



Jiang, Y., Nawala, J. T., Zhang, F., & Bull, D. R. (2024). Compressing Deep Image Super-Resolution Models. In *2024 Picture Coding Symposium (PCS)* (Picture Coding Symposium, PCS). Institute of Electrical and Electronics Engineers (IEEE).
<https://doi.org/10.1109/PCS60826.2024.10566374>

Peer reviewed version

License (if available):
CC BY

Link to published version (if available):
[10.1109/PCS60826.2024.10566374](https://doi.org/10.1109/PCS60826.2024.10566374)

[Link to publication record on the Bristol Research Portal](#)
PDF-document

This is the accepted author manuscript (AAM) of the article which has been made Open Access under the University of Bristol's Scholarly Works Policy. The final published version (Version of Record) can be found on the publisher's website. The copyright of any third-party content, such as images, remains with the copyright holder.

University of Bristol – Bristol Research Portal

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/brp-terms/>

Compressing Deep Image Super-resolution Models

Yuxuan Jiang, Jakub Nawala, Fan Zhang, and David Bull
Visual Information Laboratory, University of Bristol, Bristol, BS1 5DD, UK
{yuxuan.jiang, jakub.nawala, fan.zhang, dave.bull}@bristol.ac.uk

Abstract—Deep learning techniques have been applied in the context of image super-resolution (SR), achieving remarkable advances in terms of reconstruction performance. Existing techniques typically employ highly complex model structures which result in large model sizes and slow inference speeds. This often leads to high energy consumption and restricts their adoption for practical applications. To address this issue, this work employs a three-stage workflow for compressing deep SR models which significantly reduces their memory requirement. Restoration performance has been maintained through teacher-student knowledge distillation using a newly designed distillation loss. We have applied this approach to two popular image super-resolution networks, SwinIR and EDSR, to demonstrate its effectiveness. The resulting compact models, SwinIRmini and EDSRmini, attain an 89% and 96% reduction in both model size and floating-point operations (FLOPs) respectively, compared to their original versions. They also retain competitive super-resolution performance compared to their original models and other commonly used SR approaches. The source code and pre-trained models for these two lightweight SR approaches are released at <https://pikapi22.github.io/CDISM/>.

Index Terms—Image super-resolution, complexity reduction, model compression, knowledge distillation

I. INTRODUCTION

Image super-resolution (SR) has attracted growing research interest over the past few decades. It represents the task of generating a high spatial resolution image from a low-resolution version, with the aim of reconstructing with optimal perceptual quality, accurately recovering spatial detail. It has been widely employed in various image and video processing applications including medical imaging, image restoration and enhancement, and picture coding [1–4]. SR can be conventionally achieved by using various linear and non-linear filters [5–7], while learning-based SR has become more popular recently due to its superior reconstruction performance.

Learning-based SR algorithms can be divided into two major categories: CNN-based [8–13] and Transformer-based approaches [14–17]. The former commonly leverages a Convolutional Neural Network (CNN), which typically comprises a number of consecutive convolutional layers connected with activation functions. Typical examples include SRCNN [8], VDSR [9], EDSR [10] and RT4KSR [18]. More recently, with the invention of Vision Transformer (ViT) networks [19] which exploit self-attention mechanisms to capture more context interaction information, reconstruction performance has been further improved by integrating ViT into the SR

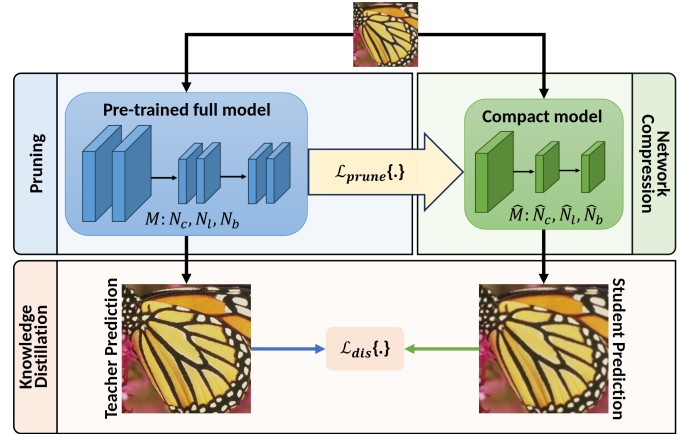


Fig. 1: The proposed workflow for compressing SR networks. The pruning process obtains the corresponding compact model \hat{M} from its intricate counterpart M . Here N_c , N_l and N_b represent the channel number, the layer number and the block number, respectively. \hat{N}_c , \hat{N}_l and \hat{N}_b denote the same but for the model after compression. $\mathcal{L}_{prune}\{\cdot\}$ and $\mathcal{L}_{dis}\{\cdot\}$ are the loss functions used in pruning and knowledge distillation processes.

framework (notable contributions include SwinIR [16] and Swin2SR [17]).

Although the aforementioned learning-based SR algorithms have significantly improved performance compared to conventional filters, the former is often associated with high computational complexity, leading to slow run-time and large memory requirements. For example, one of the best performers, SwinIR, requires approximately 11.8M parameters [16], while the widely used EDSR approach is based on a model with 43M parameters [10]. These highly complex models demand significant computational resources during their development and inference, restricting their adoption in practical applications. Hence, there is a pressing necessity to perform complexity reduction for these approaches while maintaining their excellent SR performance.

In recent reports, complexity reduction has been performed through sophisticated manual modifications of network architectures [18, 20, 21]. It has also been achieved using model compression techniques [22, 23] such as model pruning [24, 25] and knowledge distillation (KD) [26]. Pruning approaches eliminate redundancies within the original model and compress it into a more compact form, resulting in a substantial reduction in model size. KD, on the other hand, involves transferring knowledge from a large teacher model (e.g., the original version) to a smaller student model (e.g.,

The authors appreciate the funding from Netflix Inc., University of Bristol, and the UKRI MyWorld Strength in Places Programme (SIPF00006/1).

the compact one). These techniques have been proven to be effective when adopted separately for various image processing applications, such as image classification [27–29], image restoration [30] and video compression [31]. However, only a few works (mainly for video frame interpolation [32, 33]) have investigated the combination of model pruning and knowledge distillation.

In this context, with the focus on image super-resolution, we propose a new network compression framework, illustrated in Fig. 1, which integrates both model pruning and knowledge distillation to significantly reduce the model complexity while maintaining SR performance. This approach first applies sparsity inducing optimisation to the original network before compressing it into a compact model based on a new parameter distribution analysis method. The performance of the compressed model is then further improved through knowledge distillation with a modified loss. To the best of our knowledge, this is the first attempt that combines both pruning and distillation techniques for SR model compression. We have applied this to two popular SR models, EDSR and SwinIR, and their resulting compact models achieve significant model size and FLOPs (up to 96%) reductions, significantly outperforming other SR methods with similar complexity figures.

II. PROPOSED ALGORITHM

Fig. 2 illustrates a commonly employed high-level network architecture for image super-resolution. It comprises three integral components: a shallow feature extraction module, a deep feature extraction module, and an image reconstruction module. The shallow feature extraction module usually employs a small number of convolutional layers to extract shallow features containing essential low-frequency information. The core of the network lies within the deep feature extraction module, which meticulously obtains intricate and high-level features, collectively holding a pivotal role in shaping the system’s overall performance and capabilities. Ultimately, both shallow and deep features converge within the reconstruction module to facilitate the creation of high-quality image reconstructions. Since the network structure (mainly in the deep feature extraction module) consists of a stack of basic processing blocks, its complexity and performance are intimately tied to the number of channels N_c , the layer counts (excluding the convolutional layer before the output) N_l within each block, and the total number of such blocks N_b .

The proposed complexity reduction workflow for deep SR models is shown in Fig. 1, which consists of three primary stages: model pruning, network compression, and knowledge distillation. The algorithm underpinning this is detailed in the following subsections.

A. Model Pruning

In order to obtain a condensed version of a model, we start with the original pre-trained model, and fine-tune it using the following loss function,

$$\mathcal{L}_{prune} = \sqrt{(I_{SR} - I_{gt})^2 + \epsilon^2} + \lambda \|\theta\|_1, \quad (1)$$

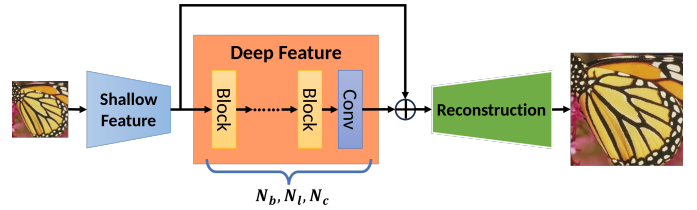


Fig. 2: The basic blueprint of a modern image SR network.

where I_{gt} represents the ground-truth target image, I_{SR} corresponds to the super-resolved output, ϵ is set to 10^{-3} , and $\lambda > 0$ is the regularisation constant, set to 10^{-4} following [33]. The parameters of the initial model are represented by θ , and $\|\cdot\|_1$ refers to the $L1$ norm regularisation term, which serves to promote network sparsity, as discussed in [34]. Such sparsity information can be used as a guide to removing redundant layers in the network. We adopt the OBProx-SG [34] solver to perform the optimisation. Eventually, a density ratio d , the ratio of non-zero parameters, is obtained. In contrast to the approach taken in [32, 33], where they computed the density ratio for each layer, we calculate d for the whole deep feature module. We then use this ratio in the next step of complexity reduction.

B. Network Compression

Contrary to the method described in [33], where the compression only focuses on the channel level, here we further analyse three hyperparameters N_c (the number of channels), N_l (the number of layers), and N_b (the number of deep feature extraction blocks). Specifically, we consider the total number of model parameters for the deep feature extraction module, P_{DF} , which can be approximately¹ written as:

$$P_{DF} \approx k N_b (N_l + 1) N_c^2. \quad (2)$$

Here, k is an approximate constant for a specific model structure. With the density ratio d calculated during the model pruning stage, we can obtain a compact model by updating these three key hyperparameters to satisfy the following:

$$\frac{\hat{P}_{DF}}{P_{DF}} \approx d, \quad (3)$$

in which \hat{P}_{DF} represents the total number of parameters of the compact deep feature extraction module,

$$\hat{P}_{DF} \approx k \hat{N}_b (\hat{N}_l + 1) \hat{N}_c^2. \quad (4)$$

If we substitute eq. (2) and (4) into eq. (3), we get

$$\frac{\hat{N}_b (\hat{N}_l + 1) \hat{N}_c^2}{N_b (N_l + 1) N_c^2} \approx d. \quad (5)$$

It is noted that the influence of these three hyperparameters on the model performance and size is not identical—the reduction of N_b and N_l leads to a greater model performance reduction,

¹If the block structure is unique, it needs to be analysed on a case-by-case basis.

while N_c has a greater influence on model size. Based on these observations, we achieve model compression by empirically assigning

$$\frac{\hat{N}_b}{N_b} \approx \sqrt[6]{d} \text{ and } \frac{\hat{N}_l + 1}{N_l + 1} \approx \sqrt[6]{d}, \quad \hat{N}_b, \hat{N}_l \in \mathbb{N}^+ \quad (6)$$

After obtaining \hat{N}_b and \hat{N}_l , \hat{N}_c can be derived from

$$\frac{\hat{N}_c}{N_c} \approx \sqrt[3]{d \frac{N_b(N_l + 1)}{\hat{N}_b(\hat{N}_l + 1)}}, \quad \hat{N}_c \in \mathbb{N}^+ \quad (7)$$

After that, we accordingly adjust the shallow feature extractor and reconstruction module based on \hat{N}_c to maintain the network integrity.

C. Knowledge Distillation

After obtaining the compact model, using a knowledge distillation approach similar to [33], we further improve the pruned model’s performance by employing the pre-trained original model as a “teacher” to instruct the training process. Specifically, the total loss \mathcal{L}_{total} for knowledge distillation is given as follows,

$$\mathcal{L}_{total} = \alpha \mathcal{L}_{stu}(I_{stu}, I_{gt}) + \mathcal{L}_{dis}(I_{stu}, I_{tea}), \quad (8)$$

where \mathcal{L}_{stu} denotes the original loss between the ground truth I_{gt} and the student model’s prediction I_{stu} (which was used for training the original full SR model), α is a tunable weight, and \mathcal{L}_{dis} stands for the loss between the student I_{stu} and the teacher’s predictions I_{tea} . In this work, the distillation loss, \mathcal{L}_{dis} , is calculated as below, which is inspired by [17, 33, 35]:

$$\begin{aligned} \mathcal{L}_{dis}(I_{stu}, I_{tea}) &= \mathcal{L}_{Lap}(I_{stu}, I_{tea}) \\ &+ \mathcal{L}_{Lap}(HF(I_{stu}), HF(I_{tea})), \end{aligned} \quad (9)$$

where \mathcal{L}_{Lap} is the Laplacian loss [35] and $HF(\cdot)$ represents the high-frequency features extracted by a 5×5 Gaussian blur kernel function. The high-frequency feature loss is included to further enhance the sharpness and overall quality of the output.

III. RESULTS AND DISCUSSION

In order to demonstrate the effectiveness of our complexity reduction workflow, we applied it to two popular image super-resolution models: EDSR [10] and SwinIR [16]. The former is a widely used CNN-based model, while SwinIR is Transformer-based and offers the state-of-the-art SR performance. The compact models are obtained from their existing lightweight versions, *EDSR_baseline* and *SwinIR_lightweight (LW)*, as reported by their original authors. We refer to our compact models as *EDSRmini* and *SwinIRmini*, respectively.

A. Experimental Setup

We use the same training set, DIV2K [36], for both model pruning and knowledge distillation, as in [10, 16]. Specifically, similarly to [32], 100 images from the DIV2K [36] dataset are used to conduct model pruning, and the whole DIV2K is employed in the knowledge distillation step. During training, we use the AdaMax optimizer with $\beta_1 = 0.9$ and $\beta_2 = 0.99$. The hyperparameter α in eq. (8) is set to 0.1, following [33].

TABLE I: Performance comparison between our approaches and several other methods on two benchmark datasets. PSNR/SSIM values on the Y channel are reported on each dataset. #Ps and FLOPs stand for the total number of network parameters and floating-point operations, respectively. FLOPs are measured under the setting of upscaling SR images to 1280×720 resolution on the $\times 2$ scale.

| Model | #Ps(M) | FLOPs(G) | Set5 | Set14 |
|--------------------------|--------------|-------------|----------------------------------|---------------------|
| | | | PSNR \uparrow /SSIM \uparrow | PSNR/SSIM |
| SRCNN [8] | 0.057 | 13.2 | 36.66/0.9542 | 32.45/0.9067 |
| LapSRN [40] | 0.251 | 29.9 | 37.52/0.9590 | 32.99/0.9124 |
| CARN [41] | 1.59 | 222.8 | 37.76/0.9590 | 33.52/0.9166 |
| LatticeNet [42] | 0.756 | 169.5 | 38.15/0.9610 | 33.78/0.9193 |
| IMDN [43] | 0.694 | 158.8 | 38.00/0.9605 | 33.63/0.9177 |
| RLFN-S [25] | 0.454 | 68.1 | 38.05/0.9607 | 33.68/0.9172 |
| FALSR-B [21] | 0.326 | 74.7 | 37.61/0.9585 | 33.29/0.9143 |
| RT4KSR-XL [18] | 0.092 | n/a | 36.83/0.9545 | 33.46/0.9197 |
| EDSR [10] | 43 | 9387.0 | 38.11/0.9601 | 33.92/0.9195 |
| EDSR_baseline [10] | 1.37 | 316.3 | 37.99/0.9604 | 33.57/0.9175 |
| EDSRmini (ours) | 0.049 | 11.7 | 37.67/0.9597 | 33.21/0.9141 |
| SwinIR [16] | 11.8 | 2301.0 | 38.35/0.9620 | 34.14/0.9227 |
| SwinIR_LW [16] | 0.878 | 195.6 | 38.14/0.9611 | 33.86/0.9206 |
| SwinIRmini (ours) | 0.099 | 22.0 | 37.88/0.9601 | 33.60/0.9187 |

For compressing the SwinIR model, by pruning the SwinIR_LW network for 100 epochs, a density of approximately 0.089 is obtained and considered as the compression rate. Based on this, \hat{N}_c , \hat{N}_l , and \hat{N}_b for SwinIRmini are calculated according to eq. (6) and (7), and their values are 24, 4, and 3, respectively. The resulting total number of parameters for the SwinIRmini is 98.8K (878K for SwinIR_LW). Similarly, by optimising the EDSR_baseline network, a density of approximately 0.03 is achieved, and \hat{N}_c , \hat{N}_l and \hat{N}_b for EDSRmini are calculated as 16, 1, and 8, respectively, with 49.6K parameters in total (1.37M for EDSR_baseline).

The evaluation is performed on Set5 [37] and Set14 [38] databases, which are commonly employed to benchmark super-resolution models. Two widely used quality metrics, peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) [39], are used to measure the model performance. All experiments are conducted using an NVIDIA RTX 3090 GPU.

B. Quantitative Evaluation

The results of the quantitative comparison between our approach and a number of existing deep SR approaches are summarised in TABLE I. It is noted that the performance results and complexity figures for all the benchmark results are taken from their original publications. It can be observed that the resulting compact models, EDSRmini and SwinIRmini, offer much smaller model sizes, 4% and 11% compared to their baseline models, EDSR_baseline and SwinIR_LW, respectively. They also require fewer FLOPs, 4% and 11% of their baselines. However, the average performance losses are minimal, 0.34 dB and 0.26 dB across both databases for EDSRmini and SwinIRmini, compared to their original models. Fig. 4 provides a more intuitive illustration, plotting

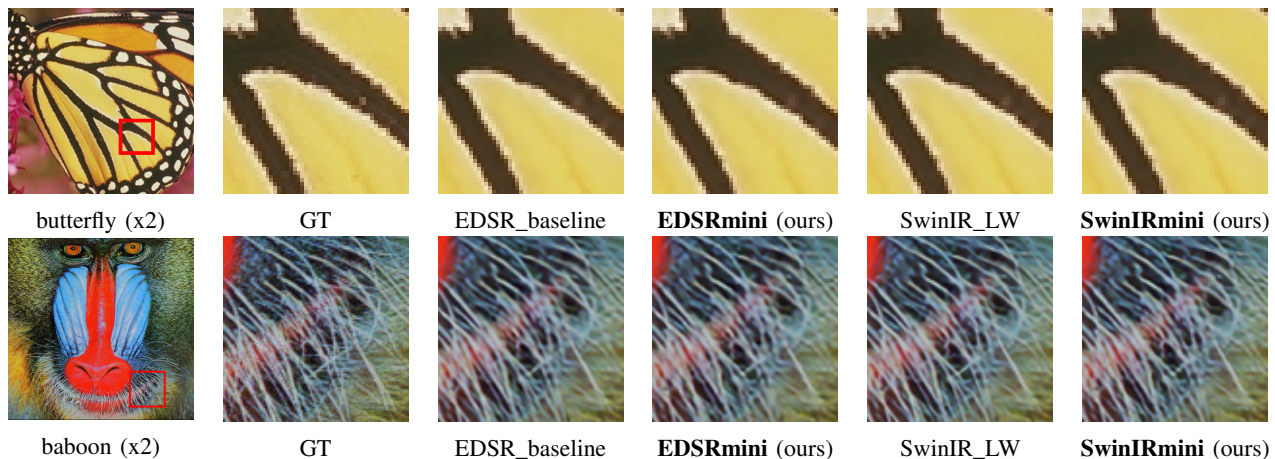


Fig. 3: Example of visual comparison between our compact models and their original counterparts.

the average PSNR values of all the benchmarked SR models against the number of model parameters and FLOPs. It can be observed from both subfigures that the EDSRmini and SwinIRmini achieve a superior trade-off between complexity and performance. For example, SwinIRmini outperforms RT4KSR-XL with a comparable number of parameters by nearly 0.6 dB.

C. Qualitative Evaluation

Examples of SR outputs using both our compact models and their corresponding original versions are shown in Fig. 3 for visual comparison. The results generated by both the compact models and their counterparts are largely indistinguishable, demonstrating the effectiveness of our complexity reduction approach. The proposed method not only significantly reduces the number of parameters and FLOPs, but also preserves the excellent interpolation performance of the original models.

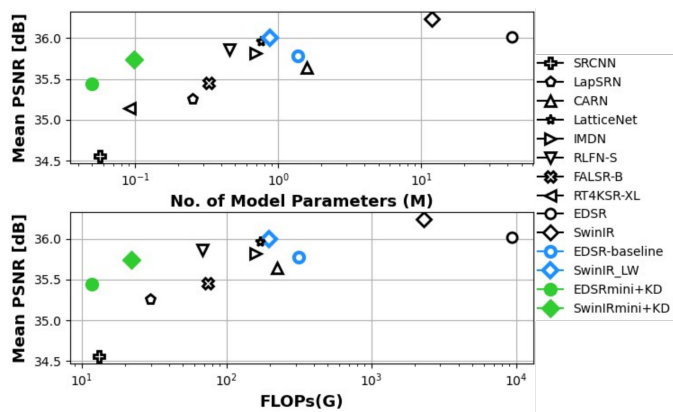


Fig. 4: (Top) Average PSNR scores on Set5 and Set14 for the models presented in Table I versus their corresponding numbers of model parameters. Results for our two compact models are marked with solid symbols. (Bottom) The plot between the performance and FLOPs based on Set5 and Set14.

IV. CONCLUSION

This paper presents a new workflow for compressing image super-resolution (SR) models based on model compression

and knowledge distillation. The proposed approach has been applied to two different SR approaches, EDSR (CNN-based) and SwinIR (Transformer-based), achieving consistent and significant complexity reduction results: up to 96% in terms of model size and FLOPs. At the same time, the competitive performance of their original models has been maintained alongside the reduced complexity, which demonstrates the effectiveness of the proposed workflow. Future endeavours will extend this work to other low level computer vision tasks.

REFERENCES

- [1] Z. Wang, J. Chen, and S. C. Hoi, “Deep learning for image super-resolution: A survey,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 43, no. 10, pp. 3365–3387, 2020.
- [2] D. Bull and F. Zhang, *Intelligent image and video compression: communicating pictures*. Academic Press, 2021.
- [3] M. Afonso, F. Zhang, and D. R. Bull, “Video compression based on spatio-temporal resolution adaptation,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29, no. 1, pp. 275–280, 2018.
- [4] F. Zhang, M. Afonso, and D. R. Bull, “ViSTRA2: Video coding using spatial resolution and effective bit depth adaptation,” *Signal Processing: Image Communication*, vol. 97, p. 116355, 2021.
- [5] J. Yang, J. Wright, T. S. Huang, and Y. Ma, “Image super-resolution via sparse representation,” *IEEE transactions on image processing*, vol. 19, no. 11, pp. 2861–2873, 2010.
- [6] S. Schuler, C. Leistner, and H. Bischof, “Fast and accurate image upscaling with super-resolution forests,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3791–3799, 2015.
- [7] M. Afonso, F. Zhang, A. Katsenou, D. Agrafiotis, and D. Bull, “Low complexity video coding based on spatial resolution adaptation,” in *2017 IEEE International Conference on Image Processing (ICIP)*, pp. 3011–3015, IEEE, 2017.
- [8] 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, 2015.
- [9] 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*, pp. 1646–1654, 2016.

- [10] B. Lim, S. Son, H. Kim, S. Nah, and K. Mu Lee, "Enhanced deep residual networks for single image super-resolution," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pp. 136–144, 2017.
- [11] 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)*, pp. 286–301, 2018.
- [12] D. Ma, F. Zhang, and D. R. Bull, "CVEGAN: a perceptually-inspired gan for compressed video enhancement," *arXiv preprint arXiv:2011.09190*, 2020.
- [13] D. Ma, F. Zhang, and D. R. Bull, "MFRNet: a new CNN architecture for post-processing and in-loop filtering," *IEEE Journal of Selected Topics in Signal Processing*, vol. 15, no. 2, pp. 378–387, 2020.
- [14] J. Liang, J. Cao, Y. Fan, K. Zhang, R. Ranjan, Y. Li, R. Timofte, and L. Van Gool, "VRT: A video restoration transformer," *arXiv preprint arXiv:2201.12288*, 2022.
- [15] Z. Wang, X. Cun, J. Bao, W. Zhou, J. Liu, and H. Li, "Uformer: A general u-shaped transformer for image restoration," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 17683–17693, 2022.
- [16] J. Liang, J. Cao, G. Sun, K. Zhang, L. Van Gool, and R. Timofte, "SwinIR: Image restoration using swin transformer," in *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 1833–1844, 2021.
- [17] M. V. Conde, U.-J. Choi, M. Burchi, and R. Timofte, "Swin2SR: Swin2 transformer for compressed image super-resolution and restoration," in *European Conference on Computer Vision*, pp. 669–687, Springer, 2022.
- [18] E. Zamfir, M. V. Conde, and R. Timofte, "Towards real-time 4k image super-resolution," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1522–1532, 2023.
- [19] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.
- [20] J. Guo, X. Zou, Y. Chen, Y. Liu, J. Liu, Y. Yan, and J. Hao, "AsConvSR: Fast and lightweight super-resolution network with assembled convolutions," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1582–1592, 2023.
- [21] X. Chu, B. Zhang, H. Ma, R. Xu, and Q. Li, "Fast, accurate and lightweight super-resolution with neural architecture search," in *2020 25th International conference on pattern recognition (ICPR)*, pp. 59–64, IEEE, 2021.
- [22] Y. Cheng, D. Wang, P. Zhou, and T. Zhang, "A survey of model compression and acceleration for deep neural networks," *arXiv preprint arXiv:1710.09282*, 2017.
- [23] H. Kirchhoffer, P. Haase, W. Samek, K. Müller, H. Rezazadegan-Tavakoli, F. Cricri, E. B. Aksu, M. M. Hannuksela, W. Jiang, W. Wang, *et al.*, "Overview of the neural network compression and representation (nnr) standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 5, pp. 3203–3216, 2021.
- [24] M. Zhu and S. Gupta, "To prune, or not to prune: exploring the efficacy of pruning for model compression," *arXiv preprint arXiv:1710.01878*, 2017.
- [25] F. Kong, M. Li, S. Liu, D. Liu, J. He, Y. Bai, F. Chen, and L. Fu, "Residual local feature network for efficient super-resolution," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 766–776, 2022.
- [26] G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," *arXiv preprint arXiv:1503.02531*, 2015.
- [27] P. Chen, S. Liu, H. Zhao, and J. Jia, "Distilling knowledge via knowledge review," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5008–5017, 2021.
- [28] Y. Jin, J. Wang, and D. Lin, "Multi-level logit distillation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 24276–24285, 2023.
- [29] L. Beyer, X. Zhai, A. Royer, L. Markeeva, R. Anil, and A. Kolesnikov, "Knowledge Distillation: A good teacher is patient and consistent," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 10925–10934, 2022.
- [30] H. Fang, X. Hu, and H. Hu, "Cross knowledge distillation for image super-resolution," in *Proceedings of the 2022 6th International Conference on Video and Image Processing*, pp. 162–168, 2022.
- [31] H. M. Kwan, G. Gao, F. Zhang, A. Gower, and D. Bull, "HiNeRV: Video compression with hierarchical encoding based neural representation," *arXiv preprint arXiv:2306.09818*, 2023.
- [32] T. Ding, L. Liang, Z. Zhu, and I. Zharkov, "CDFI: Compression-driven network design for frame interpolation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 8001–8011, 2021.
- [33] C. Morris, D. Danier, F. Zhang, N. Anantrasirichai, and D. R. Bull, "ST-MFNet Mini: Knowledge distillation-driven frame interpolation," *arXiv preprint arXiv:2302.08455*, 2023.
- [34] T. Chen, T. Ding, B. Ji, G. Wang, Y. Shi, J. Tian, S. Yi, X. Tu, and Z. Zhu, "Orthant based proximal stochastic gradient method for ℓ_1 ℓ_1 -regularized optimization," in *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2020, Ghent, Belgium, September 14–18, 2020, Proceedings, Part III*, pp. 57–73, Springer, 2021.
- [35] S. Niklaus and F. Liu, "Context-aware synthesis for video frame interpolation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1701–1710, 2018.
- [36] E. Agustsson and R. Timofte, "NTIRE 2017 challenge on single image super-resolution: Dataset and study," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pp. 126–135, 2017.
- [37] M. Bevilacqua, A. Roumy, C. Guillemot, and M. L. Alberi-Morel, "Low-complexity single-image super-resolution based on nonnegative neighbor embedding," 2012.
- [38] R. Zeyde, M. Elad, and M. Protter, "On single image scale-up using sparse-representations," in *Curves and Surfaces: 7th International Conference, Avignon, France, June 24-30, 2010, Revised Selected Papers 7*, pp. 711–730, Springer, 2012.
- [39] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image Quality Assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.
- [40] W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang, "Deep laplacian pyramid networks for fast and accurate super-resolution," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 624–632, 2017.
- [41] N. Ahn, B. Kang, and K.-A. Sohn, "Fast, accurate, and lightweight super-resolution with cascading residual network," in *Proceedings of the European conference on computer vision (ECCV)*, pp. 252–268, 2018.
- [42] X. Luo, Y. Xie, Y. Zhang, Y. Qu, C. Li, and Y. Fu, "LatticeNet: Towards lightweight image super-resolution with lattice block," in *Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXII 16*, pp. 272–289, Springer, 2020.
- [43] Z. Hui, X. Gao, Y. Yang, and X. Wang, "Lightweight image super-resolution with information multi-distillation network," in *Proceedings of the 27th acm international conference on multimedia*, pp. 2024–2032, 2019.