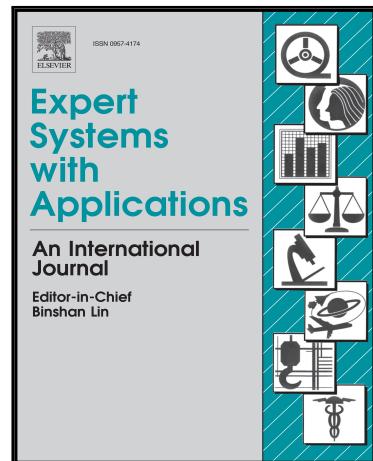


Accepted Manuscript

Fast Deep Parallel Residual Network for Accurate Super Resolution Image Processing

Feng Sha M.D. , Seid Miad Zandavi , Yuk Ying Chung Ph.D

PII: S0957-4174(19)30194-0
DOI: <https://doi.org/10.1016/j.eswa.2019.03.032>
Reference: ESWA 12558



To appear in: *Expert Systems With Applications*

Received date: 15 September 2018
Revised date: 17 March 2019
Accepted date: 17 March 2019

Please cite this article as: Feng Sha M.D. , Seid Miad Zandavi , Yuk Ying Chung Ph.D , Fast Deep Parallel Residual Network for Accurate Super Resolution Image Processing, *Expert Systems With Applications* (2019), doi: <https://doi.org/10.1016/j.eswa.2019.03.032>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Highlights

- New deep 35 convolutional layers structure without loss explosion
- More accurate SR performance than current state-of-the-art methods
- Real time application and human-eye recognition at execution speed of 27fps

ACCEPTED MANUSCRIPT

Title of the paper:

Fast Deep Parallel Residual Network for Accurate Super Resolution

Image Processing

First Author: Feng Sha

Affiliation: School of Computer Science, The University of Sydney, J12/1 Cleveland St, Darlington, NSW 2008, Australia

Email: feng.sha@sydney.edu.au

Co-author: Seid Miad Zandavi

Affiliation: School of Computer Science, The University of Sydney, J12/1 Cleveland St, Darlington, NSW 2008, Australia

Email: miad.zandavi@sydney.edu.au

Co-author: Yuk Ying Chung

Affiliation: School of Computer Science, The University of Sydney, J12/1 Cleveland St, Darlington, NSW 2008, Australia

Email: vera.chung@sydney.edu.au

Corresponding Author: Mr. Feng Sha (M.D.)

Present Address: School of Computer Science, The University of Sydney, J12/1 Cleveland St, Darlington, NSW 2008, Australia

Phone: +61-437063987

Email: feng.sha@sydney.edu.au

<PE-AT>**Fast Deep Parallel Residual Network for Accurate Super Resolution Image Processing**

Author Names: Feng Sha¹, Seid Miad Zandavi¹, Yuk Ying Chung¹

Affiliations: ¹School of Computer Science, The University of Sydney, J12/1 Cleveland St, Darlington, NSW 2008, Australia

Emails: {feng.sha, miad.zandavi, vera.chung}@sydney.edu.au

Corresponding Author: Mr. Feng Sha (M.D.)

Present Address: School of Computer Science, The University of Sydney, J12/1 Cleveland St, Darlington, NSW 2008, Australia

Phone: +61-437063987

Email: feng.sha@sydney.edu.au

Abstract

Recently, Convolutional Network (CNN) has become an excellent solution to solve single image super resolution (SISR) problems. In this paper, a novel Deep Parallel Residual Network (DPRN) has been proposed which has been proved to be a fast and efficient algorithm to solve SISR problems using residual learning. It demonstrated significant advantages in accuracy and speed factors than other approaches with its deeper convolutional network structure, more accurate image reconstruction and real-time model execution. The proposed approach categorizes layers in branches and increases the number of layers to 35 with parallel local residual learning. Also, it adapts the Adam optimizer which can make the proposed system to achieve faster training speed and better image quality. Our experiments have been conducted using standard benchmark datasets such as Set5, Set14, BSDS100 and Urban100 and compared with current state-of-the-art approaches. Our results show that the proposed DPRN can provide higher super resolution quality with real time execution (27.18 fps average) when compare with different state-of-the-art algorithms.

Keywords: Super-resolution, Deep convolutional network, Residual learning

<PE-FRONTEND>

1. Introduction

Many researchers have participated in working on single image super resolution (SISR) problem (Glasner, Bagon, & Irani, 2009) in computer vision domain since 1970s (Duchon, 1979). This typical issue has a wide range of applications including medical imaging, astronomy, machine vision, and auto-driving, etc. There is a common objective in these application areas that more image information is required for further processing. Hence the basic operation for SISR is to gain high-resolution (HR)

images based on low-resolution (LR) images.

Early implementations for SISR offer multiple ideas and algorithms include local linear regression (Timofte, De Smet, & Van Gool, 2014), sparse coding (Jianchao, Wright, Huang, & Ma, 2008), dictionary learning (Yang, Wright, Huang, & Ma, 2010) and random forest (Schulter, Leistner, & Bischof, 2015). These shallow methods have been demonstrated quite successful in both image super resolution theory and application.

As convolutional neural network models blossom in computer vision community, Super-Resolution Convolutional Neural Network abbreviated as SRCNN (Dong, Loy, He, & Tang, 2016) starts a new era for faster and much more accurate outcomes that inspires new algorithms and technologies. It applied fully convolutional structure to perform no-linear mapping from LR to HR, and without further per-engineered features, SRCNN provided significant improvements compared to non-deep learning models. However, as a representative model with small size deep learning network, the learning ability of SRCNN cannot provide satisfied results when managing large and complicated data.

After SRCNN inspires new research in SISR study, many new state-of-the-art methods have appeared in recent years. There are two main methods for enhancement. On one side, some researchers have designed network structure to increase network depth and prevent gradient explosion and disappears. Kim et al. (Kim, Kwon Lee, & Mu Lee, 2016a) proposed that based on Global Residual Learning, deeper network called Very Deep Super-Resolution (VDSR) with 20 convolution layers gains better accuracy and convergence speed in image super resolution. At the same time, Kim et al. (Kim, Kwon Lee, & Mu Lee, 2016b) proposed another network structure as “Deeply-Recursive Convolutional Network” (DRCN) with weight-balanced residual learning in recursive layer design which also demonstrated excellent results. In the most recent years, Deep Laplacian Pyramid Networks for Super-Resolution (LapSRN) (Lai, Huang, Ahuja, & Yang, 2017) introduced an innovative idea which used a cascade learning (pyramid structure) to get output step by step. This network has shown great outcome in upscale $8\times$ condition and it proposed a new loss function. On the other side, some other methods such as Enhanced Deep Residual Networks for Super-Resolution (EDSR) (Lim, Son, Kim, Nah, & Lee, 2017) and a deep convolutional neural network with selection units for super-resolution (SelNet) (Choi & Kim, 2017) had achieved a breakthrough in some existing restrictions by training much higher quality data (Agustsson & Timofte, 2017) which can produce a better resulting system. However, EDSR, SelNet and other methods from NTIRE (Timofte et al., 2017) challenge perform average quality with recent benchmarks when using same training datasets.

In this paper, a novel deep network structure design using parallel residual learning (DPRN) has been proposed to achieve the high quality of image super resolution and provide new layout for deep network with 35 layers. In this new approach, convolutional layers have been grouped as residual combination and put into branches. Each layer gains input information from the first two layers and each branch performs local residual learning to pass branch output to next branch. After up-sampling process, the original data information conducts global residual learning with branches' output to provide the final output. The structure of DPRN can avoid information deterioration during the layer training process while increasing the depth of network, and more information can be learned for convolutional layers in residual combinations. The new network also applied Adam optimizer (Kinga & Adam, 2015) instead of common Stochastic Gradient Descent (SGD) to provide adaptive learning rate for different parameters and reduce resource consuming during training process. The experiment results have demonstrated that proposed DPRN increased 1.08 dB, 0.21 dB, 0.22 dB than SRCNN, VDSR and LapSRN with upscale $2\times$ on set5 (Bevilacqua, Roumy, Guillemot, & Alberi-Morel, 2012) testing

dataset. Furthermore, the model execution time of DPRN is faster than most of the existing methods, the new approach achieves 27.18 fps average value for all 4 test datasets and it demonstrates the capability for real time applications.

This paper is organized in the following five sections. Firstly, it gives the brief introduction of single image super resolution with current state-of-the-art methods and general information about this paper. Section 2 describes the related work. We review Deep residual network (ResNet) concept (He, Zhang, Ren, & Sun, 2016) and some of the recent state-of-the-art methods such as VDSR (Kim et al., 2016a), DRCN(Kim et al., 2016b) and LapSRN (Lai et al., 2017). Section 3 explains the technical parts for the proposed DPRN and explains how our new network can achieve good quality of super resolution images when compared to other existing methods. Finally, the experimental results using benchmark datasets and conclusion for our proposed method have been shown in section 4 and 5 respectively.

2. Related work

Learned from biological processes (Matsugu, Mori, Mitari, & Kaneda, 2003), CNN based models have showed to have a great effect and usability in computer vision applications. In this section, we review recent state-of-the-art methods in SISR area and the superior idea of ResNet (He et al., 2016) which have provided the basis for DPRN. Fig. 1 provides the network structure for ResNet, VDSR and DRCN where ReLU (Nair & Hinton, 2010) and batch normalization (Ioffe & Szegedy, 2015) layers have been hided for clear view purpose.

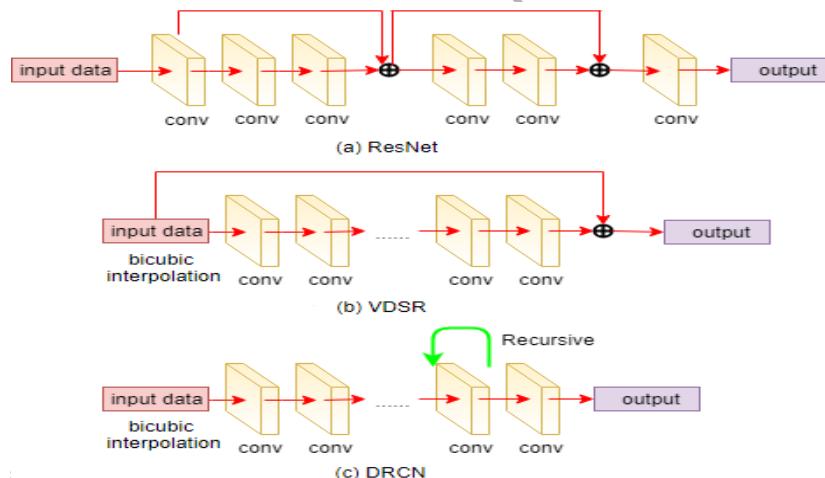


Fig. 1. (a) Simple example for ResNet Structure with two residual units implemented inside; (b) VDSR network with 20 convolutional layers and Global Residual Learning; (c) DRCN with recursive layer to perform recursions.

2.1. ResNet Concept

The excellent concept of ResNet (He et al., 2016) is to combine information from previous layer with information of current layer and deliver them together to next layer, which called residual learning framework. This operation enhances the quality of information learned in each layer and ensures that the information from early training layer can be passed to deeper layers, hence to provide better output in classification challenge through very deep network (e.g. 152 layers).

Denoting x as the input, y_d as output for each residual unit where $d = 1, 2, \dots, D$ refer to number of residual units, and \hat{y} as the output, every convolutional mapping as $H(x)$. The equation

representation for residual unit is as following:

$$y_d = \sigma(f(x, W) + h(x)) \quad (1)$$

where $h(x)$ is an identity mapping (He et al., 2016), W is weights set and function σ represent ReLU operation (Nair & Hinton, 2010).

2.2. VDSR

Instead of creating small residual unit for the network, VDSR (Kim et al., 2016a) proposed Global Residual Learning with common convolutional layers structure. The original data will pass to the end of training network and combine with convolutional residual to provide an output that learned by the input interpolated low-resolution and HR output. VDSR has three contributions: it designed much larger receptive field (41×41) than SRCNN (13×13) (Dong, Loy, He, et al., 2016) which provided wider vision for learning and it used 3×3 convolutional kernels with 64 filters in 20 layers; In order to optimize the training time and convergence speed, VDSR used adjustable gradient clipping and very large learning rate (0.1) so that each training can be finished in 4 hours on their environment; VDSR had trained different size and different scale of image data in one process, thus a single model can be output to support multi-scale super resolution. The excellent results from VDSR have brought deep network research for SISR to next stage rather than early SRCNN.

2.3. DRCN

DRCN (Kim et al., 2016b) proposed another approach to satisfy SISR evaluation. Aim to learn information within the deep network as much as possible, authors tried to add outputs from all previous convolutional layers into last layer called recursive layer and assign different weight to distinguish output, so that the network would not overfit itself. DRCN contains three components: the embedding net represents convolution feature maps from original image; the inference net process the T recursions in a recursive layer where $T=16$ (Kim et al., 2016b); the reconstruction net learned from the output of inference net to generate HR image. Authors also introduced recursive-supervision and skip-connection to optimize the network performance.

2.4. LapSRN

Deep Laplacian Pyramid Networks (LapSRN) (Lai et al., 2017) developed an innovative network structure for accurate and fast Super-Resolution. This method provides two useful contributions. Firstly, it uses cascade learning (pyramid structure) for residual output and reconstruction result from different scales and then gain final output step by step. Secondly, LapSRN conducts a new loss function inspired by Holistically-Nested Edge Detection (Xie & Tu, 2015) rather than traditional L2 loss function.

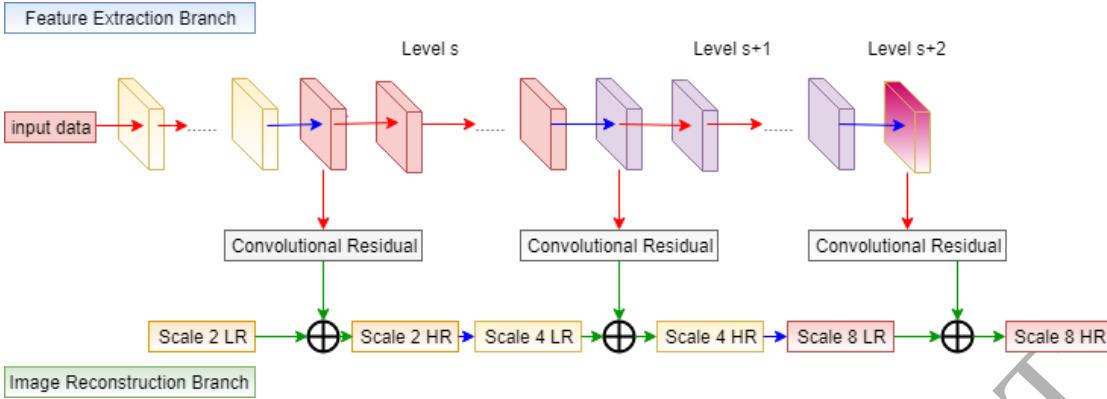


Fig. 2. LapSRN network structure where blue arrow denotes up-sampling to feature extraction in next level, green arrow denotes sum operation in eltwise layer.

Fig. 2 provides a structure view for both feature extraction branch and image reconstruction branch in LapSRN. There are two branches for LapSRN: Feature extraction and Image reconstruction. When input a LR image and start training for S scale super resolution in LapSRN, the network will initialize different levels s according to $\log_2 S$ (e.g. 3 levels for scale 8). Each level in feature extraction branch will build d convolutional layers and one transposed convolutional layer. The output of transposed convolutional layer will pass to two directions, one for residual image reconstruction at current level and the other for feature extraction in $s+1$ level. In image reconstruction branch, transposed convolutional layer will be initialized and jointly optimized with all other layers (Lai et al., 2017). The final output comes from the combination with residual image from feature extraction branch and up-sampled image from image reconstruction branch.

2.5. Summary of current state-of-the-art methods

Utilize the power of convolutional neural network, the current state-of-the-art methods achieve much higher PSNR value to represent their accuracy or amazing speed to represent their ability for industrial application on standard testing dataset. Table 1 briefly introduces some current state-of-the-art methods with their characteristics, advantages and some shortcomings.

Table 1 This table provides 6 current state-of-the-art methods' characteristics, advantages and some shortcomings

Methods	Characteristics	Advantage	Deficiency
A+	Traditional methods, Complicated methodology;	-	low image reconstruction performance, Slow speed;
SRCNN	First adoption of CNN, 3 convolutional layers, Train dataset: Set91;	Fast training speed, Fast execution speed, Fit performance;	Small number of neural network layers;
FSRCNN	Parameter optimization, Train dataset: Set91+general100, Train data zooming and rotation;	Fastest training and model execution speed, Fit performance;	Small number of neural network layers;
VDSR	“the deeper the better”, Residual learning, Increase CNN layer number to 20, Adjustable Gradient Clipping;	Dramatically increase the performance, One model can apply for different scales;	Residual learning only performs once, Can adopt ResNet's advantages;
DRCN	Use Recursive Neural Network and	Similar outcome as VDSR;	Slow training speed,

	Skip-Connection for image super resolution;		Slow execution speed, Consume more computing resource;
LapSRN	Use own loss function instead of L2 loss function, Cascade learning structure;	Fast train speed with adjustable layer definition, Good at reconstruction for large scale like upscale x8;	The performance is similar with VDSR for normal upscale operations like upscale 2, 3 and 4;

3. Deep Parallel Residual Network

In this section, detailed design and technical explanation for DPRN will be presented. We introduce a new approach for connecting multiple residual branches together. Each branch has initial feature mapping at first convolution, the information will pass to Residual Combinations for parallel convolutional training. The first convolutional layer H^0 will conduct local residual learning with output from residual combinations. The result of this branch will deliver to batch normalization layer (Ioffe & Szegedy, 2015) for improving training speed and model accuracy purposes. The output information of each branch will send to next branch, or transposed layers and element-wise operator for Global Residual Learning. The number of residual combinations and branches can be defined when initialize the whole network. Fig. 3 shows the structure of DPRN.

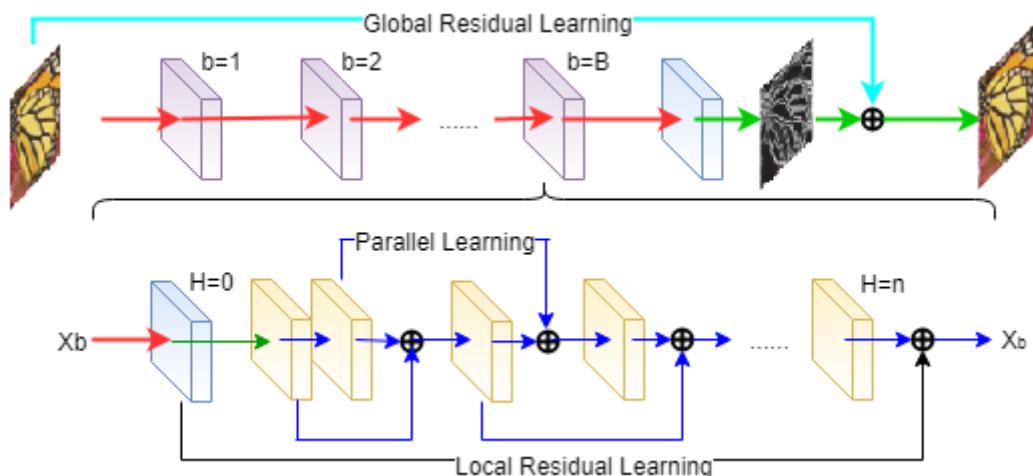


Fig. 3. Proposed DPRN structure with residual branches ($b=1, 2, \dots, B$) located on the top. Green arrows represent up-sampling and Light blue arrow represents Global Residual Learning while black arrow represents local residual learning of each branch. Blue arrows denote parallel convolutional layers that n^{th} convolutional layer gain input from element-wise operation conducts with $(n-1)$ and $(n-2)$ layers.

3.1. Residual Combination

Different from skip-connection in DRCN (Kim et al., 2016b), our residual combination uses parallel residual learning to build the structure for convolutional layers. Fig. 4 shows an example of layer structure. This new approach enhances information adhesiveness because it uses double layers to pass information to the next layer. Each residual combination is formulated as:

$$\mathcal{H}^n = f(\mathcal{H}^{n-2} + \mathcal{H}^{n-1}, \mathcal{W}^n) \quad (2)$$

where $n = 0, 1, 2, \dots, N, N$ represents the number of convolutional layers in one residual branch,

$\mathcal{H}^n, \mathcal{H}^{n-2}, \mathcal{H}^{n-1}$ represent processing operation of three consecutive convolution layers, and f is the residual combine function.

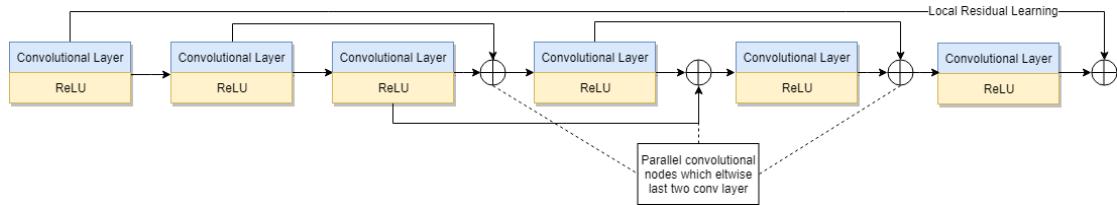


Fig.4. Residual Combination example with 6 convolutional layers: first layer denotes initial feature mapping, second layer passes its output as input of the third layer, also element-wise with output of third layer to generate the input for fourth layer. The same operation applies to fifth and sixth layer.

3.2. Residual Branch

The **Residual Branch** is similar to a group of residual combinations but with a convolutional layer as first layer to perform feature mapping at the beginning and conduct local residual learning later. Fig. 5 shows a typical 3 branches DPRN network structure.

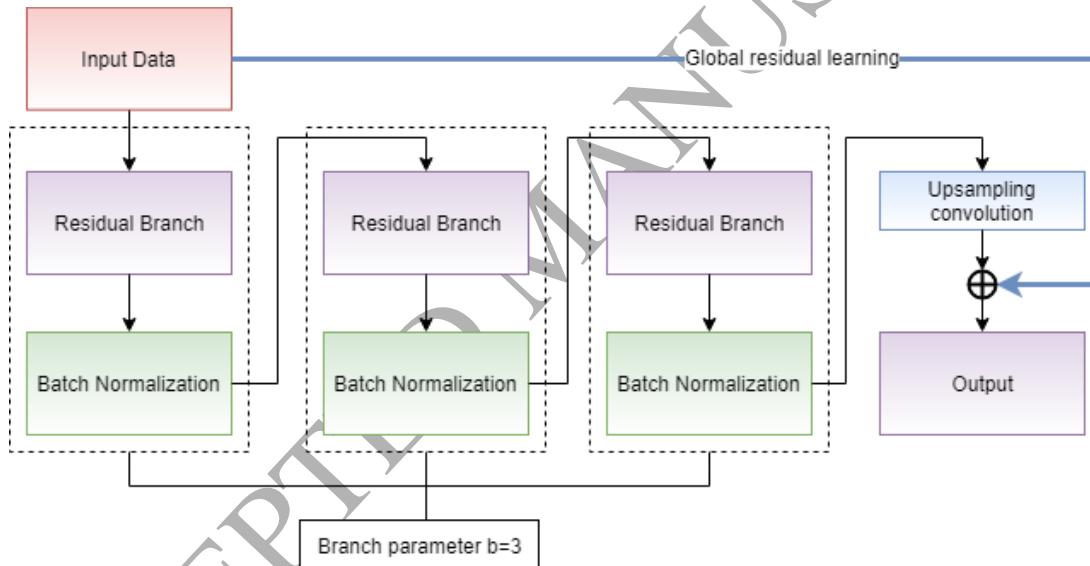


Fig.5. Residual Branch example contains 3 branches.

With multiple branches, we can add residual branch concept and improve equation (2) as following:

$$\mathcal{H}_b^n = f(\mathcal{H}_b^{n-2} + \mathcal{H}_b^{n-1}, \mathcal{W}_b^n) \quad (3)$$

where number of residual branches $b = 1, 2, 3, \dots, B$. After parallel residual learning and local residual learning, the output of a residual branch is represented as

$$\mathcal{X}_b^{n+1} = f(\mathcal{H}_b^n + \mathcal{H}_b^0, \mathcal{W}_b^n) \quad (4)$$

where \mathcal{H}_b^0 denotes the first layer in each branch and \mathcal{X}_b^{n+1} represents the output of current branch and input for next branch or transposed convolutional layer.

3.3. Network Structure

The integral DPRN network consists of residual branches and residual combinations with transposed convolutional layer and Global Residual Learning. There are two main parameters number of branches

b and total layer number n for residual combinations. There can be multiple variety such as one branch structure or complicated networks with large number of b and n definition. The final HR image output can be formulated as

$$\hat{y}_b^i = f(\mathcal{X}_B^{n+1}, \mathcal{W}_B^n) + x^i \quad (5)$$

where \hat{y}_b^i denotes the output of the network and x represents Global Residual Learning.

As the new loss function developed in LapSRN (Lai et al., 2017) is more suitable for its architecture design, we use standard L2 loss function to compute the Euclidean distance between the reconstructed image \hat{y}_b^i and ground truth in order to provide equivalent evaluation support.

Suppose there is a training dataset $\{x^i, \tilde{x}^i\}_{i=1}^M$ where M is the number of training patches. The L2 loss function is try to distinguish the difference between ground truth \tilde{x}^i and the output DPRN network \hat{y}_b^i which uses x^i as input LR image. And the equation to represent loss function for DPRN is

$$L = \frac{1}{2M} \sum_{i=1}^M \| \tilde{x}^i - \hat{y}_b^i \|^2 \quad (6)$$

3.4. Adam Optimizer

Traditional SGD (Bottou, 2012) is a common method used for deep neural network optimization. In recent years, it generally refers to mini-batch gradient descent (Ruder, 2016). Apply this optimizer may confront two serious drawbacks: one is difficulty in the selection of suitable learning rate, and the other drawback is that it may be converged to the local optimum easily.

In order to avoid drawbacks and stabilize parameters, Adaptive Moment Estimation (Adam) optimizer (Kinga & Adam, 2015) has been applied to optimize DPRN objective function. This optimizer combined the advances of AdaGrad (Duchi, Hazan, & Singer, 2011) and RMSprop (Tieleman & Hinton, n.d.), and requires less memory resource. Adam performs very well for DPRN which process large dataset with deeper network structure.

In Adam optimization process, m_t and v_t represent first Moment Estimate and Second Moment Estimate which can help to compute adaptive learning rate for different parameters. According to Caffe implementation (Jia et al., 2014), the updated formula is shown in equation (7) as below:

$$(W_{t+1})_i = (W_t)_i - \alpha \frac{\sqrt{1-(\beta_2)_i^t}}{1-(\beta_1)_i^t} \frac{(m_t)_i}{\sqrt{(v_t)_i} + \epsilon} \quad (7)$$

where $\beta_1=0.9$, $\beta_2=0.999$ and $\epsilon=10^{-8}$ respectively.

4. Experiment

In this section, we compare the proposed DPRN with several current state-of-the-art methods on standard testing datasets. The results include quantitative and qualitative evaluation of the accuracy performance, runtime comparison and model convergence trend. The experiment shows our DPRN approach achieves the best accuracy performance for image reconstruction, while the average model execution time reaches the requirement for real-time human eye recognition (>24fps). We also discuss the limitation of our new approach base on the experiment results.

4.1. Dataset

In our experiments, we have used Yang et al. (Yang et al., 2010) and Berkeley Segmentation Dataset (Martin, Fowlkes, Tal, & Malik, 2001) in the training phase of our proposed DPRN model. These two datasets have 91 and 200 images respectively and they have also been used in VDSR (Kim et al., 2016a) DRCN (Kim et al., 2016b) and LapSRN (Lai et al., 2017) training for comparison purpose. For testing phase, we considered four widely used benchmark datasets which include Set5 (Bevilacqua et al., 2012) Set14 (Zeyde, Elad, & Protter, 2010), BSDS100 (Martin et al., 2001) and Urban100 (Huang, Singh, & Ahuja, 2015). The first three datasets focus on natural scenes while Urban100 dataset contains challenged city scenes.

4.2. Implementation Details

The DPRN model has been developed, trained and tested on Ubuntu 14.04 system with NVIDIA GTX 1060 6GB memory and 8GB RAM. We trained the DPRN network and integrate upscale 2, 3 and 4 together in one model in caffe (Jia et al., 2014) environment (.caffemodel file). We used MATLAB 2017 programs to test the model and generate testing results like PSNR values, super-resolution pictures and .mat files in matlab environment.

For proposed DPRN parameters setup, branch number is defined as three with four additional convolutional layers after residual branches and one transposed layer; and there are 10 convolutional layers in one branch; the batch size for input training data and testing data are 64 and 2 respectively; each convolutional layer consists 64 filters with 3×3 kernel size and 1 pad, the weight filter is defined as msra instead of Gaussian; all convolutional layers are followed by rectified linear unit (He, Zhang, Ren, & Sun, 2015) except up-sampling layer. The basic learning rate is 10^{-4} with fixed policy; we use Adam for optimization type. An epoch has been defined to contain 5000 iterations of back-propagation, and 200 epochs training had been performed for our proposed DPRN network.

Base on the idea in (Timofte, Rothe, & Van Gool, 2016), we perform data argumentation on 291 training images for improving data diversity purpose. The augmentation processes have been performed on the original images as samples for flipping with 0.5 probability horizontally or vertically, randomly rotation in 90° , 180° or 270° , and randomly downscaling between 0.5 and 1.0, hence to generate new images but with similar feature information.

4.3. Comparisons with the State-of-the-Arts

In this paper, both qualitative and quantitative evaluation between DPRN and multiple state-of-the-art methods had been conducted. We have compared five recent benchmarks which include: A+ (Timofte et al., 2014), SRCNN (Dong, Loy, He, et al., 2016), VDSR (Kim et al., 2016a), DRCN (Kim et al., 2016b) and LapSRN (Lai et al., 2017) with our proposed DPRN. Among these five methods, A+ is a good representative for traditional methods while SRCNN bring deep learning age to SISR, furthermore, VDSR, DRCN and LapSRN are the most superior super resolution benchmarks in recent years.

Fig. 6, 7, 8 and 9 consists of four expected images from different testing datasets with x2, x3, x4 scales requirements. We compare our DPRN with most recent and typical state-of-the-art method VDSR and LapSRN. For qualitative evaluation, Fig. 6, 7, 8 and 9 also demonstrate that the proposed DPRN can restore higher quality of HR images compare to multiple benchmark methods under specific scales and image samples.



Fig.6. Scale x2, x3, x4 test results with VDSR, LapSRN and our DPRN on monarch.png from Set14.

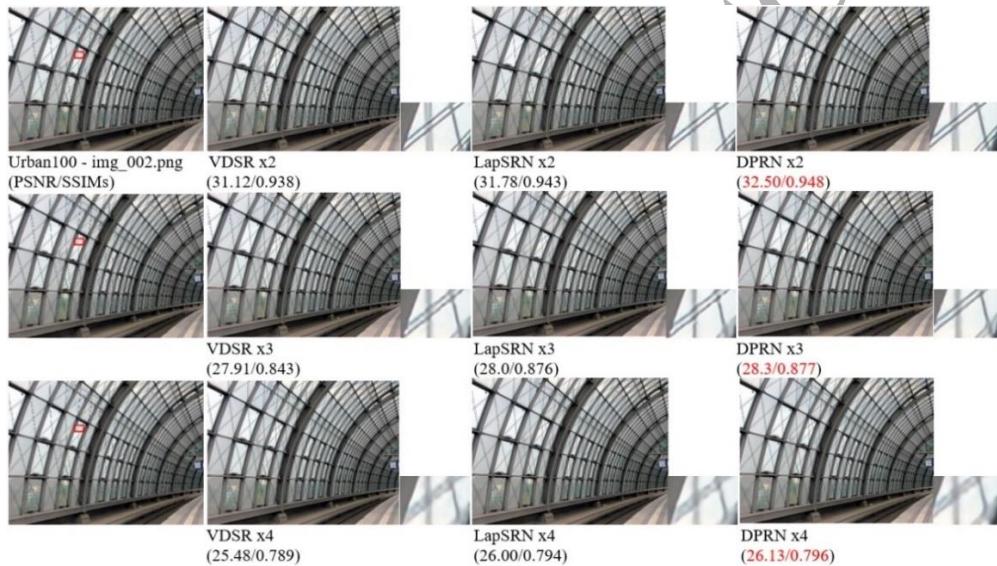


Fig.7. Scale x2, x3, x4 test results with VDSR, LapSRN and our DPRN on img002.png from Urban100.



Fig.8. Scale x2, x3, x4 test results with VDSR, LapSRN and our DPRN on 108005.png from BSDS100.



Fig.9. Scale x2, x3, x4 test results with VDSR, LapSRN and our DPRN on 253027.png from BSDS100.

By citing results in (Dong, Loy, He, et al., 2016) for A+ and DRCN, Table 2 and Table 3 provide quantitative comparison among different super resolution methods using common image quality metrics on standard test datasets.

Table 2 shows our DPRN has demonstrated successful results in both PSNR and SSIMs (Wang, Bovik, Sheikh, & Simoncelli, 2004). According to the results, DPRN outperformed most of the state-of-the-art methods use same training datasets and test datasets.

Table 3 indicates the excellent results of DPRN based on information fidelity criterion (IFC) (Sheikh, Bovik, & De Veciana, 2005) metric. Only for scale 2× results of LapSRN shows unusual correlation with perceptual scores.

Fig. 10, 11 and 12 show the visualization column charts which represent the test result values in Table 2 and Table 3, where PSNR, SSIM and IFC values are compared respectively.

Table 2 This table provides benchmark results base on average PSNR/SSIMs values about test dataset Set5, Set14, BSDS100 and Urban100 for scale 2×, 3× and 4×. Red color shows the best super resolution performance, Green color denotes second results and blue color indicates previous state-of-the-art results.

Dataset	Scale	Bicubic	A+	SRCNN	VDSR	DRCN	LapSRN	DPRN
Set5	x2	33.66/0.930	36.54/0.954	36.66/0.954	37.53/0.959	37.63/0.959	37.52/0.959	37.74/0.959
	x3	30.39/0.868	32.58/0.909	32.75/0.909	33.66/0.921	33.82/0.923	33.82/0.922	34.05/0.925
	x4	28.42/0.810	30.28/0.860	30.48/0.863	31.35/0.884	31.53/0.885	31.54/0.885	31.70/0.885
Set14	x2	30.24/0.869	32.28/0.906	32.45/0.907	33.03/0.912	33.04/0.912	33.08/0.913	33.23/0.913
	x3	27.55/0.774	29.13/0.819	29.30/0.822	29.77/0.831	29.76/0.831	29.87/0.832	29.99/0.834
	x4	26.00/0.703	27.32/0.749	27.50/0.751	28.01/0.767	28.02/0.767	28.19/0.772	28.22/0.770
BSDS100	x2	29.56/0.843	31.21/0.886	31.36/0.888	31.90/0.896	31.85/0.894	31.80/0.895	32.02/0.896
	x3	27.21/0.739	28.29/0.784	28.41/0.786	28.82/0.798	28.80/0.796	28.82/0.798	28.92/0.798
	x4	25.96/0.668	26.82/0.709	26.90/0.710	27.29/0.725	27.23/0.723	27.32/0.727	27.38/0.727
Urban100	x2	26.88/0.840	29.20/0.894	29.50/0.895	30.76/0.914	30.75/0.913	30.41/0.910	30.99/0.917
	x3	24.46/0.735	26.03/0.797	26.24/0.799	27.14/0.828	27.15/0.828	27.07/0.828	27.35/0.830
	x4	23.14/0.658	24.32/0.718	24.52/0.722	25.18/0.752	25.14/0.751	25.21/0.756	25.38/0.762

Table 3 This table contains performance evaluation information on benchmark results base on average IFC values.

The test dataset is Set5, Set14, BSDS100 and Urban100 for scale 2 \times , 3 \times and 4 \times . Red color shows the best super resolution performance, Green color denotes second results and blue color indicates previous state-of-the-art results.

Dataset	Scale	Bicubic	A+	SRCNN	VDSR	DRCN	LapSRN	DPRN
Set5	x2	6.166	8.715	8.166	8.190	8.326	9.010	8.671
	x3	3.596	4.979	4.682	5.088	5.202	5.194	5.234
	x4	2.337	3.260	2.997	3.496	3.502	3.559	3.584
Set14	x2	6.126	8.200	7.867	7.878	8.025	8.501	8.293
	x3	3.491	4.545	4.372	4.606	4.686	4.662	4.773
	x4	2.246	2.961	2.766	3.071	3.066	3.147	3.185
BSDS100	x2	5.695	7.464	7.242	7.169	7.220	7.715	7.526
	x3	3.168	4.028	3.879	4.043	4.070	4.057	4.159
	x4	1.993	2.565	2.412	2.627	2.587	2.677	2.706
Urban100	x2	6.319	8.440	8.092	8.270	8.527	8.907	8.790
	x3	3.661	4.883	4.630	5.045	5.187	5.168	5.271
	x4	2.386	3.218	2.992	3.405	3.412	3.530	3.572

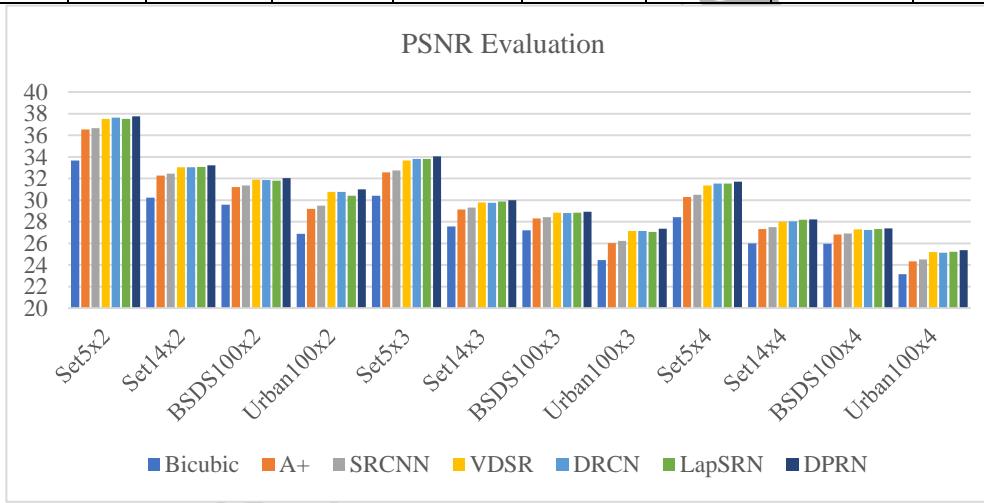


Fig.10. PSNR results comparison among proposed DPRN and other state-of-the-art methods for upscale operation x2, x3 and x4 on all 4 test datasets.

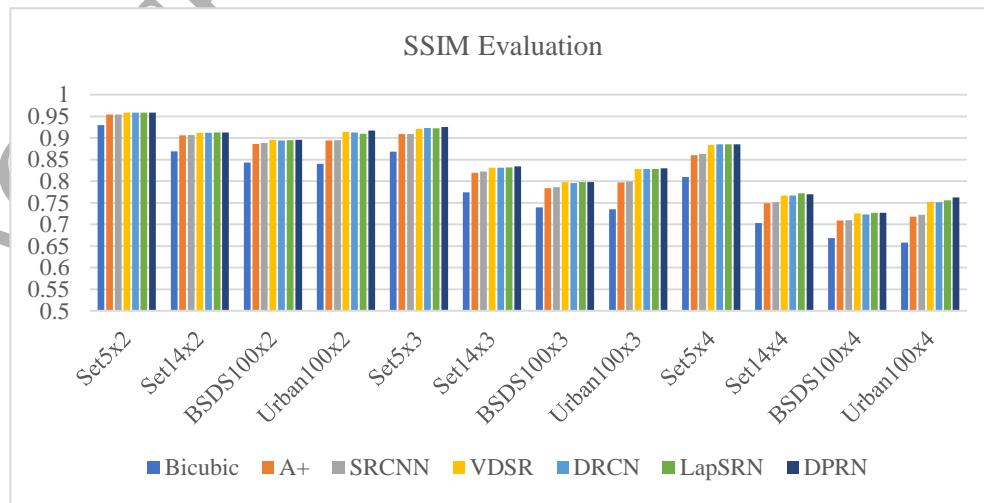


Fig.11. SSIM results comparison among proposed DPRN and other state-of-the-art methods for upscale operation x2, x3 and x4 on all 4 test datasets.

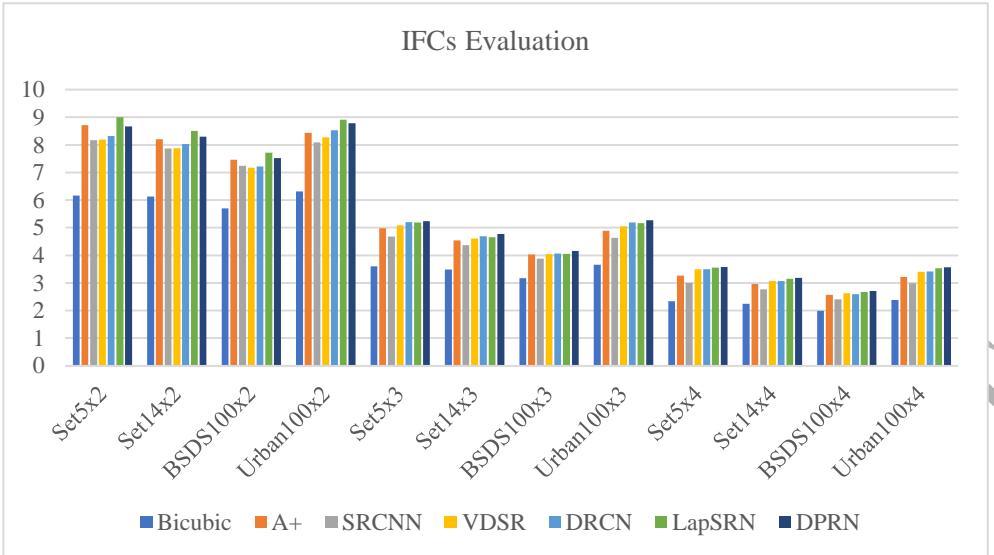


Fig.12. IFCs results comparison among proposed DPRN and other state-of-the-art methods for upscale operation x2, x3 and x4 on all 4 test datasets.

4.4. Mean Opinion Score Evaluation

The reconstructed images from Fig. 6, 7, 8 and 9 show the quality of different Super-Resolution methods, and the objective metrics like PSNR, SSIM and IFC have demonstrated the advantages of proposed DPRN approach. In order to illustrate the real performance and human feeling about the quality of upscaled images, we have performed a **Mean Opinion Score (MOS)** metric in this section.

The MOS evaluation is designed based on a 10 people survey. The rule for MOS metric (1-5 scores) is set as follows: 1. The High-Definition (HD) images are rated as 5 (top score); 2. Bicubic approach is set as an average quality (rated as 2); 3. Every participants will rate scores for all images from test datasets and for all test methods with A+, SRCNN, VDSR, DRCN, LapSRN and our DPRN; 4. Totally each participant will access 657 samples. Fig.13 shows the average MOS results base on different datasets and different upscale operations from the survey. The proposed DPRN also achieves good subjective feelings compare to other methods.

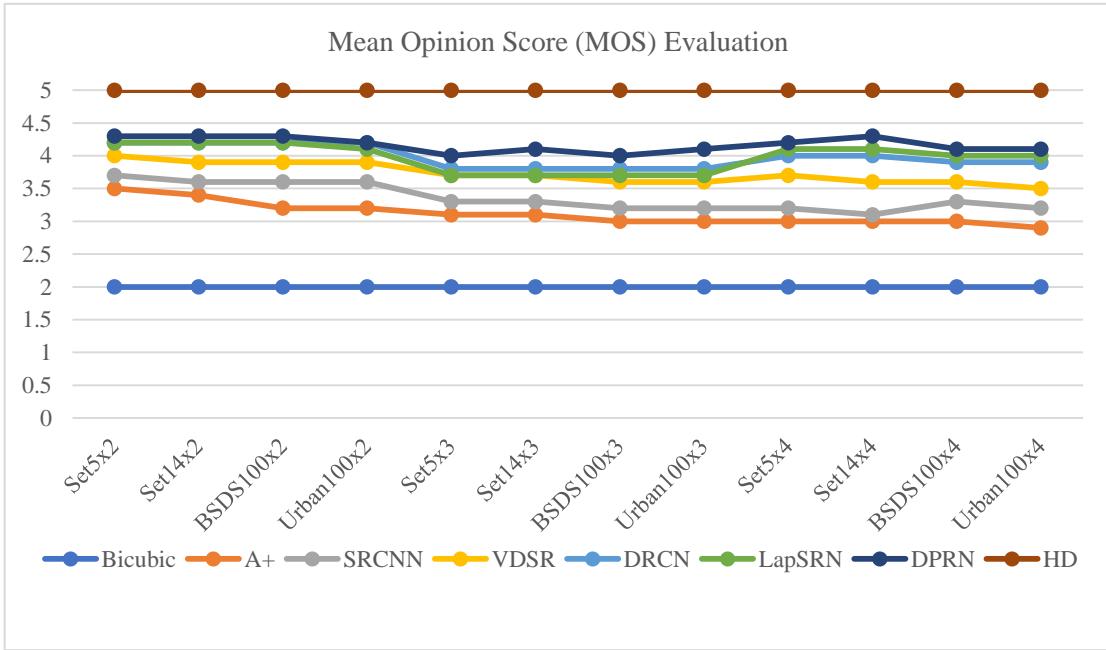


Fig.13. Average MOS results base on different datasets and different upscale operations from 10 people survey.

4.5. Runtime Evaluation and Convergence Analysis

Model execution time is an important factor for super resolution and its applications. Fig. 14 and 15 show the runtime evaluation results among proposed DPRN and other recent state-of-the-art methods for all test datasets. From these figures, our new approach has demonstrated to achieve the highest accuracy for all datasets and for different upscale operations.

According to experiment results, our proposed DPRN has achieved average 0.1705sec/29.33fps in Set5, 0.4654sec/30.08fps in Set14, 3.11sec/32.15fps in BSDS100 and 5.832sec/17.15fps in Urban100 datasets for all upscale operations (2 \times , 3 \times and 4 \times). This gives an average value of 27.18 fps overall to apply to real time human-eye recognition.

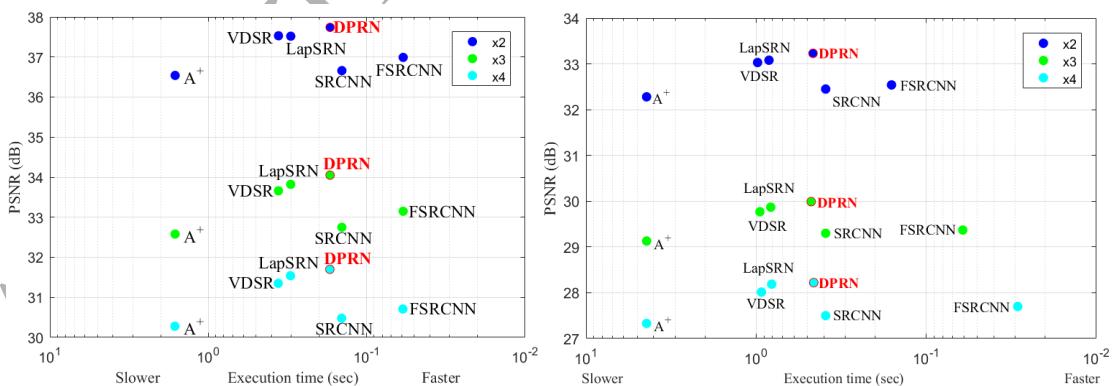


Fig.14. Speed and accuracy trade-off evaluation on our DPRN compare with other recent benchmark methods for upscale x2, x3, x4 on set5 (left) and set14 (right) datasets.

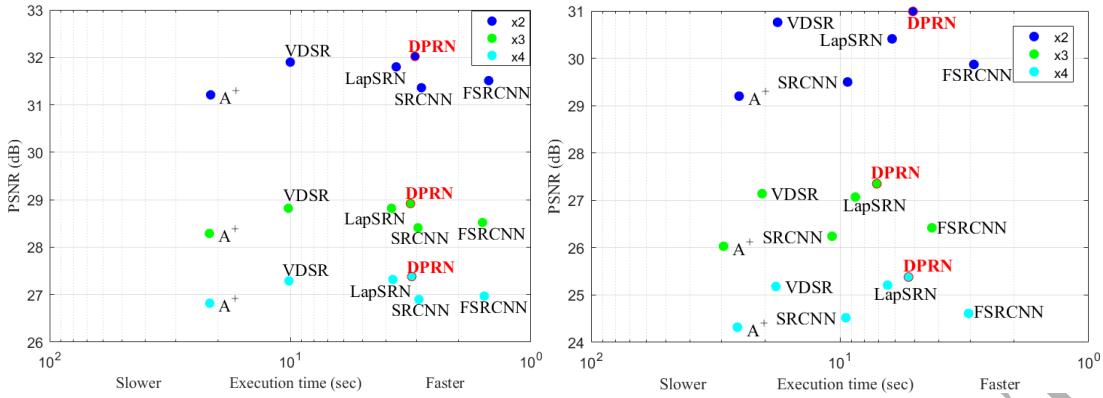


Fig.15. Speed and accuracy trade-off evaluation on our DPRN compare with other recent benchmark methods for upscale x2, x3, x4 on BSDS100 (left) and Urban100 (right) datasets.

Table 4 Experiment for execution time on all 4 test datasets with average time and frame per second value (fps).

Datasets	Scale	A+	SRCNN	FSRCNN	VDSR	LapSRN	DPRN
Set5	upscale 2x	1.6190	0.1435	0.0589	0.3607	0.3018	0.1703
Set5	upscale 3x	1.6197	0.1436	0.0589	0.3609	0.3019	0.1704
Set5	upscale 4x	1.6249	0.1440	0.0591	0.3620	0.3029	0.1709
Average time (sec)/fps		1.6212/3.08	0.1437/34.79	0.0590/84.75	0.3612/13.84	0.3022/16.55	0.1705/29.33
Set14	upscale 2x	4.4147	0.3991	0.16	0.98	0.84	0.4628
Set14	upscale 3x	4.5402	0.4102	0.061	0.95	0.82	0.4748
Set14	upscale 4x	4.4413	0.4015	0.029	0.93	0.81	0.4585
Average time (sec)/fps		4.4654/3.14	0.4036/34.69	0.083/168.67	0.963/14.54	0.823/17.01	0.4654/30.08
BSDS100	upscale 2x	21.4295	2.8531	1.4987	10.0050	3.6322	3.0331
BSDS100	upscale 3x	21.6906	2.9488	1.5901	10.1878	3.7983	3.1675
BSDS100	upscale 4x	21.6158	2.9214	1.5639	10.1354	3.7507	3.1290
Average time (sec)/fps		21.5786/4.63	2.908/34.39	1.5509/64.48	10.1094/9.89	3.727/26.83	3.11/32.15
Urban100	upscale 2x	25.4183	9.3148	2.8945	17.7979	6.1696	5.0867
Urban100	upscale 3x	29.3469	10.7545	4.2692	20.5487	8.6688	7.1093
Urban100	upscale 4x	25.8330	9.4668	3.0396	18.0883	6.4334	5.3002
Average time (sec)/fps		26.866/3.72	9.845/10.16	3.4011/29.4	18.812/5.32	7.0906/14.1	5.832/17.15

Table 4 contains detail results about running speed comparison about our proposed DPRN and other benchmark super resolution approaches. Inside, FSRCNN (Dong, Loy, & Tang, 2016) reached the fastest speed because of its less convolutional layer structure and parameter optimization. While SRCNN (Dong, Loy, He, et al., 2016) has gained the second in Set5, Set14 and BSDS100 datasets but has been beaten by LapSRN and DPRN in Urban100 dataset. The superiority of DPRN compare to VDSR and LapSRN is based on three factors: the branch network structure, the parameter selection through testing and the Adam optimizer. In order to confirm the current DPRN design, we also performed model training for pure 35 convolutional layers (DPRN-1) without branch structure, SGD optimizer (Bottou, 2012) instead of Adam optimizer (DPRN-2). The accuracy results for DPRN-1 and DPRN-2 cannot meet expectation and the runtime are similar to traditional methods.

Our experiments have included other methods' experiment restoration and comparison. We used the same environment testing conditions, the results of VDSR, LapSRN and our DPRN show similar

convergence trends, while DPRN reaches higher dB value according to Fig. 16.

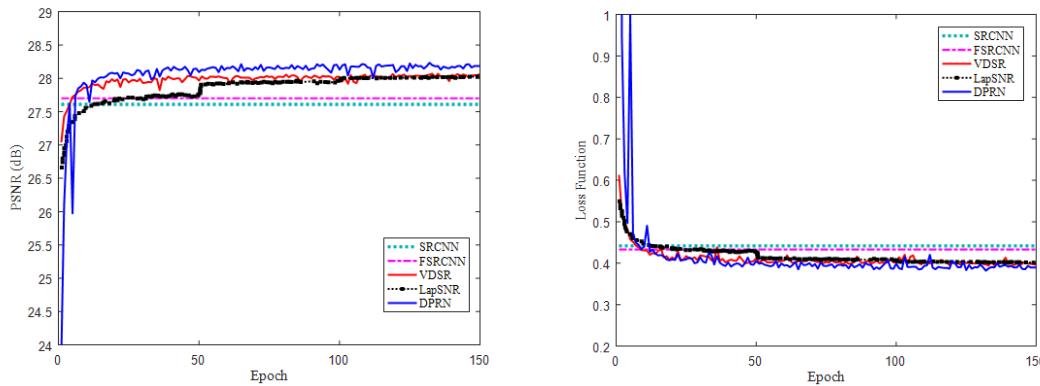


Fig.16. Convergence analysis through different epochs.

4.6. Limitations

The proposed DPRN model has achieved excellent accuracy results and sufficient runtime speed with clear reconstructed HR image, however, the training process consumes more resources than other common state-of-the-art methods. Furthermore, the detail pixels in some images meet problems, especially in large upscale condition. Currently, most of the SISR models overlearned the image and provided too smooth results or blurred results which are different from original HR image.

5. Conclusions

In this article, a novel deep convolutional neural network for single image super resolution named Deep Parallel Residual Network (DPRN) has been proposed for superior accuracy outcome and balanced real-time model execution. Our model consists of residual combination and residual branch to conduct as an efficient local residual learning and global residual learning algorithm. Each convolutional layer in this residual combination will perform as the parallel learning from previous two layers. In addition, the training process can also benefit from Adam optimizer to stabilize the parameters.

The experiment and evaluation results on benchmark datasets have shown that the proposed DPRN model outperformed several current state-of-the-art SISR methods from 0.2dB to 1.08dB in accuracy, and also successfully demonstrated 27.18fps runtime speed for real-time applications. The outcome shows DPRN is an advanced HR image reconstruction method.

In further research, we will focus on theoretical improvements and super-resolution applications. The theoretical part may include further upgrade the network with more complicated structure or adopt Generative Adversarial Network (Goodfellow et al., 2014), or on the contrary, further optimize the parameters or remove redundant layer and structure base on the test. We will also research on the model applications, such as solving blur problem in image classification and video tracking.

Author roles:

1. Feng Sha: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Validation, Writing - original draft; Writing - review & editing
2. Seid Miad Zandavi: Software, Validation
3. Yuk Ying Chung Ph.D.: Supervision, Writing - review & editing

References

- Agustsson, E., & Timofte, R. (2017). NTIRE 2017 Challenge on Single Image Super-Resolution: Dataset and Study. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*.
- Bevilacqua, M., Roumy, A., Guillemot, C., & Alberi-Morel, M. L. (2012). Low-complexity single-image super-resolution based on nonnegative neighbor embedding.
- Bottou, L. (2012). Stochastic gradient descent tricks. In *Neural networks: Tricks of the trade* (pp. 421–436). Springer.
- Choi, J.-S., & Kim, M. (2017). A deep convolutional neural network with selection units for super-resolution. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2017 IEEE Conference on* (pp. 1150–1156).
- Dong, C., Loy, C. C., He, K., & Tang, X. (2016). Image super-resolution using deep convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(2), 295–307.
- Dong, C., Loy, C. C., & Tang, X. (2016). Accelerating the super-resolution convolutional neural network. In *European Conference on Computer Vision* (pp. 391–407).
- Duchi, J., Hazan, E., & Singer, Y. (2011). Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research*, 12(Jul), 2121–2159.
- Duchon, C. E. (1979). Lanczos filtering in one and two dimensions. *Journal of Applied Meteorology*, 18(8), 1016–1022.
- Glasner, D., Bagon, S., & Irani, M. (2009). Super-resolution from a single image. In *Computer Vision, 2009 IEEE 12th International Conference on* (pp. 349–356).
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... Bengio, Y. (2014). Generative adversarial nets. In *Advances in neural information processing systems* (pp. 2672–2680).
- He, K., Zhang, X., Ren, S., & Sun, J. (2015). Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision* (pp. 1026–1034).
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770–778).
- Huang, J.-B., Singh, A., & Ahuja, N. (2015). Single image super-resolution from transformed self-exemplars. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 5197–5206).
- Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing

- internal covariate shift. In *International conference on machine learning* (pp. 448–456).
- Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., ... Darrell, T. (2014). Caffe: Convolutional architecture for fast feature embedding. In *Proceedings of the 22nd ACM international conference on Multimedia* (pp. 675–678).
- Jianchao, Y., Wright, J., Huang, T., & Ma, Y. (2008). Image super-resolution as sparse representation of raw image patches. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition* (pp. 1–8).
- Kim, J., Kwon Lee, J., & Mu Lee, K. (2016a). Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1646–1654).
- Kim, J., Kwon Lee, J., & Mu Lee, K. (2016b). Deeply-recursive convolutional network for image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1637–1645).
- Kinga, D., & Adam, J. B. (2015). A method for stochastic optimization. In *International Conference on Learning Representations (ICLR)*.
- Lai, W.-S., Huang, J.-B., Ahuja, N., & Yang, M.-H. (2017). Deep laplacian pyramid networks for fast and accurate super-resolution. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* (pp. 624–632).
- Lim, B., Son, S., Kim, H., Nah, S., & Lee, K. M. (2017). Enhanced deep residual networks for single image super-resolution. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops* (Vol. 1, p. 3).
- Martin, D., Fowlkes, C., Tal, D., & Malik, J. (2001). A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on* (Vol. 2, pp. 416–423).
- Matsugu, M., Mori, K., Mitari, Y., & Kaneda, Y. (2003). Subject independent facial expression recognition with robust face detection using a convolutional neural network. *Neural Networks*, 16(5–6), 555–559.
- Nair, V., & Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (ICML-10)* (pp. 807–814).
- Ruder, S. (2016). An overview of gradient descent optimization algorithms. *ArXiv Preprint ArXiv:1609.04747*.
- Schulter, S., Leistner, C., & Bischof, H. (2015). Fast and accurate image upscaling with super-resolution forests. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 3791–3799).
- Sheikh, H. R., Bovik, A. C., & De Veciana, G. (2005). An information fidelity criterion for image quality assessment using natural scene statistics. *IEEE Transactions on Image Processing*, 14(12), 2117–2128.
- Tieleman, T., & Hinton, G. (n.d.). *Divide the gradient by a running average of its recent magnitude. COURSERA: Neural networks for machine learning*.
- Timofte, R., Agustsson, E., Van Gool, L., Yang, M.-H., Zhang, L., Lim, B., ... others. (2017). Ntire 2017 challenge on single image super-resolution: Methods and results. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2017 IEEE Conference on* (pp. 1110–1121).
- Timofte, R., De Smet, V., & Van Gool, L. (2014). A+: Adjusted anchored neighborhood regression for

- fast super-resolution. In *Asian Conference on Computer Vision* (pp. 111–126).
- Timofte, R., Rothe, R., & Van Gool, L. (2016). Seven ways to improve example-based single image super resolution. In *Computer Vision and Pattern Recognition (CVPR), 2016 IEEE Conference on* (pp. 1865–1873).
- Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4), 600–612.
- Xie, S., & Tu, Z. (2015). Holistically-nested edge detection. In *Proceedings of the IEEE international conference on computer vision* (pp. 1395–1403).
- Yang, J., Wright, J., Huang, T. S., & Ma, Y. (2010). Image super-resolution via sparse representation. *IEEE Transactions on Image Processing*, 19(11), 2861–2873.
- Zeyde, R., Elad, M., & Protter, M. (2010). On single image scale-up using sparse-representations. In *International conference on curves and surfaces* (pp. 711–730).