JCSE Journal of Computing Science and Engineering

Regular Paper

Journal of Computing Science and Engineering,
Vol. 17, No. 1, March 2023, pp. 13-19

Knowledge Distillation for Optical Flow-Based Video Super-resolution

Jungwon Lee and Sang-hyo Park*
School of Computer Science and Engineering, Kyungpook National University, Daegu, Korea
ljw8541@knu.ac.kr, s.park@knu.ac.kr

Abstract
Recently, deep learning-based super-resolution (SR) models have been used to improve SR performance by equipping preprocessing networks with baseline SR networks. In particular, in video SR, which creates a high-resolution (HR) image with multiple frames, optical flow extraction is accompanied by a preprocessing process. These preprocessing networks work effectively in terms of quality, but at the cost of increased network parameters, which increase the computational complexity and memory consumption for SR tasks with restricted resources. One well-known approach is the knowledge distillation (KD) method, which can transfer the original model's knowledge to a lightweight model with less performance degradation. Moreover, KD may improve SR quality with reduced model parameters. In this study, we propose an effective KD method that can effectively reduce the original SR model parameters and even improve network performance. The experimental results demonstrated that our method achieved a better PSNR than the original state-of-the-art SR network despite having fewer parameters.

Category: Computer Graphics / Image Processing

Keywords: Video super-resolution; Optical flow; Knowledge distillation; Deep learning; Super-resolution

I. INTRODUCTION

Convolutional neural networks (CNN) have been successfully applied in various fields of computer vision, such as classification, object detection, and semantic segmentation, compared with traditional approaches. CNN-based methods show excellent performance in super-resolution (SR), and better performance can be achieved using deeper CNN layers [1-6]. In addition to deepening the layers, SR tasks improve performance by equipping a preprocessing network in addition to the baseline SR networks. In video SR, preprocessing networks extract additional features from images, such as optical flow [7-14]. It can facilitate the generation of high-resolution (HR) images, and these features are used as inputs for the SR networks. Despite the fact that these preprocessing networks exhibit greater SR performance, they inevitably require more model parameters. In SR, minimizing model parameters is a problem to be addressed. This implies that SR often operates on mobile devices and is not provided with sufficient computing resources. Therefore, when the number of parameters increases, problems in memory consumption and computational complexity may occur.

Knowledge distillation (KD) has been proposed as a suitable methodology for decreasing the model parameters of neural networks among the various techniques that reduce model parameters. KD is a method of distilling

Open Access **http://dx.doi.org/10.5626/JCSE.2023.17.1.13** **http://jcse.kiise.org**

Received 22 September 2022; **Accepted** 21 February 2023
*Corresponding Author

 pISSN: 1976-4677 eISSN: 2093-8020

the knowledge of a pre-trained model (teacher model) into a lightweight model (student model). KD was first proposed to distill knowledge in image classification [7, 15]. Subsequently, some studies have been conducted on the application of KD in SR and have attempted to adapt KD using information from hidden layers and the last SR image generated for a single SR network [16-18]. Although they showed that KD can be used to effectively reduce model parameters for SR models, they tried only a single SR network without considering a preprocessing network.

With regard to an SR model with a structure that incorporates a preprocessing network, we suggest an efficient KD approach in this research. The contributions of this study are as follows:

- Our method has reduced the parameters of the state-of-the-art model, supero-resolve optical flows for video SR (SOF-VSR), and improved its performance.
- Our method can be applied to other optical flow-based SR models.
- Our method can be extended to other networks with preprocessing networks.

We demonstrated the effectiveness of our proposed method by applying it to state-of-the-art video SR models, SOF-VSR [7], using an optical flow-based preprocessing network. Our experimental results showed that better peak signal-to-noise ratio (PSNR) was obtained using our method than without our method. Compared with a reduced model without any method, our method achieved better PSNR results. Moreover, the proposed method improved PSNR compared to a teacher model.

II. RELATED WORK

A. General SR

SR is a technology that produces HR images from low-resolution (LR) images. Traditional SRs restore images through interpolation-based methods, such as bicubic, bilinear, and nearest-neighbor interpolation. Interpolation-based methods estimate the pixel value of an HR image based on the surrounding pixel values. These methods are easy to implement; however, the resulting image is blurred and unrealistic.

Recently, CNN-based SR methods have achieved excellent SR performances. SRCNN was the first CNN-based model for SR tasks proposed by Dong et al. [1]. It outperformed existing SR methods with simple structures. Subsequently, many models increase the SR performance by expanding the receptive fields and increasing the depth. VDSR has a very deep CNN layer architecture, which increases SR performance by expanding the receptive field [3]. SRResNet is an SR network with ResNet structure, which shows ResNet structure is suitable for low-level SR tasks [6]. Since then, CNN networks have shown success by stacking hundreds of residual blocks, including EDSR [2], RDN [5], and RCAN [4]. However, while stacking hundreds of layers improves the SR performance, the model becomes heavier, and the number of parameters increases.

Video super-resolution (VSR) generates a central HR image using multiple consecutive frames. These multiple frames provide additional scene information. Temporary information between sequential frames can be used for motion estimation [8-14]. For motion estimation, optical flow is the scene movement between adjacent frames, which can be used to estimate the scene velocity and shift through motion compensation. Recent research has demonstrated the efficacy of CNN-based optical flow estimation techniques for motion estimation [19-22]. The number of parameters in the OFRnet is kept from rising by using recurrent modules made up of three ResBlocks and shared at all levels. SOF-VSR with a preprocessing network that estimates the HR optical flow improved the SR performance [7]. However, this also resulted in an increase in the number of model parameters.

B. SOF-VSR Model Overview

An SOF-VSR [7] is a CNN-based VSR model that uses a preprocessing network to infer an HR optical flow. It consists of three modules: the optical flow reconstruction network (OFRnet), motion compensation module, and SR network (SRnet). First, the OFRnet aims to generate an HR optical flow from adjacent LR frames. It has a multiscale architecture that can effectively estimate HR optical flows. It consists of three levels, and as it passes through each level, it expands the scale of the optical flow. It uses recurrent modules that are shared at all levels, which consist of three ResBlocks, and prevents an increase in the number of parameters in the OFRnet. For Level 3, three additional ResBlocks and SR modules are required to recover the HR optical flow. Second, the motion compensation module is part of the motion compensation of the HR optical flows estimated by OFRnet. By representing the HR optical flow at the same resolution as the LR frame, the space-to-depth transformation retrieved the LR flow cube. This flow cube is combined with an LR frame to generate a draft cube through motion compensation. Finally, the SRnet generated a central HR image from the compensated draft cube. It consisted of a 320 channels 3×3 convolutional layer, eight ResBlock modules, a subpixel layer for resolution enhancement, and 3×3 convolutional layer. Two losses were used in the SOF-VSR model: OFR and SR losses. The OFR loss was calculated using three optical flows generated at the three levels of the OFRnet. The SR loss was calculated using the mean square error (MSE) of the ground truth (GT) pixel value I_o^H and the model's SR pixel value I_o^{SR}.

Therefore, the final loss $\mathcal{L}_{SOF}$ is defined as in Eq. (1) as below:

$$\mathcal{L}_{SOF} = \mathcal{L}_{SR} + \lambda_4 \mathcal{L}_{OFR} \tag{1}$$

where $\mathcal{L}_{SOF}$ is added to $\mathcal{L}_{SR}$, as defined in Eq. (2), and $\mathcal{L}_{OFR}$, as defined in Eq. (3), multiplied by the weight of λ_4. The detailed equations are as follows:

$$\mathcal{L}_{SR} = MSE(I_o^{SR}, I_o^H). \tag{2}$$

$$\mathcal{L}_{OFR} = \sum_{i\in[-N,N],i\neq 0} \frac{\mathcal{L}_{level3,i} + \lambda_2 \mathcal{L}_{level2,i} + \lambda_1 \mathcal{L}_{level1,i}}{2N},$$

where

$$\begin{cases} \mathcal{L}_{level3,i} = \| W(I_i^H, F_{i\to 0}^H) - I_o^H \|_1 + \lambda_1 \| \nabla F_{i\to 0}^H \|_1 \\ \mathcal{L}_{level2,i} = \| W(I_i^L, F_{i\to 0}^L) - I_o^L \|_1 + \lambda_2 \| \nabla F_{i\to 0}^L \|_1 \\ \mathcal{L}_{level1,i} = \| W(I_i^{LD}, F_{i\to 0}^{LD}) - I_o^{LD} \|_1 + \lambda_3 \| \nabla F_{i\to 0}^{LD} \|_1. \end{cases} \tag{3}$$

In the original study, the value of λ_4 used was 0.01. The OFRnet and SRnet have 0.413 M and 0.587 M parameters, respectively, and the motion compensation module has no training parameters. Therefore, the SOF-VSR has a total of 1 M parameters. Fig. 1 shows the SOF-VSR architecture.

C. KD Methods for SR

Hinton first proposed KD method for model compression in image classification [23]. KD facilitates learning by transferring knowledge to a compact student network with fewer parameters from a pre-trained larger teacher network. The general loss of KD is expressed as follows:

$$\mathcal{L}_{KD} = \alpha \mathcal{L}_{Student} + (1-\alpha)\mathcal{L}_{Distillation}, \tag{4}$$

where $\mathcal{L}_{Student}$ is the student loss, which is the loss originally used for learning without KD, and $\mathcal{L}_{Distillation}$ is the distillation loss, which is an additional loss that is generated by learning the student from the outputs of the teacher in KD, and α is a hyperparameter that reflects the ratio between each loss and has a value ranging from 0 to 1.

In SR, L1 or MSE loss is generally used as the reconstruction loss for images created by teachers and students. In addition to the ground truth, many studies have shown that the results generated by the teacher, that is, the teacher's knowledge, can facilitate students learn more efficiently.

III. METHOD

A. Training Process

To efficiently distill teacher's knowledge, we propose a two-phase KD process. Fig. 2 shows the entire training process. Each phase applies KD separately to OFRnet and SRnet. It was designed to reduce performance degradation during the process of transferring teacher's knowledge.

The first phase was focused on SRnet to adapt to KD while maintaining the teachers' SR performance. The students shared a pre-trained OFRnet with the teachers. By feeding the teacher's OFRnet output to the student's SRnet, the student's SRnet generated SR images by receiving better motion compensation images from the beginning. It was expected that student's SRnet will produce better-quality images. In addition to supervising the GT, the student's SRnet also learns about the SR image of the teacher. All teacher network froze during network training.

In the second phase, we trained student's OFRnet in combination with SRnet learned in the previous step. Unlike in the first phase, the HR optical flow generated by the teacher can be used as additional knowledge to learn student. The second phase was accompanied by supervision of the HR optical flow created by the OFRnet. In addition to making the final SR image the same, we aimed for the student to make the optical flow similar to that of the teacher. Similar to the first phase, except student's OFRnet, all networks, including student's SRnet, freeze during training.

B. Distillation Loss

Different loss functions were used for each phase. We defined the loss of the first phase as $\mathcal{L}_{Phase1}$ and that of the second phase as $\mathcal{L}_{Phase2}$. In the first phase, we trained only the student's SRnet while sharing the teacher's OFRnet. Therefore, we added $\mathcal{L}_{SR\_KD}$ to make the student's SR output $\mathcal{L}_{Student}$ similar to the teacher's SR output $I_{Teacher}$.

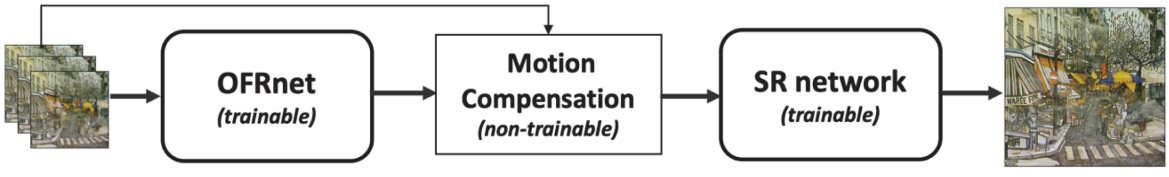

Fig. 1. Optical flow-based video super-resolution model overview.

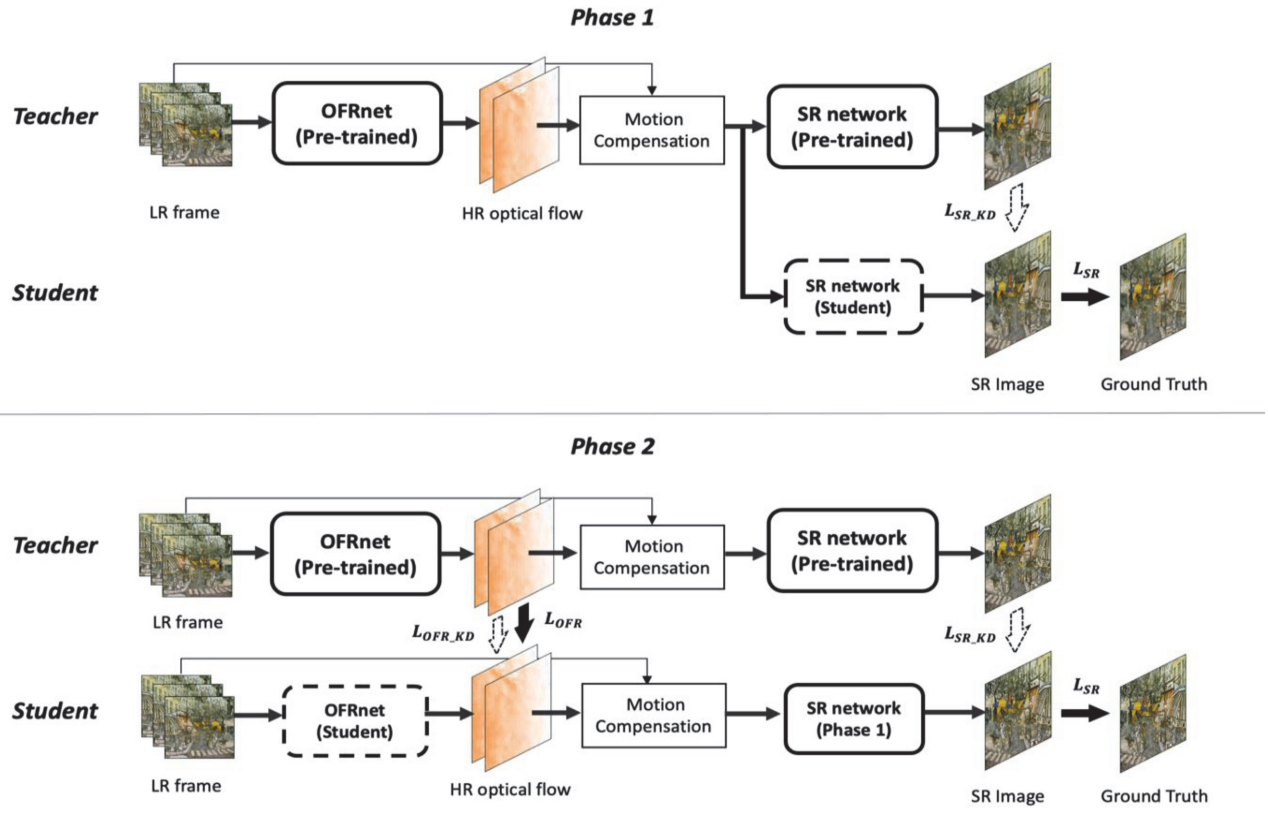

Fig. 2. Proposed Knowledge distillation method upon SOF-VSR.

$\mathcal{L}_{SR\_KD}$ is the L1 loss of the SR outputs of the student and teacher.

$$\mathcal{L}_{SR\_KD} = \| I_{Teacher} - I_{Student} \|_1. \quad (5)$$

Therefore, the first phase loss, $\mathcal{L}_{Phase1}$, is as follows:

$$\mathcal{L}_{Phase1} = \alpha\mathcal{L}_{SOF} + (1-\alpha)\mathcal{L}_{SR\_KD}. \quad (6)$$

In the second phase, we optimized only the student's OFRnet while freezing the student's SRnet. Because the student's SRnet was trained using the teacher's OFRnet in the previous step, the student's OFRnet should also generate an optical flow similar to the HR optical flow generated by the teacher's OFRnet.

Therefore, we added $\mathcal{L}_{OFR\_KD}$ to the student's OFRnet to learn the teacher's optical flow. $F_{Teacher,i}$ and $F_{Student,i}$ are denoted as the i-th level optical flow generated by the student and teacher. $\mathcal{L}_{OFR\_KD}$ was calculated as the sum of the MSE loss of the $F_{Teacher,i}$ and $F_{Student,i}$ at all levels.

$$\mathcal{L}_{OFR\_KD} = \sum_{i\in[1,3]} MSE\left(F_{Teacher,i}, F_{Student,i}\right) \quad (7)$$

As in the first phase, $\mathcal{L}_{SR\_KD}$ was used together with $\mathcal{L}_{OFR\_KD}$. Therefore, the second-phase loss is expressed as follows:

$$\mathcal{L}_{Phase2} = \alpha\mathcal{L}_{SOF} + (1-\alpha)\left(\mathcal{L}_{OFR_{KD}} + \mathcal{L}_{SR_{KD}}\right) \quad (8)$$

IV. EXPERIMENTS

In this section, we detailed the datasets we used and how the student model and the training setting were constructed. We compared the performance of the teacher baseline model with the experimental results of the general KD method and our method.

A. Experiment Details

1) Datasets and Metrics

We demonstrated the effectiveness of this method on various SR datasets. SOF-VSR showed good performance for the Vid4 and DAVIS-10 datasets. Therefore, we used DAVIS-2017 train-dataset to train the network and the Vid4 and DAVIS-2017 test dataset to evaluate the network. The down sampling operation used for generating LR frames was bicubic interpolation with a scaling factor of four.

2) Student Model

We used the original SOF-VSR as the teacher model. There was a total of 1 M parameters, 0.413 M for OFRnet

Table 1. Performance comparison of the proposed method with the anchor, SOF-VSR

	Vid4				DAVIS-2017			
	PSNR(↑)		SSIM		PSNR(↑)		SSIM	
	Average(↑)	Max	Average(↑)	Max	Average(↑)	Max	Average(↑)	Max
SOF-VSR [7]	25.6578	28.8419	0.7521	0.8808	32.1800	43.1598	0.8632	0.9801
Student w/o KD	25.1063	28.2611	0.7176	0.8682	31.7056	42.6076	0.8512	0.9766
Proposed Phase 1	25.6781	28.9139	0.7512	0.8813	32.1474	43.3542	0.8625	0.9807
Proposed Phase 2	25.6574	28.8806	0.7504	0.8808	32.1427	43.4044	0.8624	0.9807
Proposed (Phase 1+Phase 2)	25.6983	28.9384	0.7525	0.8820	32.1953	43.4571	0.8633	0.9811

Table 2. Model parameter size (unit: M)

	OFRnet	SRnet	Total
SOF-VSR [7]	0.413	0.587	1 (100%)
Student w/o KD	0.259	0.371	0.630 (63%)
Proposed Phase 1	0.413	0.371	0.784 (78%)
Proposed Phase 2	0.259	0.371	0.630 (63%)
Proposed (Phase 1+Phase 2)	0.259	0.371	0.630 (63%)

and 0.587 M for SRnet. OFRnet and SRnet contained six and eight ResBlocks, respectively, and we used the same ResBlock architecture, which was composed of 320 convolution layer channels. Our student networks reduced the number of channels in the ResBlock from 320 to 240 while maintaining the number of ResBlocks. Therefore, the student parameters decreased to 63% compared with that of the teacher. The OFRnet and SRnet had 0.259 M and 0.371 M parameters, which were 62.7% and 63.2%, respectively, of the teacher.

3) Parameter Setting

In training, we trained for a total of 400,000 iterations, 200,000 iterations, with in each phase and 64 batch sizes. The networks used Adam optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\varepsilon = 1e^{-8}$. The learning rate was initially 0.001 and was decreased to half per 80,000 iterations. The KD ratio parameter, α, was set to 0.9.

B. Performance Comparison

To demonstrate that the two-phase approach of our method is effective, we compared it with models that learn only one phase independently as shown in Table 1.

The Phase 1 model shares the teacher's OFRnet of the previous stage prior to applying Phase 2. The Phase 2 model distills both SRnet and OFRnet without freezing using $\mathcal{L}_{Phase2}$. We also compared teacher and student models without KD. For the evaluation, we measured the mean and maximum PSNR and SSIM for the Vid4 and DAVIS-2017 datasets. As shown in Table 2, the student without KD significantly reduced the PSNR and SSIM compared to the teacher model in both datasets. On the contrary, the models applied with Phases 1 and 2 independently showed similar results for both PSNR and SSIM compared with the teacher for the Vid4 datasets. However, for the DAVIS-2017 dataset, both PSNR and SSIM showed inferior results compared with the teacher. In contrast, our method with both Phase 1 and Phase 2 showed the highest PSNR and SSIM values in both datasets. The results show that our network outperformed the teacher network, even though its parameters are only 63% of those of the teacher network.

V. CONCLUSION

In this study, we proposed a two-phase KD framework for a VSR network equipped with an optical flow-based preprocessing network. We applied our method to the state-of-the-art SOF-VSR model, which is a representative optical-flow-based SR model. Compared with the KD method, which does not consider a preprocessing network, our method exhibited better PSNR results on an extensive video dataset. Our method also reduced the original network parameter to 63%, which would be beneficial to be adopted to battery-constrained devices. In addition, we believe that our approach can be easily adapted to other optical flow-based and preprocessing networks.

AKNOWLEDGMENTS

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (No. 2020R1I1A3072227).

Conflict of Interest(COI)

The authors have declared that no competing interests exist.

REFERENCES

1. C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 2, pp. 295-307, 2016.
2. B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, Honolulu, HI, 2017, pp. 136-144.
3. J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, 2016, pp. 1646-1654.
4. Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," in *Proceedings of the European Conference on Computer Vision (ECCV)*, Munich, Germany, 2018, pp. 294-310.
5. Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, 2018, pp. 2472-2481.
6. C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, et al., "Photo-realistic single image super-resolution using a generative adversarial network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, HI, 2017, pp. 105-114.
7. L. Wang, Y. Guo, Z. Lin, X. Deng, and W. An, "Learning for video super-resolution through HR optical flow estimation," in *Computer Vision–ACCV 2018*. Cham, Germany: Springer, 2019, pp. 514-529.
8. X. Wang, K. C. Chan, K. Yu, C. Dong, and C. Change Loy, "EDVR: video restoration with enhanced deformable convolutional networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, Long Beach, CA, 2019, pp. 1954-1963.
9. Y. Jo, S. W. Oh, J. Kang, and S. J. Kim, "Deep video super-resolution network using dynamic upsampling filters without explicit motion compensation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, 2018, pp. 3224-3232.
10. J. Caballero, C. Ledig, A. Aitken, A. Acosta, J. Totz, Z. Wang, and W. Shi, "Real-time video super-resolution with spatio-temporal networks and motion compensation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, HI, 2017, pp. 2848-2857.
11. X. Tao, H. Gao, R. Liao, J. Wang, and J. Jia, "Detail-revealing deep video super-resolution," in *Proceedings of the IEEE International Conference on Computer Vision*, Venice, Italy, 2017, pp. 4482-4490.
12. M. S. Sajjadi, R. Vemulapalli, and M. Brown, "Frame-recurrent video super-resolution," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, 2018, pp. 6626-6634.
13. T. H. Kim, M. S. Sajjadi, M. Hirsch, and B. Scholkopf, "Spatio-temporal transformer network for video restoration," in *Proceedings of the European Conference on Computer Vision (ECCV)*, Munich, Germany, 2018, pp. 111-127.
14. W. Li, X. Tao, T. Guo, L. Qi, J. Lu, and J. Jia, "MuCAN: multi-correspondence aggregation network for video super-resolution," in *Computer Vision–ECCV 2020*. Cham, Switzerland: Springer, 2020 pp. 335-351.
15. A. Romero, N. Ballas, S. E. Kahou, A. Chassang, C. Gatta, and Y. Bengio, "Fitnets: hints for thin deep nets," 2015 [Online]. Available: https://arxiv.org/abs/1412.6550.
16. W. Lee, J. Lee, D. Kim, and B. Ham, "Learning with privileged information for efficient image super-resolution," in *Computer Vision–ECCV 2020*. Cham, Switzerland: Springer, 2020 pp. 465-482.
17. Z. Xiao, X. Fu, J. Huang, Z. Cheng, and Z. Xiong, "Space-time distillation for video super-resolution," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Virtual Event, 2021, pp. 2113-2122.
18. Z. He, T. Dai, J. Lu, Y. Jiang, and S. T. Xia, "FAKD: feature-affinity based knowledge distillation for efficient image super-resolution," in *Proceedings of 2020 IEEE International Conference on Image Processing (ICIP)*, Abu Dhabi, UAE, 2020, pp. 518-522.
19. A. Dosovitskiy, P. Fischer, E. Ilg, P. Hausser, C. Hazirbas, V. Golkov, P. van der Smagt, D. Cremers, and T. Brox, "FlowNet: learning optical flow with convolutional networks," in *Proceedings of the IEEE International Conference on Computer Vision*, Santiago, Chile, 2015, pp. 2758-2766.
20. D. Sun, X. Yang, M. Y. Liu, and J. Kautz, "PWC-Net: CNNs for optical flow using pyramid, warping, and cost volume," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, 2018, pp. 8934-8943.
21. T. W. Hui, X. Tang, and C. C. Loy, "LiteFlowNet: a lightweight convolutional neural network for optical flow estimation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, 2018, pp. 8981-8989.
22. A. Ranjan and M. J. Black, "Optical flow estimation using a spatial pyramid network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, HI, 2017, pp. 2720-2729.
23. G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," 2015 [Online]. Available: https://arxiv.org/abs/1503.02531.

Jungwon Lee

Jungwon Lee received B.S. in computer science and engineerirng from Kyungpook National University, Daegu, South Korea in 2023. His research interests include video super resolution, knowledge distillation and deep learning.

Sang-hyo Park https://orcid.org/0000-0002-7282-7686

Sang-hyo Park received the B.S. and Ph.D. degrees in computer engineering and computer science from Hanyang University, Seoul, South Korea, in 2011 and 2017, respectively. From 2017 to 2018, he held a postdoctoral position with the Intelligent Image Processing Center, Korea Electronics Technology Institute, and a Research Fellow with the Barun ICT Research Center, Yonsei University in 2018. From 2019 to 2020, he held a postdoctoral position with the Department of Electronic and Electrical Engineering, Ewha Womans University. Since 2020, he has been an Assistant Professor with the School of Computer Science and Engineering, Kyungpook National University, Daegu, South Korea. His research interests include HEVC, VVC, encoding complexity, immersive video, and deep learning.