Selective Super-Resolution for Scene Text Images

Ryo Nakao*, Brian Kenji Iwana*, and Seiichi Uchida*
*Advanced Information Technology, Kyushu University, Japan
{brian, uchida}@human.ait.kyushu-u.ac.jp

***Abstract*—In this paper, we realize the enhancement of super-resolution using images with scene text. Specifically, this paper proposes the use of Super-Resolution Convolutional Neural Networks (SRCNN) which are constructed to tackle issues associated with characters and text. We demonstrate that standard SRCNNs trained for general object super-resolution is not sufficient and that the proposed method is a viable method in creating a robust model for text. To do so, we analyze the characteristics of SRCNNs through quantitative and qualitative evaluations with scene text data. In addition, analysis using the correlation between layers by Singular Vector Canonical Correlation Analysis (SVCCA) and comparison of filters of each SRCNN using t-SNE is performed. Furthermore, in order to create a unified super-resolution model specialized for both text and objects, a model using SRCNNs trained with the different data types and Content-wise Network Fusion (CNF) is used. We integrate the SRCNN trained for character images and then SRCNN trained for general object images, and verify the accuracy improvement of scene images which include text. We also examine how each SRCNN affects super-resolution images after fusion.**

***Keywords*-Super-Resolution, Scene Text, Super-Resolution Convolutional Neural Network, Context-wise Network Fusion**

I. Introduction

Super-resolution is a technique for enhancing a low-resolution image to a high-resolution image. Some examples of super-resolution include example-based super-resolution [1] and super-resolution using sparse coding [2] In addition, there have been attempts at super-resolution by processing the background and foreground in a single image [3] and methods doing super-resolution by distributing the luminance value of the local region to character images [4]. Furthermore, in recent years, a highly accurate super-resolution method using a neural network has been proposed, called a Super-Resolution Convolutional Neural Network (SRCNN) [5], [6]. SRCNN is a super-resolution method which uses a Convolutional Neural Network (CNN) architecture. In addition, methods have expanded SRCNNs with the use of additional layers [7]–[9]. However, most methods do not specialize for specific classes and most accuracy improvements have been used for natural scene images.

By narrowing the target to images with characters or text, it might be possible to realize super-resolution with higher accuracy for scene text images. Therefore, we experiment with SRCNN trained using only images with text and compare it to an SRCNN trained for general object super-resolution. From this comparison, we are able to analyze the characteristics of the SRCNNs with the different training regimens.

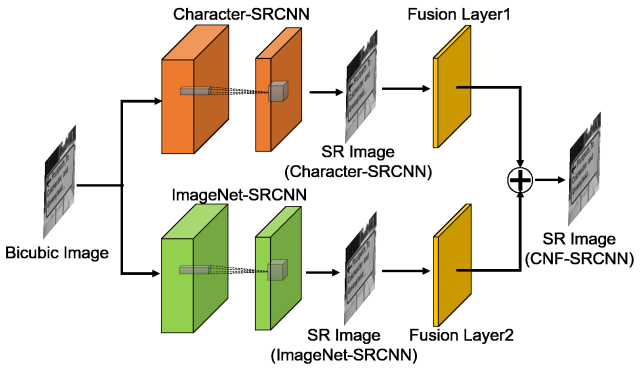


Figure 1. The combination of CNF with SRCNN.

In addition, we can attempt super-resolution for images containing both text and objects through the use of Context-wise Network Fusion (CNF) [10]. CNF integrates the outputs of individual CNNs in additional layers with fine-tuning in order to fuse heterogeneous information. Shown in Fig. 1, by fusing an SRCNN trained for super-resolution with objects and an SRCNN trained for super-resolution with text, we show and verify the accuracy improvement of images with natural scene text.

The contribution of this paper is twofold. First, we analyze the differences in characteristics with SRCNNs trained with object images and text images. Specifically, we use quantitative and qualitative evaluations of the super-resolution capabilities of the SRCNNs. Furthermore, we find the correlation between layers using Singular Vector Canonical Correlation Analysis (SVCCA) [11] and compare the filters of each SRCNN using t-Distributed Stochastic Neighbor Embedding (t-SNE) [12]. Second, we propose the use of CNF to combine SRCNNs in order to produce a more robust model.

II. Related Work

General image super-resolution is an active field in machine learning. There are many models based on CNNs, such as SRCNNs [5], [8], [9], Rapid and Accurate Image Super-resolution (RAISR) [13], Residual Channel At-

tention Networks (RCAN) [14], Deeply-Recursive Convolutional Networks (DRCN) [15], and Very Deep Super-Resolution (VDSR) [7].

The super-resolution of images with text is also a growing field. Most super-resolution models involving text use text derived from documents. For example, Capel and Zisserman [16] used a maximum likelihood estimator for the super-resolution of document-based text images. Banerjee and Jawahar [17] and Thillou and Mirmehdi [18] use super-resolution on document-based characters. There have only been few works focusing on natural scene text, most of which focus on text from video [19], [20].

III. Super-Resolution Convolutional Neural Networks

SRCNNs tackle super-resolution through the use of image-to-image CNNs. The input of an SRCNN is a small low-resolution image and the output is the corresponding large high-resolution image. It works by first enlarging the low-resolution image by Bicubic interpolation then refines the up-scaled image through three convolutional layers.

A. Total Variation Loss

The SRCNN is trained using Total Variation Loss (TV Loss) in addition to Mean Squared Error (MSE). TV Loss is the difference in luminance values between adjacent pixels in the entire image. It is useful for text due to being able to reduce noise and create smoother flat areas while maintaining sharp boundaries. TV Loss is defined as:

$$\mathrm{TV}(\mathbf{Y}) = \sum_{i=1}^{m}\sum_{j=1}^{n}(|y_{i+1,j} - y_{i,j}| + |y_{i,j+1} - y_{i,j}|), \quad (1)$$

where $\mathbf{Y}$ is the super-resolution image which is the output of CNN, and $\mathbf{X}$ is the original image. The total loss $\mathcal{L}$ is:

$$\mathcal{L}(\mathbf{Y}, \mathbf{X}) = \mathrm{MSE}(\mathbf{Y}, \mathbf{X}) + \lambda \mathrm{TV}(\mathbf{Y}), \quad (2)$$

where λ is a predefined weight.

B. Character-SRCNN and ImageNet-SRCNN

The standard use of SRCNN for general image super-resolution trained using the ImageNet dataset [21]. However, to demonstrate the inadequacy of SRCNN trained with ImageNet on images (ImageNet-SRCNN) with text, we compare SRCNN trained with generated images of characters (Character-SRCNN).

IV. Context-wise Network Fusion with SRCNNs

To combine them while maintaining the benefits of each, we propose using CNF. CNF is a method of fusing CNNs with extra convolutional layers, in order to combine networks with different characteristics. In [10], CNF is used to merge SRCNNs with different numbers of layers. However, we propose using CNF to merge the Character-SRCNN with the ImageNet-SRCNN to build a unified model. By combining the networks in this way, the Character-SRCNN can focus on the text regions and the ImageNet-SRCNN can focus on the other regions. The proposed model of using CNF with the two SRCNNs is shown in Fig. 1. In the figure, the fusion layer consists of a convolutional layer with a 3×3 convolution.

In order to train the CNF, the SRCNNs are first trained for their respective text and general image tasks. Next, the Character-SRCNN and the ImageNet-SRCNN are fused using CNF. Finally, the CNF is fine-tuned using scene text images in order to take advantage of the two SRCNNs. During the fine-tuning process, the parameters of the SRCNNs are fixed and only the fusion layers are updated.

V. Comparative Evaluation of Character-SRCNN and ImageNet-SRCNN

A. Dataset

In order to evaluate super-resolution, we prepared image datasets with low-resolution and high-resolution counterparts. The low-resolution images were constructed by down-sampling the high-resolution images by a factor of 6. Thus, the SRCNNs aim to enhance images that were magnified by 6.

1) Character dataset: To construct the character dataset, 24,000 serif, sans serif, and script text images were generated. Furthermore, in order to simulate natural scene text, the text and background colors were randomized and noise was added. The size of the characters in the generated images was randomized. Using multiple character sizes increases the generalization of the model. Using this dataset, the Character-SRCNN is trained to focus on super-resolution of text.

2) ImageNet dataset: ImageNet [21] consists of real images such as images of people and scenes. We use about 400,000 from ImageNet selected based on the conditions outlined in [5], [6]. The ImageNet dataset was used to train the ImageNet-SRCNN.

3) ICDAR 2013 dataset: For the scene text dataset, we used the International Conference on Document Analysis and Recognition Robust Reading Competition dataset (ICDAR 2013) dataset [22]. This dataset is used to evaluate the SRCNNs using images having both natural scenes similar to the ImageNet dataset and text similar to the Character dataset. Moreover, this dataset is only used for testing. 4,686 images of text areas were extracted as the test set. The size of the characters varies from 8×8 to 30×30 pixels for each image.

B. Quantitative Evaluation

In order to quantitatively evaluate the methods, we use the following four evaluations.

1) PSNR: Peak signal-to-noise ratio (PSNR) is a measure to calculate the similarity between the ground truth image and the generated super-resolution image. The evaluation is performed by determining the ratio between the maximum luminance MAX and the amount of noise. Namely, the PSNR is calculated by:

$$PSNR = 10 \log_{10} \frac{MAX^2}{MSE}. \tag{3}$$

For PSNR, the unit is in decibels (dB) and the larger the value the more similar the super-resolution image is to the ground truth image.

2) SSIM: Similar to PSNR, Structural Similarity (SSIM) is a similarity measure between the ground truth and the super-resolution image. SSIM is the perceived difference between two images and it relies on the average luminance μ, variance σ^2, and covariance σ for each local region. The equation for SSIM is:

$$SSIM = \frac{(2\mu_{\mathbf{X}}\mu_{\mathbf{Y}} + C_1) + (2\sigma_{\mathbf{XY}} + C_2)}{(\mu_{\mathbf{X}}^2 + \mu_{\mathbf{Y}}^2 + C_1)(\sigma_{\mathbf{X}}^2 + \sigma_{\mathbf{Y}}^2 + C_2)}, \tag{4}$$

where C_1 and C_2 are stabilizing constants set to $C_1 = (0.01 \times 255)^2$ and $C_2 = (0.03 \times 255)^2$. SSIM is a value from 0 to 1 with a larger value meaning more similar.

3) CRR: The Character Recognition Rate (CRR) is the evaluation of character recognition using the super-resolution images. For this purpose, a Convolutional Recurrent Neural Network (CRNN) [23] is used. CRNN is a well-known method of recognizing text in natural scene images that uses convolutional layers to extract features, recurrent layers to predict characters, and a transcription layer using Connectionist Temporal Classification (CTC). The CRR is the rate at which CRNN could recognize the text. Larger is better and it only applies to character regions.

4) NIQE: Naturalness Image Quality Evaluator (NIQE) [24] is an indicator to evaluate the image quality. Unlike PSNR and SSIM, this evaluation does not use the original ground truth images. NIQE evaluates image quality by comparing features from the evaluated image and statistical regularities observed in natural images. Smaller NIQE values are better.

C. Results

Table I shows the results of the quantitative evaluation of Character-SRCNN and ImageNet-SRCNN using the four measures. In the table, the character regions are the areas of the ICDAR 2013 scene text images which contain characters and the other regions are the areas that do not. Not surprisingly, the Character-SRCNN shows better results in the character areas and the ImageNet-SRCNN has better results in the non-character areas. However, from this result, it can be said that the SRCNNs can be trained differently to target specific types of data.

Table I
COMPARISON OF ACCURACY BETWEEN IMAGENET-SRCNN AND CHARACTER-SRCNN

	ImageNet-SRCNN	Character-SRCNN	Original Image
Character Regions			
PSNR (dB) ↑	23.9	**24.0**	∞
SSIM ↑	0.718	**0.735**	1.0
CCR (%) ↑	55.9	**57.2**	77.1
NIQE ↓	8.77	**7.49**	6.09
Other Regions			
PSNR [dB] ↑	**29.4**	29.3	∞
SSIM ↑	**0.891**	0.883	1.0
NIQE ↓	**5.06**	6.05	3.82

(a) Low-resolution image

(b) Results of ImageNet-SRCNN

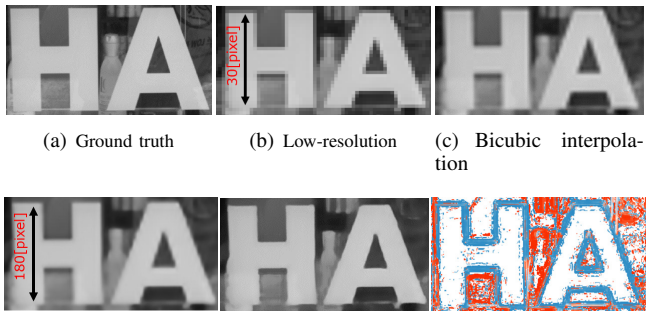
(c) Results of Character-SRCNN

(d) Accuracy comparison

Figure 2. Comparison of super-resolution on scene text areas. In (d), blue is where the SSIM value was higher for the Character-SRCNN and red was for the Image-SRCNN.

D. Qualitative Analysis

In order to inspect the characteristics of each SRCNN, we perform pixel specific accuracy comparisons between the ground truth and the super-resolution images produced by Character-SRCNN and ImageNet-SRCNN. Specifically, we use SSIM to visualize the structure of each image from each SRCNN. The SSIM is calculated in local regions and the output is colored corresponding to a higher SSIM score. Red local regions are where the ImageNet-SRCNN performed



(a) Ground truth (b) Low-resolution (c) Bicubic interpolation

(d) Results of ImageNet-SRCNN (e) Results of Character-SRCNN (f) Accuracy comparison

Figure 3. An example of text zoomed in using ImageNet-SRCNN and Character-SRCNN.

(a) Low-resolution (b) Results of Character-SRCNN (c) Accuracy comparison

Figure 4. An example of a character region in which the Character-SRCNN did poorly.

better and blue local regions are where the Character-SRCNN performed better.

Comparative results for several images are shown in Fig. 2. From the figure, it can be seen that the clarity of the text is much higher on the images created by Character-SRCNN. The improvement of the character regions can be explained by the accuracy comparison in Fig. 2 (d). In the areas surrounding the text, the Character-SRCNN results are improved, as indicated by blue. Conversely, the background areas are dominantly red pixels in the accuracy comparison.

Fig. 3 shows an example of a zoomed in region. In this region, the input low-resolution letters have significant aliasing. While the bicubic interpolation and the results from the ImageNet-SRCNN have improved anti-aliasing and smoothing, the results from the Character-SRCNN are sharp and almost totally reconstructed the ground truth.

While the results for the Character-SRCNN were impressive, there were instances where the super-resolution suffered. Fig. 4, for instance, shows a zoomed in a piece of an image where a white aberration effect was added to the text. In addition, as Fig. 4 (c) shows, excess noise was added to the result from the Character-SRCNN. We found problems with this generally when the luminance difference between the text and the background was not significant enough.

E. Comparison of the Networks

SVCCA [11] is a method to analyze the similarity of two networks. The inputs of SVCCA are the outputs of each layer and the calculation of SVCCA is divided into two stages. First, Singular Value Decomposition (SVD) [25] is performed to reduce redundant dimensions which are not necessary for network representation. Next, Canonical Correlation Analysis (CCA) [26] is performed using the results of SVD. By doing so, we derive the correlation on a layer-by-layer basis.

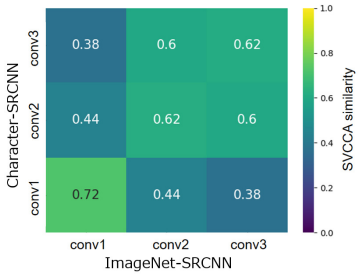


Figure 5. SVCCA between Character-SRCNN and ImageNet-SRCNN.

To compare the Character-SRCNN and ImageNet-SRCNN networks, we find the layer-wise correlation between them using SVCCA. Fig. 5 shows the correlation between the convolutional layers in the Character-SRCNN and ImageNet-SRCNN. Numerals in each cell indicate correlation coefficients between corresponding layers. From the figure, there is some degree of correlation between the corresponding layers. However, they are not exactly the same, which indicates that each SRCNN has its own characteristics.

F. Comparison of the Trained Weights

In order to compare the convolutional filters used between the two SRCNNs, t-SNE is used. t-SNE is a dimension reduction method for high-dimensional data. Using t-SNE, the dimensions can be reduced while maintaining the relation of high-dimensional features. For the comparison between the filters, t-SNE was performed on the 9×9 filters from the first layer of both SRCNNs. Fig. 6 shows the visualization. It can be seen from the figure that the filters of the ImageNet-SRCNN are distributed around the edges of the filters of the Character-SRCNN. Conversely, the filters of the Character-SRCNN are focused and do not vary as much. Therefore, it is conceivable that the filters of the Character-SRCNN are specialized toward certain tasks while the Image-SRCNN tackles many different features.

VI. Experimental Results of SRCNN Fusion

A. Quantitative Evaluation

In order to train the CNF using SRCNN networks (CNF-SRCNN), we first pre-train the Character-SRCNN and ImageNet-SRCNN with the respective datasets described in Section V-A. The second step is to fix the weights of the

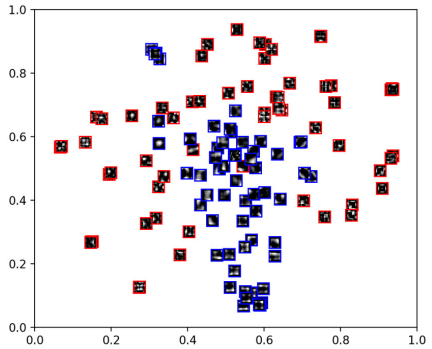


Figure 6. Visualization of the filters of the first layer of each SRCNN using t-SNE. The filters enclosed in red are from the ImageNet-SRCNN. The filters enclosed in blue are from the Character-SRCNN.

Table II
COMPARISON OF ACCURACY BETWEEN EACH SRCNN

	ImageNet-SRCNN	Character-SRCNN	CNF-SRCNN
PSNR (dB) ↑	29.1	28.6	**29.9**
SSIM ↑	0.887	0.882	**0.897**
NIQE ↓	6.25	7.16	**5.11**

SRCNNs and train the convolutional layers in the CNF. The final step is to test using the ICDAR 2013 scene text dataset.

The results for CNF-SRCNN is shown in Table II. The CNF-SRCNN performed better in every quantitative evaluation when compared to the SRCNNs separately. This shows that the CNF-SRCNN is able to use the characteristics of both SRCNNs and combine them into one robust super-resolution model.

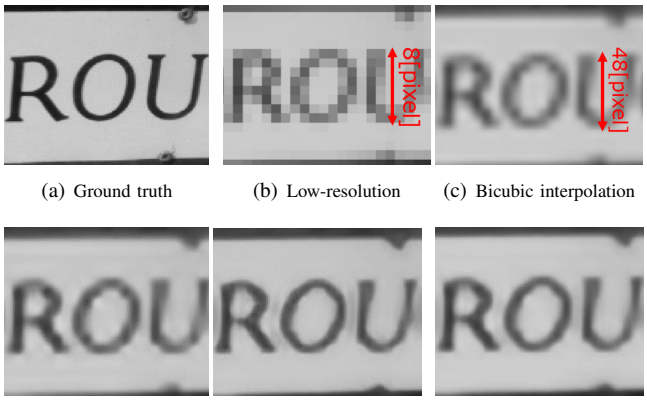
B. Qualitative Analysis

Fig. 7 shows an example of the super-resolution result by ImageNet-SRCNN, Character-SRCNN, and the proposed CNF-SRCNN. The text in Fig. 7 is especially small, so this is a good example to illustrate the differences between the methods. In Fig. 7 (d), there are still aberrations between the characters, however, in Fig. 7 (e), that noise is fixed. However, even though the results of CNF-SRCNN have less noise, the Character-SRCNN has more pronounced text.

C. Influence of Each SRCNN on CNF

We visualized the influence of Character-SRCNN and ImageNet-SRCNN on super-resolution images of CNF-SRCNN. To visualize the influence, we applied the SSIM measure for each pixel. Specifically, super-resolution images of CNF-SRCNN, Character-SRCNN, and ImageNet-SRCNN are compared.

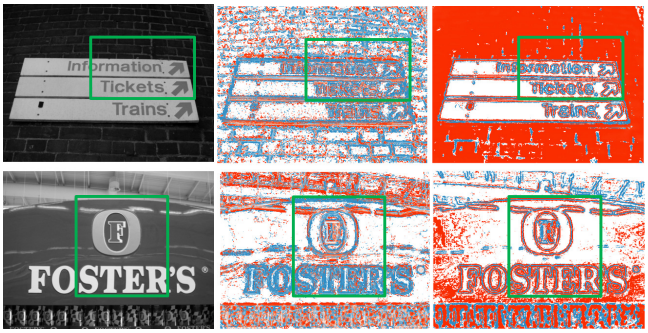
Fig. 8 shows pieces of images visualizing the influence on CNF-SRCNN that the two SRCNNs have. Blue indicates that Character-SRCNN is more influential and red indicates that ImageNet-SRCNN is more influential. From Fig. 8, ImageNet-SRCNN affects the whole image, whereas Character-SRCNN influences the CNF-SRCNN near the text. The ImageNet-SRCNN's influence is most prominent in the background of the image. However, the influence of the SRCNNs on the local regions does not necessarily correlate with accuracy.



(a) Ground truth (b) Low-resolution (c) Bicubic interpolation

(d) Results of ImageNet-SRCNN (e) Results of Character-SRCNN (f) Results of CNF-SRCNN

Figure 7. Comparison of ImageNet-SRCNN, Character-SRCNN, and CNF-SRCNN in a character region.



(a) Low-resolution (b) Accuracy comparison (c) Influence comparison

Figure 8. Influence comparison result of each SRCNN on CNF-SRCNN. Blue: influence of Character-SRCNN is larger. Red: influence of ImageNet-SRCNN is larger.

D. Using CNF with Other Super-Resolution Methods

The use of CNF for CNN-based super-resolution methods is not limited to SRCNN. We performed an additional experiment using the proposed method of combining an RCAN based on ImageNet (ImageNet-RCAN) and one based on text (Character-SRCNN) using CNF (CNF-RCAN). The results are shown in Table III. However, the results of CNF-RCAN are lower than ImageNet-RCAN by itself. The reason for this is due to the weakness of Character-RCAN. Unlike SRCNN, RCAN is aggressive at fitting features to the low-resolution image. Due to this, the Character-RCAN is over-influenced by the text training method which causes the background regions to not be modeled correctly. Thus, using Character-RCAN in CNF is not suited for scene text super-resolution. Using this knowledge, we combined ImageNet-RCAN with Character-SRCNN in one

Table III
COMPARISON OF ACCURACY BETWEEN EACH RCAN

	ImageNet-RCAN	Character-RCAN	CNF-RCAN	CNF-RCAN+SRCNN
PSNR (dB) ↑	32.8	27.4	32.5	**32.9**
SSIM ↑	**0.921**	0.863	0.918	**0.921**
NIQE ↓	**4.87**	5.53	5.02	4.93

model (CNF-RCAN+SRCNN) to get the best of both super-resolution methods. Under this scheme, we were able to improve the results.

VII. CONCLUSION

In this paper, we realized accurate super-resolution using SRCNNs for images with text. Specifically, we realized super-resolution by training an SRCNN using ImageNet and a generated character dataset separately. Then, we combined the two SRCNNs using CNF to achieve a higher accuracy. Through evaluations, we demonstrated that the CNF-SRCNN was able to produce better super-resolution results by combining the ImageNet-SRCNN and Character-SRCNN. Furthermore, we analyzed the differences between the SRCNNs trained with different datasets and found that they have different characteristics. Also, internal analysis of the networks was done using SVCCA for layer-wise comparison and t-SNE to visualize the differences in the first convolutional layer.

ACKNOWLEDGEMENT

This work was supported by JSPS KAKENHI Grant Number JP17H06100.

REFERENCES

[1] W. T. Freeman, T. R. Jones, and E. C. Pasztor, "Example-based super-resolution," *IEEE Computer Graphics and Applications*, vol. 22, no. 2, pp. 56–65, 2002.

[2] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Trans. Image Processing*, vol. 19, no. 11, pp. 2861–2873, 2010.

[3] J. Sun, Q. Chen, S. Yan, and L.-F. Cheong, "Selective image super-resolution," *arXiv preprint arXiv:1010.5610*, 2010.

[4] N. Nayef, J. Chazalon, P. Gomez-Kramer, and J.-M. Ogier, "Efficient example-based super-resolution of single text images based selective patch processing," in *IAPR Int. Workshop Document Analysis Systems*, 2014, pp. 227–231.

[5] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," in *European Conf. Computer Vision*, 2014, pp. 184–199.

[6] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 38, no. 2, pp. 295–307, 2016.

[7] J. Kim, J. Kwon Lee, and K. Mu Lee, "Accurate image super-resolution using very deep convolutional networks," in *IEEE Conf. Computer Vision and Pattern Recognition*, 2016, pp. 1646–1654.

[8] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," *arXiv preprint*, 2017.

[9] C. Dong, C. C. Loy, and X. Tang, "Accelerating the super-resolution convolutional neural network," in *European Conference Computer Vision*, 2016, pp. 391–407.

[10] H. Ren, M. El-Khamy, and J. Lee, "Image super resolution based on fusing multiple convolution neural networks," in *IEEE CVPR Workshops*, 2017, pp. 1050–1057.

[11] M. Raghu, J. Gilmer, J. Yosinski, and J. Sohl-Dickstein, "Svcca: Singular vector canonical correlation analysis for deep learning dynamics and interpretability," in *Advances in Neural Information Processing Systems*, 2017, pp. 6076–6085.

[12] L. v. d. Maaten and G. Hinton, "Visualizing data using t-sne," *J. of machine learning research*, vol. 9, no. Nov, pp. 2579–2605, 2008.

[13] Y. Romano, J. Isidoro, and P. Milanfar, "RAISR: Rapid and accurate image super resolution," *IEEE Trans. Computational Imaging*, vol. 3, no. 1, pp. 110–125, mar 2017.

[14] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," in *European Conf. Computer Vision*. Springer, 2018, pp. 294–310.

[15] J. Kim, J. K. Lee, and K. M. Lee, "Deeply-recursive convolutional network for image super-resolution," in *IEEE Conf. Computer Vision and Pattern Recognition*, 2016.

[16] D. Capel and A. Zisserman, "Super-resolution enhancement of text image sequences," in *Int. Conf. Pattern Recognition*.

[17] J. Banerjee and C. Jawahar, "Super-resolution of text images using edge-directed tangent field," in *IAPR Int. Workshop Document Analysis Systems*. IEEE, 2008.

[18] C. Mancas-Thillou and M. Mirmehdi, "Super-resolution text using the teager filter," in *Int. Workshop Camera-Based Document Analysis and Recognition*, 2005, pp. 10–16.

[19] H. Li and D. Doermann, "Superresolution-based enhancement of text in digital video," in *Int. Conf. Pattern Recognition*, 2000.

[20] K. Donaldson and G. Myers, "Bayesian super-resolution of text in video with a text-specific bimodal prior," in *IEEE Conf. Computer Vision and Pattern Recognition*, 2005.

[21] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *IEEE Conf. Computer Vision and Pattern Recognition*, 2009, pp. 248–255.

[22] D. Karatzas, F. Shafait, S. Uchida, M. Iwamura, L. G. i Bigorda, S. R. Mestre, J. Mas, D. F. Mota, J. A. Almazan, and L. P. De Las Heras, "Icdar 2013 robust reading competition," in *IAPR Int. Conf. Document Analysis and Recognition*, 2013, pp. 1484–1493.

[23] B. Shi, X. Bai, and C. Yao, "An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 39, no. 11, pp. 2298–2304, 2017.

[24] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a" completely blind" image quality analyzer." *IEEE Signal Processing Letters*, vol. 20, no. 3, pp. 209–212, 2013.

[25] J. C. Nash, "The singular-value decomposition and its use to solve least-squares problems," in *Compact Numerical Methods for Computers*. Routledge, dec 2018, pp. 30–48.

[26] D. R. Hardoon, S. Szedmak, and J. Shawe-Taylor, "Canonical correlation analysis: An overview with application to learning methods," *Neural Computation*, vol. 16, no. 12, pp. 2639–2664, dec 2004.