


Research Article

A High-Dimensional Video Sequence Completion Method with Traffic Data Completion Generative Adversarial Networks

Lan Wu ¹, Tian Gao,¹ Chenglin Wen,² Kunpeng Zhang,¹ and Fanshi Kong³

¹School of Electrical Engineering, Henan University of Technology, Zhengzhou, China

²College of Automation Engineering, Hangzhou Dianzi University, Hangzhou, China

³Zhengzhou Railway Vocational & Technical College, Zhengzhou, China

Correspondence should be addressed to Lan Wu; wulan@haut.edu.cn

Received 27 August 2020; Revised 25 February 2021; Accepted 2 March 2021; Published 15 March 2021

Academic Editor: Carles Gomez

Copyright © 2021 Lan Wu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The lack of traffic data is a bottleneck restricting the development of Intelligent Transportation Systems (ITS). Most existing traffic data completion methods aim at low-dimensional data, which cannot cope with high-dimensional video data. Therefore, this paper proposes a traffic data complete generation adversarial network (TDC-GAN) model to solve the problem of missing frames in traffic video. Based on the Feature Pyramid Network (FPN), we designed a multiscale semantic information extraction model, which employs a convolution mechanism to mine informative features from high-dimensional data. Moreover, by constructing a discriminator model with global and local branch networks, the temporal and spatial information are captured to ensure the time-space consistency of consecutive frames. Finally, the TDC-GAN model performs single-frame and multiframe completion experiments on the Caltech pedestrian dataset and KITTI dataset. The results show that the proposed model can complete the corresponding missing frames in the video sequences and achieve a good performance in quantitative comparative analysis.

1. Introduction

In recent years, with the rapid development in the field of Intelligent Transportation Systems (ITS), numerous data with rich traffic information attract the widespread attention of researchers [1–4]. Accurate and efficient real-time traffic data can not only provide travelers with a better travel plan but also assist the traffic management department to effectively manage and guide traffic operations. However, in reality, incomplete data will be collected due to the limitations of the sensor placement [5, 6], the accidental deviation of the intelligent system [7–10], and the camera occlusion [11]. These problems will affect the accuracy of traffic state analysis and the timeliness of handling traffic problems [12]. Thus, it is necessary to complete the missing data.

Most of the existing studies are carried out to complete low-dimensional data (e.g., traffic flow [13], travel time [14, 15], and trajectory [16]), which cannot cope with high-dimensional traffic video data containing more intuitive information. This can be explained for two reasons. On one hand, due to limited hardware facilities, the computing

power and processing speed of computers are restricted to capture meaningful information from high-dimensional traffic video. On the other hand, based on traditional statistical tools and proper prior knowledge, the existing data completion models are proposed to handle low-dimensional data. However, due to its high dimension and sparse representation, traffic video data is arduous to be modeled by statistical models and prior knowledge. Moreover, traffic video scenes are relatively complex, which usually include a large number of vehicles and pedestrians. This results in the difficulties of explicitly extracting semantic information in traffic scenes with low-dimensional traffic completion models.

To deal with these drawbacks, based on the generative adversarial network (GAN) [17], we proposed a traffic data completion generative adversarial network (TDC-GAN) to complete high-dimensional video sequences with the enhancement of graphics processing unit (GPU) parallel computing power. In the TDC-GAN, the Feature Pyramid Network (FPN) [18] is used to extract the multiscale features from the video frame in the generator. By learning latent representation from high-dimensional data, this paper expands

the field of traffic data completion research to high-dimensional video sequences. In addition, this paper designs global and local discriminators to capture the temporal and spatial correlation of video sequences. The two discriminators learn the time information between consecutive frames and the spatial semantic information within the frames to generate reliable frames.

The remainder of the paper is organized as follows. Section 2 introduces some related work. Section 3 describes the TDC-GAN model framework. Section 4 is the content of experiments. Finally, Section 5 summarizes the TDC-GAN model and makes a prospect for future work.

2. Related Work

In general, traditional traffic data completion methods can be classified into the following three categories: prediction, interpolation, and statistical learning [19].

The prediction method is to learn the mapping of past data to future data by establishing corresponding models. For example, both the high-order smoothing exponential model [20, 21] and the gradient boost regression tree (GBRT) model [22] complete the traffic data by modeling traffic flow. Based on the previous traffic information, Xu et al. [23] proposed a prediction model, which combines the autoregressive integrated moving average (ARIMA) model with Kalman filter. However, because the continuous data in the past period of time needs to be known, the application scenarios of the prediction model are relatively limited. In addition, compared to completion, it cannot use the subsequent adjacent data, which is not conducive to consistent representation in time series.

The interpolation method generally estimates the missing data by averaging the traffic data in adjacent time periods or using the historical data of other days that are similar to the missing data. Typical interpolation methods are k -nearest neighbor (k -NN) and local least squares (LLS) [24, 25]. Literature [26, 27] is based on an improved adaptive k -NN method, which comprehensively considers spatial neighboring points, sliding windows, spatiotemporal weights, and other spatial heterogeneity features to complete missing traffic data. The improved LLS method attempts to replace the missing traffic data with the average of the known data and iteratively obtains the weight of the nearest neighbor by using the Euclidean distance. However, the interpolation method assumes that the adjacent traffic states have strong similarities. This method is unreliable when the state is relatively random.

The statistical learning method uses the statistical characteristics to complete the missing traffic information by establishing an iterative model of the probability distribution of the data. Typical methods are Markov Chain Monte Carlo (MCMC) [28] and probabilistic principal component analysis (PPCA) [29]. However, due to the complexity of the urban road traffic system, the learning ability of the statistical learning method is limited, and its convergence is difficult to guarantee.

Recently, with the advances of modern GPU and neural networks [30, 31], deep learning-based methods have appeared to complete traffic data [32–34]. As an important technique of deep learning, GAN has been increasingly

applied in video completion due to its outstanding learning ability.

Mathieu et al. [35] showed that traditional loss functions based only on pixel loss often lead to image blurring; however, an adversarial loss can effectively solve this problem. This is the first time GAN has been applied to the modeling of video sequences. Subsequent research on the video frame with the GAN attempted to decompose it into two modules containing different information, which were studied separately. Vondrick et al. [36] separated the background and foreground of the video scene, and the GAN was used to force static background and moving foreground to predict. The Motion and Content Generative Adversarial Network (MoCoGAN) model [37] divided the potential space of video frames into content and motion. The model can generate a video that contains the same object performing different operations or different objects performing the same operation. Liang et al. [38] divided the video sequence into future frames and future streams and used two GAN models to feedback each other for training. These methods separate video frames according to different factors, which requires expensive computing power. In addition, due to complex operating procedures, they are limited to a single and simple data set. Therefore, it is difficult to effectively model complex traffic scenarios.

Different from the abovementioned methods, FutureGAN [39] and Retrospective Cycle-consistency Generative Adversarial Network (CycleGAN) [40] attempt to use the original video frames as input. The idea of not decomposing the video frame is consistent with our method, which allows the network to learn more overall information about the input frame. Inspired by this, the TDC-GAN model directly receives unlabeled raw traffic video frames. In the generator model, the FPN is used to learn the information of multiple scales of the video frame by synthesizing the feature maps of multiple levels. By combining the lower-level feature map with more target location information and the upper-level feature map with more feature semantic information, the frames generated by the generator will be more realistic and accurate. In addition, in the discriminator model, the global discriminator mainly grasps the overall information of consecutive frames in the time series, and the local discriminator can supplement the detailed information in the space, which provides a guarantee for the temporal and spatial consistency of the generated frames. Therefore, the TDC-GAN model is capable of solving the problems of missing high-dimensional traffic video frames.

3. Methodology

3.1. Generative Adversarial Network. In recent years, deep learning methods have become an important tool for video sequences modeling, especially the proposal of GAN, which is good at capturing complex features in high-dimensional data due to its outstanding learning capability. In this study, based on GAN, the TDC-GAN model is proposed to complete the missing traffic video data.

Since Goodfellow proposed the GAN, the idea of adversarial has gradually been applied to the framework of

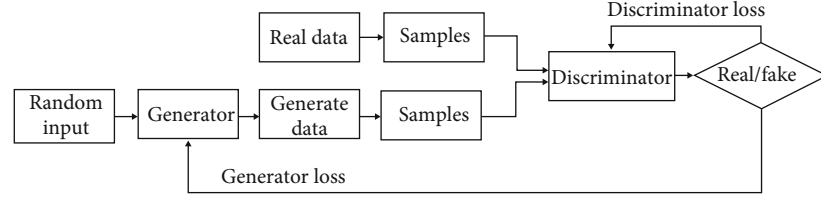


FIGURE 1: The original structure of the GAN.

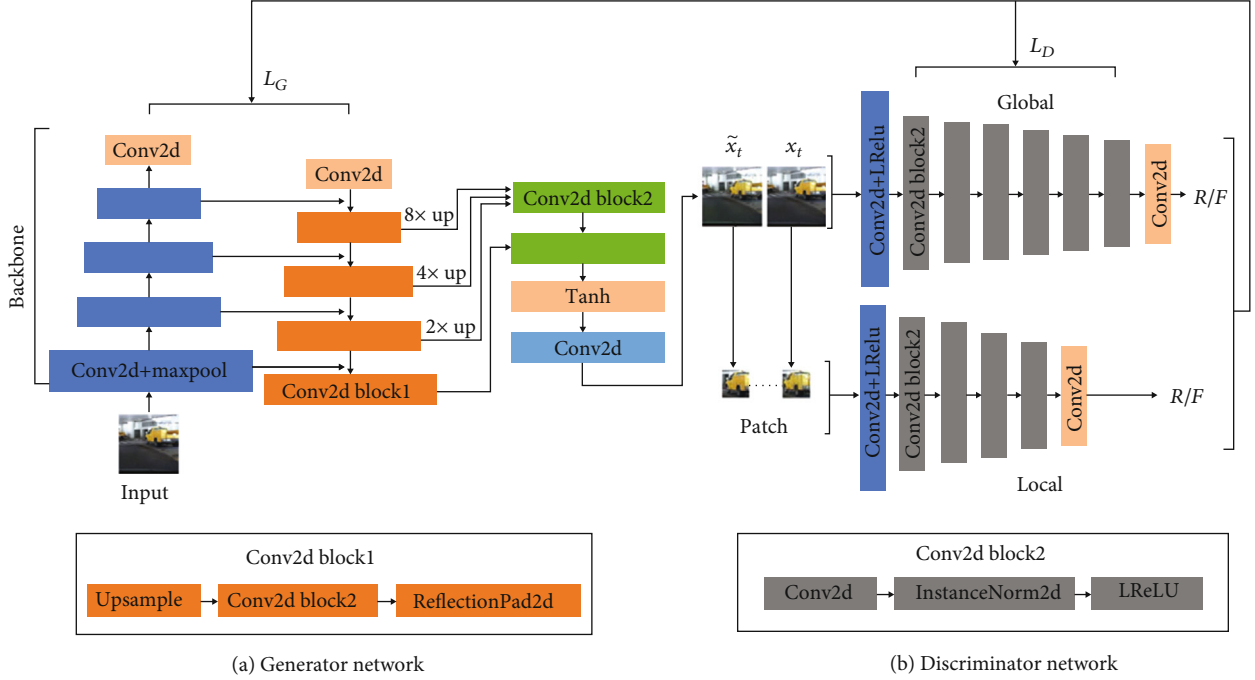


FIGURE 2: The structure of the TDC-GAN model. (a, b) are the structure diagrams of generator and discriminator, respectively. Each module in the figure is represented by a different color, and modules with the same network structure are represented by the same color. The bottom of the figure indicates the specific network layers of the two convolution modules.

generative models. As shown in Figure 1, The original GAN includes a generator and a discriminator. The generator is used to capture the distribution of sample data. By converting the distribution of the original input information into the parameters in the maximum likelihood estimation, the training deviation is finally converted into a sample of the specified distribution. During training, the generator learns to generate data samples that can confuse the discriminator, and the discriminator is used to judge the difference between the real sample and the generated sample. In constant adversarial learning, they will eventually reach a balance.

The equation for the GAN can be defined as

$$\min_G \max_D V(D, G) = \frac{1}{m} \sum_{i=1}^m \log D(x^i) + \frac{1}{m} \sum_{i=1}^m \log(1 - D(G(z^i))), \quad (1)$$

where m represents the batch size, $x^i \in P_{\text{data}}$ represents the i -th sample of m real samples, and $z^i \in P_{\text{noise}}$ represents the i -th sample of m noise samples.

Equation (1) indicates that the discriminator needs to learn to assign a higher score to the real sample data and to assign a lower score to the sample data generated by the generator. The generator needs to generate samples that confuse the discriminator as much as possible.

Video is composed of continuous images with a certain frame rate; so, the high-dimensional traffic missing data studied in this paper are video frames in continuous time. In the training process, the TDC-GAN model can learn the mapping from existing frames $x_1^T = \{x_1, x_2, \dots, x_T\}$ to complete frames $\tilde{x}_1^T = \{\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_T\}$. x_1^T represents some known frames from 1 to T , that is, the data input to the generator. \tilde{x}_1^T represents the corresponding missing frame from time 1 to T , that is, the output of the generator.

3.2. Network Architecture. As shown in Figure 2, the network structure of the TDC-GAN model includes a generator and two discriminators. The generator employs the FPN network, which includes three paths of bottom-up, top-down, and horizontal connection. The bottom-up path retains more position information through less downsampling, and the top-down upsampling path is used to obtain feature maps with more

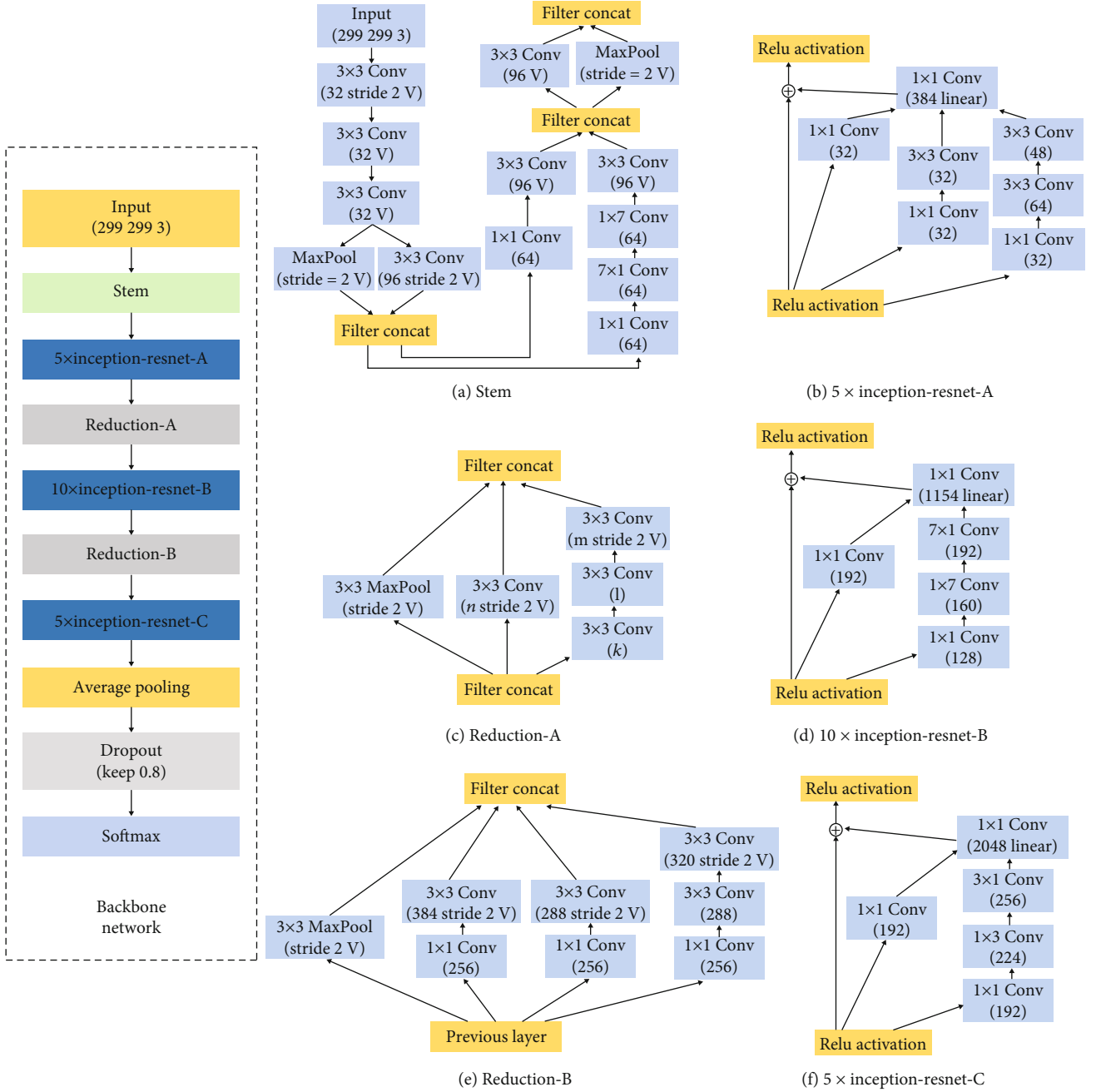


FIGURE 3: The structure diagram of InceptionResNet-v2. The left side of the picture is the backbone network, and (a)–(f) correspond to 6 important modules.

semantic information and higher resolution. In the horizontal connection, 1×1 convolution is used to fuse the two parts of position information and semantic information; so, more high-dimensional feature information can be learned through the TDC-GAN model. Upsampling and convolutional layers are added to the end of our generator network to keep the resolution of the input video frame consistent.

In the discriminator network, to obtain the completion performance with temporal and spatial consistency, the TDC-GAN model designs two discriminator models, global and local. The global discriminator integrates the complete spatial environment by alternately receiving the generated frame and the

real frame to obtain the rough motion state of the video frame in continuous time. However, the global discriminator weights the entire frame image, ignoring local spatial details. Therefore, the TDC-GAN model introduces a local discriminator, which randomly crops a certain number of patches on the whole frame and sends them to the discriminator. By performing feedback learning on each spatial local unit to obtain more details, high resolution and high details can be maintained. In addition, a Leaky Rectified Linear Unit (LReLU) is used to increase nonlinearity, and the batch normalization layer follows each LReLU.

As shown in Figure 3, to speed up training and improve network performance, the backbone module of the TDC-

GAN introduces the pretrained InceptionResNet-v2 [41] network, which can make full use of the characteristics of the training data image and reduce the feature loss in the convolution process. By introducing the residual module, the convergence can be accelerated, and the training error will not increase with the increase of the network depth.

The TDC-GAN model can complete the missing frames of the video conditioned on the incomplete frame sequences. During training, the generator network only receives pixel values of the original video frames as input and does not depend on other constraints. To complete the real and effective missing frames, the spatial and temporal components of the video sequence will be captured simultaneously. The discriminator is trained to distinguish between true and false video frames by receiving the real sequence and the generated sequence as input alternately.

3.3. Loss Function. For the problem of traffic videos completion, Wasserstein generative adversarial network-gradient penalty (WGAN-GP) [42] with a loss function of gradient penalty term is used to optimize the discriminator which is defined as

$$L_D = E_{\tilde{x} \sim P_g} [D(\tilde{x})] - E_{x \sim P_r} [D(x)] + \lambda E_{\tilde{x} \sim P_{\tilde{x}}} \left[\left(\|\nabla_{x \wedge} D(x \wedge)\|_2 - 1 \right)^2 \right], \quad (2)$$

where P_r and P_g are the real sample distribution and generator sample distribution, respectively. The gradient-penalty coefficient is represented by λ . $P_{\tilde{x}}$ is the random sampling between the two sampling points connecting P_r and P_g .

To train the generator of the TDC-GAN model to generate more realistic completion samples, our overall loss function is defined as

$$L_G = \lambda_1 \times L_p + \lambda_2 \times L_{adv} + \lambda_3 \times L_X, \quad (3)$$

where $\lambda_1, \lambda_2, \lambda_3$ is the weight coefficient. We use the mean square error (MSE) loss (L2 loss) between the real image x_t and the generated image \tilde{x}_t as the value of the first term L_p , and it is defined as

$$L_p = \sum_{t=1}^n (x_t - \tilde{x}_t)^2. \quad (4)$$

To solve the problem of image blur caused by using the L_p loss function, the adversarial loss L_{adv} is introduced in the second term, which is defined as

$$L_{adv} = \sum_{t=1}^n -D(\tilde{x}_t). \quad (5)$$

The third term L_X [43] is the loss function proposed for the general content of the restored images, and we use it for image generation. In Equation (6), $\varphi_{i,j}$ stands for the feature map, where i and j denote the convolution and the max pool-

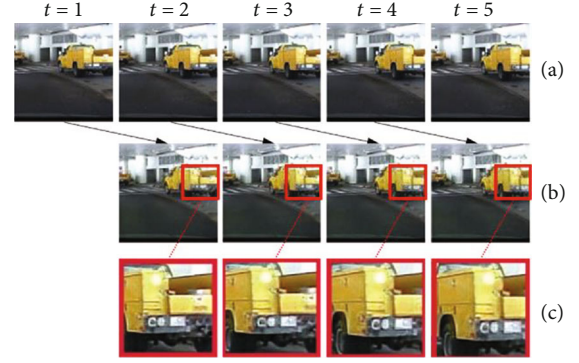


FIGURE 4: The completion results of single frame on Caltech pedestrian dataset. (a) Input/ground truth. (b) Output. (c) Details.

TABLE 1: PSNR and SSIM results.

	Caltech pedestrian		KITTI	
	PSNR	SSIM	PSNR	SSIM
Single frame	27.9	0.89	25.6	0.81
Multiple frames	26.8	0.85	24.7	0.77

ing layer, respectively. $W_{i,j}$ and $H_{i,j}$ represent the dimensions of the feature map.

$$L_X = \frac{1}{W_{i,j} H_{i,j}} \sum_{t=1}^n \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} \left(\varphi_{i,j}(x_{t+1})_{x,y} - \varphi_{i,j}(G(x_t))_{x,y} \right)^2. \quad (6)$$

4. Results and Discussion

4.1. Datasets. In the experimental part, we noticed a large-scale urban traffic dataset Caltech pedestrian dataset [44], which consists of about 10 hours of 640×480 pixel video. The video was captured by the onboard camera of a vehicle traveling through normal traffic in an urban environment. Because it contains comprehensive traffic information, many video-related pedestrian detection, target recognition, and other tasks use this dataset [45, 46]. In addition, the open source and easy-to-download attributes guarantee a fair comparison and analysis of research performance in subsequent research. Experimenting on this public dataset makes the TDC-GAN model more convincing. Since our research aims to complete the missing traffic data and no other information is needed in the training data, the annotation information of pedestrians in this dataset is ignored. Moreover, to verify the versatility of the proposed TDC-GAN model, we verified it on the KITTI dataset [47], which contains real video data collected in scenes such as urban areas, rural areas, and highways, and each frame can contain up to 15 cars and 30 pedestrians. To adapt to the network structure of TDC-GAN, we changed the video pixel to 480×480 .

4.2. Training Details. The TDC-GAN model is implemented in PyTorch, and the computer is configured as a single NVIDIA GTX 2080ti GPU under Linux. The ADAM optimizer is

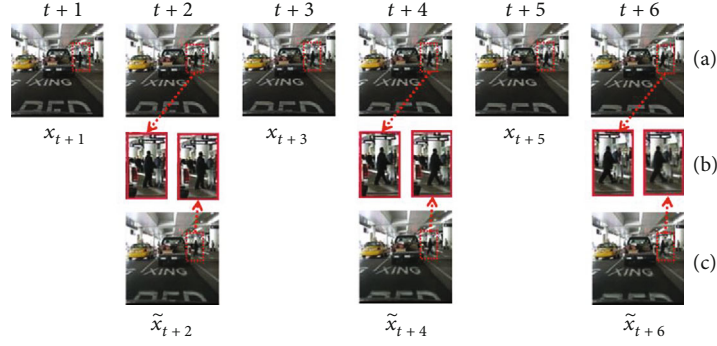


FIGURE 5: The completion results of multiple frames on Caltech pedestrian dataset. (a) Input. (b) Details. (c) Output.

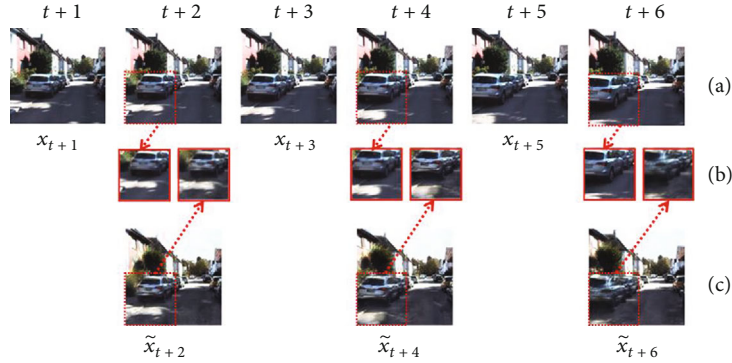


FIGURE 6: The completion results of multiple frames on the KITTI dataset. (a) Input. (b) Details. (c) Output.

used to optimize our algorithm, and the relevant parameters were set to $\beta_1 = 0.1$, $\beta_2 = 0.999$, and $l = 0.001$. The weights of the loss function are set to $\lambda = 10$, $\lambda_1 = 0.1$, $\lambda_2 = 0.5$, and $\lambda_3 = 0.02$. To quantitatively evaluate the network, we provided the values of peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) between the ground truth and the completed video frames. And they are defined as

$$PSNR = 10 \times \log_{10} \left[\frac{(2^n - 1)^2}{MSE} \right], \quad (7)$$

where n is the bit of each sampled value, and MSE is the corresponding mean square error.

$$SSIM(x, \tilde{x}) = \frac{(2\mu_x \mu_{\tilde{x}} + c_1)(2\sigma_{x\tilde{x}} + c_2)}{(\mu_x^2 + \mu_{\tilde{x}}^2 + c_1)(\sigma_x^2 + \sigma_{\tilde{x}}^2 + c_2)}, \quad (8)$$

where μ and σ^2 are the average value and variance of the real frame x or the generated frame \tilde{x} , respectively. And the covariance of x and \tilde{x} is represented by $\sigma_{x\tilde{x}}$. The value of SSIM is between 0 and 1, and the generated frame close to 1 is what we expect.

4.3. Experimental Results and Analysis

4.3.1. Single Frame Completion. We first complete the next frame based on the previous frame. Figure 4 shows the single frame completion results of the TDC-GAN model.

The generator receives the video frame $z = (x_t)$ at time t and generates the video frame $G(z) = (\tilde{x}_{t+1})$ at time $t + 1$.

By comparing the details of 5 consecutive video frames between generated frames and the ground truth, we can see that the TDC-GAN model can effectively complete the missing video frames. And the quantitative results are shown in Table 1. PSNR and SSIM can reach 27.9 and 0.89, respectively.

4.3.2. Multiple Frame Completion. We also tried to train the generator to input multiple missing frames to test the completion effect of the TDC-GAN model.

Figures 5 and 6 are the results of multiple frame completion on the two datasets. The input sequence of the generator can be expressed as $z = (x_{t+1}, x_{t+3}, x_{t+5})$, which represents the input video frames at times 1, 3, and 5, and we want to complete the sequence of video frames $G(z) = (\tilde{x}_{t+2}, \tilde{x}_{t+4}, \tilde{x}_{t+6})$ at times 2, 4, and 6.

As can be seen from the details circled in red and green boxes in the figures, the TDC-GAN model can not only complete the missing video frames but also ensure that the video frames have temporal and spatial consistency. Moreover, the multiple frame quantitative results are given in Table 1, and the best PSNR and SSIM values of the TDC-GAN model can reach 26.8 and 0.85, respectively. It is worth noting that the TDC-GAN model has good performance on both data sets. The versatility of this model is of great significance to our further research.

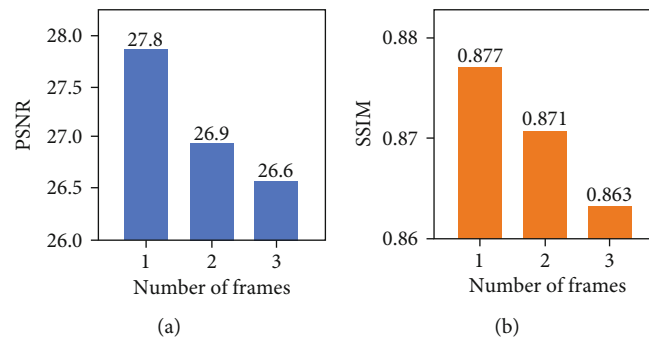


FIGURE 7: Multiple frame quantitative evaluation results on Caltech pedestrian dataset. (a, b) are the results of PSNR and SSIM, respectively. The abscissa represents the completed frame, and the ordinate is the corresponding quantitative result.

In addition, Figure 7 shows the quantitative performance of each completed frame in the multiple frame experiment on the Caltech pedestrian dataset. As the number of completion frames increases, the values of PSNR and SSIM gradually decrease. It is explained as that the past information becomes less valuable to facilitate the completion in the longer future, which causes the reduction of the performance. However, in terms of the overall effect of the completion, it is still satisfactory.

5. Conclusion

This paper proposes a TDC-GAN model for completing traffic video sequences. In the TDC-GAN model, the designed generator network learns multiscale features from video sequences with the help of the FPN. Meanwhile, the discriminator network includes two branches (i.e., the global branch and the local branch), which takes into account the time information between frames and the space information within each frame. The adversarial loss is utilized to improve the stability of training, and the perceptual loss calculates the semantic difference between the generated frame and the real frame, which enhances the performance of the proposed model. With the Caltech pedestrian dataset and KITTI dataset, the experimental results show that this TDC-GAN model can effectively complete missing frames in traffic videos. In summary, the TDC-GAN model is well suited to complete travel videos under various scenarios.

In the future, we will add technologies such as scene understanding to optimize our model to solve more complex problems (such as solving traffic video problems with more missing frames). Moreover, encouraged by the promising performance of the TDC-GAN, it is interesting to propose more GAN-based methods in the traffic field.

Data Availability

The address of our experimental datasets can be found in the link: http://www.vision.caltech.edu/Image_Datasets/CaltechPedestrians, http://www.cvlibs.net/datasets/kitti/raw_data.php.

Conflicts of Interest

We declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

This study was supported by the National Natural Science Foundation of China, No. 61973103, Henan Province Central Plains Thousand Talents Plan: Top Young Talents, Key Scientific Research Project of Henan University with No. 19A120002.

References

- [1] Y. Duan, Y. Lv, Y.-L. Liu, and F.-Y. Wang, "An efficient realization of deep learning for traffic data imputation," *Transportation Research Part C: Emerging Technologies*, vol. 72, pp. 168–181, 2016.
- [2] K. Zhang, Z. Liu, and L. Zheng, "Short-Term prediction of passenger demand in multi-zone level: temporal convolutional neural network with multi-task learning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 4, pp. 1480–1490, 2020.
- [3] X. Chen, Z. He, and L. Sun, "A Bayesian tensor decomposition approach for spatiotemporal traffic data imputation," *Transportation Research Part C: Emerging Technologies*, vol. 98, pp. 73–84, 2019.
- [4] K. Zhang, N. Jia, L. Zheng, and Z. Liu, "A novel generative adversarial network for estimation of trip travel time distribution with trajectory data," *Transportation Research Part C: Emerging Technologies*, vol. 108, pp. 223–244, 2019.
- [5] V. Akbarzadeh, C. Gagne, M. Parizeau, M. Argany, and M. A. Mostafavi, "Probabilistic sensing model for sensor placement optimization based on line-of-sight coverage," *IEEE Transactions on Instrumentation and Measurement*, vol. 62, no. 2, pp. 293–303, 2013.
- [6] A. Bagula, L. Castelli, and M. Zennaro, "On the design of smart parking networks in the smart cities: an optimal sensor placement model," *Sensors*, vol. 15, no. 7, pp. 15443–15467, 2015.
- [7] A. Alhussain, H. Kurdi, and L. Altoaimy, "A neural network-based trust management system for edge devices in peer-to-peer networks," *Computers, Materials & Continua*, vol. 59, no. 3, pp. 805–816, 2019.

- [8] J. Zhang, S. Zhong, T. Wang, H. C. Chao, and J. Wang, "Blockchain-based systems and applications: a survey," *Journal of Internet Technology*, vol. 21, no. 1, pp. 1–14, 2020.
- [9] A. Maamar and K. Benahmed, "A hybrid model for anomalies detection in AMI system combining K-means clustering and deep neural network," *Computers, Materials & Continua*, vol. 60, no. 1, pp. 15–39, 2019.
- [10] H.-J. Song, B.-D. Oh, J.-D. Kim, C.-Y. Park, and Y.-S. Kima, "Predicting concentration of PM10 using optimal parameters of deep neural network," *Intelligent Automation and Soft Computing*, vol. 25, no. 2, pp. 343–350, 2019.
- [11] J. Jung, I. Yoon, and J. Paik, "Object occlusion detection using automatic camera calibration for a wide-area video surveillance system," *Sensors*, vol. 16, no. 7, p. 982, 2016.
- [12] Z. Zhang, Q. He, J. Gao, and M. Ni, "A deep learning approach for detecting traffic accidents from social media data," *Transportation Research Part C: Emerging Technologies*, vol. 86, pp. 580–596, 2018.
- [13] Y. Tian, K. Zhang, J. Li, X. Lin, and B. Yang, "LSTM-based traffic flow prediction with missing data," *Neurocomputing*, vol. 318, pp. 297–305, 2018.
- [14] K. Zhang, Z. He, L. Zheng, L. Zhao, and L. Wu, "A generative adversarial network for travel times imputation using trajectory data," *Computer-Aided Civil and Infrastructure Engineering*, vol. 36, no. 2, pp. 197–212, 2021.
- [15] J. Cheng, G. Li, and X. Chen, *Research on travel time prediction model of freeway based on gradient boosting decision tree*, IEEE access, 2018.
- [16] G. Xie, H. Gao, L. Qian, B. Huang, K. Li, and J. Wang, *Vehicle trajectory prediction by integrating physics-and maneuver-based approaches using interactive multiple models*, IEEE Transactions on Industrial Electronics, 2017.
- [17] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza et al., "Generative adversarial networks," 2014, <https://arxiv.org/abs/1406.2661>.
- [18] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2117–2125, Honolulu, HI, USA, 2017.
- [19] Y. Li, Z. Li, and L. Li, "Missing traffic data: comparison of imputation methods," *IET Intelligent Transport Systems*, vol. 8, no. 1, pp. 51–57, 2014.
- [20] M. Qin, G. F. Yang, M. J. Deng, W. Q. Zhang, and B. Feng, "Short-term traffic flow forecasting based on exponential smoothing and Kalman filter, Journal of Beihua University," *Natural Science*, vol. 16, no. 6, pp. 814–817, 2015.
- [21] Y. Chen, G. Y. Liu, H. Z. Tang, and Y. Li, "The prediction method for traffic flow data based on the balance of exponential smoothing," *Modern Computer*, vol. 20, pp. 45–48, 2014.
- [22] X. Zhou, "Research on the traffic flow forecasting algorithm based on GBRT," *Modern Computer*, vol. 7, pp. 38–41, 2019.
- [23] D.-W. Xu, Y.-D. Wang, L.-M. Jia, Y. Qin, and H.-H. Dong, "Real-time road traffic state prediction based on ARIMA and Kalman filter," *Frontiers of Information Technology & Electronic Engineering*, vol. 18, no. 2, pp. 287–302, 2017.
- [24] L. Zhang, Q. Liu, W. Yang, N. Wei, and D. Dong, "An improved K -nearest neighbor model for short-term traffic flow prediction," *Procedia-Social and Behavioral Sciences*, vol. 96, pp. 653–662, 2013.
- [25] S. Cheng, F. Lu, P. Peng, and S. Wu, "Short-term traffic forecasting: an adaptive ST-KNN model that considers spatial heterogeneity," *Computers, Environment and Urban Systems*, vol. 71, pp. 186–198, 2018.
- [26] X. Zhang, X. Song, H. Wang, and H. Zhang, "Sequential local least squares imputation estimating missing value of microarray data," *Computers in Biology and Medicine*, vol. 38, no. 10, pp. 1112–1120, 2008.
- [27] G. Chang, Y. Zhang, and D. Yao, "Missing data imputation for traffic flow based on improved local least squares," *Tsinghua Science and Technology*, vol. 17, no. 3, pp. 304–309, 2012.
- [28] R. E. Kass, B. P. Carlin, A. Gelman, and R. M. Neal, "Markov chain Monte Carlo in practice: a roundtable discussion," *The American Statistician*, vol. 52, no. 2, pp. 93–100, 1998.
- [29] C. Chen, J. Kwon, J. Rice, A. Skabardonis, and P. Varaiya, "Detecting errors and imputing missing data for single loop surveillance systems," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1855, no. 1, pp. 160–167, 2002.
- [30] J. Wang, Y. Yang, J. Zhang, X. Yu, O. Alfarraj, and A. Tolba, "A data-aware remote procedure call method for big data systems," *Computer Systems Science and Engineering*, vol. 35, no. 6, pp. 523–532, 2020.
- [31] J. Wang, Y. Yang, T. Wang, R. S. Sherratt, and J. Zhang, "Big data service architecture: a survey," *Journal of Internet Technology*, vol. 21, no. 2, pp. 393–405, 2020.
- [32] S. Choi, Y. Hwasoo, and J. Kim, "Network-wide vehicle trajectory prediction in urban traffic networks using deep learning," *Transportation Research Record*, vol. 2672, no. 45, pp. 173–184, 2018.
- [33] N. C. Petersen, F. Rodrigues, and F. C. Pereira, "Multi-output bus travel time prediction with convolutional LSTM neural network," *Expert Systems with Applications*, vol. 120, pp. 426–435, 2019.
- [34] S. Das, R. N. Kalava, K. K. Kumar et al., "Map enhanced route travel time prediction using deep neural networks," 2019, <https://arxiv.org/abs/1911.02623>.
- [35] M. Mathieu, C. Couprie, and Y. Lecun, "Deep multi-scale video prediction beyond mean square error," 2015, <https://arxiv.org/abs/1511.05440>.
- [36] C. Vondrick, H. Pirsiavash, and A. Torralba, "Generating videos with scene dynamics," 2016, <https://arxiv.org/abs/1609.02612>.
- [37] S. Tulyakov, M. Y. Liu, and X. Yang, "MoCoGAN: decomposing motion and content for video generation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1526–1535, Salt Lake City, UT, USA, 2018.
- [38] X. Liang, L. Lee, and W. Dai, "Dual motion gan for future-flow embedded video prediction," in *Proceedings of the IEEE international conference on computer vision*, pp. 1744–1752, Venice, Italy, 2017.
- [39] S. Aigner and M. Körner, "Futuregan: anticipating the future frames of video sequences using spatio-temporal 3d convolutions in progressively growing gans," 2018, <https://arxiv.org/abs/1810.01325>.
- [40] Y. H. Kwon and M. G. Park, "Predicting future frames using retrospective cycle gan," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1811–1820, Long Beach, CA, USA, 2019.
- [41] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, inception-ResNet and the impact of residual connections on learning," in *Thirty-first AAAI conference on artificial intelligence*, San Francisco, CA, USA, 2017.

- [42] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville, "Improved training of wasserstein gans," 2017, <https://arxiv.org/abs/1704.00028>.
- [43] O. Kupyn, V. Budzan, M. Mykhailych, D. Mishkin, and J. Matas, "Deblurgan: blind motion deblurring using conditional adversarial networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 8183–8192, Salt Lake City, UT, USA, 2018.
- [44] P. Dollar, C. Wojek, B. Schiele, and P. Perona, "Pedestrian detection: an evaluation of the state of the art," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 4, pp. 743–761, 2012.
- [45] X. Du, M. El-Khamy, J. Lee, and L. Davis, "Fused DNN: A deep neural network fusion approach to fast and robust pedestrian detection," in *2017 IEEE winter conference on applications of computer vision (WACV)*, pp. 953–961, Santa Rosa, CA, USA, 2017.
- [46] S. Zhang, R. Benenson, M. Omran, J. Hosang, and B. Schiele, "Towards reaching human performance in pedestrian detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 973–986, 2018.
- [47] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? The KITTI vision benchmark suite," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3354–3361, Providence, RI, USA, 2012.