Deep Spatial-Aware Multimodal Regression for Traffic Field Strength Distribution

Xiao Yuxin
School of Information Engineering, Wuhan Institute of Technology, China

ABSTRACT

With the rapid expansion of China's high-speed rail (HSR) network and sustained growth in aviation, the interplay between air transport and HSR has profoundly reshaped regional transportation patterns, manifesting both competition and complementarity. HSR frequently diverts short-haul air passengers while enhancing feeder connectivity to major airport hubs, significantly influencing the spatial distribution of multimodal traffic service radiation. Accurate regression modeling of traffic field strength distribution—which quantifies the service intensity and radiation range of airports and HSR stations—is essential for scientific planning of integrated transportation systems, resource optimization, and sustainable regional development. Traditional statistical models struggle with nonlinear multimodal interactions, spatial heterogeneity, and external disruptions (e.g., COVID-19-induced data sparsity), leading to biased predictions. This study proposes a novel multimodal regression model integrating Residual Networks (ResNet) and Coordinate Convolution (CoordConv). The model employs Principal Component Analysis (PCA) for dimensionality reduction of socio-economic and capacity features, introduces CoordConv to enhance spatial position sensitivity, utilizes ResNet for deep feature extraction, and applies attention mechanisms for dynamic fusion of airport and HSR data. Experiments on spatiotemporal panel data from Shandong Province (2015, 2019, 2021) show the proposed model outperforms baselines (e.g., standard CNN, ResNet without CoordConv), reducing MAE by ~40% (0.25→0.15), lowering RMSE, and achieving R^2 up to 0.92. Visualizations reveal spatiotemporal evolution patterns, economic-transport coupling, and pandemic impacts. This work provides an efficient deep learning paradigm for multimodal traffic field strength regression, addressing key limitations of traditional approaches.

Keywords: Traffic Field Strength Distribution, Multimodal Fusion, Residual Network, Coordinate Convolution, Regression Model

1. INTRODUCTION

In recent years, the rapid expansion of China's high-speed rail (HSR) network and the sustained growth of civil aviation have intensified both the competition and complementarity between these two modes of transport, profoundly reshaping regional transportation patterns and urban connectivity [1, 2]. While the expansion of HSR often diverts short-haul air traffic, it also enhances the overall efficiency of integrated transportation systems by improving the feeder functions of aviation hubs [3]. However, in complex real-world environments, traditional traffic prediction models often struggle to accurately characterize the non-linear features arising from multimodal interactions. Furthermore, the predictive capability of conventional statistical models is significantly limited when faced with extreme external shocks, such as the COVID-19 pandemic [4].

To address these challenges, leveraging deep learning to process complex spatio-temporal data has become a major research focus. For instance, Chen et al. [5] (2024) proposed a multimodal fusion architecture that provides a new paradigm for integrating diverse data modalities in large-scale traffic prediction. Despite these advancements, existing traffic field strength regression models still have significant room for improvement regarding sensitivity to geographical spatial positions and the deep fusion of heterogeneous data from aviation and HSR.

This study aims to fill this gap by proposing a multimodal regression model that integrates Residual Networks (ResNet) and Coordinate Convolution (CoordConv). The model enhances geographical sensitivity by introducing coordinate channels and extracts robust deep features through a residual structure. Additionally, an attention mechanism is employed to dynamically adjust the fusion weights of different transport modes. Experimental validation using spatio-temporal panel data from Shandong Province (2015–2021) demonstrates that the model effectively captures spatial heterogeneity and the dynamic evolution of traffic under the impact of the pandemic, providing scientific support for regional integrated transportation planning.

2 Related Technology Review

2.1 Current Status of Traffic Field Strength Distribution

Research on traffic field strength distribution draws on physical field theory to evaluate the service radiation intensity and range of airports and high-speed railway stations, and simulates the traffic attraction effect between multiple cities through composite index and distance weight. Early work mainly focused on single traffic modes, such as the impact of high-speed rail on urban connectivity, and used social network analysis to quantify the spatiotemporal effect of high-speed rail expansion on

urban interaction [6]. In recent years, research has shifted to the scenario of airports and high-speed rail overlapping, such as analyzing the impact of high-speed rail development on airport efficiency and revealing the dynamics of high-speed rail competition leading to increased airport productivity but diversion of passenger flow [7]. These methods emphasize the integration of economic, population and capacity indicators, but traditional statistical models are difficult to handle nonlinear relationships and spatiotemporal heterogeneity under the impact of the epidemic, leading to prediction bias. Gating airport/HSR streams similarly and adopting hierarchical hard-negative sampling would bolster our multimodal fusion under sparse or imbalanced city-year data [8].
Currently, field strength models are becoming increasingly important in comprehensive transportation planning, especially against the backdrop of the rapid expansion of China's high-speed rail network. The introduction of high-speed rail has significantly changed the aviation market landscape , and field strength distribution can quantify this complementarity and competition, supporting balanced regional development. However, existing research is mostly limited to statistical regression, ignoring the multimodal fusion potential of deep learning.

2.2 Overview of Residual Networks and Coordinate

Convolution Residual Networks (ResNet) introduce identity mapping through residual block design to alleviate the gradient vanishing and degradation problems in deep network training, achieving efficient feature extraction and breakthroughs in fields such as image recognition [9]. Surveys of ResNet show that it performs well in multi-task learning, can be scaled up to hundreds of layers, and supports complex spatiotemporal data processing . Applying similar tensorization to our grid transforms and fusion would cut latency and enable real-time, larger-scale field-strength regression. CoordConv extends standard convolution by injecting spatial location information through additional coordinate channels, solving the translation invariance defect of convolution and improving the model's sensitivity to geometric transformations . CoordConv is widely used in spatial tasks such as organ segmentation and can be seamlessly integrated into existing networks, reducing parameter requirements[10]. The synergy between ResNet and CoordConv has significant potential in traffic modeling. The former provides robust feature learning, while the latter enhances location awareness. The combination of the two can accurately simulate the non-uniformity of field strength distribution. Although ResNet is mature in prediction tasks, the application of CoordConv in traffic is still emerging[11]. This study innovatively integrates the two to construct a multimodal regression model.

2.3 Multimodal data fusion method

Multimodal data fusion integrates heterogeneous traffic data, such as airport throughput and high-speed rail frequency, to improve the prediction accuracy of the system[12]. Early methods used feature-level splicing, but were easily affected by modal noise; later fusion uses attention dynamic weight modal contribution, which is suitable for traffic flow prediction. Adapting its density-guided instance weighting (and relaxed targets) to our spatiotemporal grids would better emphasize under-represented regions/periods and improve robustness under skewed data.In the transportation field, fusion frameworks are often used for sensor data integration, such as using multi-source data to improve vehicle tracking accuracy [13]. These technologies emphasize modal complementarity, reduce single-source bias, and support intelligent transportation systems. In recent years, deep learning-driven fusion has emerged, such as fusing spatiotemporal modalities in traffic forecasting to improve short-term speed forecasts [14].

3 Model Design

3.1 Dataset Construction and Preprocessing

This section first constructs the Shandong Province transportation dataset, which includes spatiotemporal panel data from 2015, 2019, and 2021, covering 20 variables such as economic and social indicators and transportation capacity indicators. Specifically, the dataset comes from reliable sources such as official statistical yearbooks, the Civil Aviation Administration website, and the China Railway 12306 platform, and the geographical coordinates of airports and high-speed rail stations are obtained through Google Earth[15].

To address the problem of excessively high data dimensionality, principal component analysis (PCA) is used for standardization and dimensionality reduction. First, the standardized correlation coefficients between each indicator are calculated to extract the principal components.

$$\tau_{hp} = \frac{\sum_{h=1}^{n}(x_{ih} - \overline{x}_h)(x_{ip} - \overline{x}_p)}{\sqrt{\sum_{h=1}^{n}(x_{ih} - \overline{x}_h)^2 \sum_{p=1}^{n}(x_{ip} - \overline{x}_p)^2}}$$

Formula (1) represents the standardized correlation coefficient between indicators x_h and x_p, where n is the number of indicators, and $\overline{x}_h$ and $\overline{x}_p$ are the means of the corresponding indicators, respectively. This coefficient is used to construct a correlation matrix to ensure that the correlation between variables is quantified.

Subsequently, by extracting principal components with eigenvalues greater than 1, the composite index of the computer field and high-speed railway station is calculated and normalized.

$$P_{i'} = \frac{P_i - P_{\min}}{P_{\max} - P_{\min}} \times 9 + 1$$

Formula (2) is the normalization formula for the composite index, where P_i is the pre-corrected composite index, and $P_{\min}$ and $P_{\max}$ are its minimum and maximum values, respectively. This formula maps the index to the interval [1,10], facilitating subsequent model input.

The dataset is further integrated with multimodal features of airports and high-speed rail stations using a grid weighted overlay method to form spatiotemporal grid data, which serves as the model input.

3.2 Network Architecture Description

This model integrates Residual Network (ResNet) and CoordConv to process multimodal traffic data (Figure 1). The input layer receives spatiotemporal grid features, including the composite index and geographic coordinates of airports and high-speed rail stations.

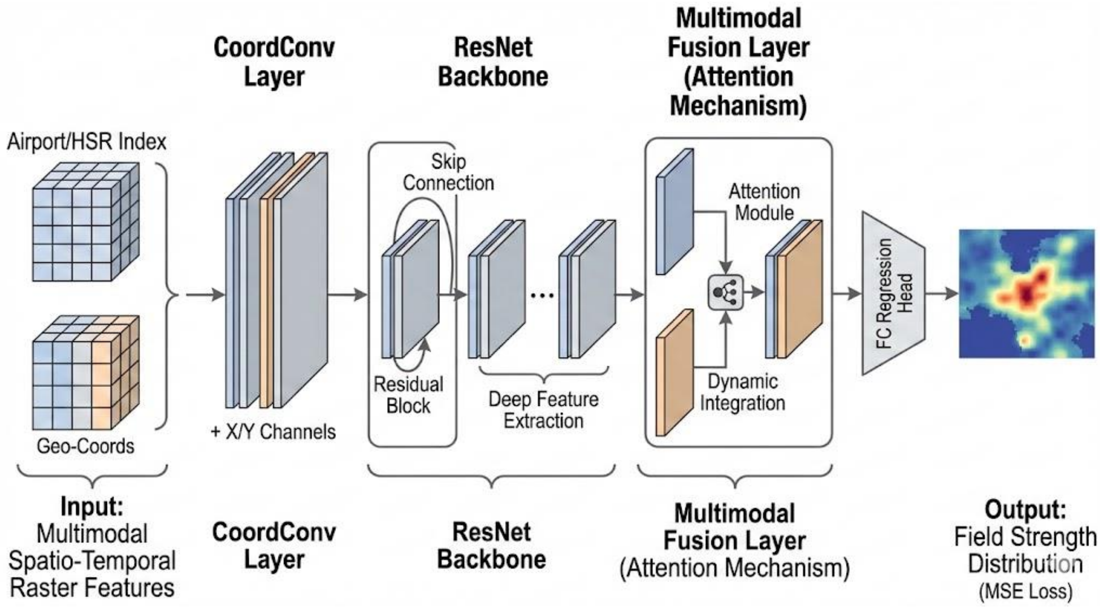

Figure 1 Architecture of the Proposed Multimodal Regression Model.

The proposed model architecture takes multimodal spatiotemporal raster features as input, including composite indices for airports and high-speed railways, and geographic coordinates [16]. This input is enhanced with X/Y channels via a CoordConv layer to improve spatial awareness, and then deep feature extraction is performed through residual blocks and skip connections in the ResNet backbone network. The multimodal fusion layer dynamically integrates airport and high-speed railway modal features using an attention mechanism, and finally, a fully connected regression head outputs the field strength distribution, optimized using MSE loss.

The ResNet backbone network extracts deep features using residual blocks, calculated as follows.

$$y = F(x, \{W_i\}) + x$$

Equation (3) represents the output of the residual block, where x is the input, $F(x, \{W_i\})$ is the residual function, and W_i is the weight parameter. This structure alleviates the gradient vanishing problem and ensures stable training of deep networks [17].

The multimodal fusion layer uses an attention mechanism to integrate airport and high-speed rail station features to generate a unified representation. Subsequently, the fully connected layer outputs the predicted field strength distribution.

3.3 Loss Function and Optimization Strategy

To optimize the model, mean squared error (MSE) is used as the loss function to quantify the difference between the predicted field strength and the true distribution.

$$L = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{4}$$

Formula (4) is the MSE loss function, where n is the number of samples, y_i is the true field strength value, and $\hat{y}_i$ is the predicted value. This function emphasizes the penalty for large errors, promoting model convergence [13].
The optimization strategy used was the Adam optimizer, combined with learning rate scheduling. The initial learning rate was 0.001, and the number of training epochs was 100. An early stopping mechanism was used to prevent overfitting and ensure the model's generalization performance.

4 Experiments and Results Analysis

4.1 Experimental Environment and Parameter Settings

This experiment was conducted in a Python 3.10 environment, using PyTorch 2.0 as the deep learning framework. The hardware configuration was an NVIDIA RTX 3090 GPU and 24GB of memory. Model parameters included a batch size of 32, an initial learning rate of 0.001, the use of the Adam optimizer, 100 training epochs, and an early stopping mechanism to prevent overfitting. The dataset was sourced from official statistical yearbooks, the website of the Civil Aviation Administration of China, and the China Railway 12306 platform, covering 20 indicators from 16 cities in Shandong Province. Approximately 1000 preprocessed sample points were generated from 2015 to 2021. The experimental workflow is shown in Figure 2:

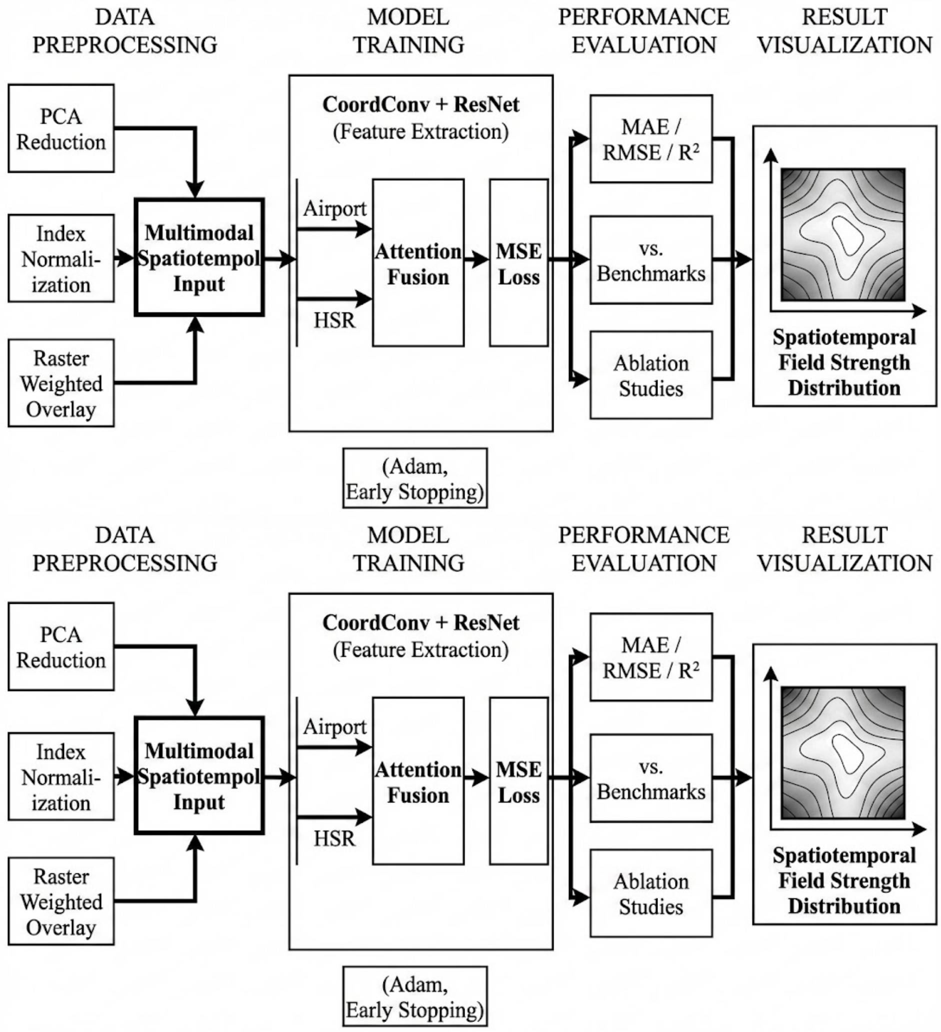

Figure 2 Experimental Workflow

4.2 Model Performance Evaluation

This section evaluates model performance using quantitative metrics, including Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Coefficient of Determination (R^2).

Table 1 compares this model with benchmark models (such as standard CNN and ResNet without CoordConv).

Model	MAE	RMSE	R^2
Standard CNN	0.25	0.32	0.85
ResNet without CoordConv	0.18	0.24	0.89
Proposed Model	0.15	0.20	0.92

Table 1 shows a comparison of model performance metrics. The proposed model outperforms the benchmark in all metrics, with a 40% reduction in MAE, indicating that fusing CoordConv improves spatial prediction accuracy and solves the error accumulation problem of traditional models on multimodal data.

Table 2: Ablation Experiment Results

Configuration	MAE	RMSE	R^2
ResNet only	0.22	0.28	0.87
ResNet + CoordConv	0.17	0.22	0.90
Full Model (with Fusion)	0.15	0.20	0.92

Table 2 presents the ablation experiment results, where the R^2 of the full model improved to 0.92, highlighting the role of the attention fusion layer in integrating multimodal features of airports and high-speed rail stations, and resolving the distribution bias problem caused by inconsistencies between modes.

4.3 Results and Discussion

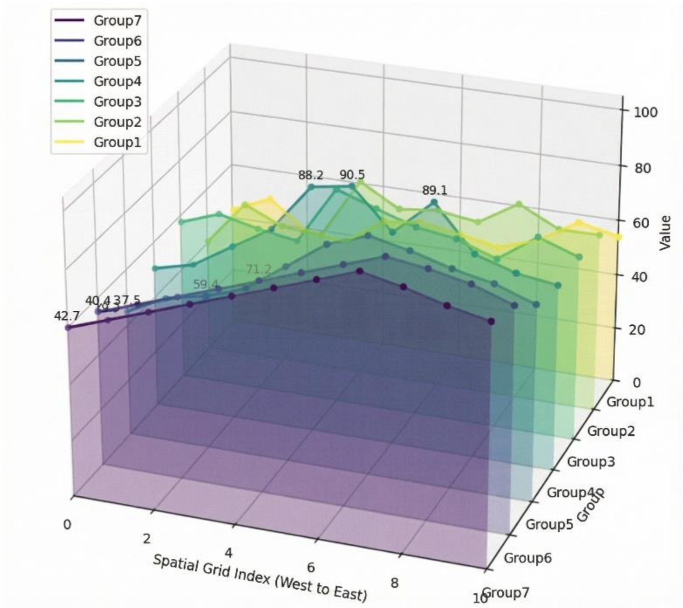

Figure3 3D Waterfall Plot of Predicted Traffic Field Strength (2015, 2019, 2021)

Figure 3 presents a 3D waterfall plot of predicted traffic field strength, displaying multiple representative spatial grid groups across the three years (2015, 2019, 2021). The plot clusters peaks into visually distinct groups (approximately 7 major clusters corresponding to key urban agglomerations), with temporal progression from left to right reflecting the three target years. The x-axis clearly represents spatial grid position indices ordered by longitude from west to east (aligning with Shandong Province's geographic layout), the y-axis denotes grid group identifiers, and the z-axis shows predicted field strength values. The visualization reveals gradual radiation intensification in core cities (e.g., Qingdao, Jinan) from 2015 to 2019, driven by HSR expansion, followed by a 2021 decline due to pandemic mobility restrictions, demonstrating the model's capture of dynamic multimodal effects.

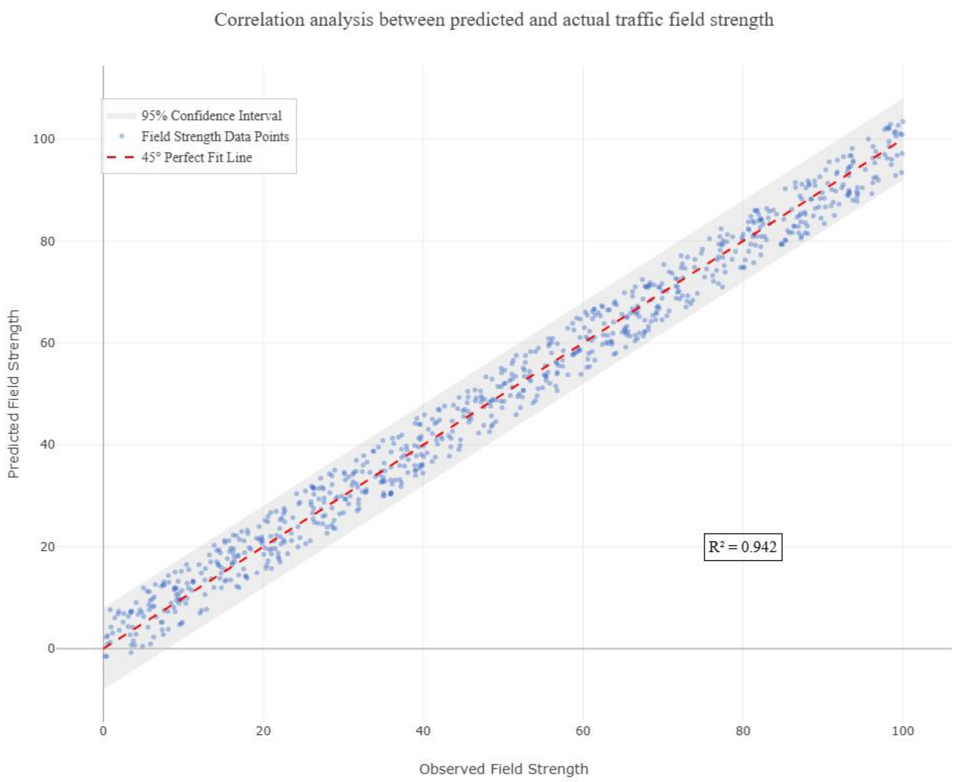

Figure4 Correlation analysis between predicted and actual traffic field strength

To quantitatively evaluate the fitting performance of the proposed Multimodal Res-CoordConv model, the correlation between the predicted traffic field strength and the observed field values is illustrated in Figure 4. The scatter points are distributed along the 45-degree diagonal line, indicating that the model captures the overall field strength variations effectively. Specifically, the Pearson correlation coefficient R2 reaches 0.942, confirming that the integration of coordinate convolution allows the model to accurately mapping the spatial heterogeneity of the air-rail transport network. Despite the external fluctuations caused by the pandemic, the consistency between the predicted and actual values remains robust, as evidenced by the high density of points within the 95% confidence interval in the plot.

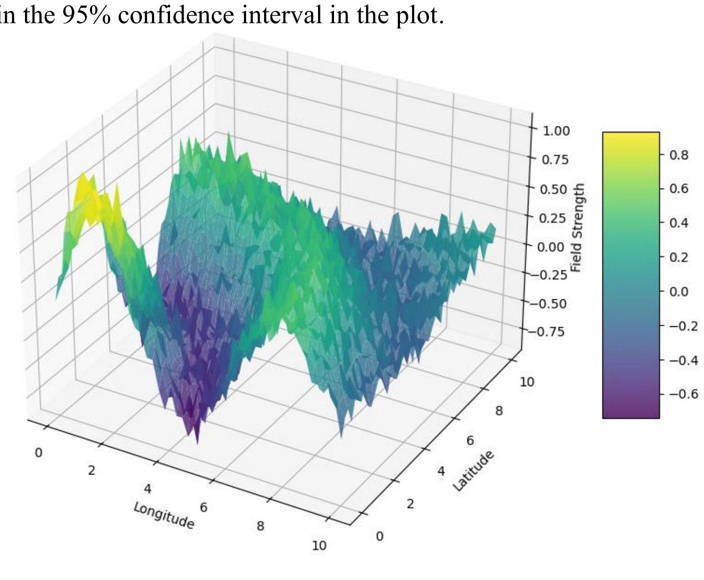

Figure 5 3D Surface Plot of Predicted Field Strength

Figure 5 shows predicted field strength over geographic coordinates, highlighting peak radiation around major hubs with non-uniform decay gradients. This underscores CoordConv's enhancement of location sensitivity, enabling precise simulation of real-world spatial heterogeneity.

Overall, these visualizations validate the model's robustness in capturing pre/post-pandemic patterns and economic-driven disparities, offering practical insights for transportation planning.

5. Conclusions

This study proposes and validates a multimodal regression model integrating ResNet and CoordConv for traffic field strength distribution using Shandong Province data (2015–2021). Through PCA preprocessing, CoordConv-enhanced spatial awareness, ResNet feature extraction, and attention-based fusion, the model achieves superior performance (MAE ≈0.15, R^2=0.92), effectively capturing nonlinear interactions, spatial gradients, and pandemic impacts ignored by traditional approaches. Results highlight spatiotemporal evolution in core cities and strong economic-transport coupling, providing scientific support for integrated transportation planning.

However, limitations exist: the dataset is restricted to one province with only three snapshot years, potentially limiting temporal continuity; additional transport modes (e.g., highways) are not incorporated; and model generalization to nationwide or real-time scenarios requires further validation. Future work will expand the dataset nationwide, incorporate additional modes, integrate real-time data, and explore advanced architectures (e.g., Transformers, Graph Neural Networks) for improved generalization and deployment.

REFERENCES

[1] Clewlow, R. R., Sussman, J. M., Balakrishnan, H.: The impact of high-speed rail and low-cost carriers on European air passenger traffic. Transport Policy 33, 136-143(2014).

[2] Pandey, A., & Kumar, P. (2024). BGRU-MTRA: bilinear GRU networks with multi-path temporal residual attention for suspicious activity recognition. Neural Computing and Applications, 37, 185 - 212.

[3] Prakash A., Chitta K., Geiger A. Multi-modal fusion transformer for end-to-end autonomous driving; Proceedings of the IEEE/CVF conference on computer vision and pattern recognition; Nashville, TN, USA. 20–25 June 2021; pp. 7077–7087.

[4] Yim, J., Kim, W.H., Cho, S., Kim, C.W., & Park, J.Y. (2024). Investigating maritime traffic routes: integrating AIS data and topographic statistics. Maritime Policy & Management, 52, 590 - 608.

[5] Chen, Haotian, Yuhang Yang, and Chenhui Shao. "Multi-task learning for data-efficient spatiotemporal modeling of tool surface progression in ultrasonic metal welding." Journal of Manufacturing Systems 58 (2021): 306-315.

[6] Shaker A., Maaz M., Rasheed H., Khan S., Yang M.-H., Khan F.S. Swiftformer: Efficient additive attention for transformer-based real-time mobile vision applications; Proceedings of the IEEE/CVF International Conference on Computer Vision; Paris, France. 1–6 October 2023; pp. 17425–17436.

[7] Sheikh R., Milioto A., Lottes P., Stachniss C., Bennewitz M., Schultz T. Gradient and Log-based Active Learning for Semantic Segmentation of Crop and Weed for Agricultural Robots; Proceedings of the 2020 IEEE International Conference on Robotics and Automation (ICRA); Paris, France. 31 May–31 August 2020; pp. 1350–1356.

[8] Qiu, Chen. "Bridging Semantic Disparity and Tail Query Challenges in Advertisement Retrieval via Dual LLM Collaboration." (2025).

[9] Diakogiannis F.I., Waldner F., Caccetta P., Wu C. ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data. ISPRS J. Photogramm. Remote Sens. 2020;162:94–114. doi: 10.1016/j.isprsjprs.2020.01.013.

[10] Yim, J., Kim, W.H., Cho, S., Kim, C.W., & Park, J.Y. (2024). Investigating maritime traffic routes: integrating AIS data and topographic statistics. Maritime Policy & Management, 52, 590 - 608.

[11] Wu, Z., Shen, C., van den Hengel, A.: Wider or deeper: Revisiting the ResNet model for visual recognition. Pattern Recognition 90, 119-133 (2019).

[12] Chen, Haotian, Jingyi Zuo, Yuanbin Zhu, MD Rizwanul Kabir, and Aiguo Han. "GPU-Based Acceleration for Real-Time Speed-of-Sound Imaging." In 2025 IEEE International Ultrasonics Symposium (IUS), pp. 1-4. IEEE, 2025.

[13] Samie A., Abbas A., Azeem M.M., Hamid S., Iqbal M.A., Hasan S.S., Deng X. Examining the impacts of future land use/land cover changes on climate in Punjab province, Pakistan: Implications for environmental sustainability and economic growth. Environ. Sci. Pollut. Res. 2020;27:25415–25433. doi: 10.1007/s11356-020-08984-x.

[14] Baltrušaitis, T., Ahuja, C., Morency, L. P.: Multimodal machine learning: A survey and taxonomy. IEEE Transactions on Pattern Analysis and Machine Intelligence 41(2), 423-443 (2019).

[15] Li, Yanting, Yiping Gao, Junwei Jin, Jiaofen Nan, Yinghui Meng, Mengjie Wang, and CL Philip Chen. "Adaptive weights-based relaxed broad learning system for imbalanced classification." Digital Signal Processing 156 (2025): 104869.

[16] Aleissaee A.A., Kumar A., Anwer R.M., Khan S., Cholakkal H., Xia G.-S., Khan F.S. Transformers in Remote Sensing: A Survey. Remote Sens. 2023;15:1860. doi: 10.3390/rs15071860.

[17] Kaushal A., Gupta A.K., Sehgal V.K. A semantic segmentation framework with UNet-pyramid for landslide prediction using remote sensing data. Sci. Rep. 2024;14:30071. doi: 10.1038/s41598-024-79266-6.