DPCNN: Efficient Implementation of Single Image Super-Resolution with A Dense Projection Networks

Anonymous CVPR submission

Paper ID ****

Abstract

A deep learning based method is proposed in this paper which is used for single image super-resolution (SISR). Image resolution can be improved when the relative displacement in the image sequence are known accurately. In this method, we build a mapping between the low and high-resolution images. Computing with deep convolution neural network (DCNN), we transfer the low-resolution image to the high-resolution one. The traditional sparse-coding-based SISR methods is a different method of DCNN that each component is handled separately. Our method obfuscates single component and optimizes them together. The benefits of the new method are apparently. It has brilliant restoration quality and fast speed which is suit for online usage, while the structure is extremely light. For better performance and speed, we try many kinds of network structures and parameter settings. Furthermore, we expend the network to handle three color channels at the same time, and get high restoration quality.

1. Introduction

Single image super-resolution (SISR) [8] refers to the task to reconstruction performance on the resolution or improving the details within an image, transform a low-resolution (LR) image to a visually pleasing high-resolution (HR) image. This is a classical problem in the field of image and vision computing. SISR is now used in various area of computer vision tasks, such as security and surveillance imaging [16], image generation [6] and or even life saving on medical applications [10]. While SISR problem is ill-posted inverse problem and there is multitude of solutions for any LR input. Obtaining more sample of an image may be a option to reconstruct HR image by combining the more similar sample data, however it is impossible in some cases. To solve this problem plenty of SISR algorithms are proposed like learning-based [5], reconstruction-based [14] and interpolation-based [15] methods.

Various method have been proposed for super-resolution to further regularize the problem. For instance MAP solution can be choose under generic image prior such as Huber Markov Random Field (MRF) and Bilateral Total Variation (BTV) [7]. However, reconstruction-based algorithms doesn't work well if the factor become too large and if the datasets don't have enough low resolution images [2]. There have another method of SR by calculation the co-occurrence prior to predict there corresponding distance between LR image to HR image patches [3]. The sparse-coding based strategy [13] is one of the representative external example-based SR strategies. This method includes a few steps in its arrangement solution pipeline. In this methodology, densely trimmed used to take input and the overlapping reconstructed patches are aggregated to produce the final output.

The approach we described in the paper is based on to construct by learning a non-linear mapping between LR-to-HR which is implemented as a deep neural network and achieve a significant improvement over convolutional method. Our method perplexes single component and optimizes them together by up-scaling and down-scaling sampling layers. Our main goal is to construct a network to handle three color channels at the same time, and get high restoration quality, so are involved in optimization. Our dense projection network successfully perform perfectly in large scaling factors. Our main contributions can be summarized in the following four aspects:

(1) Error Feedback: We have introduced a iterative error correction feedback method for super-resolution. It will calculated both up ad down projection errors in order to guide the rebuild for better results. Projection errors are uses to constraint the features in the previous or early stages which helps to make improvement in the next layers.

(2) Up and down sampling stages: Feed-forward network system which has known as a one way mapping, only map rich representations of the input to a output space. But in the larger scale this approach is not successful to map LR and HR image as there is a limited features available in the LR spaces. But in our proposed methods it not only focus

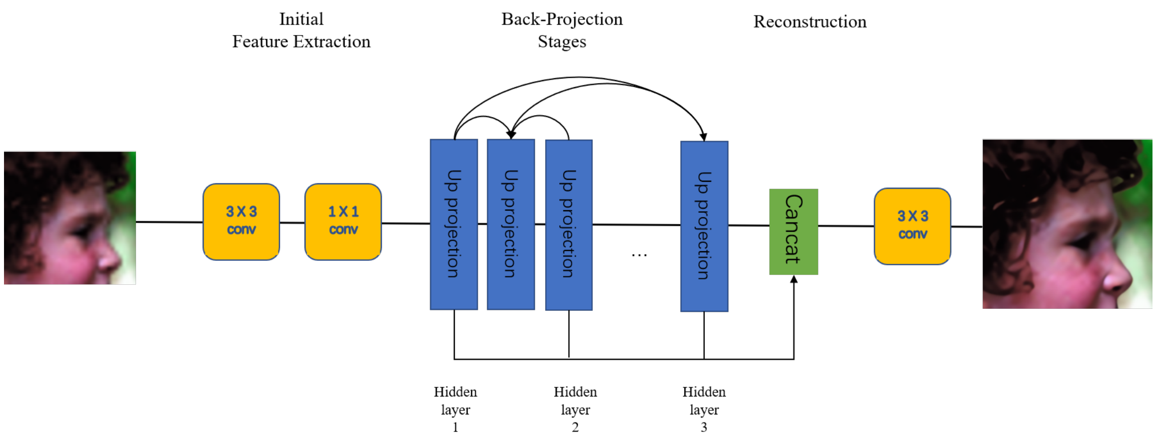

Figure 1. An implementation of DPCNN for super-resolution.

on collecting different variants for up-sampling the layers but also projecting it again to the LR spaces using down-sampling.

(3) Deep concatenation: There are distinct forms of image loss and HR components in our networks. This ability helps the networks to recreate the HR image from all the up-sampling steps using deep concatenation of the HR feature maps. Unlike other networks, without propagating them through the sampling layers, our reconstruction directly utilizes various types of LR-to-HR functionality.

2. Related Works

Deep Learning-based strategies are presently active and showing vital performance on SISR tasks. Earlier research on super-resolution was carried out by *Tsai et al.*, who used frequency domain method to disregarded the blur in the imaging process by using translated images to handle the loss of the data due to discrimination [12]. *H. Chang et al.* maps the local geometry of the LR image patch space to HR image patch space and generate a high-resolution patch as a linear combination of neighbors [3]. By using this methods many patch pattern can be represented in smaller database. *C. Dong et al.* introduce Super-Resolution using deep convolutional networks which has simple structure and higher accuracy and storage quality [5]. This paper establish a relation between SR method and traditional sparse-coding based SR method which is a guidance for design network structure. The SRCNN is learn LR-to-HR non-linear mapping with simple convolutional layers as predefined unsampling commonly uses interpolation upsampling to produce a middle stage of the image. Later improvement was done by *Y. Tai et al.* by using residual learning and recursive layers [11]. However this approach produce new noise from the MR image. In the cognition theory, feedback connections which link the cortical visual areas can transmit response signals from higher-order areas to lower-order areas [9]. The feedback mechanism works top-down manner and carry a high level information back to previous layers and refining low-level encoded information.

Xiangxiang Chu et al. is presented a new multi-objective oriented algorithm called MoreMNAS (Multi-Objective Reinforced Evolution in Mobile Neural Architecture Search) by leveraging good virtues from both evolution algorithms(EA) and reinforced learning methods (RL) [4]. Reinforced control is blended in with a characteristic transforming cycle to manage subjective transformation, keeping a delicate balance among exploration and exploitation. Consequently, it does not just do the technique keep the looked through models from degrading during the development cycle, however, it can also make better use of learned knowledge. The MoreMNAS in Super Resolution deliver rivalling models compared to some state-of-the-art methods with fewer FLOPS. *Zhang et al.* proposed a fast, accurate and lightweight Super-Resolution with Neural Architecture architecture which can pursue to maximize a trade-off between the restoration capacity and the simplicity of models [1]. The technique embraces a flexible search strategy at both the miniature and large scale level, based on a hybrid controller that benefits from evolutionary computation and reinforcement learning. Quantitative trials of the models rule the vast majority of the state-of-art methods with respect to the individuals FLOPS.

3. Model Architecture

First of all the model we proposed DPCNN is shown in Fig. 1. is separated into three sections: Initial feature extraction, Projection, and Reconstruction as described below. Here, let $conv(f;n)$ be a convolutional layer, where f is the filter size and n is the number of filters.

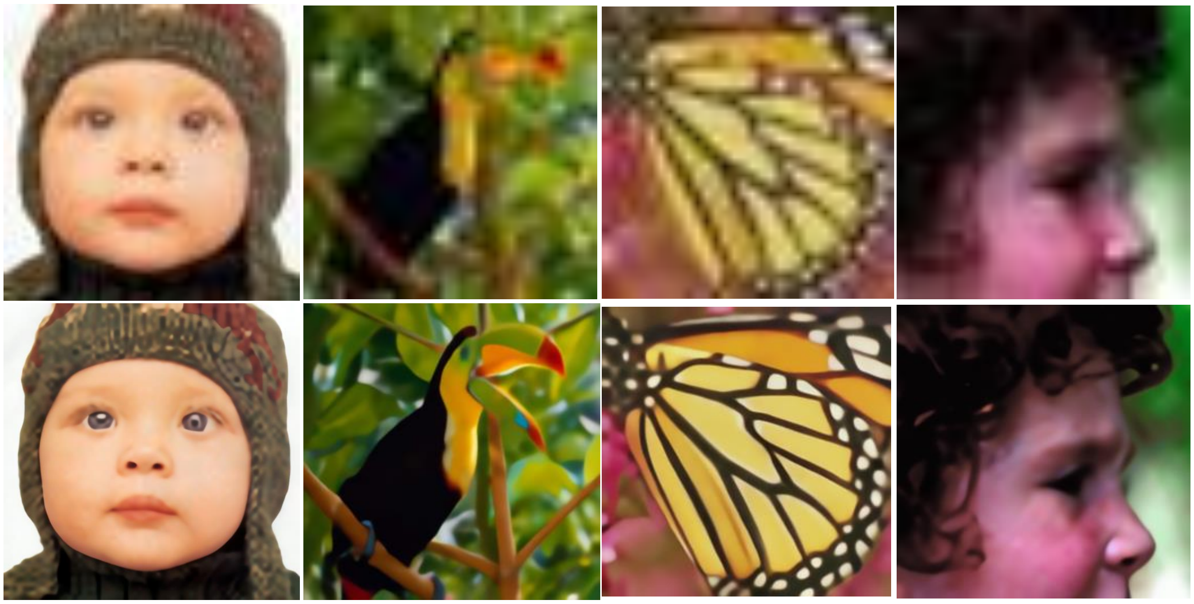

Figure 2. Result of experiment (Pictures in the first row is the row picture with low resolution, and the pictures in the second row is the output of the DPCNN, which achieves high resolution.

3.1. Model Analysis

In this paper, our goal it to build a mapping between LR image to HR image. To do this projection is really important and it is an well known effective iterative method to minimizing reconstruction error rate. we proposed DPCNN architecture for projection unit. It will trained the model to map either an LR feature map to an HR map (up-projection) or and HR map to and LR map (down-projection). The updating strategy of given just a single LR input picture can be acquired by upsampling operators and calculate the reconstruction error iteratively.Timofte et al. mentioned that backprojection can improve the quality of SR image. Zhao et al. proposed a method to refine high-frequency texture details with an iterative projection process. In this paper, we construct an end-to-end trainable architecture based on the idea of iterative up-and down-sampling: Dense Projection Networks for Super Resolution (DPCNN). Our main contributions can be summarized in the following four aspects:

- Initial feature extraction: We construct the initial LR feature maps L_0 from the input using $conv(3; n_0)$. Then $conv(1; n_R)$ is used to minimize the dimension from n_0 to n_R before entering into the projection step. Where n_0 is the number of filters used in the initial LR features extraction and n_R is the number of filters used in each and every projection unit.

- Back-projection stages: The next stage is the back-projection process. In the following initial feature extraction is a sequence of projection units and it is alternating between construction of LR and HR image feature maps H_t, L_t; each unit has access to the outputs of all previous units.

- Reconstruction: Finally, the target HR image is reconstructed as $I_{sr} = f_{Rec}([H_1, H_2, ..., H_t])$ where f Rec use $conv(3, 3)$ as re- construction and $[H_1, H_2, ..., H_t]$ refers to the concatenation of the feature-maps produced in each up-projection unit of the model. Reconstruction helps to improve the image quality.

4. Experiments

4.1. Training details

We initialize the weights based on the paper [4]. We conduct SR experiments on real-world images. In this case, the original HR images are not available and degradation model in unknown either. Here, std is calculated by using $(\sqrt{(2 \div (n_l))}$ where $n_l = f_t^2 n_t$, f_t is the filter size,and n_t is the number of filters. For example, with $f_t = 3$ and $n_t = 8$, the std is 0.111. All convolutional and deconvolutional layers are followed by parametric rectified linear units (PReLUs). The learning rate is initialized to $(1e)^{-4}$ for all layers and decrease by a factor of 10 for every $5 * 10^5$ iterations for total 10^6 iterations.For optimization, we use Adam with momentum to 0.9 and weight decay to $(1e)^{-4}$. All experiments were conducted using Pytorch on NVIDIA TITAN X GPUs. The model is trained with 20,000 epochs and 500 snapshots.

		Set5		Set14		RAISE_1k	
Algorithm	Scale	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Bicubic	2	33.65	0.930	30.34	0.870	31.04	0.88
A+	2	36.54	0.954	32.40	0.906	31.22	0.90
SRCNN	2	36.65	0.954	32.29	0.903	31.34	0.92
FSRCNN	2	36.99	0.955	32.73	0.909	31.27	0.91
VDSR	2	37.53	0.958	32.97	0.913	31.55	0.90
DRCN	2	37.63	0.959	32.98	0.913	31.66	0.93
DRRN	2	37.74	0.959	33.23	0.913	31.77	0.92
DPCNN	***2***	***38.04***	***0.96***	***33.83***	***0.919***	***32.55***	***0.93***
Bicubic	4	28.42	0.810	26.10	0.704	27.02	0.720
A+	4	30.30	0.859	27.43	0.754	27.04	0.737
SRCNN	4	30.49	0.862	27.61	0.754	27.22	0.755
FSRCNN	4	30.71	0.865	27.70	0.754	27.31	0.768
VDSR	4	31.35	0.882	28.03	0.774	27.94	0.773
DRCN	4	31.53	0.884	28.04	0.774	29.01	0.774
DRRN	4	31.68	0.888	28.21	0.774	27.64	0.776
DPCNN	***4***	***32.44***	***0.895***	***28.81***	***0.774***	***27.67***	***0.775***
Bicubic	8	24.39	0.657	23.19	0.568	23.19	0.568
A+	8	25.52	0.692	23.98	0.597	23.22	0.582
SRCNN	8	25.33	0.689	23.85	0.593	23.27	0.594
FSRCNN	8	25.41	0.682	23.93	0.592	24.01	0.596
VDSR	8	25.72	0.711	24.21	0.609	24.04	0.610
LapSRN	8	26.14	0.738	24.44	0.623	24.56	0.633
EDSR	8	26.97	0.775	24.94	0.640	24.79	0.650
DPCNN	***8***	***27.20***	***0.783***	***25.12***	***0.647***	***25.03***	***0.667***

Table 1. Quantitative evaluation of state-of-the-art Super-Resolution algorithms.

Scale	Criterion	SRCNN	VDSR	ESPCN	SRResNet	***DPCNN***
2	PSNR	24.1528	32.8283	18.4822	16.5440	***25.3120***
	SSIM	0.9998	1.0000	0.9997	0.9995	***0.9998***
4	PSNR	18.4505	27.2014	17.6023	17.3498	***24.4753***
	SSIM	0.9997	1.0000	0.9997	0.9996	***0.9997***
8	PSNR	17.6915	24.8827	16.2012	15.9221	***21.4753***
	SSIM	0.9994	0.9999	0.9996	0.9993	***0.9997***

Table 2. Quantitative evaluation of state-of-the-art NN Algorithms on BSDS300.

4.2. Experiments analysis

In order to test our proposed method we need to test our network with other developed Deep Learning-based SR algorithms. To confirm the ability of the proposed network, we compare our network with six state-of-the-art SR algorithms: A+, SRCNN, FSRCNN, VDSR, DRCN, and DRRN. We carry out extensive experiments using three datasets:Set5, Set14 and RAISE_1K. PSNR and structural similarity (SSIM) were used to quantitatively evaluate the proposed method. Note that higher PSNR and SSIM values indicate that the quality of the image is better beacuse PSNR and SSIM are normalized based method perception and it work with saliency-based error. Furthermore, we carry out the experiments of the proposed network on a larger dataset (BSDS300), and compared the results with four state-of-the-art implementations with neural network.

We showed the quantitative results in the Table 1. It shows the frames per second comparison of the mentioned algorithm. In Table 2 is scale factor. When the scale factor is larger, the larger the PSNR and SSIM of models, the better the model effect. In Table 1, the PSNR and SSIM of the DPCNN exceed all models. For example, when the scale factor is 8, PSNR and SSIM of DPCNN are 0.23dB and 0.008dB higher than EDSR in the set5; PSNR and SSIM of DPCNN are 0.18dB and 0.007dB higher than EDSR. in the set14. In Fig2, we show the results of our method, and the original picture, and we can see that DPCNN achieve high performance.

5. Conclusions

Image quality plays a vital role in digital image processing. In this paper, we propose a Dense Projection Networks to enhance quality of low resolution and make it better quality and represent as high-level ones. Unlike previous approaches that forecast the SR image in a feed-forward process, our proposed networks tend to directly enhance the SR features by using various up-sampling and down-sampling levels and feeding the error estimates through each depth in the networks to revise the sampling performance, and then aggregate self-correcting features from each up-sampling stage to create SR images. The method we use for up-scaling and down-scaling error input to direct the network to achieve a better outcome.The results demonstrated the effectiveness of the suggested framework relative to many other state-of-the-art technologies. Moreover, our proposed network successfully outperforms other state-of-the-art methods on large scaling factors such as 8X enlargement.

References

[1] Namhyuk Ahn, Byungkon Kang, and Kyung-Ah Sohn. Fast, accurate, and, lightweight super-resolution with cascading residual network. *ArXiv*, 2018. 2

[2] S. Baker and T. Kanade. Limits on super-resolution and how to break them. *Proceedings IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000 (Cat. No.PR00662)*, 2000. 1

[3] H. Chang, D. Yeung, and Y. Xiong. Super-resolution through neighbor embedding. In *CVPR 2004*, 2004. 1, 2

[4] Xiangxiang Chu, Bo Zhang, R. Xu, and Hailong Ma. Multi-objective reinforced evolution in mobile neural architecture search. *ArXiv*, 2019. 2, 3

[5] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Learning a deep convolutional network for image super-resolution. 2014. 1, 2

[6] Allan T Evans and Bruce T Bernacki. Image generation systems and image generation methods, 2015. 1

[7] Sina Farsiu, D. Robinson, Michael Elad, and P. Milanfar. Fast and robust multi-frame super-resolution. 1996. 1

[8] Michal Irani and Shmuel Peleg. Improving resolution by image registration. *CVGIP: Graphical models and image processing*, 1991. 1

[9] Z. Li, J. Yang, Z. Liu, X. Yang, G. Jeon, and Wei Wu. Feedback network for image super-resolution. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 2

[10] O. Oktay, Wenjia Bai, M. J. Lee, Ricardo Guerrero, K. Kamnitsas, Jose Caballero, Antonio de Marvao, S. Cook, D. O'Regan, and D. Rueckert. Multi-input cardiac image super-resolution using convolutional neural networks. 2016. 1

[11] Ying Tai, Jian Yang, and X. Liu. Image super-resolution via deep recursive residual network. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017. 2

[12] R. Tsai and T. Huang. Multiframe image restoration and registration. 1984. 2

[13] Jianchao Yang, J. Wright, T. Huang, and Y. Ma. Image super-resolution via sparse representation. *IEEE Transactions on Image Processing*, 2010. 1

[14] K. Zhang, X. Gao, D. Tao, and X. Li. Single image super-resolution with non-local means and steering kernel regression. *IEEE Transactions on Image Processing*, 2012. 1

[15] Lei Zhang and X. Wu. An edge-guided image interpolation algorithm via directional filtering and data fusion. *IEEE Transactions on Image Processing*, 2006. 1

[16] W. W. W. Zou and P. C. Yuen. Very low resolution face recognition problem. *IEEE Transactions on Image Processing*, 2012. 1