Nanotechnology Perceptions
ISSN 1660-6795
www.nano-ntp.com

Super-Fuse: Fuzzy Input Enhancement Using Efficient Video Super-Resolution Deep Fusion Network for Public Safety

Renuka Sambhaji Sindge, Maitreyee Dutta

National Institute of Technical Teachers' Training and Research (NITTTR), Chandigarh, India
Email: renuka.cse20@nitttrchd.ac.in

To ensure public safety, video surveillance is essential in several areas, such as law enforcement, transportation, and critical infrastructure. However, using fuzzy video footage frequently makes it difficult to analyse and identify people or things accurately. Consequently, using the data from the surveillance footage efficiently is challenging. This problem is addressed by the proposed video super-resolution, with a deep fusion network, which uses Recurrent Residual Networks (RRNs) with texture details to their full potential to discover the underlying structures and patterns in fuzzy and low-resolution video frames. Firstly, it improves the quality of video frames by successfully inferring missing high-frequency information by network and creating aesthetically pleasing, super-resolved high-resolution video frames to identify people, cars, or other exciting items in surveillance footage. Secondly, the proposed network makes monitoring busy places or public locations easier to make crowds safer for people; with improved visual quality, the surveillance systems can better identify and monitor any questionable activity or behaviour. Additionally, the experimental results demonstrate that the proposed method has much potential for improving the quality of video frames and can effectively and efficiently reconstruct HR video frames for surveillance datasets.

Keywords: Video surveillance, video super-resolution, super-resolution, recurrent residual network, deep learning.

1. Introduction

Video surveillance systems have become increasingly critical in recent years, especially in improving public safety and security measures. These systems are widely used in many different places, including public spaces, highways, business buildings, residential neighbourhoods, banking institutions, and transportation hubs like airports and train stations (Socha & Kogut, 2020; Sreenu & Saleem Durai, 2019). They play vital roles in various fields, such as transportation, law enforcement, and infrastructure protection. Through video surveillance systems, criminal activity risk can be successfully reduced by quickly identifying potential threats and suspicious activity and closely monitoring it. These systems are helpful to law enforcement because they allow them to obtain important information by looking at surveillance footage and asking about the presence of questionable people and vehicles (Chin, 2022). Furthermore, during the investigative stage of criminal investigations, surveillance film frequently functions as objective evidence. As a result of developments in science, network research, and criminal technology, video surveillance has become a critical component of investigative technology, ranking as the fourth most important domain.

Developing a network that utilises video surveillance for safety aids in preserving national security, and social order in the current environment. However, only real-time monitoring and manual case analysis using surveillance footage can be done with a traditional video surveillance system. It results in a low rate of video surveillance data utilisation. Low-resolution (LR) video frequently makes it difficult to analyse and identify people or things clearly due to the limits of camera technology, which presents significant difficulty for surveillance systems. This problem is addressed by Video Super-Resolution (VSR) methods, which increase the clarity and detail of surveillance footage, resulting in better situational awareness and preventative safety measures. Deep learning (DL) based techniques are used to produce high-resolution (HR) frames from LR input (X. Liu, Chen, et al., 2022; X. Liu, Fu, et al., 2022).

The use of VSR in public safety has several benefits. It enhances the quality of video frames for the identification of people, cars, or other interesting items in surveillance footage. HR videos offer characteristics that are crisper and easier to distinguish, allowing for more precise analysis and identification. Investigating crimes, responding to incidents, and taking preventative action are all aided by this. Numerous advanced video surveillance systems have been made available by large Internet corporations. On the other hand, the area of surveillance video limits easy-to-use auxiliary tools due to the expense of software product acquisition, operation, and maintenance. Video surveillance systems can intelligently analyse video information, identify odd behaviours, and uncover potentially hazardous behaviours with the use of artificial intelligence and DL technology (Guo et al., 2020; Zhang et al., 2021). On the other side, the existing security video suffers from LR and blurry visual perception due to equipment cost, hardware technology, and environmental limitations.

In summary, there are two issues with the use of surveillance video in the context of public security: (1) Using LR video frames is challenging since high-frequency information is easily lost while zooming in to examine an object, leading to blurring and difficulty in recognising it. (2) It is frequently necessary for viewers of surveillance footage to manually identify the object, which is wasteful and makes it simple to lose the item.

Traditional VSR algorithms rely mainly on interpolation-based (Parihar et al., 2022)and frequency domain-based methods. Complex and varied patterns in video sequences are sometimes challenging for traditional VSR algorithms to manage. They might not perform as well as they might in diverse scenarios, lighting circumstances, and object movements, resulting in artefacts in the super-resolved video frames. Fine-grained features, such as texture and sharp edges, may be difficult for traditional VSR algorithms to capture and restore, especially when working with heavily compressed or subpar input movies. This restriction may lead to output that is fuzzy or less aesthetically pleasing.

To address these issues, this work uses the proposed DL-based VSR method to its full potential to discover the underlying structures and patterns in LR images. This proposed method can successfully infer missing high-frequency information and create aesthetically pleasing, super-resolved frames by training on the dataset.

In essence, the primary contributions of this study are:

1. Proposed a VSR with Deep Fusion Network (VSRDFNet), which uses the recurrent residual framework to learn details from the ground-truth HR frame, and increases the performance of the model with high-quality visuals. The effectiveness and clarity of video analysis are enhanced by using recurrent residual learning.

2. Additionally, developed a texture details framework from the ground-truth HR frame, and fused it with the recurrent residual framework to provide additional information for the Video SR process that allows it to achieve high performance and visually pleasing results.

3. Constructed and optimized an End-to-End deeply fused network for surveillance video analysis. This proposed method used surveillance data to produce sharper visuals with improved resolution.

The remaining section of the paper is organised as follows: Related work is summarised in Section 2. In addition, Section 3 provides a step-by-step description of the VSRDFNet architecture. Section 4 presents the experiment results; an overview of the findings is offered in Section 5 to conclude.

2. Related Work

To produce the matching HR image, the SR approach might use one LR image or a collection of LR images (H. Liu et al., 2022; Z. Wang et al., 2021). This technique functions at a low level in computer vision and is the basis of algorithms that work at a higher level. SR algorithms utilise Deep Convolutional Neural Networks (CNNs) to generate high-quality super-resolved images with unique texture features and rich high-frequency information. This sets a standard for research methodology. The majority of conventional Video Super-Resolution (VSR) techniques were surpassed when Kappelar et al.(Kappeler et al., 2016) created the VSR neural network (VSRNet). Among other things, VSRNet stands out for using a three-layer convolutional neural network, a revolutionary technique in the VSR field. VSRNet's architecture is similar to that of SRCNN (Dong et al., 2016), but the main difference is how many input frames are processed by VSRNet as opposed to SRCNN's single frame.

Handling motion within video frames incorporates motion estimation and compensation (MEMC) in addition to CNN. Compared to example-based SR techniques, this innovation can significantly improve the results of SR reconstruction by backpropagation, resulting in higher quality and efficiency. The challenge of building using convolutional neural networks with deep layers to enhance the extraction of features and characterisation capabilities is one of SRCNN's main drawbacks. ResNet, a residual network, solves this by using shortcut links to connect data from input to output (He et al., 2016). The convolutional layer, fundamental to building a deep neural network, basically entails understanding the distinctions between input and output data. This basic framework depends on maximising high-frequency information essential to SR reconstruction. The learning process is very effective, especially in collecting rich high-frequency information, as many of these differences are almost zero. As a result, the residual network functions as the base network for VSR techniques, significantly improving the effectiveness and calibre of VSR reconstruction. The introduction of the residual network into the VSR field and the construction of a VSR network increases the receptive field and speeds up convergence (S. Li et al., 2019; Yang et al., 2018; Zhu et al., 2019). According to Li et al. (D. Li et al., 2018), a bidirectional recurrent neural network called the residual recurrent convolutional network (RRCN) learns a residual frame. The complete recurrent convolutional network proposed by RRCN is unsynchronized, that it receives input from numerous successive video frames, with only the middle frame being super-resolved.

Recurrent residual VSR methods reconstruct video frames using a single network, and it is substantially faster and of higher quality than VSRNet. Therefore, many researchers prefer RRN algorithms to outperform high-quality results. Sajjadi et al. (Sajjadi et al., 2018) recommended performing warp and motion estimation procedures between the prior and present frames, and then recurrent super-resolving the aligned frame. However, incorrect motion estimations run the risk of producing unwanted artefacts and increasing the chance of error accumulation. To provide past information in feature space without explicit motion estimations, Fouli et al. introduced RLSP (Fuoli et al., 2019). This work method also disseminates previous information in the feature space and is related to RLSP. Additionally, each hidden state was given three RLSP frames in a row. With additional input frames, the hidden state is likely to encounter error accumulation, particularly if there is considerable motion between neighbouring frames. To retain the intricate characteristics over layers, identity mapping in the hidden state is employed in this work. In contrast to previous approaches, this proposed method uses a Recurrent Residual Network (RRN) to recover more information with great accuracy.

.

3. Video Super-Resolution with Deep Fusion Network

An overview of the process flow and particular setups for the VSRDFNet approach are provided in this section. The system has a deep fusion network, which consists of simultaneous frameworks: RRN and texture detailing, that uses implicit motion information to integrate the reference and succeeding frames as input.

3.1 VSRDFNet Network Model

This work proposed a VSR approach to video surveillance based on a deep fusion network to address the issue that public security video surveillance systems lack basic and clever administration and analysis techniques. It aids in the tracking, identifying, and analysis of video for police as well as other surveillance video analysts. It involves improving the spatial resolution of low-quality video frames while reducing artefacts and maintaining significant features. By optimizing the proposed VSR algorithm, this work improves the quality of video frames, which assists the person in analyzing the video content. To reconstruct the HR video frame, this work employed the VSR approach depending upon a deep fusion network. It aids in raising the resolution of the LR frames, assisting the person watching the surveillance video to carefully examine the important object's features, and raising the standard of the content analysis of the surveillance video.

In the domain of public security, the proposed approach for surveillance video developed in this research can help security professionals monitor and analyse surveillance videos. To restore HR video frames from several LR frames, this work uses the VSR approach, which is an extension of image SR. However, the differences between video and image SR methods are very clear; usually the former makes use of inter-frame data.

Figure.1. displays the work process of the proposed DL-based VSR architecture, which consists of real-time video surveillance footage and significantly changing video frames and the VSR method to reconstruct the HR frames of the important video. So, for that model gather several matched LR and HR video sequences to create a huge dataset and ensure that the dataset includes various scenarios, motion styles, and content variants.

To aid in training and enhance model generalisation, preprocess the data by cropping, resizing, and enhancing the video frames. This gives HR visuals that can help for monitoring and analysis purposes. The approach put forth in this study has the potential to resolve the issues with conventional video surveillance, offer viewers of surveillance videos effective support, and improve public safety (Ren et al., 2021).

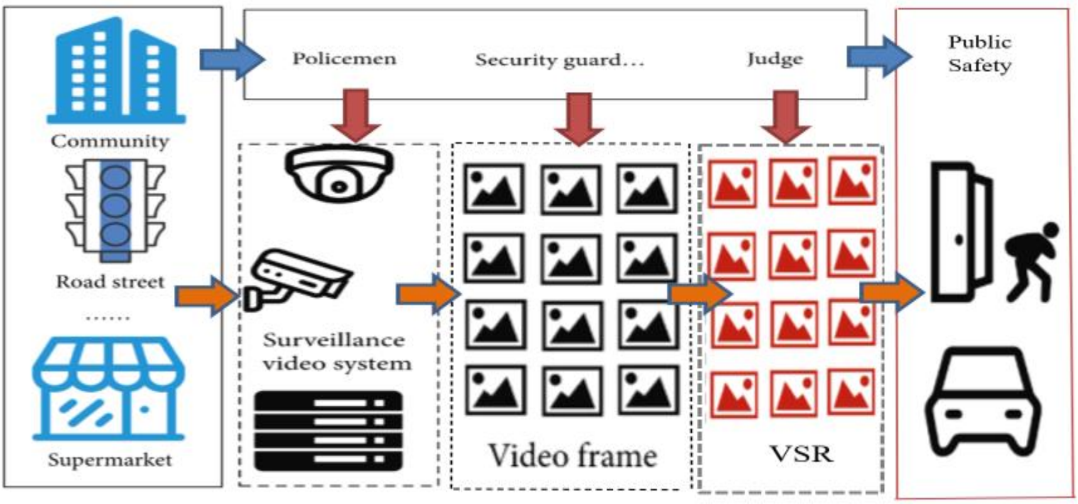

Figure 1. The work process of the proposed VSR method (Ren et al., 2021)

3.2 Network Model Algorithm

This architecture is made up of several interconnected modules that are intended to extract and combine data from various input sources. To efficiently capture spatial and temporal dependencies, this model used convolutional layers, skip connections, recurrent units and fusion techniques for optimal efficiency while maintaining computational performance. Longer videos can potentially be handled more easily by RNNs due to the smooth information flow provided by RRN architecture (Gao et al., 2021; Isobe, Jia, et al., 2020; Isobe, Zhu, et al., 2020), which also lowers the chance that a gradient may vanish during training. Therefore, network model built in this paper consists of four steps: degradation, feature extraction, RRN learning and reconstruction.

Step 1: Use the degradation process to enhance the centre frame's reconstruction effect by utilising the information of nearby LR frames. A factor of k first scales down the HR image (IHR) to obtain the LR image (ILR). It is then downsampled to diminish its resolution with noise.

$$I_{LR} = (I_{HR} * k) \downarrow_s + n \qquad (1)$$

Eq.1. represents the degradation process of a video frame, where ILR stands for LR frame, IHR for HR frame, ↓s for downsampling, k for blur, and n for noise.

The VSR technique typically employs bicubic interpolation to HR frames to enhance its quality and downsampling it to reduce the resolution, which results in paired frames (LR, HR) for learning the mapping connection. This approach facilitates supervised DL. Eq.2., represents the VSR interpolation procedure.

$$I_{LR}=\text{bicubic}(I_{HR})\downarrow_s \qquad (2)$$

Step 2: Use the feature extraction module after the video frames are degraded. Using the LR frames, this stage involves extracting relevant spatial and temporal data. For every time step t, the RRN generates two outputs, ht and ot, for the subsequent time step, t+1, using a series of equations.:

$$x_o = \sigma(W_{conv2D}\{[I_{t-1}, I_t, o_{t-1}, h_{t-1}]\}) \qquad (3)$$

$$x_k = g(x_{k-1}) + F(x_{k-1}), \qquad (4)$$

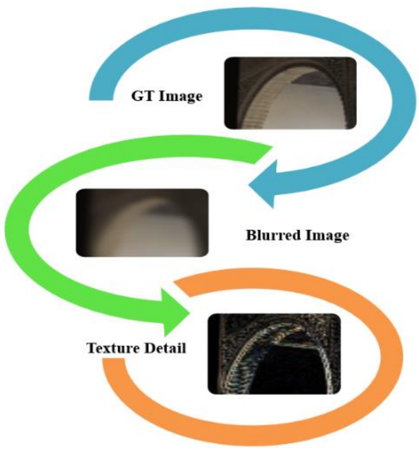

Figure 2. Illustration of texture detailing

where,k ∈ [1, K], ht = σ(Wconv2D{xK}), ot = Wconv2D{xK} Eq.3 uses the ReLU function represented by σ(•). The term g(xk−1) denotes an identity mapping in the k-th residual block, which means g(x k−1) = x k−1 in Eq.4. The term F(x k−1) learned the residual mapping.

Due to the HR image's missing details problem, recovering it from its LR counterpart might be challenging.

$$\text{Feat}(ILR\_R) = xk \qquad (5)$$

Eq.5 represents feature extraction by the RRN framework. This research work used the texture detailing framework, as depicted in Figure 2, to get around this problem. This technique separates high-frequency information, such as texture, to improve the quality of the final image. By comparing the original image to a blurred version, it is possible to recover the high-frequency details within an image, such as textures. This technique has been applied to simple image processing applications including boundary detection and image quality evaluation.

The network has two input frames, current and previous, as I_t and I_{t-1}, respectively shown in Eq.6.

$$F_{oi} = H_{Decomp}(I_t, I_{t-1}) \qquad (6)$$

$$F_{o1} = H_f(I_{t\_C}), \qquad F_{o2} = H_f(I_{t-1\_C})$$

Where H_{Decomp} (•) denotes the decomposition operation, I_{t_C} and I_{t-1_C} denote the decomposed inputs.

Step 3: Through the RRN learning module, mapping among LR and HR features is obtained. Instead of feeding a CNN recurrent network a single frame at a time step, the RNN utilised video frames with residual learning. By including information from every frame in their hidden states, recurrent networks can handle sequential data. Therefore, the extracted details Foi (i = 1, 2) are put forward to Eq.7.

$$F_{Feat\_T} = H_{RRN}(F_{oi}) \quad (i=1,2) \qquad (7)$$

H_{RRN} (•) denotes the residual RNN feature extraction module, consisting of residual learning to extract deep features (Feat). Furthermore, the extracted detail feature FFeat is then upscaled via the upsampling module to eliminate the pixelation effect and estimate extra image details.

$$Feat(I_{LR\_T}) = Feat\_T \qquad (8)$$

Eq.8 represents feature extraction by texture detail framework. Typically, there are three stages in the process. First, we extract the LR features using the layer of convolution (Feature). LR to HR nonlinear mapping to be obtained by the residual network (Res). Afterwards, the same process is done using texture detail input. Figure 3 shows the framework for RRN with residual block presentation, where I_{t-1}, I_t represents the previous and current input frames,h_{t-1} represents the hidden state, and O_t represents the previous output, respectively.

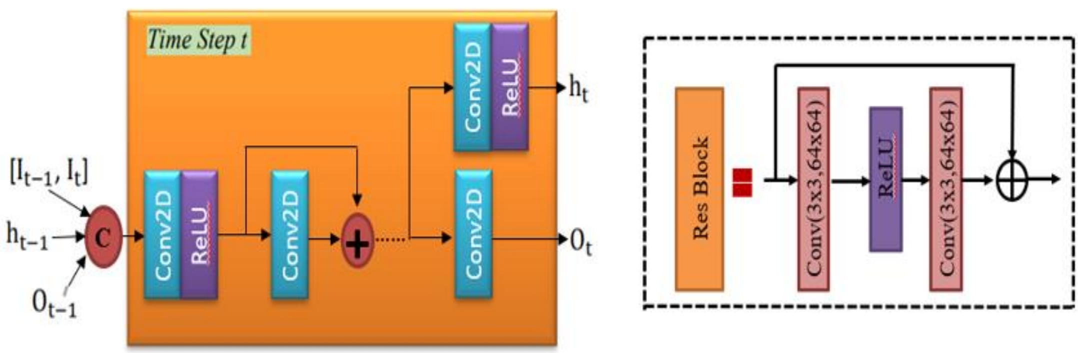

Figure.3. Recurrent residual framework

Step 4: Finally, reconstruct HR frames by combining predicted features with the upsampled LR frames. Eq.9 represents the fusion of these two framework features. The proposed network reconstructs the super-resolved HR video frames by fusing these results and scaling up the LR using the upsampling layer (Upsample) as per Eq.10.

$$Feat(I_{LR}) = (Feat(I_{LR\_R}) , Feat(I_{LR\_T})) \qquad (9)$$

$$I_{SR} = Upsample(Res(Feat(I_{LR}))) \qquad (10)$$

To deal with observed issues, this work presents a novel technique to improve captured characteristics and produce high-quality frames. To achieve high accuracy and thorough reconstruction, the proposed network is built to learn the original frame's textural details and minute details of the ground-truth HR frame.

Therefore, DL technology used with the VSR approach may significantly increase the accuracy of video content of video surveillance systems. It can rebuild high-quality frames by fusing the benefits of the two frameworks in a deep fusion network.

4. Experimental Setup

4.1 Datasets

In this experimental investigation, authors trained their models using the publicly available Vimeo-90k (Xue et al., 2019) dataset. This dataset contains 90k high-quality video scenes. This technique used a dataset with a 64 x 64 patch size and a Gaussian blur with = 1.6. A 4x scale factor was used to further execute the downsampling. The proposed method was evaluated on the VID4 [34] dataset as well as on the institute campus gate surveillance dataset. The quantitative outcomes are compared using peak-signal-noise-ratio (PSNR) and structure similarity index metrics (SSIM) (Sara et al., 2019).

Table 1. Quantitative comparison for 4x VSR

METHOD	BICUBIC	TOFLOW (Xue et al., 2019)	FRVSR (Sajjadi et al., 2018)	DUF (Jo et al., 2018)	RBPN (Haris et al., 2019)	PFNL (Yi et al., 2019)	IPRRN (S. Wang et al., 2023)	GRRN (Ashoori & Amini, 2023)	PROPOSED
PARAM[M]	N/A	1.4	5.1	5.8	12.8	9.5	6.1	8.9	3.7
RUNTIME [MS]	N/A	1658	129	1393	3482	295	57	123	46
VID4 DATASET	23.78 / 0.634	25.89 / 0.765	26.70 / 0.812	27.13 / 0.826	27.41 / 0.838	27.40 / 0.838	28.36 / 0.858	27.36 / 0.827	28.47/ 0.859

4.2 Implementation Details

The proposed method consists of residual blocks. Two convolutional layers with ReLU activation are placed between each residual block. The convolutional layer has a 3 x 3 filter size and 128 channels. The resolution of the LR features is increased to HR using sub-pixel convolution in this technique (Ledig et al., 2017). At the initial step t0, the prior estimation is zero. To train models, the learning rate starts at $1 \times 10{-}4$ and drops by 0.1 per 60 epochs until 70 epochs.

L1 loss function (pixel-wise) with parameters $\beta 1 = 0.9$, $\beta 2 = 0.999$, are used. Adam (Cai et al., 2022) optimizer and a weight decay of $5 \times 10{-}4$ are used to train the models. The luminance

(Y) channel is used to measure the outcomes, and all pixels within the original frames and the network's output have had the L1 loss applied to them. Throughout each experiment, Pytorch 1.1 and Python 3.6.4 were used.

4.3 Comparisons with State-of-the-Arts

This section validates the proposed method with other state-of-the-art VSR methods to demonstrate its effectiveness, including Bicubic, TOFLOW(Xue et al., 2019), FRVSR(Sajjadi et al., 2018), DUF(Jo et al., 2018), RBPN(Haris et al., 2019), RLSP (Fuoli et al., 2019), PFNL(Yi et al., 2019), IPRRN (S. Wang et al., 2023), GRRN (Ashoori & Amini, 2023).

- Quantitative comparison:

Table 1 compares the proposed method quantitatively with state-of-the-art VSR methods with and without alignment methods, like CNN and RNN. This work utilized the VID4 dataset (C. Liu & Sun, 2014) and institute campus surveillance datasets with a scale factor of ×4. Specifically, comparing the recovered results at ×4 scale on the given dataset, the proposed method shows an improvement of 0.11 dB, in terms of PSNR.

Table 2 shows that the performance gets better with an increase in the number of residual blocks (RBs). The effects of increasing RB numbers reflect on reconstruction efficiency.

Table 2. Study on the number of RBs in VSRDFNet on Surveillance dataset

RB NUMBER	PSNR(dB)/SSIM
5L	26.30/ 0.712
10L	27.28/ 0.764

- Qualitative comparison:

The qualitative comparison of surveillance data with relevant frames consisting of persons and vehicles is further analysed in Figure 4, using different residual blocks, such as 5L and 10L. More intriguingly, the performance of the surveillance data is becoming better with the help of the proposed strategy. Visualisations demonstrate that residual block with 10L provides more precise details. To restore missing details, information from a previously hidden state is complementary. The suggested method established recursive architecture without increasing parameters to learn important semantic elements and explore further into the network level.

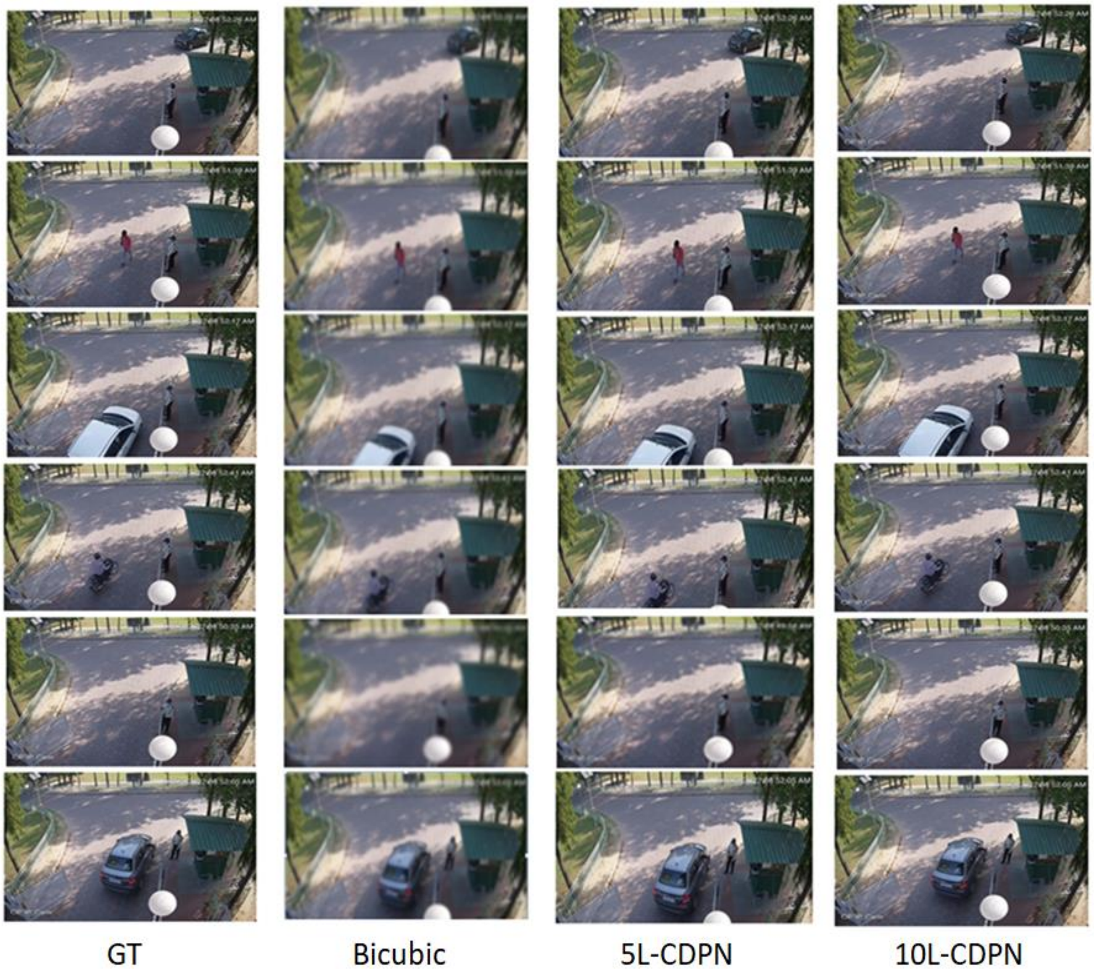

Figure 4. Qualitative comparison of surveillance dataset for 4x VSR

4.4 Model Parameter Size Comparison

Table 3 compares the proposed and state-of-the-art method's parameters. Among these popular methods, the proposed method consists of fewer parameters. Therefore, the proposed method produces better results with fewer parameters, resulting in the highest efficiency in terms of parameters compared to other methods, demonstrating exceptional performance with less computational overload.

Table 3. Model size and performance with scaling 4x.

Model	TOFLOW (Xue et al., 2019)	FRVSR (Sajjadi et al., 2018)	DUF (Jo et al., 2018)	RBPN (Haris et al., 2019)	PFNL (Yi et al., 2019)	IPRRN (S. Wang et al., 2023)	GRRN (Ashoori & Amini, 2023)	Proposed
Parameter(M)	1.4	5.1	5.8	12.8	9.5	6.1	8.9	3.7

5. Conclusion

Safeguarding and ensuring public safety is the cornerstone of intelligent and safe cities. Throughout investigations, surveillance equipment widely placed in public areas can quickly identify suspicious indicators, use surveillance footage to learn more about suspicious vehicles and people and provide unbiased evidence for court cases. Traditional systems need help achieving effective video surveillance because they rely on human interpretation of indistinct video footage. Consequently, this work merges deep learning technology with the video super-resolution system. It consists of real-time video surveillance footage, significantly changing video frames, and the VSR technique for reconstructing the HR frames for the essential video. This provides HR visualisations that might be useful for monitoring and analysing data. The strategy proposed in this study has the potential to address the problems associated with traditional video surveillance, provide appropriate support to those who watch surveillance footage, and improve public safety. Firstly, the RRN framework with residual learning focuses on learning abundant local features in frames. Then, the texture detailing framework enables the network to focus on information detail learning of the original frame, which can be supervised by the ground-truth HR frames and obtain satisfying results. Further, the proposed network fused these frameworks to focus on surveillance data to recover sufficient sharpness and archive high-quality visuals. Besides, this work can be applied to track objects in real-time applications for security and detection purposes in the future. However, video frames contain misalignment, significant motion, and occlusion, which are hard to handle compared to a single frame. Future researchers need to work on these challenges to improve the outcomes for real-time video surveillance applications.

References

1. Ashoori, M., & Amini, A. (2023). Video Super-Resolution Using a Grouped Residual in Residual Network (arXiv:2310.11276; Version 1). arXiv. http://arxiv.org/abs/2310.11276
2. Cai, Q., Li, J., Li, H., Yang, Y.-H., Wu, F., & Zhang, D. (2022). TDPN: Texture and Detail-Preserving Network for Single Image Super-Resolution. IEEE Transactions on Image Processing, 31, 2375–2389. https://doi.org/10.1109/TIP.2022.3154614
3. Chin, N. T. L. and C. (2022, April 7). Police surveillance and facial recognition: Why data privacy is imperative for communities of color. Brookings. https://www.brookings.edu/research/police-surveillance-and-facial-recognition-why-data-privacy-is-an-imperative-for-communities-of-color/
4. Dong, C., Loy, C. C., & Tang, X. (2016). Accelerating the Super-Resolution Convolutional Neural Network (arXiv:1608.00367). arXiv. https://doi.org/10.48550/arXiv.1608.00367
5. Fuoli, D., Gu, S., & Timofte, R. (2019). Efficient Video Super-Resolution through Recurrent Latent Space Propagation (arXiv:1909.08080). arXiv. http://arxiv.org/abs/1909.08080
6. Gao, T., Xiong, R., Zhao, R., Zhang, J., Zhu, S., & Huang, T. (2021). Recover The Residual Of Residual: Recurrent Residual Refinement Network For Image Super-Resolution. 2021 IEEE International Conference on Image Processing (ICIP), 1804–1808. https://doi.org/10.1109/ICIP42928.2021.9506149
7. Guo, K., Hu, B., Ma, J., Ren, S., Tao, Z., & Zhang, J. (2020). Toward Anomaly Behavior Detection as an Edge Network Service Using a Dual-Task Interactive Guided Neural Network. IEEE Internet of Things Journal, PP, 1–1. https://doi.org/10.1109/JIOT.2020.3015987
8. Haris, M., Shakhnarovich, G., & Ukita, N. (2019). Recurrent Back-Projection Network for Video

Super-Resolution (arXiv:1903.10128). arXiv. http://arxiv.org/abs/1903.10128
9. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770–778. https://doi.org/10.1109/CVPR.2016.90
10. Isobe, T., Jia, X., Gu, S., Li, S., Wang, S., & Tian, Q. (2020). Video Super-Resolution with Recurrent Structure-Detail Network (arXiv:2008.00455). arXiv. http://arxiv.org/abs/2008.00455
11. Isobe, T., Zhu, F., Jia, X., & Wang, S. (2020). Revisiting Temporal Modeling for Video Super-resolution (arXiv:2008.05765). arXiv. http://arxiv.org/abs/2008.05765
12. Jo, Y., Oh, S. W., Kang, J., & Kim, S. J. (2018). Deep Video Super-Resolution Network Using Dynamic Upsampling Filters Without Explicit Motion Compensation. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 3224–3232. https://doi.org/10.1109/CVPR.2018.00340
13. Kappeler, A., Yoo, S., Dai, Q., & Katsaggelos, A. K. (2016). Video Super-Resolution With Convolutional Neural Networks. IEEE Transactions on Computational Imaging, 2(2), 109–122. https://doi.org/10.1109/TCI.2016.2532323
14. Ledig, C., Theis, L., Huszar, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., & Shi, W. (2017). Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. 4681–4690. https://openaccess.thecvf.com/content_cvpr_2017/html/Ledig_Photo-Realistic_Single_Image_CVPR_2017_paper.html
15. Li, D., Liu, Y., & Wang, Z. (2018). Video Super-Resolution Using Non-Simultaneous Fully Recurrent Convolutional Network. IEEE Transactions on Image Processing: A Publication of the IEEE Signal Processing Society. https://doi.org/10.1109/TIP.2018.2877334
16. Li, S., He, F., Du, B., Zhang, L., Xu, Y., & Tao, D. (2019). Fast Spatio-Temporal Residual Network for Video Super-Resolution. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 10514–10523. https://doi.org/10.1109/CVPR.2019.01077
17. Liu, C., & Sun, D. (2014). On Bayesian Adaptive Video Super Resolution. IEEE Transactions on Pattern Analysis and Machine Intelligence, 36(2), 346–360. https://doi.org/10.1109/TPAMI.2013.127
18. Liu, H., Ruan, Z., Zhao, P., Dong, C., Shang, F., Liu, Y., Yang, L., & Timofte, R. (2022). Video super-resolution based on deep learning: A comprehensive survey. Artificial Intelligence Review, 55(8), 5981–6035. https://doi.org/10.1007/s10462-022-10147-y
19. Liu, X., Chen, S., Song, L., Woźniak, M., & Liu, S. (2022). Self-attention negative feedback network for real-time image super-resolution. Journal of King Saud University - Computer and Information Sciences, 34(8, Part B), 6179–6186. https://doi.org/10.1016/j.jksuci.2021.07.014
20. Liu, X., Fu, L., Chun-Wei Lin, J., & Liu, S. (2022). SRAS-net: Low-resolution chromosome image classification based on deep learning. IET Systems Biology, 16(3–4), 85–97. https://doi.org/10.1049/syb2.12042
21. Parihar, A. S., Varshney, D., Pandya, K., & Aggarwal, A. (2022). A comprehensive survey on video frame interpolation techniques. The Visual Computer, 38(1), 295–319. https://doi.org/10.1007/s00371-020-02016-y
22. Ren, S., Li, J., Tu, T., Peng, Y., & Jiang, J. (2021). Towards Efficient Video Detection Object Super-Resolution with Deep Fusion Network for Public Safety. Security and Communication Networks, 2021, 1–14. https://doi.org/10.1155/2021/9999398
23. Sajjadi, M. S. M., Vemulapalli, R., & Brown, M. (2018). Frame-Recurrent Video Super-Resolution (arXiv:1801.04590). arXiv. http://arxiv.org/abs/1801.04590
24. Sara, U., Akter, M., & Uddin, M. S. (2019). Image Quality Assessment through FSIM, SSIM, MSE and PSNR—A Comparative Study. Journal of Computer and Communications, 7(3), Article 3. https://doi.org/10.4236/jcc.2019.73002
25. Socha, R., & Kogut, B. (2020). Urban Video Surveillance as a Tool to Improve Security in Public

Spaces. Sustainability, 12, 6210. https://doi.org/10.3390/su12156210
26. Sreenu, G., & Saleem Durai, M. A. (2019). Intelligent video surveillance: A review through deep learning techniques for crowd analysis. Journal of Big Data, 6(1), 48. https://doi.org/10.1186/s40537-019-0212-5
27. Wang, S., Yu, M., Xue, C., Guo, Y., & Yan, G. (2023). Information Prebuilt Recurrent Reconstruction Network for Video Super-Resolution (arXiv:2112.05755). arXiv. https://doi.org/10.48550/arXiv.2112.05755
28. Wang, Z., Chen, J., & Hoi, S. C. H. (2021). Deep Learning for Image Super-Resolution: A Survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 43(10), 3365–3387. https://doi.org/10.1109/TPAMI.2020.2982166
29. Xue, T., Chen, B., Wu, J., Wei, D., & Freeman, W. T. (2019). Video Enhancement with Task-Oriented Flow. International Journal of Computer Vision, 127(8), 1106–1125. https://doi.org/10.1007/s11263-018-01144-2
30. Yang, W., Feng, J., Xie, G., Liu, J., Guo, Z., & Yan, S. (2018). Video super-resolution based on spatial-temporal recurrent residual networks. Computer Vision and Image Understanding, 168, 79–92. https://doi.org/10.1016/j.cviu.2017.09.002
31. Yi, P., Wang, Z., Jiang, K., Jiang, J., & Ma, J. (2019). Progressive Fusion Video Super-Resolution Network via Exploiting Non-Local Spatio-Temporal Correlations. 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 3106–3115. https://doi.org/10.1109/ICCV.2019.00320
32. Zhang, W., Li, H., Yongqin, L., Liu, H., Chen, Y., & Ding, X. (2021). Application of deep learning algorithms in geotechnical engineering: A short critical review. Artificial Intelligence Review, 54, 1–41. https://doi.org/10.1007/s10462-021-09967-1
33. Zhu, X., Li, Z., Zhang, X.-Y., Li, C., Liu, Y., & Xue, Z. (2019). Residual Invertible Spatio-Temporal Network for Video Super-Resolution. Proceedings of the AAAI Conference on Artificial Intelligence, 33(01), Article 01. https://doi.org/10.1609/aaai.v33i01.33015981