analysis concerns temporal correlations of PMaps, it is critical to ensure that the time sequence of samples remains unchanged, so shuffling was not carried out. Nevertheless, the missing data were noted to protect from extracting inconsecutive data as the analysis material.

Initially, a basic CNN was built, as shown in *Figure 4a*. There were no measures to prevent overfitting, to evaluate the basic model's ability to classify. We employed a 3×3 convolution kernel, ReLU (Rectified Linear Unit) activation in the convolution layers, and a 2×2 max-pooling kernel in the pooling layers. Other hyperparameters were set, including the epochs=30, batch size=64, learning rate=1e-4 by previous experimentations. RMSProp (Root Mean Square Prop) was used as the optimiser and categorical cross-entropy as the loss function. Interested, readers should refer to [21] for the information of the above parameters.

The accuracy of the basic model reached about 85%. However, there was severe overfitting since the accuracy of the training set was constantly increasing, and the accuracy of the test set stopped improving since the fifth epoch. Hence, measures



a) GIS map

b) Pixelized map

c) PMap example

Figure 3 – The transition from the GIS map to the PMap

Layer Name	Output Shape	Layer Type
input	(82,82,1)	
model1_conv1	(80,80,32)	Conv2D+ReLU
model1_pool1	(40,40,32)	MaxPooling2D
model1_conv2	(38,38,64)	Conv2D+ReLU
model1_pool2	(19,19,64)	MaxPooling2D
model1_conv3	(17,17,64)	Conv2D+ReLU
model1_flatten	(18496)	Flatten
model1_fc	(64)	Dense+ReLU
model1_predictions	(5)	Dense+softmax

a) Basic CNN structure

	Layer Name	Input/Output Shape	Layer Type	
	input	(82,82,3)		
convolutional base of VGG16	block 1~4	(5,5,512)		Feature Extraction
	block 5	(2,2,512)	Fine Tuning Levels	
	input	(2,2,512)		
self-defined classifier	model1_fc	(256)	Dense + ReLU (L2 regularization)	
	dropout	(256)	Dropout rate = 0.5	
	model1_predictions	(5)	Dense + softmax	

b) Optimised CNN model

Figure 4 – Structures of two CNNs

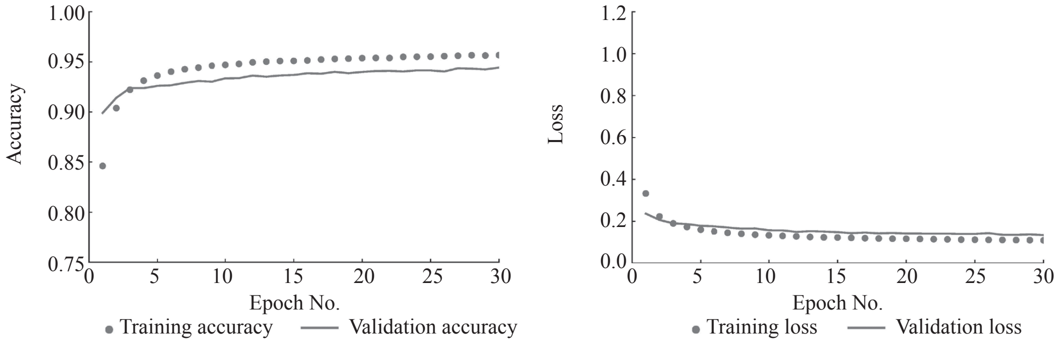
to prevent overfitting are necessary. After several experiments and optimisations, an optimised CNN model structured as *Figure 4b* was constructed.

In the optimised model: (1) Convolutional base of VGG16 worked as a pretrained network. (2) To suit the pretrained network, every PMap copied itself twice to reconstruct as an 82×82×3 image in the input layer. (3) For each PMap, the pretrained network extracted its features, and the block5 layers were particularly fine-tuned to fit the problem at hand. These are the crucial steps to use the pretrained network, which helped a lot to prevent overfitting. (4) The pretrained network was followed by a self-defined classifier, in which L2 regularisation (λ=0.001) and a dropout layer (dropout rate = 0.5) were combined to protect overfitting further. (5) Other hyperparameters stated above remained unchanged. Accuracy and loss curves of the optimised model in *Figure 5* illustrate that the model significantly reduced the index gaps between the training set and the test set, which means it succeeded in avoiding overfitting.

Given the performance of this model, one will be interested in how it can make decisions so effectively. Understanding which regions weigh the most when the model decides NPL will be enlightening to decide the priority of management strategy. Visualisation can help with displaying the model's focus intuitively.

3.4 Visualising the optimised model

To make full use of data, the training set and the test set were both input into the optimised model to get a predicted class for each sample. Samples were divided into two sets according to the result correct-



a) Accuracy curve

b) Loss curve

Figure 5 – Accuracy and loss curves of the optimised mode

ness: the correctly classified samples formed a set named CS, and the misclassified samples formed a set named WS. With them, four visualisation experiments were carried out:

Experiment 1: We studied all NPLs in chronological order to find their changing rules. Misclassified samples were marked to find when the model failed to make the right decisions, and they were studied further in Experiment 4.

Experiment 2: For each sample from CS, we calculated its CAM of the block5_conv3 layer activated by its real class, which would help the most to understand which part of the PMap drove the model to make the decision. For the same reason, all CAMs in the following experiments were corresponding to this layer.

Experiment 3: Considering most samples were consecutive in time, it was feasible to obtain sequential chronological PMaps in CS. We picked up these PMap sequences and their chronological CAMs and analysed the changing trend of CAMs.

Experiment 4: For WS, CAMs for the real and misclassified classes were calculated, respectively. We compared the two CAMs and determined the parts the model neglected when making decisions. The comparison indicated the drawbacks of the model.

Given that city managers are more concerned about poor road network performance, the analysis focused on NPL II, III, and IV, which correspond to inefficient traffic.

4. RESULTS AND DISCUSSION

4.1 Change of NPLs

For Experiment 1, the NPL changing in chronological order was analysed. There were no leaping changes from one level to the levels with a distance of more than 1, consistent with the intuitive knowledge about road network performance change. Moreover, the model seemed prone to confuse NPL II and NPL III, NPL III and IV. It was also noticed that the errors were particularly likely to occur when NPL transitions and during the period when NPL volatiles.

Table 1 is the confusion matrix describing the CNN's wrong decisions.

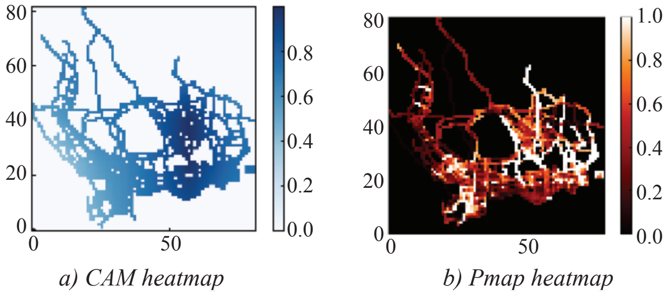
With the above result, the interval sequences of interest were selected, and their CAMs were analysed respectively in Experiments 2~4.

Following the conventional practice, heatmaps were used to represent CAMs to show the varied activation intensity of different regions in PMaps. The larger the thermal value of the pixel, the more significant the role it plays in CNN's decision-making. In this paper, the pixel with a higher thermal value in CAMs means the links located in this pixel drive the model more strongly to classify the road network in the interval to the current NPL.

In addition, to compare with and distinguish from CAM, we also displayed PMaps, represented initially as a grayscale image (see *Figure 3c*), as a heatmap. *Figure 6* presents the colormaps for the CAM and PMap heatmaps.

Table 1 – Numbers of misclassified samples according to their real and predicted NPL

Real NSL	Predicted NPL				
	O	I	II	III	IV
O	--	102	0	0	0
I	34	--	126	0	0
II	0	26	--	290	0
III	0	0	22	--	217
IV	0	0	0	33	--



a) CAM heatmap

b) Pmap heatmap

Figure 6 – Colormaps of CAM and Pmap used in this pape

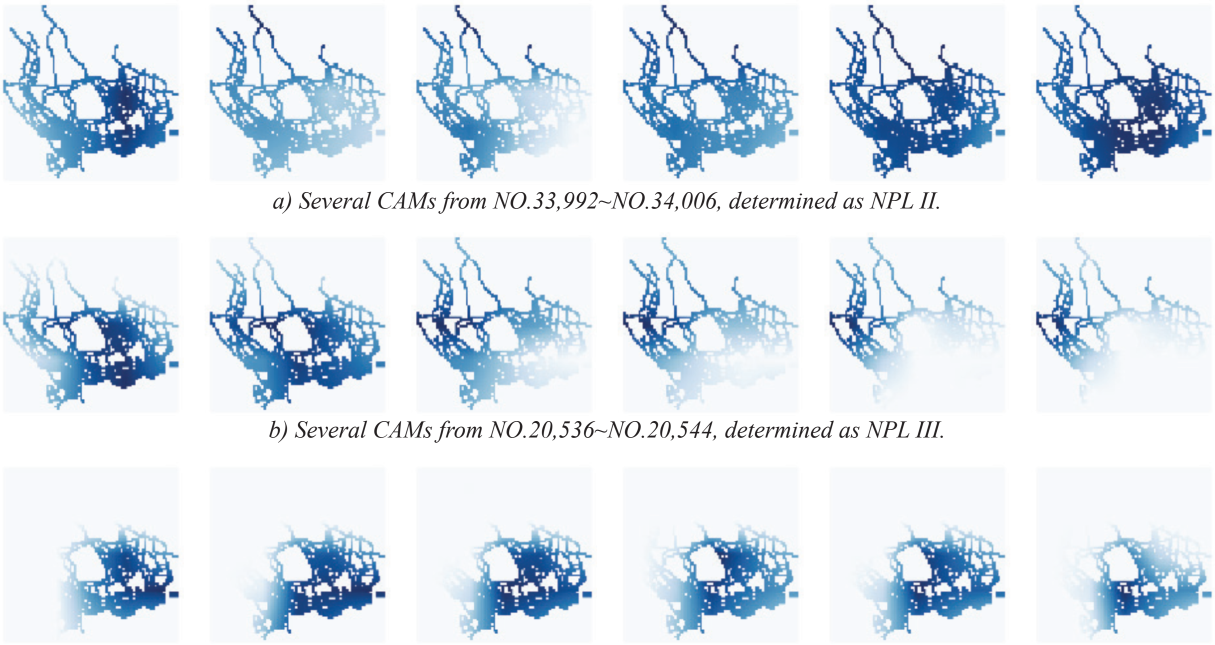
4.2 Highly activated areas in CAMs

The Experiment 2 results showed that for different NPLs, the CNN observed different areas in PMaps. In other words, some areas were found to be worth more reference than others in the CNN classifying decisions for each NPL. *Figure 7* displays the sample CAMs for NPL II, III and IV, respectively.

As seen in *Figure 7a*, the entire road network was considered essential in most cases when PMaps were classified as NPL II. It means all pixels were taken into account, for they shared a similar PMap pattern that the CNN determined for NPL II. In contrast, the modes became diverse when NPL III was determined. In *Figure 7b*, in some cases, the road network was activated by a particular PMap pattern, while only the roads in the northwest were activated in others. When it comes to NPL IV, the activated areas in CAM were fixed. Besides the CAMs displayed in *Figure 7c*, almost all other CAMs determined as NPL IV were activated in the southeast area, with different activation intensities.

With the intensively activated areas in CAMs, it is easy to determine what patterns the CNN searched for in PMaps when it decided each NPL. Therefore, CAMs and the corresponding PMaps were paired to be observed.

Figure 8 illustrates that for NPL II, the detector did not pay special attention to the isolated grids with extremely high TTI, but it was easily attracted by

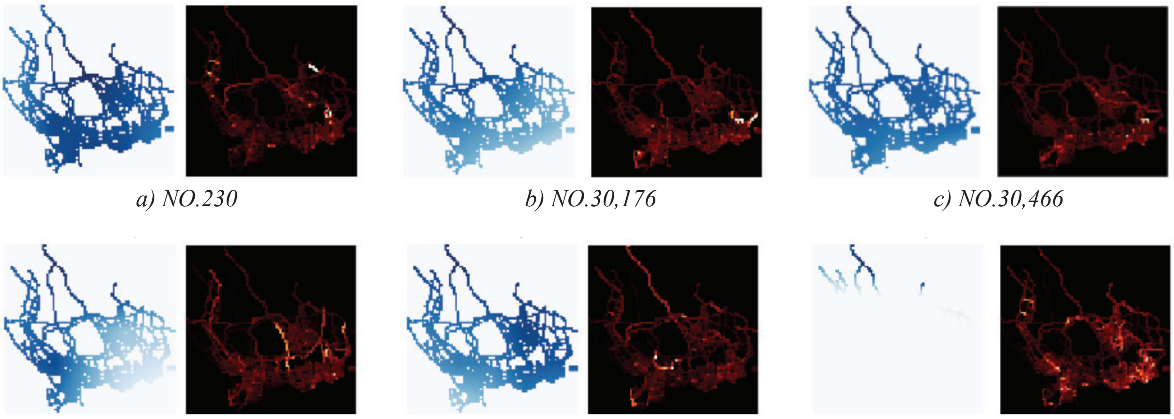


a) Several CAMs from NO.33,992~NO.34,006, determined as NPL II.

b) Several CAMs from NO.20,536~NO.20,544, determined as NPL III.

c) Several CAMs from NO.10,024~NO.10,037, determined as NPL IV.

Figure 7 – CAMs for various NPLs, different areas were activated



a) NO.230

b) NO.30,176

c) NO.30,466

d) NO.16,549

e) NO.27,104

f) NO.4,869

Figure 8 – Selected CAMs and corresponding PMaps identified as NPL II

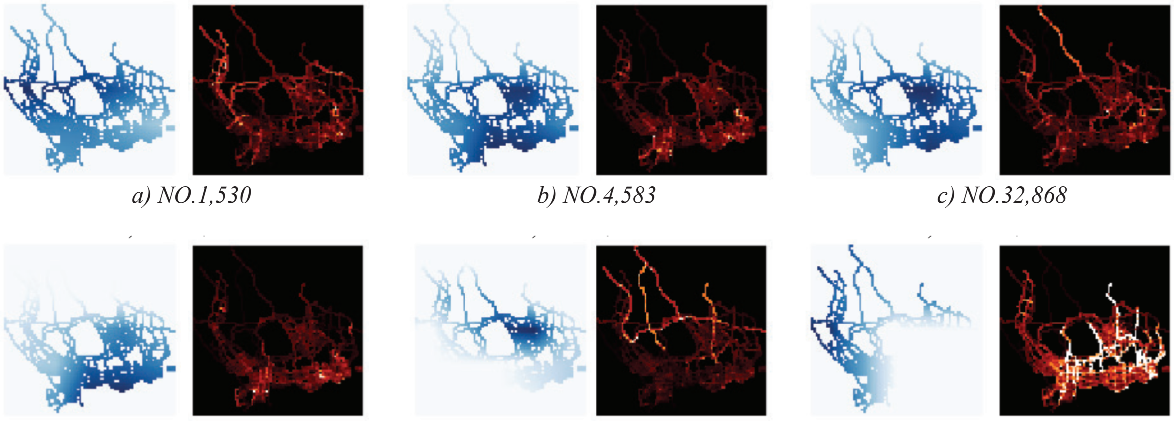
clusters of grids with low TTIs, between 0.30 and 0.45 (*Figures 8a-8e*). Reviewing the conclusions from *Figure 7a*, one can see that the CNN tended to classify a road network as NPL II when most links held a good service level, which was supposed higher than grade C according to a later calculation. It is a relatively harsh requirement.

Moreover, just like NPL III, the CAM of NPL II occasionally presented a distinctive mode (*Figure 8f*). Subsequent investigation in Experiment 3 showed that this was due to the upcoming change of NPL. It is a noteworthy sign of the impending degradation of network performance. Section 4.3 explains more about this phenomenon.

For NPL III, although the activated mode of CAM diversified, the PMap features that the CNN was interested in were pronounced. It ignored those links with low TTIs and was not sensitive to scattered high-TTI grids but paid particular attention to the areas composed of pixels with medium TTIs. In detail, the activation mode in *Figure 9a-9c* was the most common for NPL III. The TTIs corresponding to the activated pixels of NPL III were more significant than NPL II, most between 0.41 to 0.54. This result is acceptable. For the road network identified as NPL III, its TTIs are expected higher than that of NPL II. However, unlike the paralyzed road network, it hardly sees clustered pixels with extreme-TTI, which warn of terrible congestion or large-scale performance decline. Although in some cases, PMaps may possess a lot of high-TTI (0.84 to 0.94) grids in the southeast area (*Figure 9f*), the counterpart of CAM was dormant, which means that the CNN considered the pattern here had little significance in the current interval. However, the high-TTI area may become too big to ignore later without timely management. Actually, this is the critical state between NPL III and NPL IV.

One may notice that PMaps in *Figure 8f* and *Figure 9d* are similar in that the southeast area is a mixture of low and medium-TTI grids. However, the CNN determined them as NPL II and NPL III, respectively. As stated above, the performance of the road network in *Figure 8f* was approaching NPL III, but the medium-TTI grids were not as concentrated as in *Figure 9d*. It seems that the CNN considered both the number of grids with higher TTI and the concentration of these grids when making decisions. It is inferred that the CNN had learned a principle that isolated higher-TTI grids do not imply performance decline of the entire road network.

With the activations of NPL IV in *Figure 10*, it seems the CNN was sensitive to the densely interweaving grids with different levels of TTIs, especially when some of these grids held extremely high TTIs (*Figure 10a and 10b*). It suggested that inefficient links not far apart were to do significant damage to the road network performance. Moreover, the NPL IV detector did not seem interested in an isolated link with high TTI (*Figure 10c-10e*). The corresponding pixels in CAM were not highlighted at all. It seems that CNN had learned from the data fed to it that the overall performance of the road network cannot be determined by occasional events, even if they may have a destructive impact on a few links. Extreme and occasional traffic jams can be evacuated quickly in a robust regional road network without pushing the CNN to degrade the performance of the whole network. This finding suggests that combining CAM and PMap is conducive to comprehensively understanding the road network's operation.


a) NO.1,530

b) NO.4,583

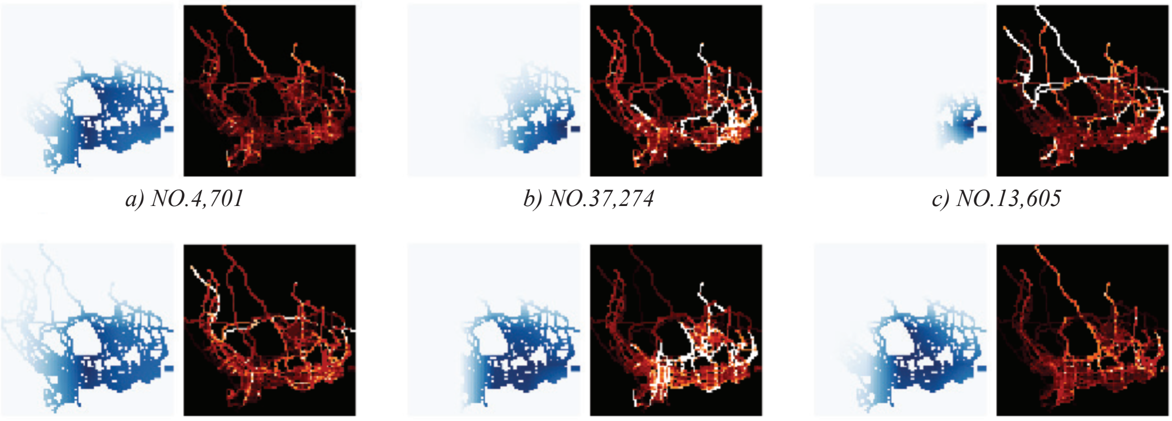
c) NO.32,868

d) NO.5,157

e) NO.13,526

f) NO.24,365

Figure 9 – Selected CAMs and corresponding PMaps identified as NPL III


a) NO.4,701 *b) NO.37,274* *c) NO.13,605*

d) NO.37,284 *e) NO.12,725* *f) NO.17,474*

Figure 10 – Selected CAMs and corresponding PMaps identified as NPL IV

PMap observes the road network from a micro-level and reports the decline of each link, and CAM provides a better view from a macro respective.

4.3 CAMs for NPL transition

Figures 7–10 illustrate that the activation modes of the same NPL were not always consistent. Besides, under some circumstances, the road networks with similar PMaps were identified as different NPLs. To explore these inconsistencies, Experiment 3 tracked CAMs of consecutive intervals. It was found that the PMaps determined as the same NPL saw gradual changes in CAMs, and there were unique patterns when CAMs transited from one NPL to another. Consecutive CAMs over several periods are displayed in *Figure 11*. During each period, NPL remained the same, and at the next interval, NPL changed. Given that some consecutive CAMs resemble each other, several similar CAMs were omitted. See the interval numbers.

As *Figure 11* presents, when the activated area of NPL II gradually moved from the entire road network to the northwest area, the road network performance may be deteriorating. If there was no proper management, performance degradation was likely to occur (*Figure 11a*). The same rule worked for the transition from NPL III to IV (*Figure 11b*). In addition, whether from NPL II to III or III to IV, this transformation seemed to start from the southeast road network. Keeping in mind the implication of CAM, the transfer of the activated area meant that the areas no longer activated had lost the modes the original NPL detectors were interested in. Therefore, we can say that the southeast region may have been degraded in priority. The corresponding region of PMap took the lead in losing the pattern that the low-NPL detector focused on, and then other regions followed up later. However, the early deterioration had not been strong enough to impel the CNN to change its rating decisions.

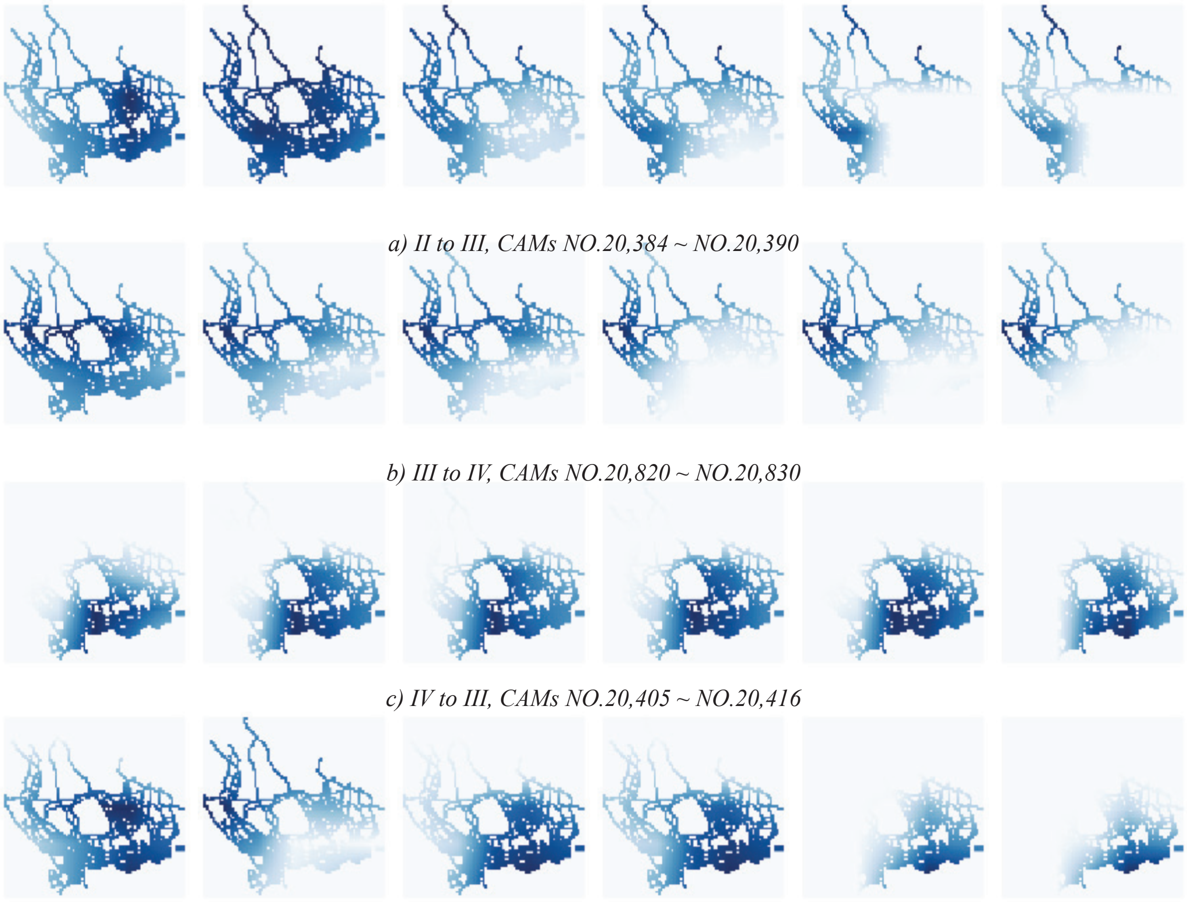
In comparison, as the road network performance recovered (IV to III and III to II), the activated area of CAM showed a reversed transferring process compared with that when the NPL rose. The southeast area was still the hot zone. However, the activation scope shrank gradually, indicating that more regions no longer maintained the current NPL's interested mode. Until the activation of the current NPL was not intensive enough, the NPL changed.

From the above results, CAM sequences reveal a lot about the change of road network performance. It makes it possible to predict NPL and makes it easier for managers to find out the reasons for the performance deterioration of the road network and take measures accordingly.

4.4 The misclassified samples

As *Table 1* suggests, the model was inclined to misclassify PMaps as higher NPLs. To figure out the causes, we selected several misclassified samples that were wrongly identified as performing worse to compare the CAMs activated by real and predicted NPL. *Figure 12* shows the results. The three images grouped in one sub-figure are the CAM of the real NPL, the CAM of the predicted NPL, and the PMap.

As the figures imply, the real and predicted NPLs focused on different areas. They both found 'their pixels' on the PMaps, thus convincing the CNN that the current PMap should be at their respective levels. The CAM of the predicted NPL was activated more intensively, so the model failed to allocate



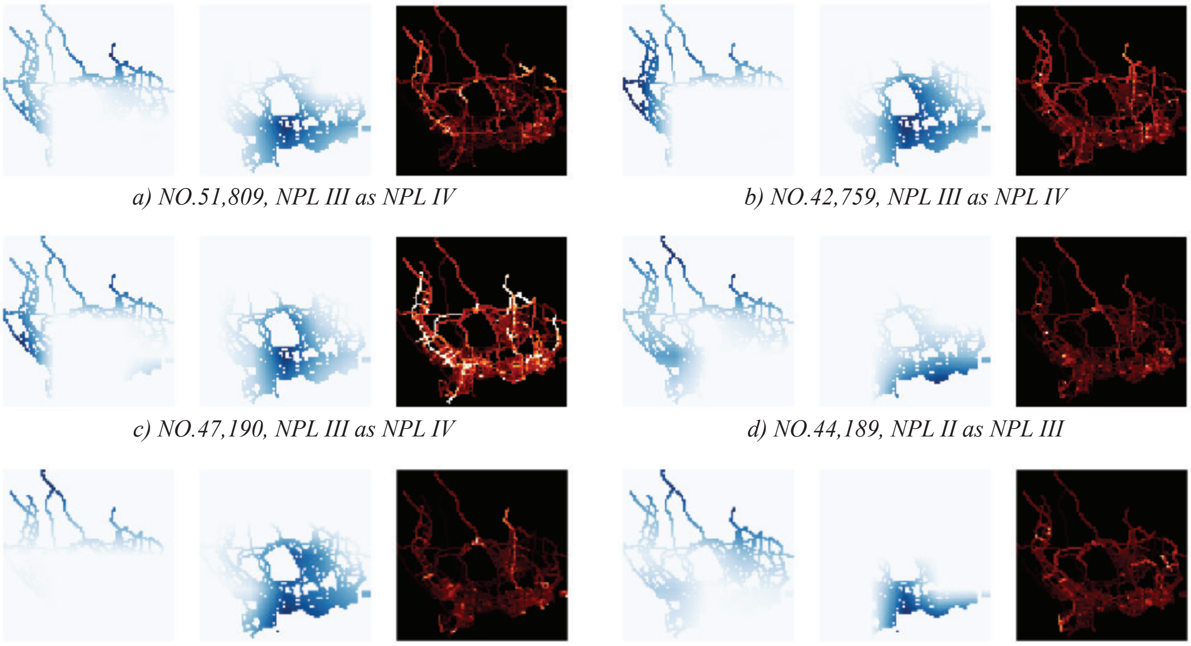
a) II to III, CAMs NO.20,384 ~ NO.20,390

b) III to IV, CAMs NO.20,820 ~ NO.20,830

c) IV to III, CAMs NO.20,405 ~ NO.20,416

d) III to II, CAMs NO.27,430 ~ NO.27,437

Figure 11 – CAM changes during its transition from an NPL to another.



a) NO.51,809, NPL III as NPL IV

b) NO.42,759, NPL III as NPL IV

c) NO.47,190, NPL III as NPL IV

d) NO.44,189, NPL II as NPL III

e) NO.47,976, NPL II as NPL III

f) NO.40,166, NPL II as NPL III

Figure 12 – CAMs and PMaps of some misclassified samples.

the appropriate attention to the significant area. In general, the CNN presented a stationary behaviour when making mistakes. Firstly, as discussed in Section 4.1, all misclassifications occurred in the NPL transition process. Secondly, although the CNN had learned unique patterns of NPLs, it seemed to pay excessive attention to the southeast area and underestimated the impact of the northwest area. It brought up some inspirations to the model improvement, such as weighting the pixels in the northwest area when training the CNN.

4.5 Discussion

From the results, both the single CAM paired with PMap and consecutive CAM sequences can provide insights into the NPL degradation and recovery to help improve the traffic operation. Some applications are listed as follows.

Manual NPL grading based on deep road network understanding: given the PMap patterns that CNN cares about when making decisions, managers can grade NPLs based on real-time PMaps with significant specific characteristics by themselves. Taking Shenzhen's road network as an example, it is easy to determine a PMap as NPL II when most pixels are filled with 0.30~0.45 TTI. When high-TTI and extremely high-TTI grids are mixed in the southeast region, the road network is prone to perform poorly, no matter the north area's situation. Sporadic high-TTI grids in the road network do not mean declining performance, and only regional management is required.

Near-future NPL trend forecasting: PI of each link is constantly changing. However, NPL may maintain the same, improve or deteriorate. The trend forecast is of great significance for traffic management. CNN visualisation can provide valuable knowledge and even illustrate the decisive area. According to Experiment 3, for Shenzhen's road network, the shrinking of the activation area, especially from the entire network to the northwest region, by the original NPL is a warning signal of approaching performance deteriorating. Hence, measures can be taken to manage the links and nodes corresponding to the high-TTI pixels displayed by PMap to alleviate traffic congestion.

Improving the CNN models that help management: the samples misclassified by CNN allow people to understand the model's deficiencies. Generally speaking, these deficiencies are related to the inappropriate weights the model assigned to pixels, which may be caused by insufficient training or data skew. Without visualisation, the parameter adjustment task is highly dependent on experience and repeated trial and error. Now that these pixels can be located by visualisation, one can enhance the model accordingly. Visualisation improves the efficiency of model improvement, and a better model can be obtained with ease to fulfil traffic analysis tasks.

Since managers will be benefited much from visualisation, a method with higher resolution is encouraged. However, the coarse granularity of the visualisation in this paper may be noticed. The following three measures can be implemented for improvement: firstly, this paper selected 82 × 82 images as input considering the calculation speed and road network scale. This low resolution allows computing to be performed online. Actually, the GIS map can be segmented into more pixels so that each link and even their lanes in different directions can be distinguished on the PMaps. Secondly, only the highest levels are visualised in this paper; however, the lower layers can also be visualised with adequate computing power, which may bring up a heatmap with finer granularity. Lastly, with the dramatic development of XAI, many visualisation methods are emerging to explain the mechanisms of CNN. Many of them achieve higher resolution, among which layer-wise relevance propagation (LRP) [22], deep learning important features (DeepLIFT) [23], and saliency maps [24] are three the authors consider worth exploring. Although they have not been introduced into traffic analysis, more satisfactory performances are expected given their finer granularity than Grad-CAM.

5. CONCLUSION

A CNN was trained to evaluate road network performance online based on all links' PIs. Compared with previous studies, this paper introduced the emerging CNN visualisation technology to discover what the CNN learned from the data set and visually present how the CNN made decisions. Four Experiments were conducted to show the insights CNN visualisation reveals about the road network and the CNN model. Firstly, it helps to understand the NPL transition from a macro perspective. It allows managers to see the hidden correlations between road network regions, which may have been neglected before. Knowing the transfer mode of crucial areas in the performance degradation process, one can effectively predict each area's performance to take measures beforehand. Secondly, visualisation can also help understand the model's failures to make the right decisions in certain circumstances, making it possible to optimise the models accordingly.

Improving CNN visualisation granularity is still an issue that must be tackled. Three ways are worth exploring, i.e. producing images with more pixels, visualising more layers of CNN, and applying more visualisation methods will high-resolution instinct to explain CNN. Besides CNN, XAI also provides solutions for explaining other AI models. Given that the interpretation of Long Short-Term Memory (LSTM) and Gate Recurrent Unit (GRU) models will open more black boxes for explaining time series in traffic analysis, it will be addressed in future studies.

ACKNOWLEDGMENTS

The authors are grateful to GAIA Open Dataset Initiative (https://gaia.didichuxing. com) for offering the data for the experiments. This research was financially supported by the National Key Research and Development Program of China under Grant 2018YFB1600805.

李君羡，博士研究生[1]，（通讯作者）
邮箱：1911549@tongji.edu.cn
吴志周, 博士, 副教授[1]
邮箱: wuzhizhou@tongji.edu.cn
沈宙彪, 硕士[2]
邮箱: shenzhoubiao@sucdri.com
[1] 同济大学，道路与交通工程教育部重点实验室
中国 上海市嘉定区曹安公路4800号,
邮编: 201804
[2] 上海市城市建设设计研究总院（集团）有限公司
中国 上海市浦东新区东方路3447号,
邮编: 200125

打开黑盒： 以可视化手段理解CNN在路网表现分级中的决策

摘要

可视化技术可以帮助理解深度学习模型的作用机理，但少见于交通分析领域。本文搭建了一个卷积神经网络*(CNN)*评估路网表现级别*(NPL)*，然后将其可视化以理解其决策机制。首先，使用某城市路网一年的数据生成路网表现地图，用他们去训练*CNN*。这一过程借鉴了预训练网络技术，解决交通研究领域常见的数据不足问题。然后，用梯度加权分类激活映射*(Grad-CAM)*将此*CNN*可视化，开展了四组实验。结果显示，*CNN*将路网认定为不同级别时关注的区域有明显差别，由此可找到在路网表现降级过程中的关键区域。当路网表现升、降级时，可视化模式有可循的变化规律，这有利于预测路网表现。分级错误的样本可视化结果则暴露出模型的缺陷。本文证明了可视化技术在辅助制定交通管理策略和交通智能模型提升方面的潜力。

关键词

可视化；卷积神经网络*(CNN);* 梯度加权分类激活映射*(Grad-CAM);* 预训练网络；路网表现.

REFERENCES

[1] Zhang J, et al. Data-driven intelligent transportation systems: A survey. *IEEE Transactions on Intelligent Transportation Systems*. 2011;12(4): 1624-1639. doi: 10.1109/TITS.2011.2158001.

[2] Pamula T. Road traffic conditions classification based on multilevel filtering of image content using convolutional neural networks. *IEEE Intelligent Transportation Systems Magazine*. 2018;10(3): 11-21. doi: 10.1109/MITS.2018.2842040.

[3] Duan Y, Lv Y, Liu Y, Wang F. An efficient realization of deep learning for traffic data imputation. *Transportation Research Part C - Emerging Technologies*. 2016;72: 168-181. doi: 10.1016/j.trc.2016.09.015.

[4] Zhao Z, et al. LSTM network: A deep learning approach for short-term traffic forecast. *IET Intelligent Transport Systems*. 2017;11(3): 68-75. doi: 10.1049/iet-its.2016.0208.

[5] Cheng Z, Wang W, Lu J, Xing X. Classifying the traffic state of urban expressways: A machine-learning approach. *Transportation Research Part A - Policy and Practice*. 2020;137: 411-428. doi: 10.1016/j.tra.2018.10.035.

[6] Hoang N, Le-Minh K, Tao W, Chen C. Deep learning methods in transportation domain: A review. *IET Intelligent Transport Systems*. 2018;12(9): 998-1004. doi: 10.1049/iet-its.2018.0064.

[7] Karlaftis MG, Vlahogianni EI. Statistical methods versus neural networks in transportation research: Differences, similarities and some insights. *Transportation Research Part C - Emerging Technologies*. 2011;19(3): 387-399. doi: 10.1016/j.trc.2010.10.004.

[8] Selvaraju RR, et al. Grad-CAM: Visual explanations from deep networks via gradient-based localization. *International Journal of Computer Vision*. 2020;128(2): 336-359. doi: 10.1007/s11263-019-01228-7.

[9] Du S, Li T, Gong X, Horng S. A hybrid method for traffic flow forecasting using multimodal deep learning. *International Journal of Computational Intelligence Systems*. 2020;13(1): 85-97. doi: 10.2991/ijcis.d.200120.001.

[10] Bogaerts T, et al. A graph CNN-LSTM neural network for short and long-term traffic forecasting based on trajectory data. *Transportation Research Part C - Emerging Technologies*. 2020;112: 62-77. doi: 10.1016/j.trc.2020.01.010.

[11] Zeiler MD, Fergus R. Visualizing and understanding convolutional networks. *Proceedings of ECCV, 6-12 Sep. 2014, Zurich, Switzerland*. Cham, Switzerland: Springer International Publishing; 2014. p. 818-833. doi: 10.1007/978-3-319-10590-1_53.

[12] Springenberg JT, Dosovitskiy A, Brox T, Riedmiller M. Striving for simplicity: The all convolutional net. *Proceedings of ICLR, 7-9 May 2015, San Diego, CA, USA*. 2015. arXiv:1412.6806.

[13] Mahendran A, Vedaldi A. Visualizing deep convolutional neural networks using natural pre-images. *International Journal of Computer Vision*. 2016;120(3): 233-255. doi: 10.1007/s11263-016-0911-8.

[14] Dosovitskiy A, Brox T. Inverting visual representations with convolutional networks. *Proceedings of IEEE Conf. on CVPR, 27-30 June 2016, Las Vegas, NV, USA*. Los Alamitos, CA, USA: IEEE Computer Society; 2016. p. 4829-4837. doi: 10.1109/CVPR.2016.522.

[15] Zhou B, et al. Learning deep features for discriminative localization. *Proceedings of IEEE Conf. on CVPR, 27-30 June 2016, Las Vegas, NV, USA*. Los Alamitos, CA, USA: IEEE Computer Society; 2016. p. 2921-2929. doi: 10.1109/CVPR.2016.319.

[16] Lin M, Chen Q, Yan S. Network in network. *Proceedings of ICLR, 14-16 Apr. 2014, Banff, Canada.* 2014. arXiv:1312.4400, 2014.

[17] Simonyan K, Zisserman A. Very deep convolutional networks for large scale image recognition. *Proceedings of ICLR, 7-9 May 2015, San Diego, CA, USA*. 2015. arXiv:1409.1556.

[18] Jogin M, et al. Feature extraction using convolution neural networks (CNN) and deep learning. *Proceedings of IEEE Int. Conf. on RTEICT, 18-19 May 2018, Bangalore, India*. Piscataway, NJ, USA: IEEE; 2018. p. 2319-2323. doi: 10.1109/RTEICT42901.2018.9012507.

[19] Agrawal P, Girshick R, Malik J. Analyzing the performance of multilayer neural networks for object recognition. *Proceedings of ECCV, 6-12 Sept. 2014, Zurich, Switzerland.* Cham, Switzerland: Springer International Publishing; 2014. p. 329-344. doi: 10.1007/978-3-319-10584-022.

[20] Chollet F. *Deep Learning with Python*. Shelter Island, NY, USA: Manning Publications; 2017.

[21] Hinton G. *Coursera Course Lectures*. 2012. http://www.cs.toronto.edu/~hinton/coursera_slides.html [Accessed 21st Dec. 2021].

[22] Bach S, et al. On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. *PLoS ONE*. 2015;10(7): e0130140. doi: 10.1371/journal.pone.0130140.

[23] Shrikumar A, Greenside P, Kundaje A. *Learning important features through propagating activation differences.* 2019. arXiv:1704.02685v2.

[24] Simonyan K, Vedaldi A, Zisserman A. Deep inside convolutional networks: Visualising image classification models and saliency maps. *Proceedings of ICLR, 14-16 Apr. 2014, Banff, AB, Canada.* 2015. arXiv: 1312.6034.