**ANSWER TO REVIEWER COMMENTS**

**TRD\_2016\_462\_R1**

The authors would first of all like to thank the reviewers for the positive and constructive feedback on our original manuscript. We believe that your comments have enabled us to more clearly state our scope, results and conclusions and we hope we have reflected this to your satisfaction in our revised manuscript. The tables below address the specific comments. Correspondingly, some changes to the text of the original submission have been made and such changes are marked with red color.

**REVIEWER #1**

|  |  |  |
| --- | --- | --- |
| **#** | **Comment** | **Answer** |
| **1** | #1. This paper does not provide very convinced literature review to indicate the background why the CNN model is suitable to forecast the traffic flow, is only that the CNN right now a popular tools? Please provide more convinced relevant literature review to support the very reason why CNN is suitable to be employed to traffic flow forecast. | Thank you for pointing out this issue. CNN is widely used in computer vision and image classification applications, Besides, in the field of transportation, Ma et al. (2017) has demonstrated that convolutional neural network (CNN) is suitable for traffic prediction, and after that, Du et al. (2018) uses CNN as a first layer to capture the features of different modality traffic data. |
| **2** | #2. Section 2.1 should move to the relevant literature review section, due to it divides the traffic flow conditions have temporal and spatial characteristics. | Thank you for pointing out this issue. Section 2.1 had moved to the relevant literature review section and some of statements have been improved. |
| **3** | #3. In addition, in page 6, please provide more statements for the contributions of this paper, at least 150 words for each contribution point. | Thank you for pointing out this issue.  In light of the comments of other reviewers, we restated the contributions of this paper as follows:   1. Traffic flow spatio-temporal relationship analysis and proposed a spatio-temporal feature selection algorithm.   正如文章的摘要部分所述，交通流量中蕴含了丰富的时空特性，传统的只针对时间维度的预测算法丢失了交通流量的空间分布和转移关系，这限制了预测准确度的提高。大部分最新的研究中只是对于预测算法模型的建立和更新，很少有研究输入数据的选取对于预测结果的影响。本文通过对于输入时空数据和预测结果的影响，提出一种对于输入时空数据选取的算法，能够尽可能大地提高模型预测精度，同时减少输入数据的大小，降低模型拟合参数的数量，提高在线预测时间。   1. A CNN prediction model based on spatio-temporal correlations.   深度学习方法能够很好地应对交通数据的多维度、大量、复杂非线性关系的特点因此，近年来在交通领域引起了很高的关注度。卷积神经网络及其变体在计算机视觉和图像分类领域得到了持续的关注和足够的应用，它能够对于array-like数据提取其中的空间相互关系，在交通领域中，从不同检测线圈中采集的交通数据经过时间推移也是一种array-like数据，因此我们尝试将交通采集线圈采集的数据转化为二维矩阵形式，并通过卷积神经网络对于这个二维矩阵提取其中的相互关联关系称之为时空相关性，并将其用于预测。   1. Optimized objective function construction method for noisy traffic flow prediction.   从检测线圈或者是路侧设备、红外线装置等采集的数据都不是绝对准确的，这是由于交通流量演变过程具有一定的不确定性，而且由于道路中的意外事件、天气、交通管制和测量误差等等原因给交通流量预测带来了阻碍，这种阻碍体现在数据上就是交通流量数据的变化较为剧烈，在调优预测模型参数的过程中，我们一般需要对于具体的目标函数进行优化，本文对于具体的目标函数构造提出了一些看法，并在最后的实施过程中取得了良好的预测结果。 |
| **4** | #4. By the way, for the citation problem, line 31 on page 3, Yang and Zhu (1999) is lost in the reference list, please check it carefully. For the reference list problems, (1) please avoid citing working papers, such as Ermagun and Levinson (2016); (2) please also avoid citing meeting conclusion, such as Zhang et al. (2016); (3) please provide complete citation information, such as Kingma and Ba (2014), Wu et al. (2015), and Wu and Tan (2016); (4) please provide relevant DOI for those papers in press status, such as Xu et al. (2018) and Zhang et al. (2018). | Thank you for pointing out this issue. We have fixed all the issues mentioned and checked the full text. |
| **5** | #5. Authors should provide the very details illustrating how the proposed model is working in the experimental results section, i.e., lacking of some essential brief explanation vis-à-vis the text to indicate how the proposed methodology (Figure 1 in page 10) is working in the experimental results section. In addition, for Figure 3 in page 14, please also provide some necessary wordings to guide readers to understand what authors have done and obtained from Figure 3. For Eq. (11) in page 18, please explain why this equation should be involved to be as the loss function; similarly, for Eq. (12) in page 19. Algorithm 1 in pages 26 and 27 is not introduced in text, please provide some necessary illustrations for Algorithm 1. | Thank you for pointing out this issue.  We have added the necessary explanations for the figures, tables, etc. in the articles mentioned.  Figure 1 is a flowchart of the traffic flow prediction framework proposed in this paper. 在数据预处理部分，不同路段的检测线圈被提取后以矩阵的形式表示，这部分内容在文章的3.3节被详细介绍；在数据选择部分，时空相关性分析在3.1和3.2章详细介绍；在预测方法部分，神经网络的内部构造方法在4.1章详细介绍。  Figure 3 is a data extraction and matrix representation process 我们在图3上增加了更多的说明性信息，我们按照时间维度和空间维度将不同位置采集的线圈数据整合为一个大的二维矩阵，然后通过STFSA选取这个大的二维矩阵的子矩阵的大小作为构造模型的一个输入数据实例的大小。  For Eq. (11) in page 18, where *λ* denotes L2 regularization coefficient, *wj* represents the weight of layers which uses L2 regularization. The L2 regularization method reduces the over-fitting risk of the model by penalizing the large weighting coefficients between the neuron connections, thereby improving the generalization ability of the model.  For Eq. (12) in page 19, in the process of tuning up the weight coefficient of the model using the gradient descent algorithm, the use of mini-batch reduces the influence of a single sample instance with large errors on the entire optimization process by weighting the gradient of a sample set, and speeds up the entire convergence process. |
| **6** | #6.  For Table 5 in page 31, authors should try to conduct some statistical test to verify the significance of the forecasting performance from the proposed approach. Without the significant test, this paper only has minor contribution. Please refer Diebold and Mariano (1995) and Derrac et al. (2011). F. X. Diebold and R. S. Mariano, “Comparing predictive accuracy,” Journal of Business & Economic Statistics, vol. 13, No. 3, pp. 134-144, 1995. Derrac, J.; García, S.; Molina, D.; Herrera, F. A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms. Swarm Evolutionary Computation 2011, 1, 3–18. | Thank you for indicating this issue.  Our foregoing analysis dose not answer the question of how much statistical confidence we have in saying our proposed method is “better” than other methods on the basis of MAPE. So we perform a two-tailed binomial test pairwise comparison between the proposed algorithm and other algorithms. An α=0.05 significance level was used for hypothesis testing. The two-tailed binomial test was carried out by the statistical software R. |

**REVIEWER #2**

|  |  |  |
| --- | --- | --- |
| **#** | **Comment** | **Answer** |
| **1** | #1. First, the paper is unnecessarily lengthy, and the authors should consolidate the presentations greatly for concisely demonstrating the studies. | Thank you for pointing out this issue.  We have simplified the language of the full text, and deleted the redundant part, adding the necessary explanation for the diagram and table. |
| **2** | **2.   #**2. Second, the proposed methods and the comparative methods should be given in separate sections for clear presentation. | Thank you for pointing out this issue.  A brief introduction to each comparative model and detailed parameter settings have separately given in Section 5. |
| **3** | 3.     #3. Third, computational efficiency is important for online traffic prediction, and this should be investigated in this paper. | Thank you for pointing out this issue.  We analyzed the training time and online prediction time of the proposed model and comparison models, and presented in Figure 8 and Figure 9 of the revised manuscript. |
| **4** | 4.    #4. Finally, seasonal time series model and k-nearest neighbor model are two conventional short term traffic prediction models, and should be selected in this paper as comparative methods. | Thank you for pointing out this issue.  SARIMA is a linear model that is most widely used in the forecasting field, and KNN is also a representative non-parametric model. Both models are widely used in traffic prediction. We added these two models as a comparison model in the revised submission. The construction process of the model is referred to Williams and Hoel (2003) and Habtemichael and Cetin (2016).  Williams, B. M., & Hoel, L. A. (2003). Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results. *Journal of transportation engineering, 129*(6), 664-672.  Habtemichael, F. G., & Cetin, M. (2016). Short-term traffic flow rate forecasting based on identifying similar traffic patterns. *Transportation Research Part C: Emerging Technologies, 66*, 61-78. |

**REVIEWER #3**

|  |  |  |
| --- | --- | --- |
| **#** | **Comment** | **Answer** |
| **1** | #1. The contributions of the work in the context of existing literature is not clear. Presently, the contributions stated are:  *(1) A framework for traffic flow prediction, (2) an optimal input data selection algorithm, and (3) a CNN prediction model based on spatial-temporal correlations.*  (1) & (2) are not contributions to knowledge. (3) can be a contribution provided it is established through literature. So, please rephrase and be specific about the contributions. | Thank you for pointing out this issue.  Contributions are updated as below:  (1) Traffic flow spatio-temporal relationship analysis and proposed a spatio-temporal feature selection algorithm.  (2) A CNN prediction model based on spatio-temporal correlations.  (3) Optimized objective function construction method for noisy traffic flow prediction.  (将第一个评委的意见改完之后再粘贴到这边)增加适当的文献指引，例如在文献增加部分许多文章中都没有给出时间和空间数据选择的方法。数据的选取缺乏依据例如： |
| **2** | #2. The matrix presentation methods described for choosing the temporal and spatial input points seems heuristic /arbitrary. The ACF plot presented is for non-stationary data and 0.68 cannot be arbitrarily chosen as a cut-off! The spatial correlation method described is quite vague. How are the datasets are actually matched? Pearson’s correlation? What about lagged spatial correlation?? | Thank you for pointing out this issue.  我们对于检测线圈的自相关性分析和不同检测线圈之间的互相关分析是为了说明交通流量数据中包含着时间相关性和空间相关性，但是以往的模型并没有着重于对于交通输入数据的分析，而是注重于对于交通预测模型的设计。0.68的设置是为了构造一个对比样本，没有经过数据选择的预测模型和经过STFSA选择的模型之间的区别。  在修改后的稿件中我们将这个看似不合理选取的0.68选取规则去除，而是将另外一篇文章中的时间维度选取数量作为对比模型，验证经过STFSA选取后的数据是否能够降低模型的预测误差。 |
| **3** | #3. In figure one the prediction methods contain a few boxes which do not show any sequence details of the process. It is necessary to provide more information in significantly more details on how the boxes are linked and what happens inside a box. For eg. ‘Determine network structure’ Does not give information about which network, how to structure is determined, what elements of the structures are determined at every step of prediction or during the training stage. It is also unclear whether all these steps are necessary to be followed during the production process or just during the training of the network. | Thank you for pointing out this issue.  对于CNN模型构建过程中的超参数设置很大程度上取决于设计者的经验，在模型构造过程中，我们需要对网络层数、激活函数、损失函数、mini-batch大小、优化函数等进行选择。we do need provide some necessary wordings to guide readers to understand what authors have done and obtained from. 因此我们在修改后的稿件中对于图中的每个部分都进行了阐述和读者的引导。 |
| **4** | #4. How is the initial correlation analysis for matrix presentation is linked with STFSA? Does it provide/generate Rinitial? Why two steps are necessary? Why not start with STFSA directly rather than using arbitrary correlation analysis? | Thank you for pointing out this issue.  我们确实不需要将一个arbitrary的相关性分析加入我们的STFSA中，我们在更新后的稿件中对这一部分做了重新的论述，并加入了详细的解释 |
| **5** | #5. Fig 6 presents some confusing results. It shows that with longer length of data the prediction accuracy decreases. This is counterintuitive. The MAPE should show that prediction accuracy increases when input data length increases between 5-20mins and then it levels out for further data points. Again, if the input length in increased significantly to up to 24 hrs or longer, we should see another improvement in prediction accuracy. Longer length of data should not affect the prediction accuracy. For spatial sections, it was initially described only 8 neighboring detectors were used…however the results are presented for nearly up to 30 detectors. Please explain. | Thank you for pointing out this issue.  First of all, we should explain that the first submission version does not clearly explain this part, which may lead to readers' misunderstanding. The conclusion in Figure 6 corresponds to the 5-minute traffic prediction problem for different prediction problems. (If the predicted length changes) Data selection and model training should be performed separately. As can be seen from Figure F of Chapter N, in actuality, for the 5-minute traffic flow forecast, the extended time lag does not bring about an effective improvement in improving the prediction effect. Further, if the time interval is increased, the prediction can be significantly reduced. The error (which depends on the selection of the insensitive coefficient α), STFSA can extend the prediction time lag to 24h or even longer. We know that adding a certain seasonal analysis to the time prediction problem may bring some improvement to the prediction results, but for the highway traffic problem, our research hopes to finally screen out relatively small input data to reduce the model's Complexity, so that the model can be processed on devices with small computing power.  For spatial sections, it was initially described only 8 neighboring 因为我们是想做一组比较试验，分析固定的时间和空间点数据的预测结果与经过STFSA之后的区别。第一版稿件中的图6是分析了对于五分钟的交通流量预测选择不同的检测线圈对于交通流量的影响。 |
| **6** | #6. Apart from the aforementioned issues, the rest of the results seem appropriate. But the STFSA which is claimed to be the main contribution of the work, does not seem to make a major difference to the MAPE error of the ANN or SVM accuracies published elsewhere. I understand that the model accuracy is dependent on the data characteristics to some extent, however STFSA framework is not making any significant difference. | Thank you for pointing out this issue.  It can be seen from Figure 9 in Chapter 6 that the number of different time and space detection coils does have a large impact on the final prediction error value. We analyzed the short-term traffic flow prediction for 5-20 minutes, and finally we came to In conclusion, for different forecasting tasks, the impact of changes in time and spatial dimensions on the final forecast is unbalanced. For example, we found 5 minutes of traffic in different time and space data organization methods at different intervals of 20 intersections. The main source of influence of the flow prediction process error is the amount of spatial data. However, this situation is not applicable to the 10-15 minute traffic prediction problem, so we finally found that the necessary data selection process for different prediction tasks. It is indeed effective in reducing the error of prediction. |
| **7** | #7. The input data size for ANN, SVR and CNN as it is needs to be presented and compared with ANN+STFSA, SVR+STFSA and CNN+STFSA. Computational time and efficiency needs to be discussed. | Thank you for pointing out this issue.  We conducted experiments on the traffic flow prediction of multiple road sections to verify the applicability of the proposed method to different road sections. In Chapter 4, we took the analysis coil of No. 96 as an example to give an analysis process for a specific road section. Parameter configuration and final use data for models and comparison models.  We analyzed the training time and online prediction time of the proposed model and comparison models, and presented in Figure 8 and Figure 9 of the revised manuscript. |
| **8** | #8. The paper should be organized better. STFSA should be presented before other methods. Also, no need to give so much details of CNN, ANN or SVR. Provide appropriate reference. | Thank you for pointing out this issue.  We optimized the structure of the full text in the revised manuscript, and streamlined the linguistic expression of the full text, removed some of the redundancy, and added more explanations to the charts and tables to guide the reader through the work done by the author. In addition, the SARIMA and KNN models are added as two classic models. A brief introduction and detailed parameter settings for each comparison algorithm are given in Section 5, and an appropriate reference is provided for each comparison algorithm. Besides, We believe that it is also important to deploy CNN in the transportation field, so the description of CNN is retained. |
| **9** | #7. The phrase ‘time correlation’ should be changed to ‘temporal correlation’. | Thank you for pointing out this issue.  We checked the full text and corrected the statement. |
| **10** | #10. Figure 1: There are some typos in this figure. The word person should be changed to Pearson. | Thank you for pointing out this issue.  We checked the full text and corrected the statement. |