JVCSR : VIDEO COMPRESSIVE SENSING RECONSTRUCTION WITH JOINT IN-LOOP REFERENCE ENHANCEMENT AND OUT-LOOP SUPER-RESOLUTION

Yang, Jian

(出版者 / Publisher)
法政大学大学院理工学研究科
(雑誌名 / Journal or Publication Title)
法政大学大学院紀要. 理工学研究科編
(巻 / Volume)
63
(開始ページ / Start Page)
1
(終了ページ / End Page)
7
(発行年 / Year)
2022-03-24
(URL)
https://doi.org/10.15002/00025362

JVCSR: VIDEO COMPRESSIVE SENSING RECONSTRUCTION WITH JOINT IN-LOOP REFERENCE ENHANCEMENT AND OUT-LOOP SUPER-RESOLUTION

Yang Jian
Appiled Informatics,Graduate School of Science and Engineering,
Hosei University,
Supervisor: Jinjia Zhou

Abstract— **Recently, deep learning-based video compressive sensing reconstruction (VCSR) technologies have achieved tremendous success in improving the quality of the reconstructed video. However, the reconstructed video quality at low bit rates or high compression ratios still dissatisfies with the requirement in practice. In this paper, a video compressive sensing reconstruction method with joint in-loop reference enhancement and out-loop super-resolution (JVCSR) is proposed, which focuses on removing reconstruction noises, blocking artifacts and increase the resolution simultaneously. As an in-loop part, the enhanced frame is utilized as a reference to improve the recovery performance of current frame. Furthermore, it is the first work that realizes out-loop super-resolution for VCSR to obtain high quality image at low bit rates. As a result, our JVCSR can improve average of 2.53 dB PSNR by comparing with state-of-the-art compressive sensing methods at the similar bit rate.**

Index Terms— Video compressive sensing reconstruction, low bit rate, super-resolution, reference enhancement

1. INTRODUCTION

Compressive sensing (CS) reconstructs the signal at the rate that is lower than the Nyquist-Shannon Sampling Criterion [1]. In the encoder, N-length signal x will be transformed to a specific domain in which the transformation of x is a sparse signal, and obtain $M \times 1$ measurement y by multiplying measurement matrix. Such as binary matrix, random matrix and structure matrix, numerous measurement matrices are designed to play a crucial role in compressive sensing. In the decoder, the signal x is reconstructed given the transmitted measurement y and measurement matrix by different reconstruction algorithms, e.g., OMP, SPGL and SAMP[2]. In a word, CS is an advantageous reconstruction method that can achieve the comparable performance of full sampling by using a few pieces of information. It is also gradually applied in various practical fields such as surveillance camera sensor, Magnetic Resonance Imaging (MRI), medical scanners, etc.

Suffer from the high computational complexity, traditional CS methods achieve acceptable reconstruction performance but with high time consumption. Deep neural network is well-known for its outstanding performance in feature extracting, learning and representation power when it comes to image processing, and have also been applied to CS to enhance the visual quality or accelerate the reconstruction recently. The authors of [3] propose an ISTA-Net for image CS reconstruction. Nonlinear transform is used to solve the proximal mapping associated with the sparsity-inducing regularizer and dramatically promotes the results of traditional approaches while maintaining fast run time. Nevertheless, image-based CS algorithms can be directly applied to video tasks but fail to utilize the temporal and spatial correlation, such as neighboring frames in video sequences. The research [4] presents a novel recurrent convolutional neural network (CNN) called CSVideoNet, which extracts spatial-temporal correlation features, and proposes a synthesizing LSTM network for motion estimation (ME). Compare with other iterative algorithms, they can significantly improve the recovery video quality. Huang et al.[5] introduce a learning-based CS algorithm (CS-MCNet) with multi-hypothesis motion compensation (MC) to extract correlation information, improves the reconstruction performance by reusing the similarity between adjacent frames. However, some artifacts and noises still remain in the reconstructed videos, resulting in an unpleasant visual effect, especially at low bit rates.

In order to further improve the quality of the reconstructed video, we proposes a novel VCSR framework by employing in-loop reference enhancement and out-loop super-resolution. The contributions of our work can be described as follows:

- In this work, we realize degradation-aware super-resolution for the video compressive sensing reconstruction and obtain state-of-the-art performance.
- An in-loop reference enhancement is designed in this work to remove the artifacts and noises before CS reconstructed current frame is provided reference to the next frame, which improves CS reconstruction performance significantly.
- Our proposal improves the rate-distortion performance

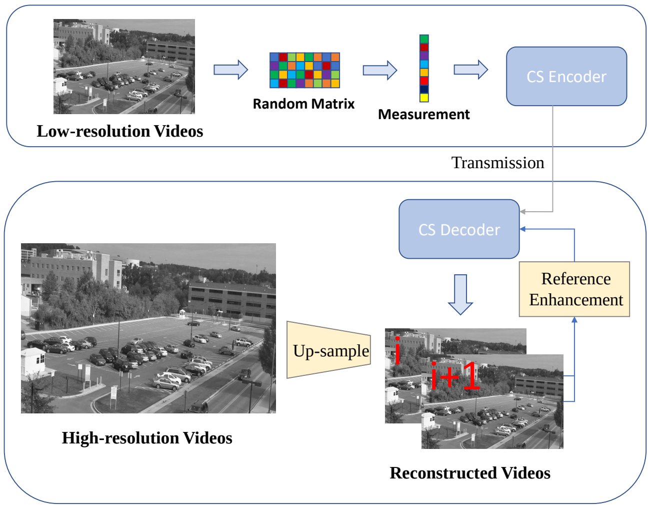

Fig. 1. The concept of our proposal. Reference enhancement improves the quality of reconstructed frame before it is referred to next frame. Up-sample is Our out-loop super-resolution, which utilized to obtain the final high-resolution video.

of the existing compressive sensing algorithms in a wide bit rate range after the out-loop super-resolution.

2. RELATED TECHNOLOGIES

Image enhancement. The paper of [6] investigates three compression architectures, including using super-resolution (SR). Their experimental results demonstrate that super-resolution can achieves superior rate-distortion performance that compares to BPG compression images. In [7], the authors present an end-to-end SR algorithm (CISRDCNN) for JPEG images that improves image resolution and reduces compression artifacts jointly. Plenty of researches indicate that it is beneficial to exploit super-resolution methods with compressed images or videos. Nevertheless, there are not many researches conduct super-resolution base on compressive sensing. On the other hand, such as transform and quantization in HEVC, AVC, and limited measurements in CS, all of them split the image into non-overlapping blocks and each block is performed by lossy compression. Images can not be fully reconstructed and lead to signal distortion. Therefore, to recover the original signal, it requires a method to achieve quality enhancement with fewer artifacts and clearer structures. In the past few years, several CNN-based algorithms present powerful potentiality of denoising and artifacts removal [8], [9]. Dong [8] demonstrate that their 3-layer convolutional neural network is efficient in reducing various compression artifacts. They also mention that reusing shallow features can help learn deeper models for artifact removal. Similar to other compression methods, some blocking artifacts and blurs are generated after compressive sensing reconstructing, especially at low sampling rates.

Pixel Downsample-based Coding. Due to the limited bandwidth and storage capacity, videos and images are downsampled at the encoder and upsampled at the decoder, which can be an effective strategy to save data in storing and transmission. On the other hand, super-resolution, where high-resolution images are obtained given the low-resolution ones, offers higher resolution for images and videos that are captured and recorded in the low-resolution. With the rapid development of CNN methods recently, not only the algorithms for normal image super-resolution, but also some CNN-based compressed image super-resolution methods are also proposed. For the compressed images or videos, directly performing super-resolution would magnify the artifacts and noises simultaneously. To address this issue, the authors in [10] and [11] present a restoration-reconstruction deep neural network (RR-DnCNN), which solves degradation from downsample and compression by using degradation-aware method. The technique of degradation-aware consists of restoration and reconstruction. Restoration is to remove the compression artifacts, and reconstruction leverages up-sampled features from restoration to generate high resolution video.

3. PROPOSED JVCSR

3.1. Overall Framework

The concept of our proposal is shown in Fig. 1. Low-resolution videos are sampled by the CS encoder and upsample after decoding. Low-resolution videos have lower bit rate in transmission, and high-resolution videos are acquired after up-sampling by our network. Our reference enhancement is added to optimize the restored frame, and thus improves the MC performance and final CS reconstruction.

Our proposed framework consists of three parts: Compressive sensing with motion compensation (CS-MC), in-loop reference enhancement (I-RE) and out-loop super-resolution (O-SR). The overall architecture is shown in Fig. 2. As the input of CS-MC, the measurement is acquired by multiplying the pixels of low-resolution videos with a random measurement matrix. After a frame is reconstructed, a buffer is used to store reconstructed results of the current frame and provide a reference for the next frame. It is noteworthy that the images restored through CS-MC would generate black spots in some specific cases, which are caused by block-based MC. To address this problem, we design an I-RE module to realize optimization for the reference frame. By this module, noises and artifacts of the recovery frame are removed effectively. The recovery frame with higher quality is both utilized for motion compensation of CS decoder to reconstruct better next frame cyclically, and also as the input of super-resolution. After in-loop section, an out-loop super-resolution module is presented to increase the resolution of sequences to achieve pleasant visual results and compare performance with orig-

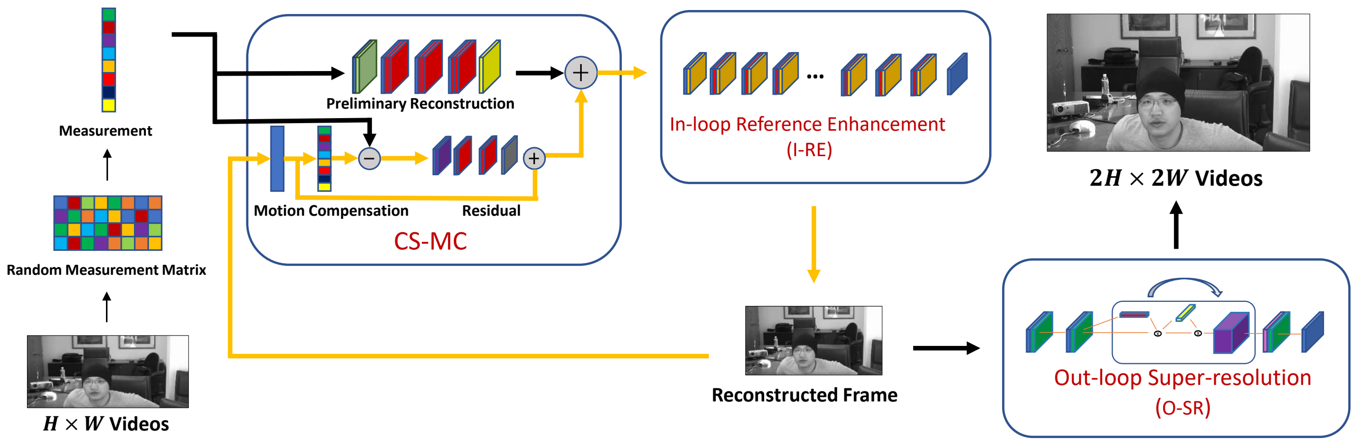

Fig. 2. The overall architecture of our framework is divided into three parts. Low-resolution Video is firstly to do the compressive sensing by CS-MC. After a frame is reconstructed, it will be enhanced by our I-RE before being provided as a reference for the next frame. The final enhanced video frame will be fed to our O-SR to obtain the high-resolution video.

inal videos. Details of our architecture are discussed in the following sections.

3.2. Network Architectures

In-loop Reference Enhancement (I-RE): In Huang's work [5], they design a buffer to store reconstructed results of the current frame and provides a reference for the recovery to the next frame. There are still some artifacts and noises that remain in the reconstructed image, especially in high CR. In our work, low-resolution video sequences are sampled by CS encoder instead of original high resolution videos. Therefore, it is more significant to conduct enhancement. As shown in Fig. 3, we design our architecture base on the work of [9], which demonstrates that combining residual learning and batch normalization can achieve the outstanding visual performance of denoise models. As we mentioned before, sampling rates, also well-known as compression ratio (CR) is defined as $CR = \frac{M}{N}$. Since smaller CR indicates fewer signals are sampled, a deep neural network can not completely play its role. To simplify the network and reduce the size of parameters, we appropriately delete some hidden layers when CR is set to small. For hidden layers, we use 64 filters of size $3 \times 3 \times 64$ to get feature maps, and batch normalization is also connected after each convolution layer. Except for the last layer, all convolutional layers are followed by rectified linear units (ReLU) layers. Experiments demonstrate that fewer layers of our enhancement can achieve acceptable performance with fewer network parameters.

Out-loop Super-resolution (O-SR): In [12], The authors introduce a convolutional block attention module (CBAM), which is extensively applied in various learning-based tasks such as image recognition and classification. This attention module has two parts: The channel attention module utilizes both average-pool and max-pool synchronously to im-

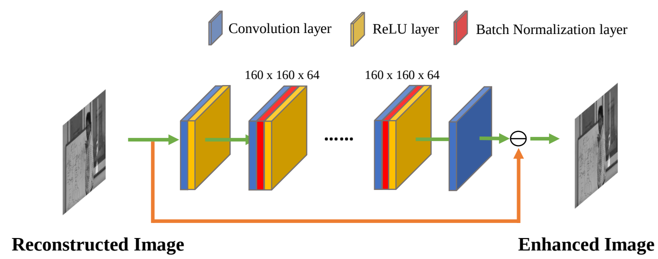

Fig. 3. The architecture of our enhancement module. The number of hidden layers depends on the value of CR. Enhanced images will be obtained after subtracting the learning result from the reconstructed image since it is residual learning.

prove the representation power of the network. On the other hand, the spatial attention module works to find an informative part to supplement channel attention. It is worth noting that this attention module also performs well in our task. [13] (SRFBN) proposes a novel feedback block module, which effectively reuses feature and feedback information to achieves state-of-the-art SR performance. Therefore, a CNN-based super-resolution module is presented in our work to obtain the ultimate high quality high-resolution video after enhancing the reconstructed ones. The architecture is shown in Fig. 4. CBAM is connected in each feedback block for adaptive feature refinement.

4. EXPERIMENTAL RESULTS

4.1. Experimental settings

Training dataset: We use Ultra Video Group (UVG)[14] to build the training dataset of training super-resolution. UVG is composed of 16 versatile 4K (3840 * 2160) video sequences and commonly used in video-based works. These natural sequences were captured either at 50 or 120 frames per second (fps) and stored online in different formats. We choose videos from UVG dataset and get around 1050 pairs of images for training and validation in total. For training compressive sensing model, as there is not standard dataset designed for video CS, we randomly pick 15% UCF-101 dataset with 100 frames for training. All the video sequences are converted to one channel and only extracted luminance signal.

Testing dataset: It is difficult for humans to distinguish quality difference between two video sequences of the same content when they have the close compressed ratio. The MCL-JCV dataset[15] is designed to measure this phenomenon for each test subject. Since surveillance is ubiquitous in practical and requires high resolution videos, we also use the VIART dataset [16] to test the robustness of our methods.

Training setting: The experiments of our framework are implemented with Pytorch 1.2.0 on Ubuntu 16.04, and NVIDIA GeForce RTX2080Ti GPUs are supported for our training. Adam optimization is used to refine the parameters while training. We separate the training into three modules, a super-resolution module, an enhancement module, and a compressive sensing module but finally connect them to an end-to-end trainable network. In order to demonstrate the robustness of our framework, four models are trained for each module with different compression ration video (0.125, 0.25, 0.5 and 0.75).

In the training of super-resolution, the scale factor is set to 2. We initialize the learning rate to 1×10^{-4}, and multiplies by $\frac{1}{2}$ every 250 epochs with total 1000 epochs. For training the enhancement module, as we mentioned before, the network depth depends on compression. The model of largest CR we used (0.75) has 17 hidden layers (as shown in Fig. 3) and CR = 0.5 has 16 layers, etc. For compressive sensing, training for 200 epochs with a batch size of 400 and 0.01 learning rate yield best reconstruction performance in our case.

4.2. Experimental results

The results of experiments are evaluated by two standards, signal-to-noise ratio (PSNR) and structural similarity (SSIM). Visual performance is also shown in the following sections.

Results on in-loop reference enhancement: Before upsampling of this work, higher quality images are generated since we have the enhancement for compressive sensing. We show some visual examples of reconstructed images and enhanced results in Fig. 5. It can be easily seen the effect of our enhancement, a number of noises are removed successfully. Table. 1 lists the average PSNR/SSIM results of our test dataset, and demonstrates that our enhancement performs meaningful effect, especially under high compression ratio.

Results on out-loop super-resolution: To perform the superiority of each module in our work, we conduct the comparison as follow: 1) Delete I-RE and O-SR module then directly up-sample the images by bicubic interpolation. 2) Delete O-SR module then up-sample the images by bicubic interpolation. 3) Replaced O-SR module with SRCNN[17]. 4) Replaced O-SR module with DRRN[18]. 5) Replaced O-SR module with DBPN[19]. 6) Both I-RE and O-SR are utilized. As shown in the Table. 2, it is easy to judge our work achieves superior performance by these ablation experiments.

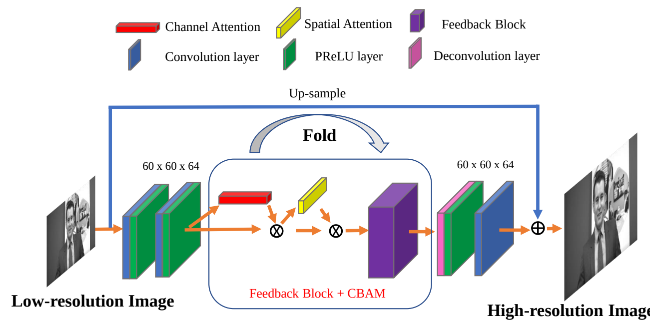

Fig. 4. The architecture of out-loop super-resolution. CBAM is connected in each feedback block for adaptive feature refinement. The final output high-resolution image is obtained by adding the upsampled low-resolution image and learning results. Upsample kernel is set to bicubic here.

Overall bit-rate reduction and comparison: To perform the coding advantage of our proposal, Fig. 6 shows the rate-distortion results at different bit rates obtained by the test dataset. We compare with SAMP, ISTA-Net[3] and CS-MCNet[5], the curves demonstrate that our proposal outperforms other compressive sensing algorithms over a wide range of bit rates. Moreover, Fig. 7 shows the visual results of some examples under the same bit rate, to further show the superiority of our framework. As we can see, our proposal retains the most details, suffers minimal block effect, and removes noises significantly. PSNR and SSIM results of some test examples (10 sequences from VIRAT and 9 sequences from MCL-JCV) at two bit rates comparison with different compressive sensing algorithms are shown in Table. 3.

5. CONCLUSIONS

In this paper, we propose a video compressive sensing reconstruction framework with joint in-loop reference enhancement and out-loop super-resolution (JVCSR). First, an in-loop enhancement module is designed to enhance the pixel information and realize the optimization of the reference frame for CS. Experimental results show that the artifacts and noises are removed effectively, leading to better CS recovery results. Furthermore, adopt the concept of downsampling-based video coding, we propose an out-loop super-resolution module to increase resolution for the low-resolution videos with

Table 1. Average PSNR and SSIM performance comparison of I-RE under different CRs on our test dataset.

CR	Metric	Reconstructed	Enhanced
0.125	PSNR	24.16	26.42
	SSIM	0.6837	0.7790
0.25	PSNR	25.96	29.01
	SSIM	0.7932	0.8744
0.5	PSNR	27.15	31.63
	SSIM	0.8719	0.9281
0.75	PSNR	28.99	33.35
	SSIM	0.9099	0.9535

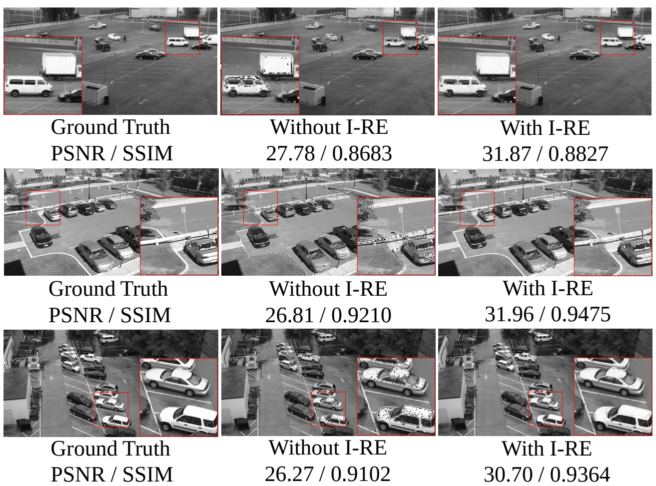

Fig. 5. The enhancement visual results. Noises and artifacts reconstructed by compressive sensing are removed by our I-RE module effectively. Please zoom in for better views and comparisons.

different compression ratio. By comparing to other state-of-the-art algorithms, our proposal achieves better performance relatively. Moreover, to show the advantage of our framework in low bit rate video coding, we obtain better rate-distortion performance in a wide range than other CS algorithms. To the best of knowledge, it is the first work to propose degradation-aware super-resolution for video compressive sensing reconstruction. Regarding to the future work, we tend to achieve better performance of CS by enhancing measurement directly instead of pixel.

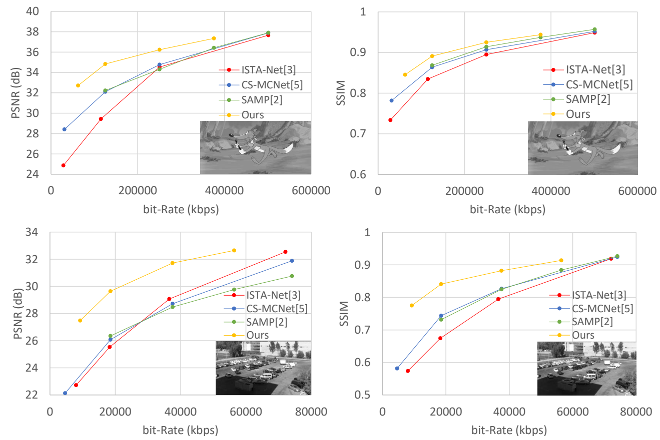

Fig. 6. The PSNR and SSIM rate-distortion curves of two test video datasets. Our proposal outperforms other compressive sensing algorithms for comparison over a wide range of bit rates.

6. REFERENCES

[1] Marco F Duarte, Mark A Davenport, Dharmpal Takhar, Jason N Laska, Ting Sun, Kevin F Kelly, and Richard G Baraniuk, "Single-pixel imaging via compressive sampling," *IEEE signal processing magazine*, vol. 25, no. 2, pp. 83–91, 2008.

[2] Shihong Yao, Qingfeng Guan, Sheng Wang, and Xiao Xie, "Fast sparsity adaptive matching pursuit algorithm for large-scale image reconstruction," *EURASIP Journal on Wireless Communications and Networking*, vol. 2018, no. 1, pp. 78, 2018.

[3] Jian Zhang and Bernard Ghanem, "Ista-net: Interpretable optimization-inspired deep network for image compressive sensing," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 1828–1837.

[4] Kai Xu and Fengbo Ren, "Csvideonet: A real-time end-to-end learning framework for high-frame-rate video compressive sensing," in *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 2018, pp. 1680–1688.

[5] Bowen Huang, Jinjia Zhou, Xiao Yan, Ming'e Jing, Rentao Wan, and Yibo Fan, "Cs-mcnet: A video compressive sensing reconstruction network with interpretable motion compensation," in *Proceedings of the Asian Conference on Computer Vision*, 2020.

[6] Zhengxue Cheng, Heming Sun, Masaru Takeuchi, and Jiro Katto, "Performance comparison of convolutional autoencoders, generative adversarial networks and super-resolution for image compression.," in *CVPR Workshops*, 2018, pp. 2613–2616.

[7] Honggang Chen, Xiaohai He, Chao Ren, Linbo Qing, and Qizhi Teng, "Cisrdcnn: Super-resolution of compressed images using deep convolutional neural networks," *Neurocomputing*, vol. 285, pp. 204–219, 2018.

[8] Chao Dong, Yubin Deng, Chen Change Loy, and Xiaoou Tang, "Compression artifacts reduction by a deep convolutional network," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 576–584.

[9] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng,

Table 2. Average PSNR and SSIM performance comparison between I-RE + O-SR and other super-resolution algorithms at different CRs.

CR	0.125	0.25	0.5	0.75
Bicubic	24.20 / 0.6793	25.90 / 0.7391	26.99 / 0.8168	27.91 / 0.8575
I-RE + Bicubic	25.64 / 0.7253	27.44 / 0.7822	28.83 / 0.8504	30.03 / 0.8873
I-RE + SRCNN[17]	25.95 / 0.7532	28.03 / 0.8141	29.47 / 0.8721	30.93 / 0.8998
I-RE + DRRN[18]	26.79 / 0.7784	28.97 / 0.8345	30.14 / 0.8823	31.68 / 0.9074
I-RE + DBPN[19]	27.24 / 0.7882	29.32 / 0.8492	31.07 / 0.8942	32.59 / 0.9169
I-RE + O-SR	27.75 / 0.8053	29.91 / 0.8610	31.98 / 0.9020	33.31 / 0.9238

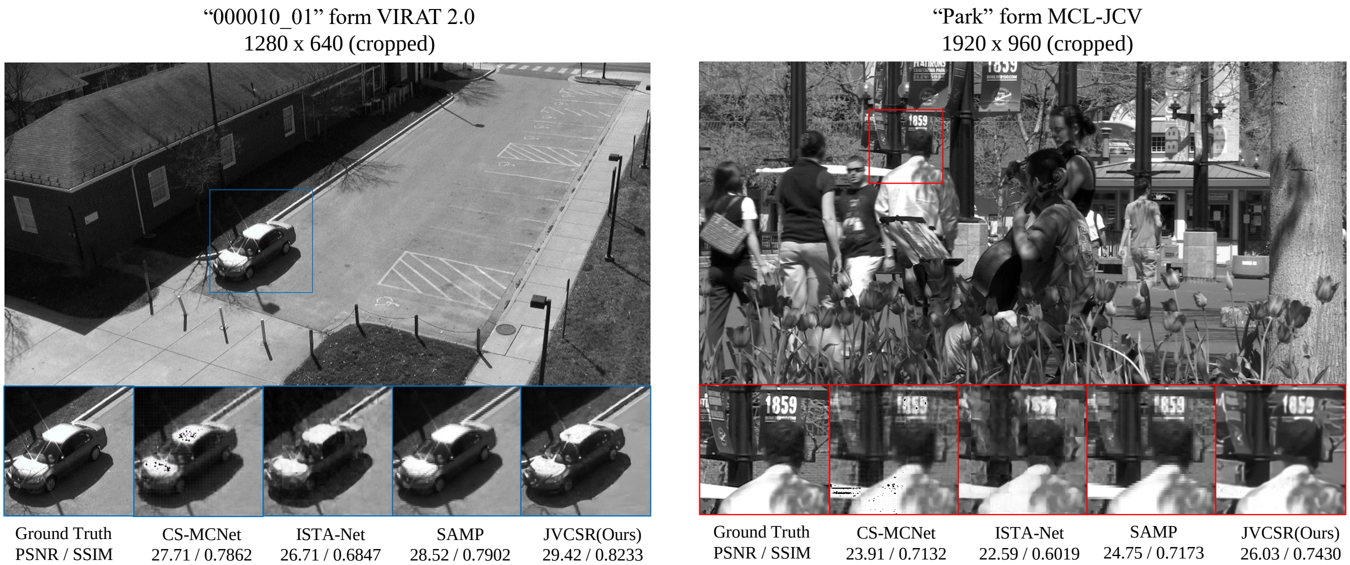

Fig. 7. The visual results of some examples around the same bit rate. The original images are also presented in the figure. Please zoom in for better views and comparisons.

and Lei Zhang, "Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising," *IEEE Transactions on Image Processing*, vol. 26, no. 7, pp. 3142–3155, 2017.

[10] Minh-Man Ho, Gang He, Zheng Wang, and Jinjia Zhou, "Down-sampling based video coding with degradation-aware restoration-reconstruction deep neural network," in *International Conference on Multimedia Modeling*. Springer, 2020, pp. 99–110.

[11] Man M Ho, Jinjia Zhou, Gang He, Muchen Li, and Lei Li, "Sr-cl-dmc: P-frame coding with super-resolution, color learning, and deep motion compensation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2020, pp. 124–125.

[12] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon, "Cbam: Convolutional block attention module," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 3–19.

[13] Zhen Li, Jinglei Yang, Zheng Liu, Xiaomin Yang, Gwanggil Jeon, and Wei Wu, "Feedback network for image super-resolution," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 3867–3876.

[14] Alexandre Mercat, Marko Viitanen, and Jarno Vanne, "Uvg dataset: 50/120fps 4k sequences for video codec analysis and development," in *Proceedings of the 11th ACM Multimedia Systems Conference*, 2020, pp. 297–302.

[15] Haiqiang Wang, Weihao Gan, Sudeng Hu, Joe Yuchieh Lin, Lina Jin, Longguang Song, Ping Wang, Ioannis Katsavounidis, Anne Aaron, and C-C Jay Kuo, "Mcl-jcv: a jnd-based h. 264/avc video quality assessment dataset," in *2016 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2016, pp. 1509–1513.

[16] "Virat description, available at," http://www.viratdata.org/.

[17] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang, "Learning a deep convolutional network

Table 3. PSNR and SSIM results of test examples (10 sequences from VIRAT and 9 sequences from MCL-JCV)at two bitrates comparison with different compressive sensing algorithms. The best performance is indicated as red color and second best as blue color.

bit rate (kbps)	Sequence	Bicubic	ISTA-Net[3]	CS-MCNet[5]	SAMP[2]	JVCSR (Ours)
1.9×10^4	000201_00	27.49 / 0.7234	27.82 / 0.6953	27.85 / 0.7682	28.18 / 0.7431	31.02 / 0.8631
	000201_01	27.42 / 0.7212	27.73 / 0.6934	27.72 / 0.7658	28.10 / 0.7411	30.91 / 0.8610
	010100_03	25.94 / 0.7336	25.78 / 0.6862	26.24 / 0.7442	26.35 / 0.7433	29.82 / 0.8343
	010100_04	25.81 / 0.7312	25.68 / 0.6824	26.17 / 0.7419	26.24 / 0.7414	29.74 / 0.8318
	000200_00	27.78 / 0.7367	26.71 / 0.6847	27.71 / 0.7862	28.52 / 0.7902	29.42 / 0.8233
	000200_01	27.85 / 0.7384	26.79 / 0.6861	27.83 / 0.7884	28.60 / 0.7931	29.42 / 0.8242
	010101_01	28.99 / 0.7724	28.01 / 0.7234	30.21 / 0.8483	30.44 / 0.8351	31.33 / 0.8722
	010101_02	28.92 / 0.7724	27.94 / 0.7217	30.15 / 0.8462	30.35 / 0.8322	31.24 / 0.8703
	050201_08	24.28 / 0.6834	24.11 / 0.6561	25.25 / 0.7344	25.40 / 0.7522	27.89 / 0.8159
	050201_09	24.32 / 0.6873	24.12 / 0.6567	25.32 / 0.7361	25.44 / 0.7532	27.81 / 0.8142
1.2×10^5	Park	29.44 / 0.7234	30.63 / 0.8041	30.85 / 0.8452	30.05 / 0.8244	33.17 / 0.9023
	Cartoon01	29.30 / 0.7202	30.51 / 0.8019	30.75 / 0.8433	29.97 / 0.8230	33.06 / 0.8994
	Cartoon02	28.14 / 0.7436	29.04 / 0.7992	28.83 / 0.8223	28.52 / 0.8217	31.78 / 0.8843
	Telescope	28.03 / 0.7401	28.96 / 0.7972	28.83 / 0.8201	28.52 / 0.8193	31.70 / 0.8817
	Kimono	29.73 / 0.7847	30.11 / 0.7947	30.67 / 0.8362	30.47 / 0.8397	31.37 / 0.8613
	Car	29.80 / 0.7869	30.21 / 0.7971	30.79 / 0.8388	30.54 / 0.8421	31.46 / 0.8637
	Building	30.92 / 0.8234	31.44 / 0.8542	31.49 / 0.8653	31.33 / 0.8632	32.41 / 0.8939
	Beach	30.87 / 0.8212	31.38 / 0.8511	31.40 / 0.8631	31.26 / 0.8617	32.33 / 0.8912
	Parrot	25.05 / 0.7934	27.24 / 0.7846	27.51 / 0.8337	26.99 / 0.8354	29.80 / 0.8731

for image super-resolution," in *European conference on computer vision*. Springer, 2014, pp. 184–199.

[18] Ying Tai, Jian Yang, and Xiaoming Liu, "Image super-resolution via deep recursive residual network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 3147–3155.

[19] Muhammad Haris, Gregory Shakhnarovich, and Norimichi Ukita, "Deep back-projection networks for super-resolution," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 1664–1673.