

SUPER-RESOLUTION USING CONVOLUTIONAL NEURAL NETWORKS WITHOUT ANY CHECKERBOARD ARTIFACTS

Yusuke Sugawara, Sayaka Shiota, and Hitoshi Kiya

Tokyo Metropolitan University, 6-6 Asahigaoka, Hino-shi, Tokyo, Japan

ABSTRACT

It is well-known that a number of excellent super-resolution (SR) methods using convolutional neural networks (CNNs) generate checkerboard artifacts. A condition to avoid the checkerboard artifacts is proposed in this paper. So far, checkerboard artifacts have been mainly studied for linear multirate systems, but the condition to avoid checkerboard artifacts can not be applied to CNNs due to the non-linearity of CNNs. We extend the avoiding condition for CNNs, and apply the proposed structure to some typical SR methods to confirm the effectiveness of the new scheme. Experiment results demonstrate that the proposed structure can perfectly avoid to generate checkerboard artifacts under two loss conditions: mean square error and perceptual loss, while keeping excellent properties that the SR methods have.

Index Terms— Super-Resolution, Convolutional Neural Networks, Checkerboard Artifacts

1. INTRODUCTION

This paper addresses the problem of checkerboard artifacts generated by some super-resolution (SR) methods using convolutional neural networks (CNNs). SR methods using CNNs have been widely studying as one of single image SR techniques, and have superior performances [1–5]. Moreover, in order to accelerate the processing speed, CNNs including upsampling layers such as deconvolution [6] and sub-pixel convolution [7] ones have been proposed [7–12]. However, it is well-known that these SR methods generate periodic artifacts, referred to as checkerboard artifacts [13].

In CNNs, it is well-known that checkerboard artifacts are generated by operations of deconvolution, sub-pixel convolution layers [14]. To overcome these artifacts, smoothness constraint [15], post-processing [13], initialization scheme [16] and different upsampling layer designs [14, 17, 18] have been proposed. Most of them can not avoid checkerboard artifacts perfectly, although they reduce the artifacts. Among them, Odena et al. [14] have demonstrated that checkerboard artifacts can be perfectly avoided by using resize convolution layers instead of deconvolution ones. However, the resize convolution layers can not be directly applied to upsampling layers such as deconvolution and sub-pixel convolution ones, so this method needs not only large memory but also high computational costs.

On the other hand, checkerboard artifacts have been studied to design linear multirate systems including filter banks and wavelets [19–22]. In addition, it is well-known that checkerboard artifacts are caused by the time-variant property of interpolators in multirate systems, and the condition for avoiding these artifacts have been given [19–21]. However, the condition to avoid checkerboard artifacts for linear systems can not be applied to CNNs due to the non-linearity of CNNs.

In this paper, we extend the avoiding condition for CNNs, and apply the proposed structure to SR methods using deconvolution and

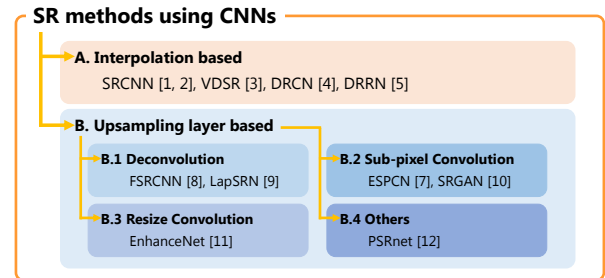


Fig. 1: Classification of SR methods using CNNs

sub-pixel convolution layers to confirm the effectiveness of the new scheme. Experiment results demonstrate that the proposed structure can perfectly avoid to generate checkerboard artifacts under two loss conditions: mean square error and perceptual loss, while keeping excellent properties that the SR methods have. As a result, it is confirmed that the proposed structure allows us to offer efficient SR methods without any checkerboard artifacts.

2. PREPARATION

Conventional SR methods using CNNs and works related to checkerboard artifacts are reviewed, here.

2.1. SR Methods using CNNs

SR methods using CNNs are classified into two classes as shown in Fig. 1. Interpolation based methods [1–5], referred to as class A, do not generate any checkerboard artifacts in CNNs, due to the use of an interpolated image as an input to a network. In other words, CNNs in this class do not have any upsampling layers.

On the other hand, when CNNs include upsampling layers, there is a possibility that the CNNs generate some checkerboard artifacts. This class, called class B in this paper, have provided numerous excellent SR methods [7–12], which can be executed faster than those in class A. Class B is also classified into a number of sub-classes according to the type of upsampling layers. This paper focuses on class B.

CNNs are illustrated in Fig. 2 for an SR problem, as in [7], where the CNNs consist of two convolutional layers and one upsampling layer. I_{LR} and $f_c^{(l)}(I_{LR})$ are a low-resolution (LR) image and a c -th channel feature map at layer l , and $f(I_{LR})$ is an output of the network. The two convolutional layers have learnable weights, biases, and ReLU [23] as an activation function, respectively, where the weight at layer l has $K_l \times K_l$ as a spatial size and N_l as the number of feature maps.

There are numerous algorithms for computing upsampling layers, such as deconvolution, sub-pixel convolution and resize convolution ones, which are widely used as typical CNNs. Besides, deconvolution [6], sub-pixel convolution [7] and resize convolution [14] layers are well-known upsampling layers, respectively.

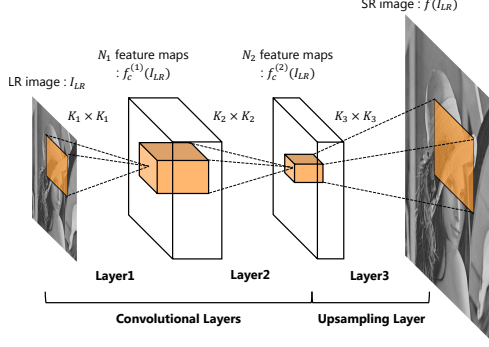


Fig. 2: CNNs with an upsampling Layer

2.2. Works Related to Checkerboard Artifacts

Checkerboard artifacts have been discussed to design multirate systems including filter banks and wavelets by researchers [19–22]. However, most of the works have been limited to in case of using linear systems, so they can not be directly applied to CNNs due to the non-linearity. Some works related to checkerboard artifacts for linear systems are summarized, here.

It is known that linear interpolators which consist of up-samplers and linear time-invariant systems cause checkerboard artifacts due to the periodic time-variant property [19–21]. Figure 3 illustrates a linear interpolator with an up-sampler $\uparrow U$ and a linear time-invariant system $H(z)$, where positive integer U is an upscaling factor and $H(z)$ is the z transformation of an impulse response. The interpolator in Fig. 3(a) can be equivalently represented as a polyphase structure as shown in Fig. 3(b). The relationship between $H(z)$ and $R_i(z)$ is given by

$$H(z) = \sum_{i=1}^U R_i(z^U) z^{-(U-i)}, \quad (1)$$

where $R_i(z)$ are often referred to as a polyphase filter of the filter $H(z)$.

The necessary and sufficient condition for avoiding the checkerboard artifacts in the system is shown as

$$R_1(1) = R_2(1) = \dots = R_U(1) = G. \quad (2)$$

This condition means that all polyphase filters have the same DC value i.e. a constant G [19–21]. Note that each DC value $R_i(1)$ corresponds to the steady-state value of the unit step response in each polyphase filter $R_i(z)$. In addition, the condition eq.(2) can be also expressed as

$$H(z) = P(z)H_0(z), \quad (3)$$

where,

$$H_0(z) = \sum_{i=0}^{U-1} z^{-i}, \quad (4)$$

$H_0(z)$ and $P(z)$ are an interpolation kernel of the zero-order hold with factor U and a time-invariant filter, respectively. Therefore, the linear interpolator with factor U does not generate any checkerboard artifacts, when $H(z)$ includes $H_0(z)$. In the case without checkerboard artifacts, the step response of the linear system has a steady-state value G as shown in Fig. 3(a). Meanwhile, the step response of the linear system has a periodic steady-state signal with the period of U , such as $R_1(1), \dots, R_U(1)$, if eq.(3) is not satisfied.

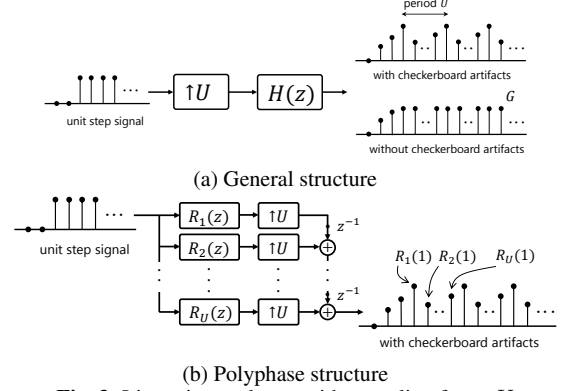


Fig. 3: Linear interpolators with upscaling factor U

3. PROPOSED METHOD

CNNs are non-linear systems, so conventional works related to checkerboard artifacts can not be directly applied to CNNs. A condition to avoid checkerboard artifacts in CNNs is proposed, here.

3.1. CNNs with Upsampling Layers

We focus on upsampling layers in CNNs, for which there are numerous algorithms such as deconvolution [6], sub-pixel convolution [7] and resize convolution [14]. For simplicity, one-dimensional CNNs will be considered in the following discussion.

It is well-known that deconvolution layers with non-unit strides cause checkerboard artifacts [14]. Figure 4 illustrates a system representation of deconvolution layers [6] which consist of some interpolators, where H_c and b are a weight and a bias in which c is a channel index, respectively. The deconvolution layer in Fig. 4(a) can be equivalently represented as a polyphase structure in Fig. 4(b), where $R_{c,n}$ is a polyphase filter of the filter H_c in which n is a filter index. This is a non-linear system due to the bias b .

Figure 5 illustrates a representation of sub-pixel convolution layers [7], where $R_{c,n}$ and b_n are a weight and a bias, and $f'_n(I_{LR})$ is an intermediate feature map in channel n . Compared Fig.4(b) with Fig.5, we can see that the polyphase structure in Fig. 4(b) is a special case of sub-pixel convolution layers in Fig. 5. In other words, Fig. 5 is reduced to Fig. 4(b), when satisfying $b_1 = b_2 = \dots = b_U$. Therefore, we will focus on sub-pixel convolution layers as the general case of upsampling layers to discuss checkerboard artifacts in CNNs.

3.2. Checkerboard Artifacts in CNNs

Let us consider the unit step response in CNNs. In Fig. 2, when the input I_{LR} is the unit step signal I_{step} , the steady-state value of the c -th channel feature map in layer 2 is given as

$$\hat{f}_c^{(2)}(I_{step}) = A_c, \quad (5)$$

where A_c is a positive constant value, which is decided by filters, biases and ReLU. Therefore, from Fig. 5, the steady-state value of the n -th channel intermediate feature map is given by, for sub-pixel convolution layers,

$$\hat{f}'_n(I_{step}) = \sum_{c=1}^{N_2} A_c \bar{R}_{c,n} + b_n, \quad (6)$$

where $\bar{R}_{c,n}$ is the DC value of the filter $R_{c,n}$.

Generally, the condition,

$$\hat{f}'_1(I_{step}) = \hat{f}'_2(I_{step}) = \dots = \hat{f}'_U(I_{step}), \quad (7)$$

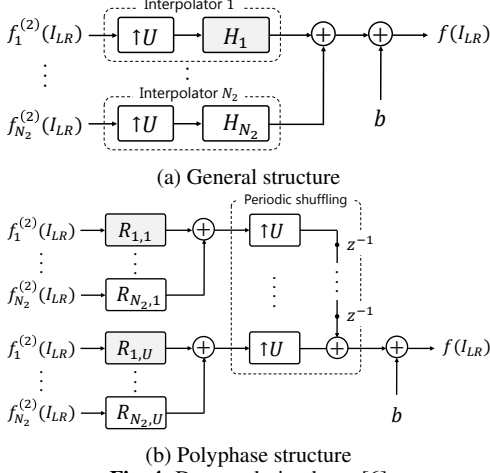


Fig. 4: Deconvolution layer [6]

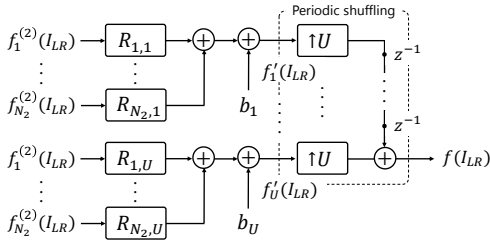


Fig. 5: Sub-pixel convolution layer [7]

is not satisfied, so the unit step response $f(I_{step})$ has a periodic steady-state signal with the period of U . To avoid checkerboard artifacts, eq.(7) has to be satisfied, as well as for linear multirate systems.

3.3. Upsampling Layers without Checkerboard Artifacts

To avoid checkerboard artifacts, CNNs must have the non-periodic steady-state value of the unit step response. From eq.(6), eq.(7) is satisfied, if

$$\bar{R}_{c,1} = \bar{R}_{c,2} = \dots = \bar{R}_{c,U}, c = 1, 2, \dots, N_2 \quad (8)$$

$$b_1 = b_2 = \dots = b_U, \quad (9)$$

Note that, in this case,

$$\hat{f}'_1(K \cdot I_{step}) = \hat{f}'_2(K \cdot I_{step}) = \dots = \hat{f}'_U(K \cdot I_{step}), \quad (10)$$

is also satisfied as for linear systems, where K is an arbitrary constant value. However, even when each filter H_c in Fig.5 satisfies eq.(3), eq.(9) is not met, but eq.(8) is met. Therefore, we have to seek for a new insight to avoid checkerboard artifacts in CNNs.

In this paper, we propose to add the kernel of the zero-order hold with factor U after upsampling layers as shown in Fig. 6. In this structure, the output signal from H_0 can be a constant value, even when an arbitrary periodic signal is inputted to H_0 . As a result, Fig. 6 can satisfy eq.(7).

There are three approaches to use H_0 in CNNs by the difference in training CNNs as follows.

A. Training CNNs without H_0

The simplest approach for avoiding checkerboard artifacts is to add H_0 to CNNs after training the CNNs. This approach allows us to perfectly avoid checkerboard artifacts generated by a pre-trained model.

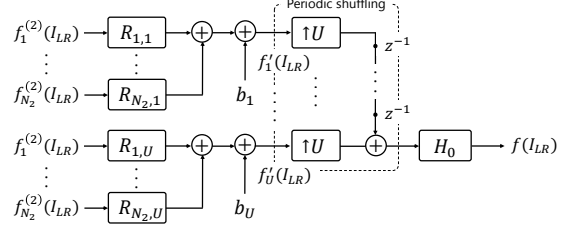


Fig. 6: Proposed upsampling layer structure without checkerboard artifacts

B. Training CNNs with H_0

In approach B, H_0 is added to CNNs before training the CNNs, and then the CNNs with H_0 are trained. This approach also allows us to perfectly avoid checkerboard artifacts as well as for approach A. Moreover, this approach provides higher quality images than those of approach A.

C. Training CNNs with H_0 inside upsampling layers

Approach C is applicable to only deconvolution layers, although approaches A and B are available for both of deconvolution layers and sub-pixel convolution ones. Deconvolution layers always satisfy eq.(9), so eq.(8) only has to be considered. Therefore, CNNs do not generate any checkerboard artifacts when each filter H_c in Fig.5 satisfies eq.(3). In approach C, checkerboard artifacts are avoided by convolving each filter H_c with the kernel H_0 inside upsampling layers.

4. EXPERIMENTS AND RESULTS

The proposed structure without checkerboard artifacts was applied to the SR methods using deconvolution and sub-pixel convolution layers to demonstrate the effectiveness. CNNs in the experiments were carried out under two loss functions: mean squared error (MSE) and perceptual loss.

4.1. Datasets for Training and Testing

We employed 91-image set from Yang et al. [24] as our training dataset. In addition, the same data augmentation (rotation and down-scaling) as in [8] was used. As a result, the training dataset consisting of 1820 images was created for our experiments. Besides, we used two datasets, Set5 [25] and Set14 [26], which are often used for benchmark, as test datasets.

To prepare a training set, we first downsampled the ground truth images I_{HR} with a bicubic kernel to create the LR images I_{LR} , where the factor $U = 4$ was used. The ground truth images I_{HR} were cropped into 72×72 pixel patches and the LR images were also cropped 18×18 pixel ones, where the total number of extracted patches was 8,000. In the experiments, the luminance channel (Y) of images was used for the MSE loss, although the three channels (RGB) of images were used for the perceptual loss.

4.2. Training Details

Table 1 illustrates CNNs used in the experiments, which were carried out based on CNNs in Fig. 2. For other two layers in Fig. 2, we set $(K_1, N_1) = (5, 64)$, $(K_2, N_2) = (3, 32)$ as in [7]. In addition, the training of all networks was carried out to minimize the mean squared error $\frac{1}{2} \|I_{HR} - f(I_{LR})\|^2$ and the perceptual loss $\frac{1}{2} \|\phi(I_{HR}) - \phi(f(I_{LR}))\|^2$ averaged over the training set, respectively, where ϕ calculates feature maps at the fourth layer of the pre-trained VGG-16 model as in [13]. It is well-known that the perceptual loss results in sharper SR images despite lower PSNR values, and generates checkerboard artifacts more frequently than under the MSE loss.

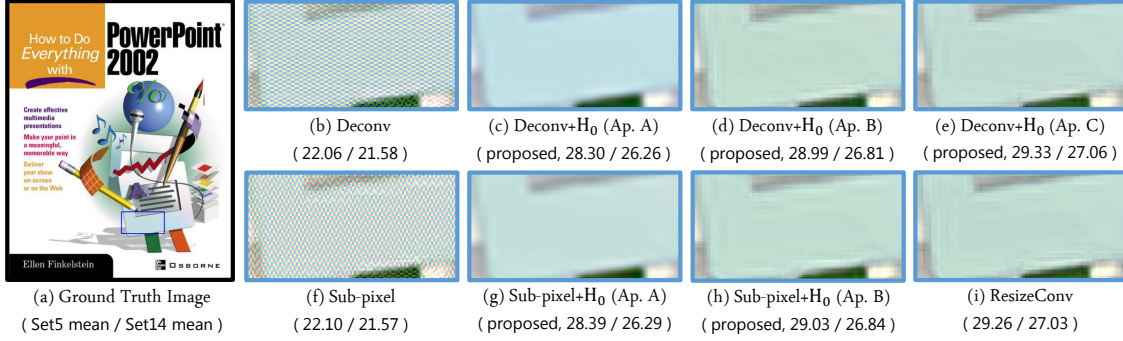


Fig. 7: Experimental results of super-resolution under perceptual loss (PSNR(dB))

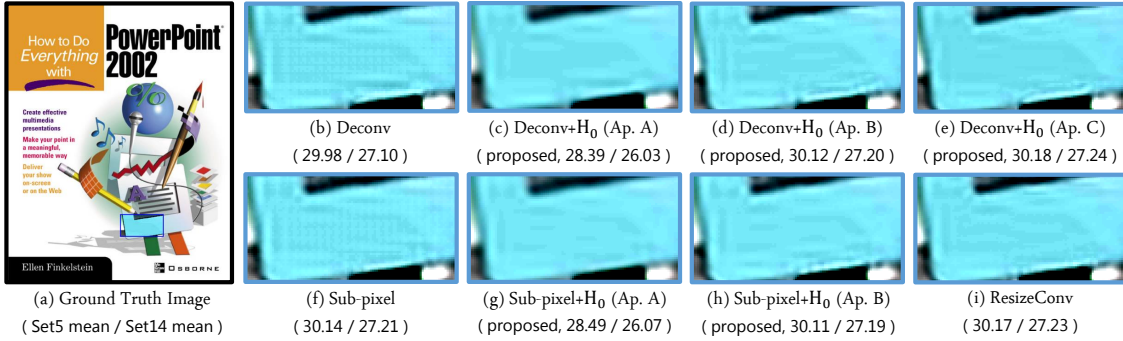


Fig. 8: Experimental results of super-resolution under MSE loss (PSNR(dB))

Table 1: CNNs used in the experiments

Network Name	Upsampling Layer	$K_3 \times K_3$
Deconv	Deconvolution [6]	9×9
Sub-pixel	Sub-pixel Convolution [7]	3×3
ResizeConv	Resize Convolution [14]	9×9
Deconv+H ₀	Deconvolution with H ₀ (Approach A or B)	9×9
Deconv+H ₀ (Ap. C)	Deconvolution with H ₀ (Approach C)	9×9
Sub-pixel+H ₀	Sub-pixel Convolution with H ₀ (Approach A or B)	3×3

For training, Adam [27] with $\beta_1 = 0.9, \beta_2 = 0.999$ was employed as an optimizer. Besides, we set the batch size to 4 and the learning rate to 0.0001. The weights were initialized with the method described in He et al. [28]. We trained all models for 200K iterations. All models were implemented by using the tensorflow framework [29].

4.3. Experimental Results

Figure 7 shows examples of SR images generated under the perceptual loss, where mean PSNR values for each dataset are also illustrated. In this figure, (b) and (f) include checkerboard artifacts, although (c), (d), (e), (g), (h) and (i) do not include any ones. Moreover, it is shown that the quality of SR images was significantly improved by avoiding checkerboard artifacts. Approach B and C also provided better quality images than approach A. In Fig. 8, (b) and (f) also include checkerboard artifacts as well as in Fig. 7, although the distortion is not so large, compared to under the perceptual loss. Note that ResizeConv does not generate any checkerboard artifacts, because it uses a pre-defined interpolation like in [1].

Table 2 illustrates the average executing time when each CNNs were carried out 10 times for some images in Set14. ResizeConv needs the highest computational cost in this table, although it does not generate any checkerboard artifacts. From this table, the pro-

Table 2: Execution time of super-resolution (sec)

Resolution of Input Image	Deconv	Deconv+H ₀ (Ap. A or B)	Deconv+H ₀ (Ap. C)
69×69	0.00871	0.0115	0.0100
125×90	0.0185	0.0270	0.0227
128×128	0.0244	0.0348	0.0295
132×164	0.0291	0.0393	0.0377
180×144	0.0343	0.0476	0.0421

Resolution of Input Image	Sub-pixel	Sub-pixel+H ₀ (Ap. A or B)	ResizeConv
69×69	0.0159	0.0242	0.107
125×90	0.0398	0.0558	0.224
128×128	0.0437	0.0619	0.299
132×164	0.0696	0.0806	0.383
180×144	0.0647	0.102	0.450

posed structures have much lower computational costs than with resize convolution layers. Note that the result was tested on PC with a 3.30 GHz CPU and the main memory of 16GB.

5. CONCLUSION

This paper addressed a condition to avoid checkerboard artifacts in CNNs including upsampling layers. The proposed structure can be applied to both of deconvolution layers and sub-pixel convolution ones. The experimental results demonstrated that the proposed structure can perfectly avoid to generate checkerboard artifacts under two loss functions: mean squared error and perceptual loss, while keeping excellent properties that the SR methods have. As a result, the proposed structure allows us to offer efficient SR methods without any checkerboard artifacts. The proposed structure will be also useful for various computer vision tasks such as semantic segmentation, image synthesis and image generation.

6. REFERENCES

- [1] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a Deep Convolutional Network for Image Super-Resolution," in *Proc. European Conference on Computer Vision (ECCV)*, Springer, 2014, pp. 184–199.
- [2] 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.
- [3] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 1646–1654.
- [4] J. Kim, J. K. Lee, and K. M. Lee, "Deeply-recursive convolutional network for image super-resolution," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 1637–1645.
- [5] Y. Tai, J. Yang, and X. Liu, "Image super-resolution via deep recursive residual network," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 2790–2798.
- [6] M. D. Zeiler, G. W. Taylor, and R. Fergus, "Adaptive deconvolutional networks for mid and high level feature learning," in *Proc. IEEE International Conference on Computer Vision (ICCV)*, 2011, pp. 2018–2025.
- [7] W. Shi, J. Caballero, F. Huszar, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 1874–1883.
- [8] C. Dong, C. C. Loy, and X. Tang, "Accelerating the super-resolution convolutional neural network," in *Proc. European Conference on Computer Vision (ECCV)*, Springer, 2016, pp. 391–407.
- [9] W. Lai, J. Huang, N. Ahuja, and M. Yang, "Deep laplacian pyramid networks for fast and accurate super-resolution," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 5835–5843.
- [10] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. P. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 105–114.
- [11] M. S. M. Sajjadi, B. Schölkopf, and M. Hirsch, "EnhanceNet: Single Image Super-Resolution through Automated Texture Synthesis," in *Proc. IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 4491–4500.
- [12] Y. Sugawara, S. Shiota, and H. Kiya, "A parallel computation algorithm for super-resolution methods using convolutional neural networks," in *Proc. Asia Pacific Signal and Information Processing Association (APSIPA) Annual Summit and Conference*, 2017, pp. 1169–1173.
- [13] J. Johnson, A. Alahi, and F. Li, "Perceptual losses for real-time style transfer and super-resolution," in *Proc. European Conference on Computer Vision (ECCV)*, 2016, pp. 694–711.
- [14] A. Odena, V. Dumoulin, and C. Olah, "Deconvolution and checkerboard artifacts," *Distill*, 2016. [Online]. Available: <http://distill.pub/2016/deconv-checkerboard>
- [15] A. Dosovitskiy, P. Fischer, E. Ilg, P. Häusser, C. Hazirbas, V. Golkov, P. Smagt, D. Cremers, and T. Brox, "Flownet: Learning optical flow with convolutional networks," in *Proc. IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 2758–2766.
- [16] A. P. Aitken, C. Ledig, L. Theis, J. Caballero, Z. Wang, and W. Shi, "Checkerboard artifact free sub-pixel convolution: A note on sub-pixel convolution, resize convolution and convolution resize," *arXiv preprint arXiv:1707.02937*, 2017.
- [17] Z. Wojna, V. Ferrari, S. Guadarrama, N. Silberman, L. C. Chen, A. Fathi, and J. Uijlings, "The devil is in the decoder," in *Proc. British Machine Vision Conference (BMVC)*, 2017.
- [18] H. Gao, H. Yuan, Z. Wang, and S. Ji, "Pixel deconvolutional networks," *arXiv preprint arXiv:1705.06820*, 2017.
- [19] Y. Harada, S. Muramatsu, and H. Kiya, "Multidimensional multirate filter without checkerboard effects," in *Proc. European Signal Processing Conference (EUSIPCO)*, 1998, pp. 1881–1884.
- [20] T. Tamura, M. Kato, T. Yoshida, and A. Nishihara, "Design of checkerboard-distortion-free multidimensional multirate filters," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E81-A, no. 8, pp. 1598–1606, 1998.
- [21] Y. Harada, S. Muramatsu, and H. Kiya, "Multidimensional multirate filter and filter bank without checkerboard effect," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E81-A, no. 8, pp. 1607–1615, 1998.
- [22] H. Iwai, M. Iwahashi, and H. Kiya, "Methods for avoiding the checkerboard distortion caused by finite word length error in multirate system," *IEICE Transactions on Fundamentals of Electronics, Communications, and Computer Sciences*, vol. E93-A, no. 3, pp. 631–635, 2010.
- [23] A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," in *Proc. ICML Workshop on Deep Learning for Audio, Speech and Language Processing*, 2013.
- [24] 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.
- [25] M. Bevilacqua, A. Roumy, C. Guillemot, and M. L. Alberi-Morel, "Low-complexity single-image super-resolution based on nonnegative neighbor embedding," in *Proc. British Machine Vision Conference (BMVC)*, 2012.
- [26] R. Zeyde, M. Elad, and M. Protter, "On single image scale-up using sparse-representations," in *Proc. Curves and Surfaces*, 2010, pp. 711–730.
- [27] D. P. Kingma, and J. Ba, "Adam: A method for stochastic optimization," in *Proc. International Conference on Learning Representations (ICLR)*, 2015.
- [28] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proc. IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 1026–1034.
- [29] M. Abadi, A. Agarwal, P. Barham, et al, "Tensorflow: Large-scale machine learning on heterogeneous systems," 2015, software available from tensorflow.org. [Online]. Available: <http://tensorflow.org/>