Weight-Sharing Method for Upsampling Layer from Feature Embedding Recursive Block

Jinwoo Hyun[1], **YunKyong Hyon**[2,*] **Mi Ra Lee**[2],
Sunju Lee[2], **Taeyoung Ha**[2], **Young Rock Kim**[1]
[1]Hankuk University of Foreign Studies, Republic of Korea
[2]National Institute for Mathematical Sciences, Republic of Korea
{hyunjw0123, rocky777}@hufs.ac.kr
{hyon, mrlee1012, sjlee, tha}@nims.re.kr

Abstract

In the field of super-resolution, the Laplacian pyramid framework-based model needs to estimate the result of the inverse convolution for upscaling layers. Generally, the transposed convolution is applied to estimate the result close to the inverse convolution. In this process, the transposed convolution can be designed efficiently to reduce the trainable weights. In this study, we propose a new model compression method that replaces the transposed convolution layer by sharing the weights of the convolution layer trained in the feature embedding recursive block. The proposed weight-sharing method effectively reduces training complexity and training time. The experiments demonstrate the results accordingly, even for relatively large image sizes.

1 Introduction

Recently, with the rapid development of deep learning-based image processing technology, research on super-resolution (SR) that converts low-resolution (LR) images into high-resolution (HR) images is actively being conducted. SR technology is gaining importance in various application fields such as medical imaging, satellite imaging, and image restoration. In particular, in practical applications, efficiency as well as performance of SR models is emerging as important issues [1, 2, 3, 4, 5, 6].

To solve the SR problem, various deep learning-based models have been proposed. Among them, models based on the Laplacian Pyramid Framework (LPF) have shown excellent performance through stepwise image resolution [7, 8]. This framework uses a method of decomposing the original image into pyramid structures of multiple resolutions, learning residuals at each level, and finally estimating the HR image. In this process, accurate estimating the inverse convolution plays an important role, and this task can generally be performed using transposed convolution (TConV). However, the inefficiency of TConV is one of the main causes of increasing the complexity and computational cost of the model. Other studies mainly tried to improve performance through deeper and more complex network structures, but this caused problems such as increasing the depth of the model and the weights to be learned [9, 10, 11]. For this reason, the need for model compression has arisen, and for this purpose, it is necessary to design an appropriate and efficient TConV layer.

In this study, we propose a new method to replace the TConV layer by sharing the weights of the convolution (ConV) layer learned in the feature embedding stage. Through this method, we can obtain an effectively compressed model while we could preserve the performance of the existing LPF-based model. In particular, by reusing the ConV layer weights learned through experiments, we

*YunKyong Hyon is the corresponding authors.

38th Conference on Neural Information Processing Systems (NeurIPS 2024).

can reduce the number of unnecessary parameters and maximize computational efficiency, thereby building an efficient model while maintaining similar performance compared to the other models.

2 Related Works

Table 1: Comparison of related works on SR models

Method	Upsampling stratege	Reconstruction	Multi-scale training	GRL	LSC	Depth	Filters	Parameters
VDSR [3]	Pre-upsampling	Direct	✓	✓		20	64	665k
DRCN [4]				✓		20	256	1775k
DRRN [5]			✓	✓	✓	52	128	297k
MDSR [6]	Post-upsampling		✓		✓	162	64	8000k
LapSRN [7]		Progressive		✓		24	64	812k
MS-LapSRN [8]			✓	✓	✓	84	64	222k
CMS-LapSRN (ours)			✓	✓	✓	84	64	185k

Models based on the LPF [12] have been proposed to solve the SR problem. This framework is advantageous in upscaling LR images through progressive upsampling. [13] proposed a model that progressively SR images by applying GANs to each scale of the Laplacian pyramid structure. This model focuses on extracting detailed information of image step by step using the post-upsampling technique, thereby generating SR images similar to HR images. In addition, [14] and [15] proposed LapSRN and MS-LapSRN models, which are deep learning networks utilizing the LPF. These models extract image details by applying Local Skip Connection (LSC) and perform progressive upscaling at each scale. In particular, MS-LapSRN increases the efficiency of learning by sharing weights between blocks as well as between scales, and improves the overall quality of SR through Global Residual Learning (GRL). There are comparison of related works on SR models on the depth, filter size, and the number of parameters for each model in the case of a $4\times$ upscaling in Table 1. Since other models for the target image resolution is available, we just restrict the comparison scale to $4\times$ upscaling.

3 Proposed Methodology

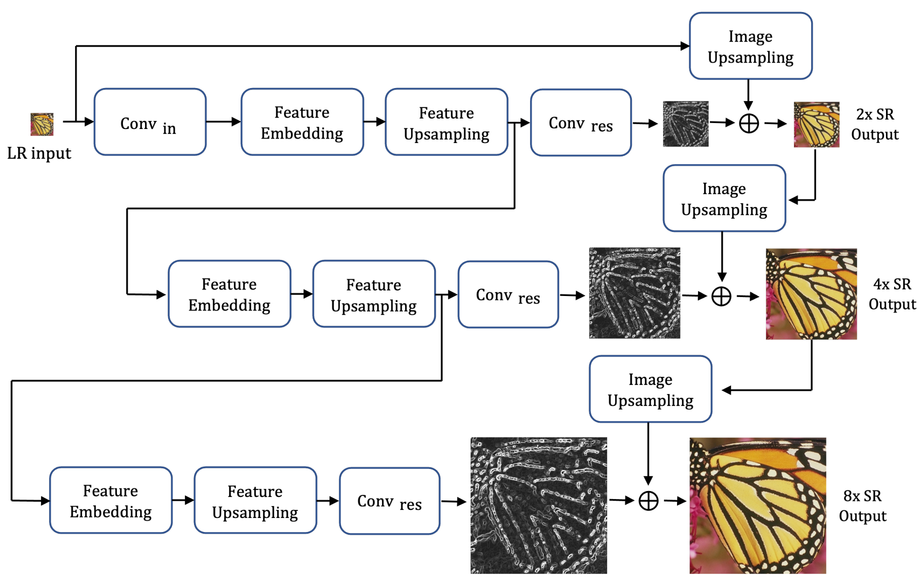

Figure 1: Architecture of the proposed CMS-LapSRN model

We propose a Compressed Multi-Scale Laplacian Pyramid Super-Resolution Network (CMS-LapSRN) model which is described in detail in this section. The proposed network is based on the MS-LapSRN model. It is a LPF-based SR model that progressively scales up LR images by scale and post-upsamples them. To overcome the limitations of existing LPF-based models with increasing weights to be learned according to the size of the SR scale, we propose a model compression method that can achieve compression of the model by reusing the weights learned in the ConV layer and sharing them with the TConV layer. The following subsections describe the model's architecture and method for obtaining model compression.

3.1 Architecture

The architecture of the proposed model consists of networks of appropriate depth according to the SR scale, and the SR task that expands the resolution of the input image to $2\times$ is performed in each layer. For example, when performing SR at the scale of $8\times$, the architecture of the proposed model is composed as shown in Figure 1. Each layer consists of the following components: a Conv_{in} sub-network that extracts high-dimensional feature maps from the input LR image, a Feature Embedding sub-network that extracts high-dimensional nonlinear feature maps, a Feature Upsampling sub-network that upsamples the extracted feature maps, a Conv_{res} sub-network that estimates sub-band residual images, and an Image Upsampling sub-network that upsamples the input image without separate feature extraction. Except for the first layer, the Conv_{in} sub-network is omitted in the remaining layers. Specifically, the feature embedding sub-network consists of r recursive blocks, and each block consists of d ConV layers. We apply GRL to the estimated sub-band residual images through the Image Upsampling sub-network and optimize information flow between scales. Also we apply LSC between each recursive block to improve gradient flow and solve the vanishing gradient problem. In this way, the stability of model learning is improved, the learning speed is accelerated, and the details of the input image are preserved to improve the SR performance. This structure is designed to effectively extract more detailed feature maps.

3.2 Weight-Sharing for model compression

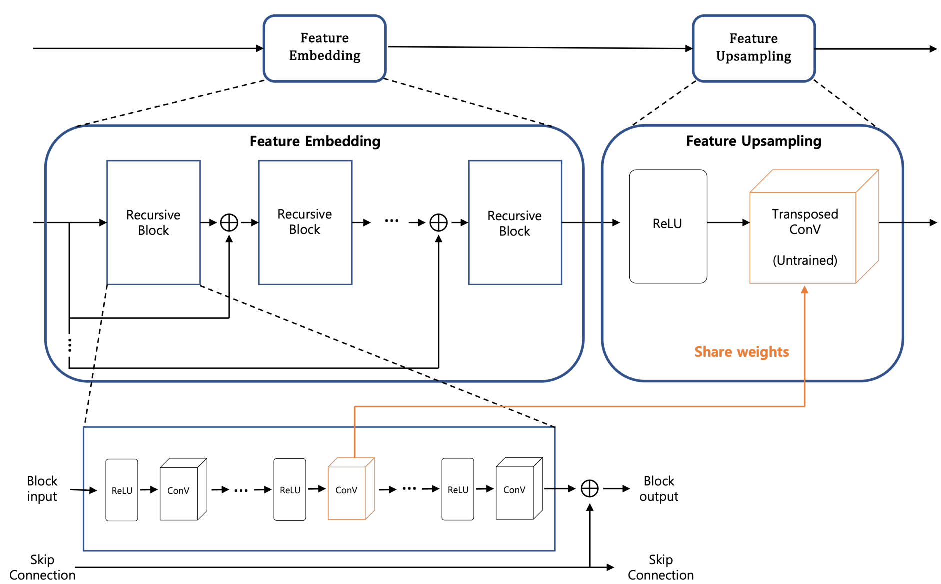

Figure 2: Weight-sharing method for model compression

The initially implemented model has a structural characteristic that the depth of the model and the amount of weights to be learned increase according to the SR scale and the number of feature embedding recursive blocks. To overcome this problem, we applied a model compression technique using a weight-sharing method to improve the efficiency of the model, as shown in Figure 2. First, by sharing the weights of the feature embedding sub-network of each layer with the sub-networks of other layers, as in the basic model, we effectively compressed the amount of weights to be learned according to the SR scale. In addition, by sharing the weights of the recursive blocks in the feature embedding sub-network with other r blocks, we also compressed the amount of learning weights according to the number of feature embedding recursive blocks.

Table 2: Quantitative evaluation of LPF-based SR models

Model	Scale	SET5	SET14	BSDS100	URBAN100	MANGA109
		PSNR / SSIM	PSNR / SSIM	PSNR / SSIM	PSNR / SSIM	PSNR / SSIM
Bicubic	$2\times$	33.69 / 0.931	30.25 / 0.870	29.57 / 0.844	26.89 / 0.841	30.86 / 0.936
LapSRN		**37.24 / 0.957**	**32.78 / 0.910**	**31.78 / 0.892**	**30.60 / 0.911**	**36.73 / 0.971**
MS-LapSRN		36.76 / 0.955	32.57 / 0.908	31.38 / 0.888	29.59 / 0.899	36.16 / 0.970
CMS-LapSRN (1st ConV)		36.65 / 0.954	32.50 / 0.906	31.31 / 0.886	29.44 / 0.895	36.04 / 0.967
CMS-LapSRN (3rd ConV)		36.88 / 0.955	32.61 / 0.908	31.44 / 0.888	29.61 / 0.898	36.31 / 0.970
CMS-LapSRN (5th ConV)		36.46 / 0.954	32.38 / 0.905	31.23 / 0.885	29.34 / 0.894	35.37 / 0.968
Bicubic	$4\times$	28.43 / 0.811	26.01 / 0.704	25.97 / 0.670	23.15 / 0.660	24.93 / 0.790
LapSRN		31.33 / 0.884	27.80 / 0.769	27.31 / 0.724	25.26 / 0.757	29.03 / 0.886
MS-LapSRN		31.48 / 0.885	28.19 / 0.771	27.30 / 0.726	25.35 / 0.761	29.27 / 0.890
CMS-LapSRN (1st ConV)		31.44 / 0.884	**28.20 / 0.771**	27.31 / 0.726	25.36 / 0.761	29.26 / 0.890
CMS-LapSRN (3rd ConV)		31.51 / 0.884	28.19 / 0.770	**27.32 / 0.726**	**25.38 / 0.762**	**29.37 / 0.891**
CMS-LapSRN (5th ConV)		**31.52 / 0.884**	28.19 / 0.769	27.31 / 0.724	25.29 / 0.758	29.17 / 0.888
Bicubic	$8\times$	24.40 / 0.658	23.10 / 0.566	23.67 / 0.548	20.74 / 0.516	21.47 / 0.650
LapSRN		26.19 / 0.750	24.29 / 0.624	**24.61 / 0.585**	21.97 / 0.589	23.72 / 0.741
MS-LapSRN		26.39 / 0.754	24.66 / 0.629	**24.61 / 0.587**	22.09 / 0.596	23.87 / 0.751
CMS-LapSRN (1st ConV)		26.34 / 0.753	24.60 / 0.627	**24.61 / 0.587**	22.08 / 0.596	23.86 / 0.751
CMS-LapSRN (3rd ConV)		**26.40 / 0.756**	**24.68 / 0.629**	24.60 / 0.587	**22.12 / 0.598**	**23.92 / 0.754**
CMS-LapSRN (5th ConV)		26.35 / 0.753	24.60 / 0.627	24.60 / 0.586	22.05 / 0.594	23.82 / 0.749

In addition to the existing model, we compressed the amount of weights to be learned in the feature upsampling sub-network by sharing some of the weights of the recursive block with the TConV layer of the feature upsampling sub-network. Specifically, by sharing the weights of one of the d Conv layers in the recursive block with the TConv layer in the feature upsampling sub-network, we were able to reduce approximately $3.7\text{e}{+}04$ trainable parameters. As a result, our proposed CMS-LapSRN model uses only $1.85\text{e}{+}05$ parameters, which is fewer than the $2.22\text{e}{+}05$ parameters used by the MS-LapSRN model. Additionally, while other models in Table 1 increase the trainable parameters with respect to the upscaling factor, the MS-LapSRN and CMS-LapSRN models preserve $2.22\text{e}{+}05$ and $1.85\text{e}{+}05$ parameters, respectively, independent to the upscaling factor.

4 Experimental Results

Data Preparation & Training Details To train the proposed model, the DIV2K dataset [16] was collected and preprocessed. In [14, 17, 18], data augmentation techniques such as scaling, rotation, and flipping were applied. The training process was performed on a system equipped with two Intel Xeon Gold 5220R CPUs (676 GB RAM) and three NVIDIA RTX A6000 GPUs with 48 GB memory each. We used batch size 16 and patch size 512×512. The architecture of the proposed model uses 64 filters in all ConV layers except the Conv_{in} layer and the image upsampling subnetwork, and applies 4×4 filters to the ConV layer and the TConV layer. In addition, following the method proposed in [15], the feature Embedding subnetwork is composed of $r = 8$ recursive blocks, and each block is efficiently composed of $d = 5$ ConV layers. The model is optimized using the Adam optimizer with weight decay $1\text{e}{-}4$, and the Charbonnier loss function is applied to progressively improve high-frequency residual prediction. This training process can be explained that the model is trained in a multi-scale manner.

Comparisons The proposed model is compared with existing LPF-based SR models using benchmark datasets (SET5 [19], SET14 [20], BSDS100 [21], URBAN100 [22], and MANGA109 [23]). The performance of each model is quantitatively compared using image quality metrics such as PSNR and SSIM. The results for $2\times$, $4\times$, and $8\times$ are presented in Table 2. In addition, the results are presented visually in Figure 3 and Figure 4 to qualitatively compare the $8\times$ performance of each model.

We conducted experiments by selecting the weights of the first, third, and fifth ConV layers among the $d = 5$ ConV layers of each block to efficiently share some of the weights of the recursive block with the TConV layer of the feature upsampling subnetwork. The experimental results are presented in Figure 3, 4, and Table 2, denoted as CMS-LapSRN (1st ConV), CMS-LapSRN (3rd ConV), and CMS-LapSRN (5th ConV), respectively.

According to the comparison results, we were able to reduce the number of trainable parameters and maximize computational efficiency by sharing the weights of the third ConV layer of the recursive

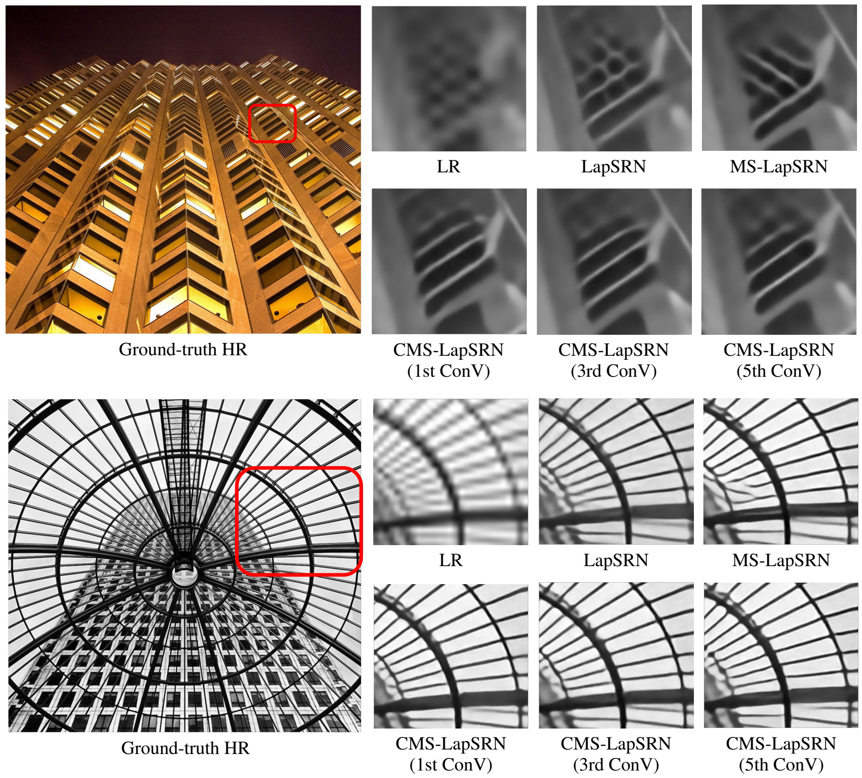

Figure 3: Qualitative evaluation of LPF-based SR models for $8\times$ SR on the URBAN100 datasets.

Figure 4: Qualitative evaluation of LPF-based SR models for $8\times$ SR on the MANGA109 datasets.

block with the TConV layer of the feature upsampling subnetwork, and also confirmed that it showed better or similar performance compared to the existing model. Furthermore, based on the results for the URBAN100 dataset in Figure 3, our proposed CMS-LapSRN model detects topological structures such as grids and orthogonal structures more precisely compared to other models. Similarly by the results for the MANGA109 dataset in Figure 4, the CMS-LapSRN model removes noise between objects more effectively and captures the structure of letters more accurately than other models. In other words, the proposed model has the smallest number of trainable parameters among the models trained with the multi-scale method introduced in the related works, and also has the smallest number of trainable parameters among the models applying GRL and LSC. Despite being a compressed model, it achieved comparable performance at the $2\times$ scale and even better results at the $4\times$ and $8\times$ scales.

Specifically, since the inference time of the model depends on its depth, it shows a similar inference time to the MS-LapSRN model. However, as training time is dependent on the number of parameters, the CMS-LapSRN model, with approximately $3.7\text{e}{+}04$ fewer parameters at $8\times$ upscaling factor, reduces training time by around 21,700 seconds (around 6 hours). Furthermore, despite the reduced training time and model complexity, it maintains comparable or even improved performance. This reduction in training time and complexity through weight-sharing not only accelerates model deployment but also reduces more computational expenses.

5 Conclusion

This paper presents a novel weight-sharing method for upsampling layers in SR models, addressing a remedy for traditional TConV layers. The proposed CMS-LapSRN model has the smallest number of trainable parameters in this study. Despite being a compressed model, CMS-LapSRN performs better or similar than the others, especially in the highest scale. The weight-sharing for upsampling layer shows an efficient method for replacing the TConV approximating inverse convolution . Experiment results have shown that the upsampling layer in the LPF for SR can be replaced by a block in a feature embedding recursive block. Nevertheless, the upsampling layer in the Laplace Pyramid network is initially proposed is an inverse convolution.

Therefore, we expected that finding a more appropriate way to replace the inverse convolution could improve the prediction performance, hence we plan to do as the next research followed by this study. Moreover, another interesting application is to use the trainable CMS-LapSRN to medical imaging datasets as its external validation or generalization.

Acknowledgements

Jinwoo Hyun and Young Rock Kim were supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2021R1A2C1011467). Young Rock Kim was supported by Hankuk University of Foreign Studies Research Fund. YunKyong Hyon, Mi Ra Lee, Sunju Lee, and Taeyoung Ha were supported by National Institute for Mathematical Sciences (NIMS) grant funded by the Korean government (No. B24910000).

References

[1] Christian Ledig 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*. 2017, pp. 4681–4690.

[2] Chao Dong et al. "Image super-resolution using deep convolutional networks". In: *IEEE transactions on pattern analysis and machine intelligence* 38.2 (2015), pp. 295–307.

[3] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. "Accurate image super-resolution using very deep convolutional networks". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 1646–1654.

[4] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. "Deeply-recursive convolutional network for image super-resolution". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 1637–1645.

[5] 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.

[6] Bee Lim et al. “Enhanced deep residual networks for single image super-resolution”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*. 2017, pp. 136–144.

[7] Wei-Sheng Lai et al. “Deep laplacian pyramid networks for fast and accurate super-resolution”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 624–632.

[8] Namhyuk Ahn, Byungkon Kang, and Kyung-Ah Sohn. “Fast, accurate, and lightweight super-resolution with cascading residual network”. In: *Proceedings of the European conference on computer vision (ECCV)*. 2018, pp. 252–268.

[9] Bee Lim et al. “Enhanced deep residual networks for single image super-resolution”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*. 2017, pp. 136–144.

[10] Kai Zhang et al. “Learning deep CNN denoiser prior for image restoration”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 3929–3938.

[11] Xintao Wang et al. “Esrgan: Enhanced super-resolution generative adversarial networks”. In: *Proceedings of the European conference on computer vision (ECCV) workshops*. 2018, pp. 0–0.

[12] Peter J Burt and Edward H Adelson. “The Laplacian pyramid as a compact image code”. In: *Readings in computer vision*. Elsevier, 1987, pp. 671–679.

[13] Emily L Denton, Soumith Chintala, Rob Fergus, et al. “Deep generative image models using a laplacian pyramid of adversarial networks”. In: *Advances in neural information processing systems* 28 (2015).

[14] Wei-Sheng Lai et al. “Deep laplacian pyramid networks for fast and accurate super-resolution”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 624–632.

[15] Wei-Sheng Lai et al. “Fast and accurate image super-resolution with deep laplacian pyramid networks”. In: *IEEE transactions on pattern analysis and machine intelligence* 41.11 (2018), pp. 2599–2613.

[16] Radu Timofte et al. “Ntire 2017 challenge on single image super-resolution: Methods and results”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*. 2017, pp. 114–125.

[17] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. “Accurate image super-resolution using very deep convolutional networks”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 1646–1654.

[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] Marco Bevilacqua et al. “Low-complexity single-image super-resolution based on nonnegative neighbor embedding”. In: (2012).

[20] Roman Zeyde, Michael Elad, and Matan 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*. Springer. 2012, pp. 711–730.

[21] Pablo Arbelaez et al. “Contour detection and hierarchical image segmentation”. In: *IEEE transactions on pattern analysis and machine intelligence* 33.5 (2010), pp. 898–916.

[22] Jia-Bin Huang, Abhishek Singh, and Narendra Ahuja. “Single image super-resolution from transformed self-exemplars”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, pp. 5197–5206.

[23] Yusuke Matsui et al. “Sketch-based manga retrieval using manga109 dataset”. In: *Multimedia tools and applications* 76 (2017), pp. 21811–21838.