

Optimised Highway Deep Learning Network for Fast Single Image Super-Resolution Reconstruction

Viet Khanh Ha¹ · Jinchang Ren^{2, 1} · Xinying Xu² · Wenzhi Liao¹ ·
Sophia Zhao¹ · Jie Ren³ · Gaowei Yan²

Received: date / Accepted: date

Abstract

With the success of the deep residual network for image recognition tasks, the residual connection or skip connection has been widely used in deep learning models for various vision tasks, including single image super-resolution (SISR). Most existing SISR approaches pay particular attention to residual learning, while few studies investigate highway connection for SISR. Although skip connection can help to alleviate the vanishing gradient problem and enable fast training of the deep network, it still provides the coarse level of approximation in both forward and backward propagation paths and thus challenging to recover high-frequency details. To address this issue, we propose a novel model for SISR by using highway connection (HNSR), which composes of a non-linear gating mechanism to further regulate the information. By using the global residual learning and replacing all local residual learning with designed gate unit in highway connection, HNSR has the capability of efficiently learning different hierarchical features and recovering much more details in image reconstruction. Experimental results have validated that HNSR can provide not only improved quality but also less prone to a few common problems during training. Besides, the more robust and efficient model is suitable for implementation in real-time and mobile systems.

Keywords Single image super-resolution · Highway connection · Residual learning · Gating mechanism.

1 Introduction

Single image super-resolution (SISR) aims to reconstruct a high-resolution (HR) image from its corresponding low-resolution (LR) version. It has attracted increasing interest from both academic and industrial communities for its broad applications in computer vision, face recognition in security and surveillance video, medical imaging, and object detection *ect*. The image reconstruc-

tion approach becomes challenging when the captured images are affected by several factors such as bandwidth, noise, light conditions, and other artifacts.

Super-resolution image reconstruction approach is an ill-posed problem since there exist multiple solutions HR for any given LR image. Super-resolution methods can be divided into three main categories, i.e., interpolation-based, reconstruction-based, and learning-based methods. Among them, interpolation-based methods are the most classical and straightforward, which interpolating missing pixels in the image from its neighborhoods. However, interpolation-based methods tend to smooth the reconstructed image; hence, the information of lost high frequencies could not be reconstructed. Reconstruction-based methods usually predefine certain knowledge priors or constraints such as local structure similarity, non-local means, or edge priors in an inverse reconstruction problem. These knowledge priors are broad and vary depending on a particular dataset, which makes them challenging in practical applications. Lastly, the learning-based methods are capable of learn-

Corresponding author: Jinchang Ren
(jinchang.ren@strath.ac.uk)

¹ Department of Electronic and Electrical Engineering,
University of Strathclyde, Glasgow, G1 1XW UK

² College of Electrical and Power Engineering, Taiyuan
University of Technology, Taiyuan, 030600, China

³ College of Electronics and Information, Xi'an Polytechnic
University, Xi'an, 710048, China

ing prior knowledge from a massive amount of data, and consequently result in complex high dimensional mapping between LR-HR patch pairs, which have shined a new light on SISR [1]. Among learning-based methods, deep learning-based algorithms, especially those based on Convolutional Neural Networks (CNNs), have recently proved extremely powerful for SISR tasks due to their capability in feature extraction and mapping.

The first CNN-based SISR model is introduced by Dong et al., named Super-resolution convolutional neural network (SRCNN) [2]. Although only three layers are used, the results significantly outperform those non-deep learning algorithms. With the success of the deep residual network, Kim et al. proposed a Very Deep Convolutional Networks (VDSR) [3] and a Deeply Recursive Convolutional Network (DRCN) [4], both using 20 convolutional layers. The introduced residual connections have indeed made these models possible to train deep SISR networks for improved quality of reconstructed images.

Tai et al. [5] introduced deeply recursive residual networks (DRRN), which used both a global residual connection and local residual connections and to reduce the number of parameters with recursive convolution inside residual blocks. Laplacian Pyramid has been used for decades, which enables to decomposes an image into a series of high-pass bands and low-pass bands. Instead of using one-step upsampling, Lai et al. [6] proposed Laplacian Pyramid super-resolution networks (Lap-SRN) to reconstruct the image progressively.

Since the success of ResNet [7] in 2015, many refinements have been proposed in SISR. DenseNet [8] is another effective architecture, which has inspired several models such as Memory Network (MemNet [9]), Cascading Residual Network (CARN [10]), and a Residual Dense Network (RDN [11]). [Another approach is to fuse dense connections and progressive learning, which has been applied in Progressive Dilated Residual Densenet \(DRDN \[12\], \[13\]\), and even in generative adversarial network-based model \(G-GANISR \[14\]\).](#) Residual connections from ResNet enable to re-use features, while dense connections from DenseNet enable us to explore new features through the concatenation of all the features from preceding blocks. The common idea shared by these Densenet-based networks is to design a model that operates on both fine time scale (short-term memory) and coarse time scale (long-term memory). To transfer information through a long-time, the Recurrent Network has shown to be great potential in capturing short-time and long-time dependencies.

Based on this new view, we propose in this paper a novel SISR architecture, called Highway Network for Super Resolution (HNSR). Unlike most previous work

on SISR, we replace all local residual connections with highway connections to use features from all previous blocks. With the same target of mitigating gradient vanishing as skip connection, the highway connection stabilizes the training especially in addressing problems such as dying ReLU, gradient exploding and our gating mechanism enable to recover fine detail in lost high-frequencies.

In summary, we highlight our main contribution in the following points:

1. We introduce highway connection based architecture for SISR, which differs most of the existing models, while achieves competitive performances in widely used benchmarks and impressive visual performance.
2. [Our introduced highway connection-based model can achieve faster and better convergence, which is less prominent to dying ReLU and gradient exploding problems than those using skip connections. Experimental results validate the outstanding discriminative learning ability of our model compared with other well-established baselines for SISR.](#)

2 Related work

2.1 Deep CNN-based network for SISR

Convolutional Neural Networks (CNNs) has widely used in various computer vision tasks. By using multiple feature maps, filters or kernels, CNNs are capable of capturing abundant features from the original input. Also, the universal approximation theorem states that neural networks with a single hidden layer can approximate any continuous function. Therefore, the capability in feature extraction and mapping make it very helpful to predict high-frequency details loss in low-resolution image. The CNN-based SISR models attempt to learn the mapping between observed LR and HR patch pairs and then applied to predict a super-resolution (SR) image from an unseen LR image. The pioneer CNN-based SISR model introduced by Dong et al. [2], abbreviated as SRCNN [2], which contained three layers presenting three steps: feature extraction, non-linear mapping, and reconstruction. Later, observing "the deeper, the better", Kim et al. proposed two models named Very Deep Convolutional Network (VDSR [3]) and Deeply Recursive Convolutional Network (DRCN [4]), both stacking 20 convolutional layers. The CNN-based SISR network can go deeper using residual learning. The Enhanced Deep Super Resolution (EDSR [15]) has been drawn research attention in due to its significant improvement in SISR accuracy. There also have several promising architectures to demonstrate different novelty such as Residual Channel Attention Network (RCAN [16]),

Wide Activation Super-resolution (WDSR [17]), Super-resolution Clique Network (SRCliqueNet [18]), Residual Dense Network (RDN [11]), Information Distillation Network (IDN [19]), and Cascading Residual Network (CARN [10]), *etc.* However, as performance improves, computation complexity and storage requirements become a significant issue in real-time applications. Also, to apply reproduced CNN-based models in real-world applications can be difficult if those models are sensitive to gradient vanishing/exploding, inappropriate learning rate. Hence, the design of the model should consider such common problems in CNN training.

2.2 Skip connection and Highway connection

Training the deep learning networks can be challenging due to several reasons, including the gradient vanishing and information morphs problems. Let x_n denote the network's input at layer n , constantly transforming at each layer $x_{n+1} = T(x_n)$ leads to information morphs, where it is difficult to exploit the best usable information in the past properly. Instead, both Residual Networks (ResNet [7]) and Highway Network [20] can be regarded as an application of LSTM, following the similar way of any state change: $x_{n+1} = x_n + \Delta x_{n+1}$. ResNet indeed does exactly, which utilizes extra identity connections to enhance information flow such that very deep neural networks can be effectively optimized. Such skip connections guarantee the direct propagation of signals among different layers, thereby avoid gradient vanishing and also information morphs. Given the input x_n at layer n , after being transformed $F_n(x_n, W_{fn})$, the output at layer $n + 1$ can be represented as:

$$x_{n+1} = x_n + F_n(x_n, W_{fn}). \quad (1)$$

where $F_n(x_n, W_{fn})$ is equivalent to Δx_{n+1} , the residual between x_n and x_{n+1} .

In practice, although gradient vanishing has solved, the subsequent change in distribution through the network can still lead to dying ReLU or gradient exploding problems. For example, if F function is as $Conv_1 - RELU - Conv_2$, and the incoming neurons to ReLU are entirely on the negative range, the backpropagation gradients through ReLU will vanish, thus $Conv_1$ is hardly to learn. Over time, a large part of the network will possibly leave unused if such neurons are unable to recover from negative. In other words, the ReLU is always dying for those neurons. The converse of the range could lead to gradient exploding. The ideal distribution of the input to ReLU should be symmetric with a zero mean. For that reason, by using of Batch-normalization [21] helps to normalize the layer before ReLU activation or if possible to find another way to achieve that.

The Highway Network [20] is another approach to solve gradient vanishing.

$$x_{n+1} = \sigma_n \odot x_n + (1 - \sigma_n) \odot \Delta x_{n+1} \quad (2)$$

where σ_n is a sigmoid function ($0 \leq \sigma_n \leq 1$) with trainable parameters, and \odot is a Hadamard product or element-wise product. The best usable neurons in the past can be exploited by adaptively setting a particular σ_i to 1, avoiding gradient vanishing.

By using such highway connection in (2), the layer distribution hardly shift to the extreme range in the network, since the output of a layer is always a convex combination of the input and the transformation. This property could allow the model to further increase the learning rate, speeding up the training with minimizing gradient exploding or vanishing. We use a learning rate of $4e-4$ with the baseline model but different connections. We measure the percentage of the positive responses by ReLU activation. Assuming that x_i in \mathbf{x} is independent of each other but shares the same mean and variance. The linear transform before ReLU activation, named pre-activation $\mathbf{z} = \mathbf{w}^T \mathbf{x}$, will approach normal distribution, according to the central limit theorem. The distribution after ReLU activation is skew with $mode = 0, mean \geq 0$, and $standard\ variance \geq 0$. Pearson's coefficient of skewness can be calculated as:

$$sk_p = \frac{mean - mode}{standard\ variance} = \frac{mean}{standard\ variance} \quad (3)$$

The lower sk_p is, the more skewed the distribution is. It is also can interpret $\frac{1}{sk_p}$ as a Coefficient of Variation (CV), which shows how much variance is around the mean in the data. It is well-known that the network training converges faster if the average of each input variable over the training set has a zero mean [22]. From Fig. 1, we observe that the ReLUs activation in skip connection-based model have extremely skewed distribution ($sp_k \approx 0.2$) in most ReLUs activations. This is caused by a significant number of 0 values outputted by ReLU. In other words, the mean of the pre-activation \mathbf{z} is more negative, whose then most values were zeroed out by ReLU activation. A few last ReLUs activations continuously aim to correct the bias shift introduced by previous ReLUs. At this learning rate, the skip connection-used model can not learn effectively. In contrast, the highway connection-based model keeps all ReLU's $sp_k \approx 0.4$ in a stable range with less correction. This advantage of highway connections can be explained by the convex combination giving an upper estimation of expectation of all Frobenius p-norm lower than that of skip connections.

$$\begin{aligned} \mathbb{E}[\|\mathbf{x}_{n+1}\|_p] &\leq \mathbb{E}[\|c_n \odot \mathbf{x}_n\|_p] + \mathbb{E}[\|(1 - c_n) \odot \mathbf{h}_n\|_p] \\ &< \mathbb{E}[\|\mathbf{x}_n\|_p] + \mathbb{E}[\|\mathbf{h}_n\|_p] \end{aligned} \quad (4)$$

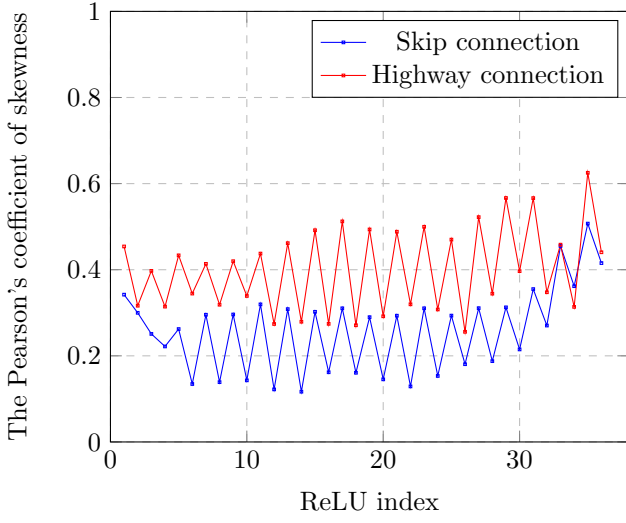


Fig. 1: The Pearson's coefficient of skewness (sp_k) of ReLU's activation vs the network depth.

where the first inequality hold by Minkowski inequality and the second inequality hold since c_n is a sigmoid function ($0 \leq c_n, 1 - c_n \leq 1$.)

The network input, x_0 , is typically normalized to have a zero mean and a variance of 1, hence there is less chance for highway connection-based model to push the mean of consequent layer far away from 0 in comparison with those using skip connections.

3 The proposed model

In this section, we describe our proposed Highway Network for Super Resolution (HNSR). We construct a *carry* gate that is inspired by Gated Recurrent Unit (GRU [23]). The only global residual connection was used in our model. For each HNSR block, assume that x_n is the network's representation of network input x_0 at layer n . Let $h_n = F_n(x_n, W_{fn})$ be the intermediate transform function of input x_n , then the $c_n = C_n([x_n, h_n], W_{cn})$, $t_n = T_n([x_n, h_n], W_{tn})$ are *carry* and the *transform* gate, typically utilize a sigmoid nonlinear function. The transform gate t_n is set to $1 - c_n$. The *carry* gate bias, b_{cn} as following is set to +1 at the start of training. Given the input x_n , our HNSR model is defined as follow:

$$h_n = F_n(W_{fn}x_n + b_{fn}). \quad (5)$$

$$c_n = C_n(W_{cn}[x_n, h_n] + b_{cn}). \quad (6)$$

$$x_{n+1} = c_n \odot x_n + (1 - c_n) \odot h_n. \quad (7)$$

where W denote trainable weights, and b is trainable biases.

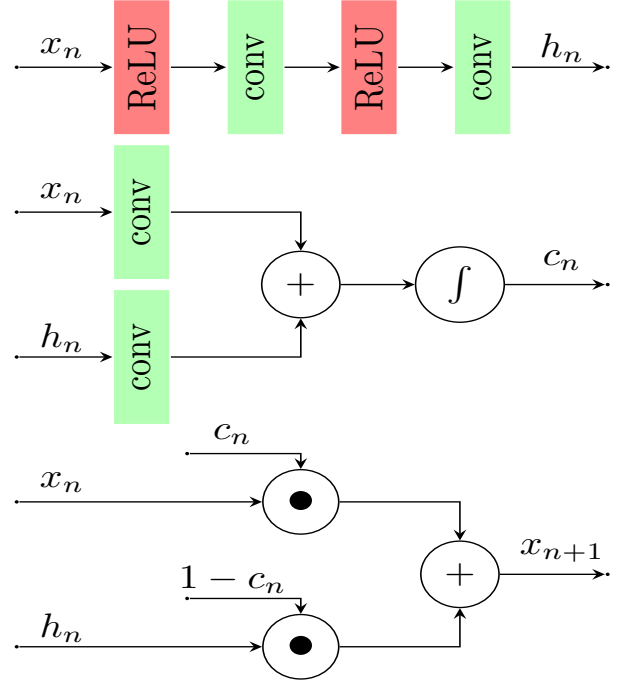


Fig. 2: The structure of each HNSR block, transforming input x_n to output x_{n+1} .

The motivation behind this model is two-fold. First, we use highway connection to constraint the distribution of the output at the end of each block not going out of the optimal range for ReLU's activation, which helps speed up the training. The significant change of the distribution will cause dying ReLU's, gradient exploding, or the ReLU's activation has continuously to correct the *bias shift*, which leads to slower training [24]. Second, we form our attention mechanism, combining both the input x_n and intermediate output h_n to enhance the discriminative learning ability. Since x_n and h_n have many identical features, combining both of them will facilitate c_n to decide whether to disregard or retain features in x_n . This design is the difference between ours and the original Highway Network [20], where the latter blinds regulate information based on x_n only. The transform function F_n has the same design as in a Pre-activation Residual block [25], i.e. ReLU-Conv-ReLU-Conv without Batch normalization.

4 Experiments

4.1 Experiment settings

We use Tensorflow as a framework to implement our model. Our model is evaluated by comparing the test accuracy with other architectures on the same dataset.

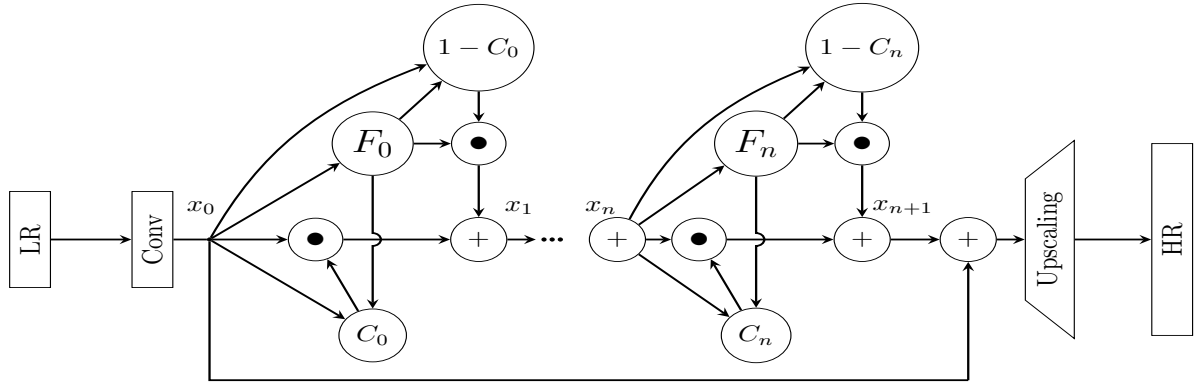


Fig. 3: A proposed HNSR model

Nvidia GeForce GTX 1080 is utilized to conduct our experiments.

4.2 Dataset

We use 800 training images from the DIV2K dataset [26] as a training set. To fairly compare with other state-of-the-art methods, we use four benchmark datasets for testing: Set5 [27], Set14 [28], B100 [29], and Urban100 [30].

We use input image as three color channels RGB rather than luminance (Y) channel in SRCNN [2]. For each training mini-batch, we crop 32 random LR patches and their corresponding HR patches as the input and the ground truth, respectively. To augment the training data, we randomly rotated by 90, 180, 270 and horizontally flipped. We normalize the input by subtracting the mean from each pixel and then dividing the result by the standard deviation. We also detect and remove noisy patches from the training dataset. We use mean squared error (MSE) as the loss function. Given a training set $\{I_{LR}^i, I_{HR}^i\}_{i=1}^N$, which contains N pairs of LR inputs and their HR counterparts. The goal of training is to minimize the loss function:

$$L(\theta) = \sum_{i=1}^N \|I_{SR}^i - I_{HR}^i\|^2. \quad (8)$$

where θ denotes the parameter set of our network. We choose subpixel up-sampling as a method that is similar to previous work.

4.3 Hyperparameters

The batch size is set to 32. The initial learning rate is $4e-4$, which is decreased if validation loss does not decrease after two epochs. We stopped training when the loss ceased to fall after three successive decreases in the

learning rate. We use a checkpoint of the best validation accuracy to evaluate the test accuracy. Adam optimization with default parameters is utilized for training.

4.4 Network depth

We set up a model with 18 highway blocks between the first layer and the last up-scaling layer. Referring to Fig. 2, we use 3×3 convolutional filters in all F_n , and 1×1 convolutional filters in all C_n .

5 Results

5.1 Benchmark results

Here, we provide quantitative comparisons for x2, x3, and x4 SR results in Table 1. We compare our proposed method with bicubic interpolation and the following the state-of-the-art SR method SRCNN [2], FSRCNN [31], VDSR [3], DRCN [4], LapSRN [6], DRRN [5], Memnet [9], SelNet [32], and CARN [10], which were considered as small-size models. Following [2], we evaluated the performance with the Peak Signal-to-Noise Ratio (PSNR) and structural similarity index (SSIM [33]) on the Y channel (luminance) after transforming the images to the YCbCr space. As can be seen from Table 1, our proposed model achieved better performance in terms of SSIM, especially with large testing datasets such as Urban 100 and B100. Our PSNRs are slightly lower than those of CARN [10]. It may be linked to underfit problems in gates, with a small number of the parameter (around $64 \times 1 \times 1 \times 64 = 4096$ parameters) each block functioning with the data of around 1,000k patches. Therefore, it is important to increase expressive transformations for *carry* gate in our networks.

Table 1: Average PSNR/SSIMs for scale 2x, 3x and 4x. **Red** color indicates the best, **blue** color indicates the second best performance, and missing information that was not provided by the authors is marked by [-/-].

Scale	Model	Params	Set5 PSNR/SSIM	Set14 PSNR/SSIM	B100 PSNR/SSIM	Urban100 PSNR/SSIM
2	SRCNN	57K	36.66/0.9542	32.42/0.9063	31.36/0.8879	29.50/0.8946
	FSRCNN	12K	37.00/0.9558	32.63/0.9088	31.53/0.8920	29.88/0.9020
	VDSR	665K	37.53/0.9587	33.03/0.9124	31.90/0.8960	30.76/0.9140
	DRCN	1,774K	37.63/0.9588	33.04/0.9118	31.85/0.8942	30.75/0.9133
	LapSRN	813K	37.52/0.9590	33.08/0.9130	31.80/0.8950	30.41/0.9100
	DRRN	297K	37.74/0.9591	33.23/0.9136	32.05/0.8973	31.23/0.9188
	MemNet	677K	37.78/0.9597	33.28/0.9142	32.08/0.8978	31.31/0.9195
	SelNet	974K	37.89 /0.9598	33.61 /0.9160	32.08/0.8984	-/-
	IDN	796K	37.83/0.9600	33.30/0.9148	32.08/0.8985	31.27/0.9196
	CARN	1,592K	37.76/0.9590	33.52 /0.9166	32.09 /0.8978	31.92 /0.9256
	HNSR (ours)	1,487K	37.89 /0.9603	33.33/0.9158	32.13 /0.8999	31.49 /0.9322
3	SRCNN	57K	32.75/0.9090	29.28/0.8209	28.41/0.7863	26.24/0.7989
	FSRCNN	12K	33.16/0.9140	29.43/0.8242	28.53/0.7910	26.43/0.8080
	VDSR	665K	33.66/0.9213	29.77/0.8314	28.82/0.7976	27.14/0.8279
	DRCN	1,774K	33.82/0.9226	29.76/0.8311	28.80/0.7963	27.15/0.8276
	DRRN	297K	34.03/0.9244	29.96/0.8349	28.95/0.8004	27.53/0.8378
	MemNet	677K	34.09/0.9248	30.00/0.8350	28.96/0.8001	27.56/0.8376
	SelNet	974K	34.27 /0.9257	30.30 /0.8399	28.97/0.8025	-/-
	IDN	796K	34.11/0.9253	29.99/0.8354	28.95/0.8013	27.42/0.8359
	CARN	1,592K	34.29 /0.9255	30.29 /0.8407	29.06 /0.8034	28.06 /0.8493
	HNSR(ours)	1,487K	34.27 /0.9262	30.06/0.8391	29.04 /0.8053	28.04 /0.8515
4	SRCNN	57K	30.48/0.8628	27.49/0.7503	26.90/0.7101	24.52/0.7221
	FSRCNN	12K	30.71/0.8657	27.59/0.7535	26.98/0.7150	24.62/0.7280
	VDSR	665K	31.35/0.8838	28.01/0.7674	27.29/0.7251	25.18/0.7524
	DRCN	1,774K	31.53/0.8854	28.02/0.7670	27.23/0.7233	25.14/0.7510
	LapSRN	813K	31.54/0.8850	28.19/0.7720	27.32/0.7280	25.21/0.7560
	DRRN	297K	31.68/0.8888	28.21/0.7720	27.38/0.7284	25.44/0.7638
	MemNet	677K	31.74/0.8893	28.26/0.7723	27.40/0.7281	25.50/0.7630
	SelNet	974K	32.00 /0.8931	28.49 /0.7783	27.44/0.7325	-/-
	IDN	796K	31.82/0.8903	28.25/0.7730	27.41/0.7297	25.41/0.7632
	CARN	1,592K	32.13 /0.8937	28.60 /0.7806	27.58 /0.7349	26.07 /0.7837
	HNSR(ours)	1,487K	31.98/0.8927	28.34/0.7775	27.53 /0.7359	25.97 /0.7840

5.2 Visual performance

1). Comparison with skip connections

Since the learning rate of $4e-4$ is not optimal for skip connection-based model as being discussed previously, we choose it as $1e-4$ for a fair comparison. As can be seen from Fig. 4, the highway-based model show marginally faster convergence than skip-based model right from the beginning, achieving a 27 dB accuracy at a step of 13.6k compared with 23.4k in the model with skip connections. At the step of 604k, the skip-based model is unable to improve further, while highway-based methods continue to learn more before stopping at the step of 676k. This observation is compatible with the results in Fig. 5, where we evaluate the loss of training and validation at the end of each epoch.

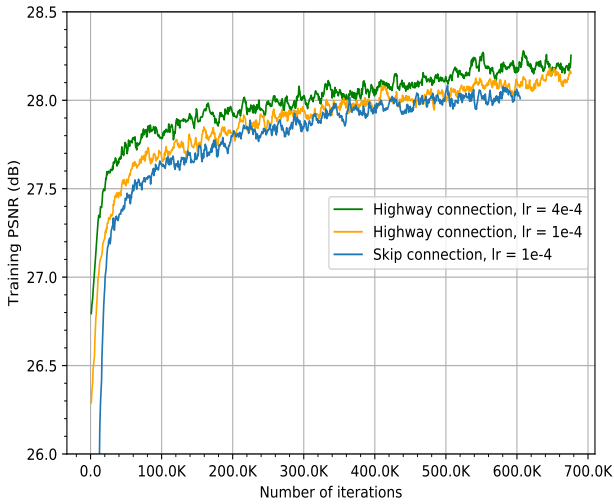


Fig. 4: Training PSNR of the model with different types of connections. All parameters were initialized with the same seed values.

As can be seen from Fig. 5, the highway-based models outperform those using skip-connections in both train and validation evaluation. Increasing the learning rate to $4e-4$ enables our model to converge faster for choosing a better optimal solution. In contrast, the performance of the skip connection-based model at that learning rate is worst, which verifies our observation in Fig. 1. With the same initialization, our model takes the benefit of convex combination property in highway connection, speeds up the training convergence. Meanwhile, the proposed attention mechanism enhances the discriminative learning ability, and enables to be further converged.

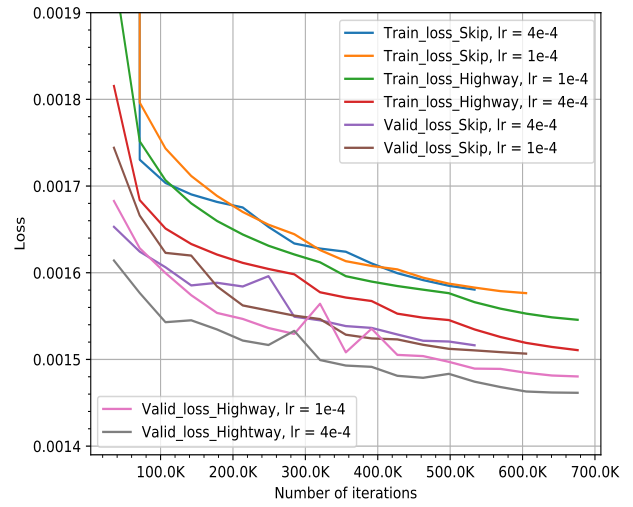


Fig. 5: Training and validation loss of models with different connections and learning rates.

2). Visual gate unit

To examine the attention of gates, we extract the *carry* gate, which is a sigmoid function. We average the feature map of all the channel dimensions, then normalized to a range from 0 to 1, before associating with a heat map. Note that the gates regulate the information; they do not capture features to feed to the next layer. As seen in Fig. 6, the carry gates do not assign the same priority to a different position but give distinct attention to some specific regions. Because the CNNs have complicated interaction, it is difficult to explain why each gate gives particular attention. However, we can see that the focus on one specific area is reducing or emphasising on the successor layer. Finally, the *carry* gate 18 achieves a balance for focused intensity.

3). Visual performance on test images

As seen from Fig. 7, for *image_067*, most of the compared methods produce blurring artifacts along the diagonal lines, while our HNSR produces more sharp, faithful details. To make a fair comparison, we use *image_083*, which achieved slightly lower PSNR/SSIM, to demonstrate the details reconstruction ability. As seen from Fig. 8, most of the compared methods cannot recover the lattices and would suffer from blurring artifacts. In contrast, our HNSR can alleviate the blurring artifacts better and recover more details. It can be explained by the help of the gating mechanism, to forget irrelevant parts instead of remembering all features from the previous layer. Such obvious comparisons demonstrate that the gating mechanism provides more a powerful

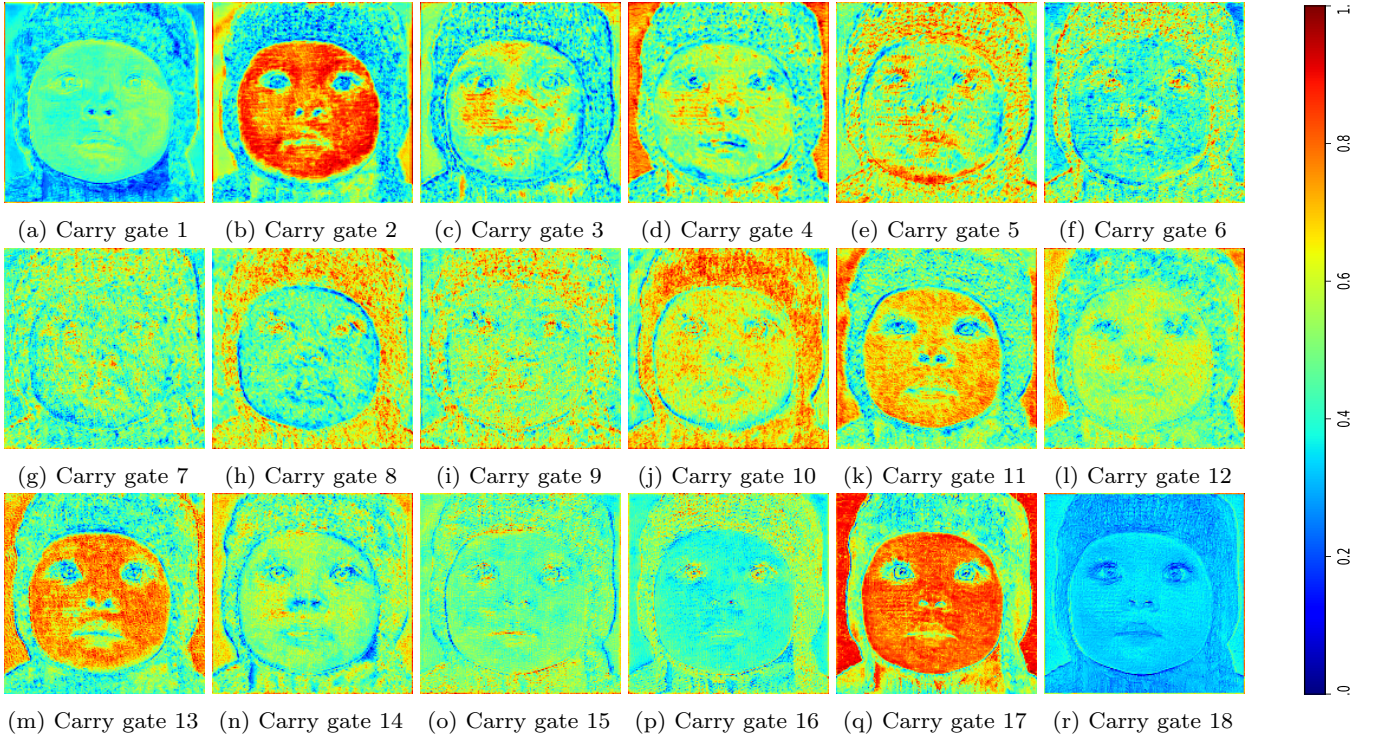


Fig. 6: The attention of 18 *carry* gates in 18 HNSR blocks on the image of a “baby” in Set5. The colormap from 0 to 1 shows the increasing level of attention on a particular area on the image.

representational ability to extract sophisticated features from the LR space. While focusing on recovering high-frequency details, however, the HNSR model pays less attention to low-frequency information. It is typically the behavior of attention-based networks to focus on selective feature maps that more important for the end

task. This may be considered a promising method for saliency detection [34, 35], where the attentional ability focusing on the region of interests (ROIs) play an important key. Besides, by using 1×1 convolutional filters, our attention mechanism regulates the information base on the cross-channel correlations. It is possible to

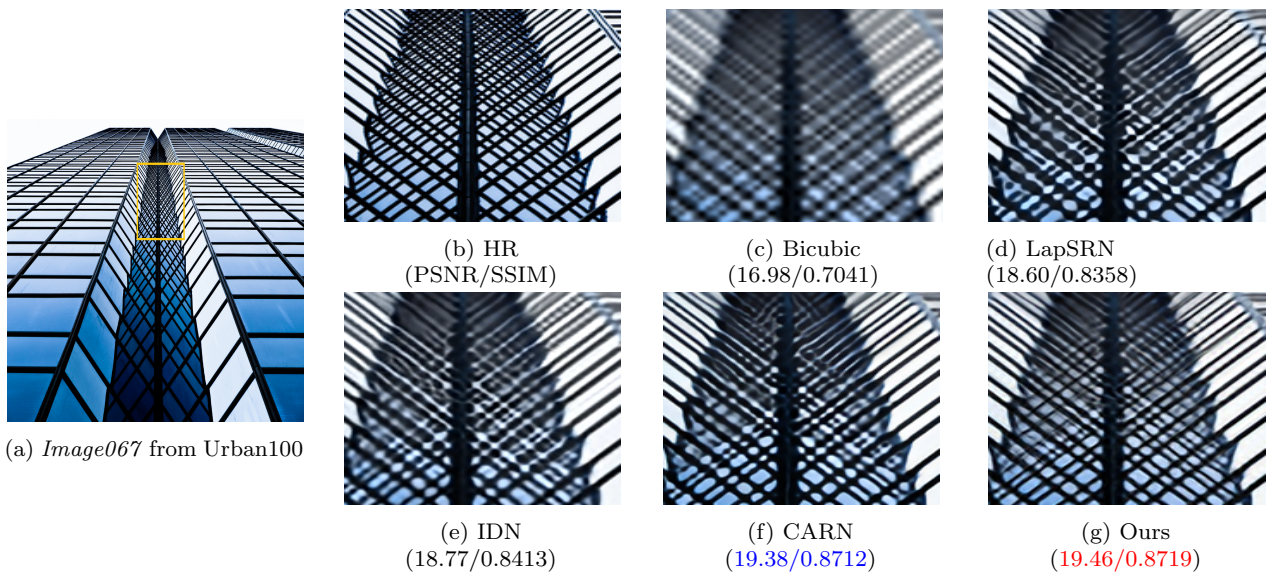


Fig. 7: Visual qualitative comparison on the *Image067*, Urban100 dataset, magnified by factor of 4.

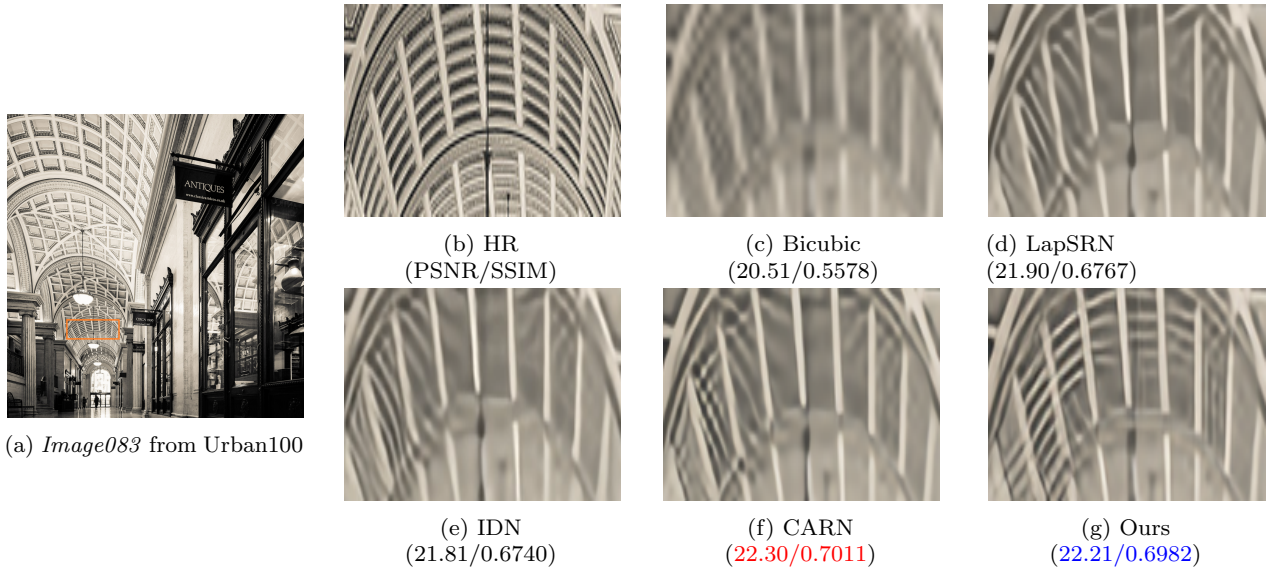


Fig. 8: Visual qualitative comparison on the *Image083*, Urban100 dataset, magnified by factor of 4.

investigate the application in channel selection or band selection [36].

6 Conclusion

Our proposed SISR model has a similar structure as ResNet except for using highway connections instead of residual connections. It achieved high performance and strong recovery capability in the textured area in comparison with recent model architectures. Moreover, it does not require implementing a complicated structure and enables stable training with less problem of gradient exploding or dying neuron. As this architecture has achieved outstanding performance, we intend to bound parameters to perform contraction mapping, which can further lead to speed up convergence. We also plan to generalize our method for other applications, such as hyperspectral images, as well as real-time super-resolution image reconstruction in mobile environments.

Acknowledgements The authors would like acknowledge the support from the Shanxi Hundred People Plan of China, the ETF Scholarship from the Faculty of Engineering, University of Strathclyde, and the Government Scholarship of Vietnam. The authors would greatly thank our colleagues from the Image Processing Group in Strathclyde University for their valuable suggestions.

Conflict of interest

The authors declare that they have no conflict of interest.

References

1. Ha, V.K., Ren, J., Xu, X. et al.: Deep Learning Based Single Image Super-resolution: A Survey. *Int. J. Autom. Comput.* 16, 413–426 (2019).
2. C. Dong, C. C. Loy, K. M. He, X. O. Tang.: Image super-resolution using deep convolutional networks. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 2, pp. 295–307 (2016).
3. J. Kim, J. Kwon Lee, K. Mu Lee.: Accurate image super-resolution using very deep convolutional networks. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1646–1654 (2016).
4. J. Kim, J. Kwon Lee, K. Mu Lee.: Deeply-recursive convolutional network for image super-resolution. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1637–1645 (2016).
5. Y. Tai, J. Yang, X. M. Liu.: Image super-resolution via deep recursive residual network. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 2790–2798 (2017).
6. W. S. Lai, J. B. Huang, N. Ahuja, M. H. Yang.: Deep laplacian pyramid networks for fast and accurate super-resolution. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, IEEE*, vol. 2, pp. 5835–5843 (2017).
7. He, K., Zhang, X., Ren, S. and Sun, J.: Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778 (2016).
8. Huang, G., Liu, Z., Van Der Maaten, L. and Weinberger, K.Q.: Densely connected convolutional networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4700–4708 (2017).
9. Y. Tai, J. Yang, X. M. Liu, C. Y. Xu.: MemNet: A persistent memory network for image restoration. In: *Proceedings of IEEE International Conference on Computer Vision, IEEE*, pp. 4539–4547 (2017).
10. N. Ahn, B. Kang, K. A. Sohn.: Fast, accurate, and lightweight super-resolution with cascading residual network. In: *Proceedings of the 15th European Conference on Computer Vision, Springer*, pp. 252–268 (2018).
11. Y. L. Zhang, Y. P. Tian, Y. Kong, B. N. Zhong, Y. Fu.: Residual dense network for image super-resolution. In: *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition, IEEE*, pp. 2472–2481 (2018).
12. Shamsolmoali, P., Li, X. and Wang, R.: Single image resolution enhancement by efficient dilated densely connected residual network. *Signal Processing: Image Communication*, 79, pp.13–23 (2019).
13. Shamsolmoali, P., Zhang, J. and Yang, J.: Image super resolution by dilated dense progressive network. *Image and Vision Computing*, 88, pp.9–18 (2019).
14. Shamsolmoali, P., Zareapoor, M., Wang, R., Jain, D.K. and Yang, J.: G-GANISR: Gradual generative adversarial network for image super resolution. *Neurocomputing*, 366, pp.140–153 (2019).
15. Lim, B., Son, S., Kim, H., Nah, S., Mu Lee, K.: 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).
16. Y. L. Zhang, K. P. Li, K. Li, L. C. Wang, B. N. Zhong, Y. Fu.: Image super-resolution using very deep residual channel attention networks. In: *Proceedings of the 15th European Conference on Computer Vision, Springer*, pp. 286–301 (2018).
17. J. H. Yu, Y. C. Fan, J. C. Yang, N. Xu, Z. W. Wang, X. C. Wang, T. Huang.: Wide Activation for Efficient and Accurate Image Super-resolution. *arXiv:1808.08718* (2019).
18. Z. S. Zhong, T. C. Shen, Y. B. Yang, Z. C. Lin, C. Zhang.: Joint sub-bands learning with clique structures for wavelet domain super-resolution. In: *Proceedings of the 32nd Conference on Neural Information Processing Systems, Curran Associates, Inc.*, pp. 165–175 (2018).
19. Z. Hui, X. M. Wang, X. B. Gao.: Fast and accurate single image super-resolution via information distillation network. In: *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition, IEEE*, pp. 723–731 (2018).
20. Srivastava, R.K., Greff, K. and Schmidhuber, J.: Highway networks. *arXiv:1505.00387* (2015).
21. Ioffe, S. and Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv:1502.03167* (2015).
22. LeCun, Y.A., Bottou, L., Orr, G.B. and Müller, K.R.: Efficient backprop. In: *Neural networks: Tricks of the trade*, pp. 9–48 (2012).
23. Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H. and Bengio, Y.: Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv:1406.1078* (2014).
24. Clevert, D.A., Unterthiner, T. and Hochreiter, S.: Fast and accurate deep network learning by exponential linear units (elus). *arXiv:1511.07289* (2015).
25. He, K., Zhang, X., Ren, S. and Sun, J.: Identity mappings in deep residual networks. In: *European conference on computer vision*, pp. 630–645 (2016).
26. R. Timofte, E. Agustsson, L. Van Gool, M.-H. Yang, L. Zhang, et al.: Ntire 2017 challenge on single image superresolution: Methods and results. In: *CVPR 2017 Workshops* (2017).
27. M. Bevilacqua, A. Roumy, C. Guillemot, M. L. Alberi-Morel.: Low-complexity single-image super-resolution based on nonnegative neighbor embedding. In: *Proceedings of British Machine Vision Conference* (2012).
28. R. Zeyde, M. Elad, M. Protter.: On single image scale-up using sparse-representations. In: *Proceedings of the 7th International Conference on Curves and Surfaces, Springer*, pp. 711–730 (2010).
29. D. Martin, C. Fowlkes, D. Tal, J. Malik.: A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: *Proceedings the 8th IEEE International Conference on Computer Vision, IEEE* (2001).
30. J. B. Huang, A. Singh, N. Ahuja.: Single image super-resolution from transformed self-exemplars. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5197–5206 (2015).
31. Dong, C., Loy, C.C. and Tang, X., 2016, October.: Accelerating the super-resolution convolutional neural network. In: *European conference on computer vision, Springer*, pp. 391–407 (2016).
32. Choi, J. S., Kim, M.: A deep convolutional neural network with selection units for super-resolution. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 154–160 (2017).
33. Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli.: Image quality assessment: from error visibility to structural similarity. In: *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612 (2004).
34. Wang, Z., Ren, J., Zhang, D., Sun, M. and Jiang, J.: A deep-learning based feature hybrid framework for spatiotemporal saliency detection inside videos. *Neurocomputing*, 287, pp.68–83 (2018).

35. Yan, Y., Ren, J., Sun, G., Zhao, H., Han, J., Li, X., Marshall, S. and Zhan, J.: Unsupervised image saliency detection with Gestalt-laws guided optimization and visual attention based refinement. *Pattern Recognition*, 79, pp.65-78 (2018).
36. Tschannerl, J., Ren, J., Yuen, P., Sun, G., Zhao, H., Yang, Z., Wang, Z. and Marshall, S.: MIMR-DGSA: Unsupervised hyperspectral band selection based on information theory and a modified discrete gravitational search algorithm. *Information Fusion*, 51, pp.189-200 (2019).