Image Super-Resolution using Deep Convolutional Neural Networks (2016)

Paper reviewed by Taegyun Jeon

Dong, Chao, et al. "Image super-resolution using deep convolutional networks."
IEEE transactions on pattern analysis and machine intelligence. 38.2 (2016): 295-307.

연구소개 | 문제 정의

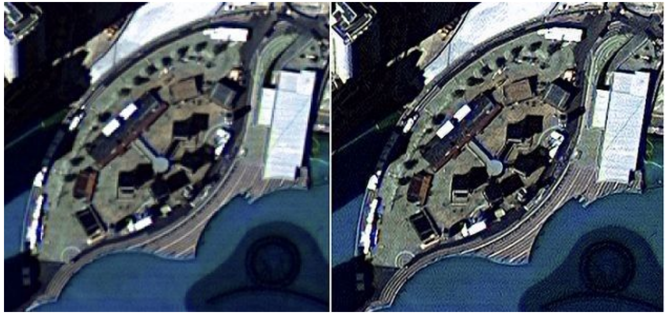
- Single Image Super-Resolution
  - ILL-posed problem (many candidate solutions)

Super-Resolution

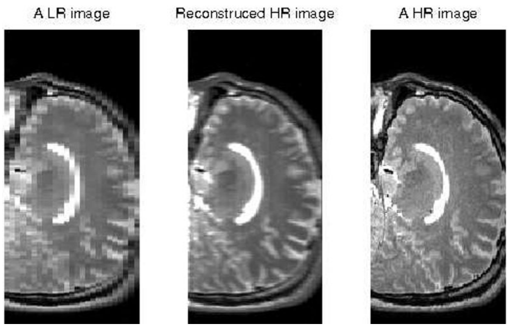
저해상도 이미지
(Low resolution)

고해상도 이미지
(High resolution)

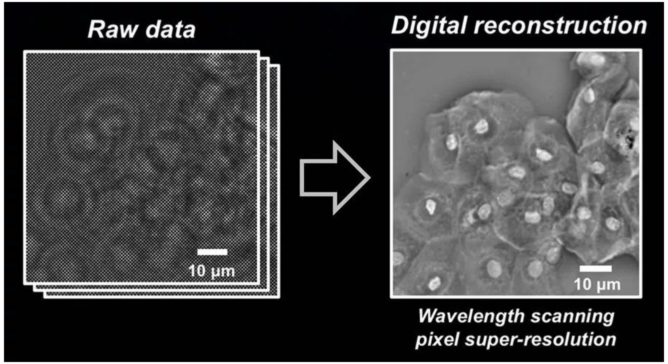
연구소개 | 파급효과 및 지향점



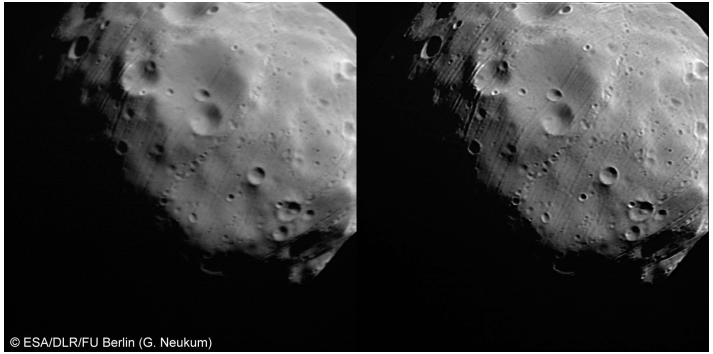
위성영상



의료 영상



현미경영상



천체영상

연구소개 | 데이터셋

Train Dataset	Image source
Yang 91	Yang et al. CVPR 2008
BSD 200	Martin et al. ICCV 2001
General 100	Dong et al. ECCV 2016
ImageNet	Olga Russakovsky et al. IJCV 2015
COCO	Tsung-Yi Lin et al. ECCV 2014

Test Dataset	Image source
Set 5	Bevilacqua et al. BMVC 2012
Set 14	Zeyde et al. LNCS 2010
BSD 100	Martin et al. ICCV 2001
Urban 100	Huang et al. CVPR 2015

배경지식 | 평가방식

- **PSNR** (Peak Signal-to-Noise Ratio)
  - 최대 신호 대 잡음비 (**영상 화질 손실정보**에 대한 평가)

$$PSNR = 10 \cdot \log_{10}\left(\frac{MAX_I^2}{MSE}\right)$$

- **SSIM** (structural similarity index)
  - “밝기, 명암, 구조”를 조합하여 **두 영상의 유사도** 평가

$$\text{SSIM}(\mathbf{x},\mathbf{y}) = [l(\mathbf{x},\mathbf{y})]^{\alpha} \cdot [c(\mathbf{x},\mathbf{y})]^{\beta} \cdot [s(\mathbf{x},\mathbf{y})]^{\gamma}$$

- **MS-SSIM** (Multi-scale SSIM)
  - **다양한 배율 (scale)**에 대해 SSIM 평가

$$\text{SSIM}(\mathbf{x},\mathbf{y}) = [l_M(\mathbf{x},\mathbf{y})]^{\alpha_M} \cdot \prod_{j=1}^{M} [c_j(\mathbf{x},\mathbf{y})]^{\beta_j} [s_j(\mathbf{x},\mathbf{y})]^{\gamma_j}$$

- **IFC** (Information fidelity criterion)
  - 서로 다른 두 확률분포에 대한 **의존도** 평가

$$\text{IFC} = \sum_{k \in \text{subbands}} I(C^{N_k,k}; D^{N_k,k} | s^{N_k,k})$$

배경지식 | SISR 관련 선행 연구

Example-Based Super-Resolution

William T. Freeman, Thouis R. Jones, and Egon C. Pasztor
Mitsubishi Electric Research Labs

To address the lack of resolution independence in most models, we developed a fast and simple one-pass, training-based super-resolution algorithm for creating plausible high-frequency details in zoomed images.

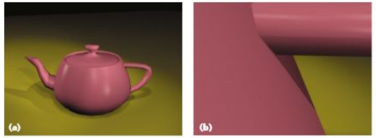
Polygon-based representations of 3D objects offer resolution independence over a wide range of scales. With this approach, object boundaries remain sharp when we zoom in on an object until very close range, where faceting appears due to finite polygon size (see Figure 1). However, constructing polygon models for complex, real-world objects can be difficult. Image-based rendering (IBR), a complementary approach for representing and rendering objects, uses cameras to obtain rich models directly from real-world data. Unfortunately, these representations no longer have resolution independence. When we enlarge a bitmapped image, we get a blurry result. Figure 2 shows the problem for an IBR version of a teapot image, rich with real-world detail. Standard pixel interpolation methods, such as pixel replication (Figures 2b and 2c) and cubic-spline interpolation (Figures 2d and 2e), introduce artifacts or blur edges. For images enlarged three octaves (factors of two) such as these, sharpening the interpolated result has little useful effect (Figures 2f and 2g).

We call methods for achieving high-resolution enlargements of pixel-based images *super-resolution* algorithms. Many applications in graphics or image processing could benefit from such resolution independence, including IBR, texture mapping, enlarging consumer photographs, and converting NTSC video content to high-definition television. We built on another training-based super-resolution algorithm[1] and developed a faster and simpler algorithm for one-pass super-resolution. (The one-pass, example-based algorithm gives the enlargements in Figures 2h and 2i.) Our algorithm requires only a nearest-neighbor search in the training set for a vector derived from each patch of local image data. This one-pass super-resolution algorithm is a step toward achieving resolution independence in image-based representations. We don't expect perfect resolution independence—even the polygon representation doesn't have that—but increasing the resolution independence of pixel-based representations is an important task for IBR.

Example-based approaches

Super-resolution relates to image interpolation—how should we interpolate between the digital samples of a photograph? Researchers have long studied this problem, although only recently with machine learning or sampling approaches. (See the "Related Approaches" sidebar for more details.)

Three complimentary ways exist for increasing an image's apparent resolution:



1 (a) When we model an object with traditional polygon techniques, it lacks some of the richness of real-world objects but behaves properly under enlargement. (b) The teapot's edge remains sharp when we enlarge it.

0272-1716/02/$17.00 © 2002 IEEE

Image Super-Resolution as Sparse Representation of Raw Image Patches

Jianchao Yang†, John Wright‡, Yi Ma‡, Thomas Huang†
University of Illinois at Urbana-Champagin
Beckman Institute† and Coordinated Science Laboratory‡
{jyang29, jnwright, yima, huang@uiuc.edu}

Abstract

This paper addresses the problem of generating a super-resolution (SR) image from a single low-resolution input image. We approach this problem from the perspective of compressed sensing. The low-resolution image is viewed as downsampled version of a high-resolution image, whose patches are assumed to have a sparse representation with respect to an over-complete dictionary of prototype signal-atoms. The principle of compressed sensing ensures that under mild conditions, the sparse representation can be correctly recovered from the downsampled signal. We will demonstrate the effectiveness of sparsity as a prior for regularizing the otherwise ill-posed super-resolution problem. We further show that a small set of randomly chosen raw patches from training images of similar statistical nature to the input image generally serve as a good dictionary, in the sense that the computed representation is sparse and the recovered high-resolution image is competitive or even superior in quality to images produced by other SR methods.

1. Introduction

Conventional approaches to generating a super-resolution (SR) image require multiple low-resolution images of the same scene, typically aligned with sub-pixel accuracy. The SR task is cast as the inverse problem of recovering the original high-resolution image by fusing the low-resolution images, based on assumptions or prior knowledge about the generation model from the high-resolution image to the low-resolution images. The basic reconstruction constraint is that applying the image formation model to the recovered image should produce the same low-resolution images. However, because much information is lost in the high-to-low generation process, the reconstruction problem is severely underdetermined, and the solution is not unique. Various methods have been proposed to further regularize the problem. For instance, one can choose a MAP (maximum a-posteriori) solution under generic image priors such as Huber MRF (Markov Random Field) and Bilateral Total Variation [14, 11, 25].

However, the performance of these reconstruction-based super-resolution algorithms degrades rapidly if the magnification factor is large or if there are not enough low-resolution images to constrain the solution, as in the extreme case of only a single low-resolution input image [2]. Another class of super-resolution methods that can overcome this difficulty are learning based approaches, which use a learned co-occurrence prior to predict the correspondence between low-resolution and high-resolution image patches [12, 26, 16, 5, 20].

In [12], the authors propose an example-based learning strategy that applies to generic images where the low-resolution to high-resolution prediction is learned via a Markov Random Field (MRF) solved by belief propagation. [23] extends this approach by using the Primal Sketch priors to enhance blurred edges, ridges and corners. Nevertheless, the above methods typically require enormous databases of millions of high-resolution and low-resolution patch pairs to make the databases expressive enough. In [5], the authors adopt the philosophy of LLE [22] from manifold learning, assuming similarity between the two manifolds in the high-resolution patch space and the low-resolution patch space. Their algorithm maps the local geometry of the low-resolution patch space to the high-resolution patch space, generating high-resolution patch as a linear combination of neighbors. Using this strategy, more patch patterns can be represented using a smaller training database. However, using a fixed number K neighbors for reconstruction often results in blurring effects, due to over- or under-fitting.

In this paper, we focus on the problem of recovering the super-resolution version of a given low-resolution image. Although our method can be readily extended to handle multiple input images, we mostly deal with a single input image. Like the aforementioned learning-based methods, we will rely on patches from example images. Our method does not require any learning on the high-resolution patches, instead working directly with the *low-resolution* training patches or their features. Our approach is motivated

방법론 | 핵심 아이디어

- Convolutional Neural Networks with 3-layers

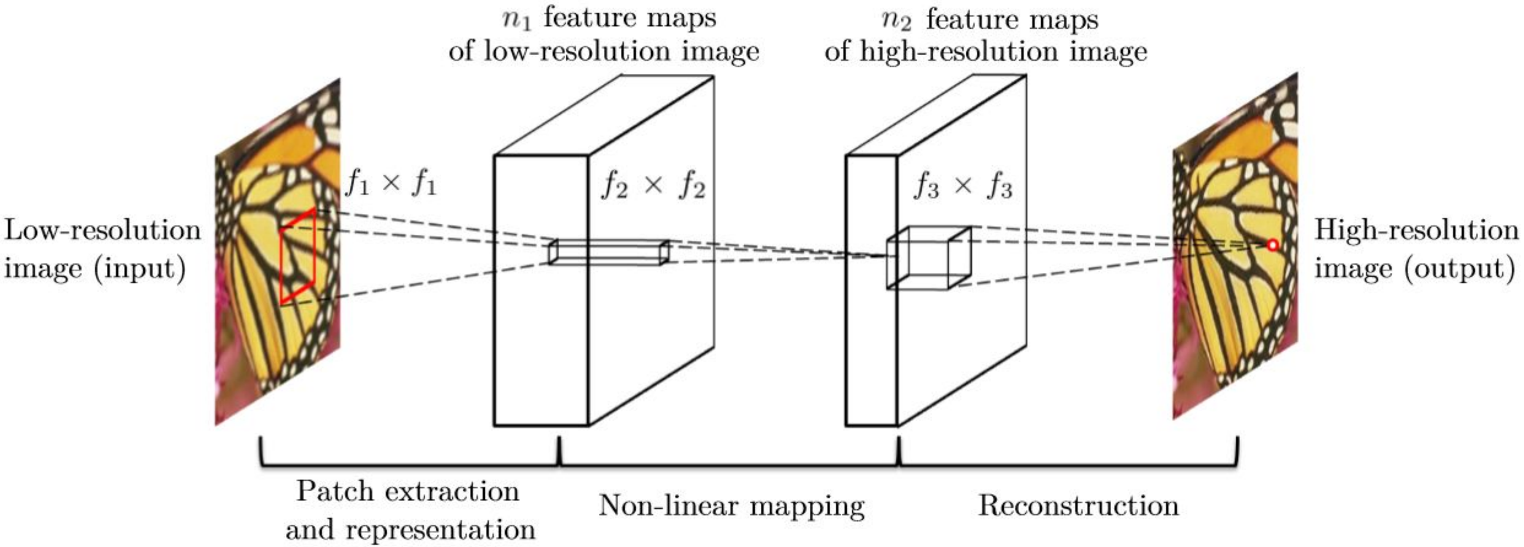
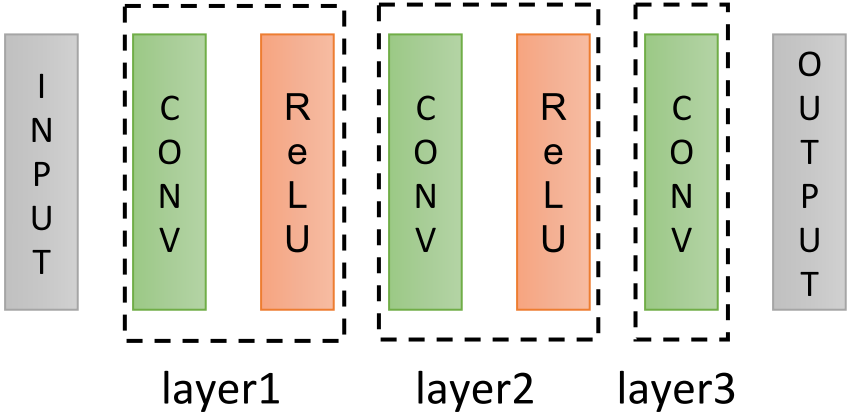


Fig. 2. Given a low-resolution image $\mathbf{Y}$, the first convolutional layer of the SRCNN extracts a set of feature maps. The second layer maps these feature maps nonlinearly to high-resolution patch representations. The last layer combines the predictions within a spatial neighbourhood to produce the final high-resolution image $F(\mathbf{Y})$.

방법론 | 핵심 아이디어



$$F_1(\mathbf{Y}) = \max\left(0, W_1 * \mathbf{Y} + B_1\right)$$

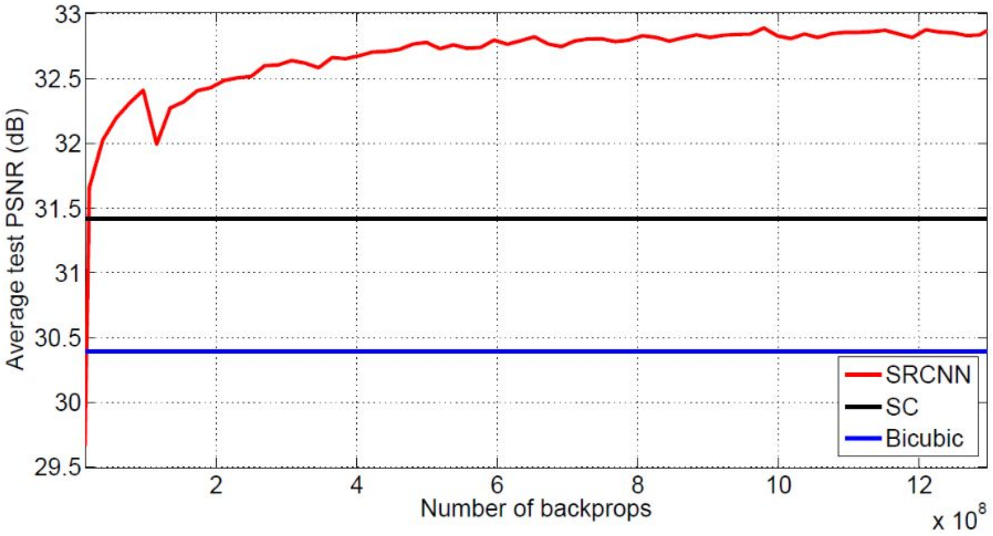
$$F_2(\mathbf{Y}) = \max\left(0, W_2 * F_1(\mathbf{Y}) + B_2\right)$$

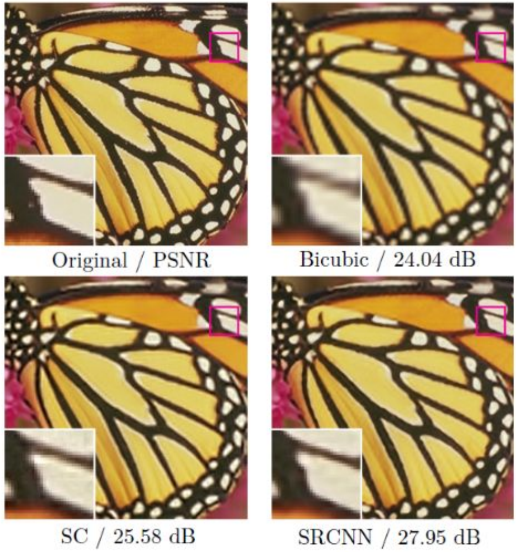
$$F(\mathbf{Y}) = W_3 * F_2(\mathbf{Y}) + B_3.$$

$$\boxed{L(\Theta) = \frac{1}{n}\sum_{i=1}^{n} \|F(\mathbf{Y}_i;\Theta) - \mathbf{X}_i\|^2,}$$

$$PSNR = 10 \cdot \log_{10}\left(\frac{MAX_I^2}{\boxed{MSE}}\right)$$

주요결과 | 실험 결과





The proposed Super-Resolution Convolutional Neural Network (SRCNN) surpasses the bicubic baseline with just a few training iterations, and outperforms the sparse-coding-based method (SC) with moderate training. The performance may be further improved with more training iterations.

주요결과 | 실험 결과

- Training data

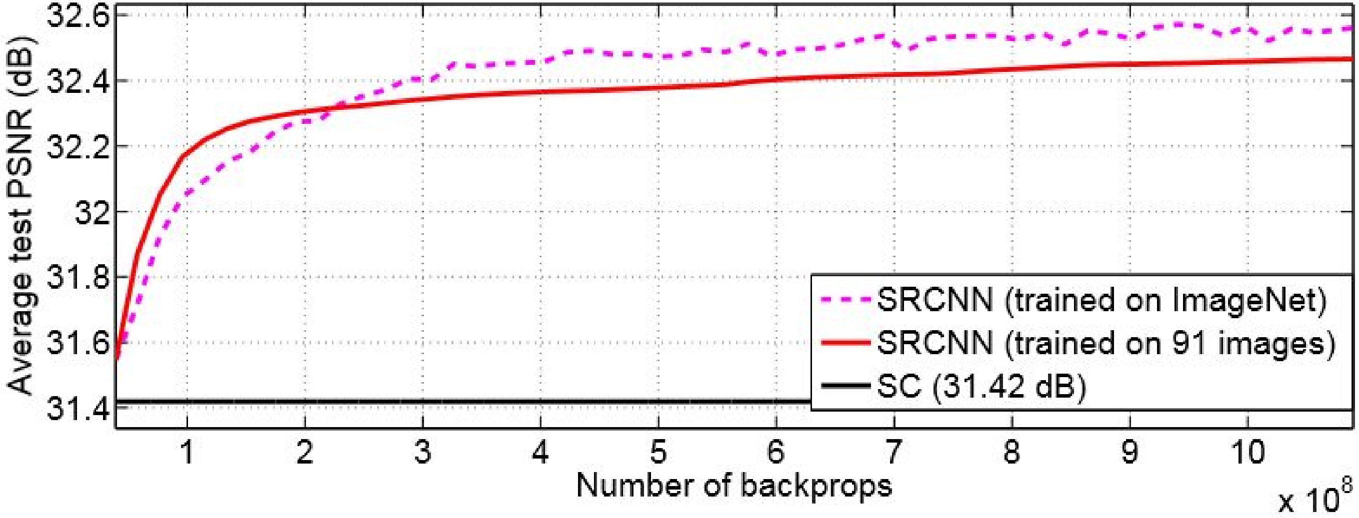


Fig. 4. Training with the much larger ImageNet dataset improves the performance over the use of 91 images.

주요결과 | 실험 결과

- Number of filters

TABLE 1
The results of using different filter numbers in SRCNN. Training is performed on ImageNet whilst the evaluation is conducted on the Set5 dataset.

$n_1 = 128$, $n_2 = 64$		$n_1 = 64$, $n_2 = 32$		$n_1 = 32$, $n_2 = 16$	
PSNR	Time (sec)	PSNR	Time (sec)	PSNR	Time (sec)
32.60	0.60	32.52	0.18	32.26	0.05

주요결과 | 실험 결과

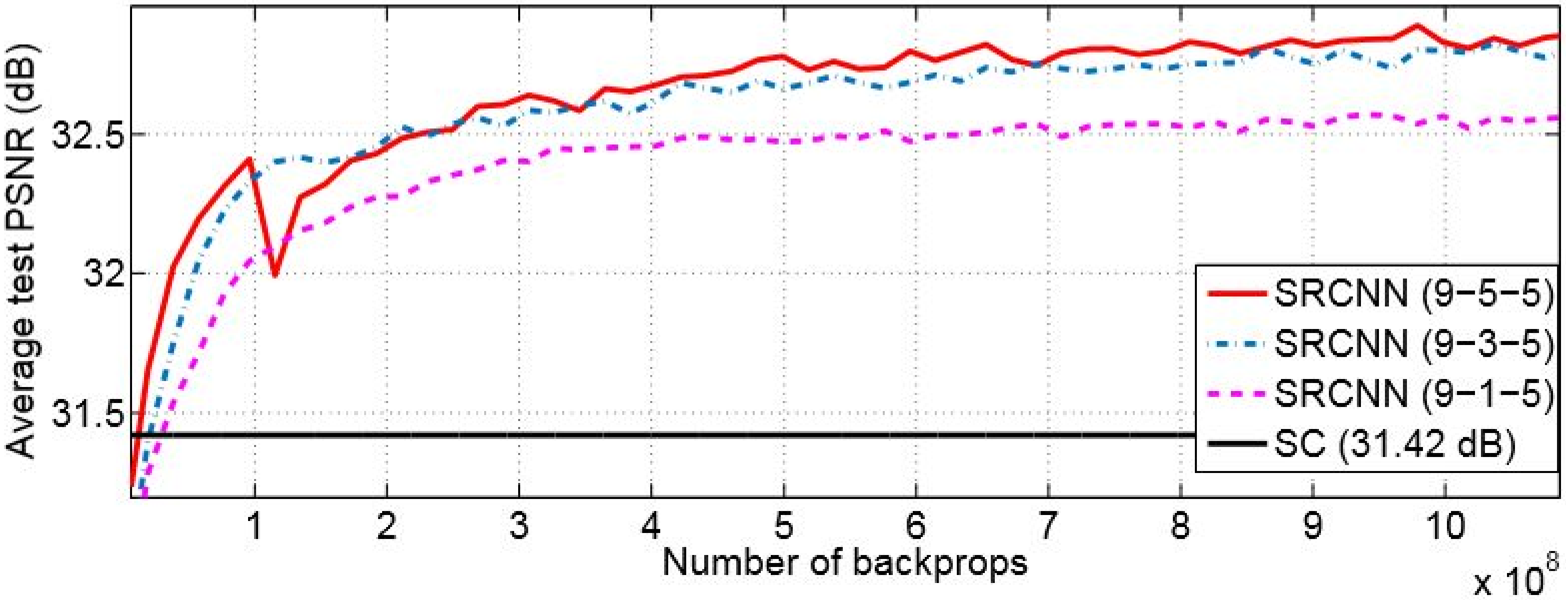


Fig. 7. A larger filter size leads to better results.

주요결과 | 실험 결과

- Color channels

TABLE 5
Average PSNR (dB) of different channels and training strategies on the Set5 dataset.

Training Strategies	PSNR of different channel(s)			
	Y	Cb	Cr	RGB color image
Bicubic	30.39	45.44	45.42	34.57
Y only	**32.39**	45.44	45.42	36.37
YCbCr	29.25	43.30	43.49	33.47
Y pre-train	32.19	**46.49**	**46.45**	36.32
CbCr pre-train	32.14	46.38	45.84	36.25
RGB	32.33	46.18	46.20	**36.44**
KK	32.37	44.35	44.22	36.32

리뷰의견 |

- 방법과 결과가 유효한가?
  - Sparse-coding과 같은 맥락의 표현을 CNN으로 구현
- 효율적인 방법인가?
  - Representation learning 관점에서 end-to-end learning이 가능하므로 효율적
- 독창적인가?
  - Single Image Super-Resolution 문제에 대해 최초로 딥러닝 적용
- 중요한 결과를 만들어냈는가?
  - 2014년 이후 거의 모든 Super-Resolution 문제에 대한 연구들이 SRCNN을 기반으로 함.
  - Super-Resolution 에 대한 breakthrough

Accurate Image Super-Resolution using Very Deep Convolutional Networks (2016)

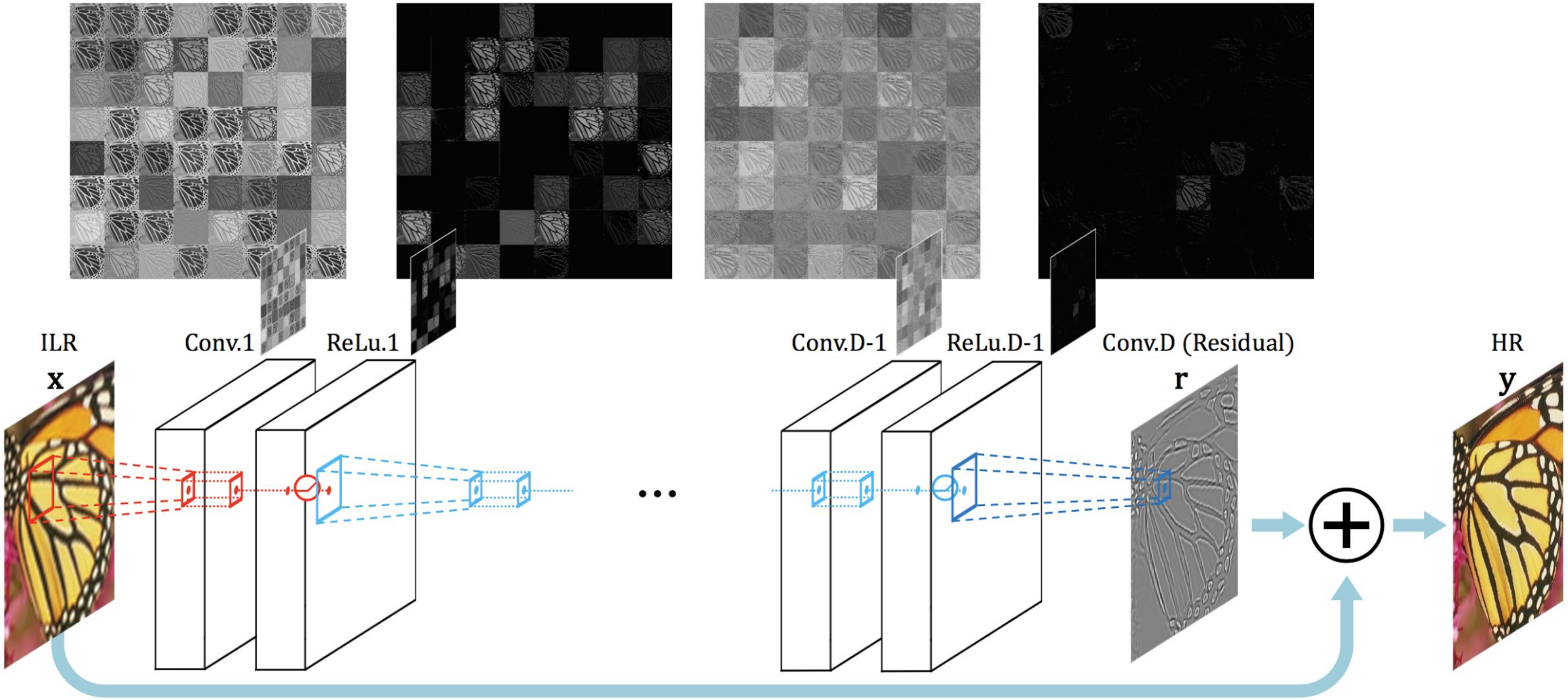
Paper reviewed by Taegyun Jeon

Kim, Jiwon, Jung Kwon Lee, and Kyoung Mu Lee. "Accurate image super-resolution using very deep convolutional networks." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016.

방법론 | 핵심 아이디어

	SRCNN	VDSR
Architecture	end-to-end learning	end-to-end learning
Receptive fields	13 x 13	41 x 41
Scale factors	x3	x2, x3, x4
Learning rate	10^{-5}	learning rate decay (10^{-2}, 10^{-6})
Depth	4	(Up to) 20

방법론 | 핵심 아이디어



방법론 | 핵심 아이디어

- Convolutional Neural Networks with 20 layers
  - Large receptive fields (41 x 41) with consecutive 3x3 convolution kernels
  - Convolutional layers with zero-padding
- Residual learning
  - Original loss: $MSE\ \frac{1}{2}\|y-f(x)\|^2$
  - Residual loss: $MSE\ \frac{1}{2}\|r-f(x)\|^2$, where $r = y - x$.

y: output
x: input
f(x): prediction

- Adjustable gradient clipping with high learning rates
  - Common gradient clipping: $[-\theta,\ \theta]$
  - Adjustable gradient clipping: $[-\frac{\theta}{\gamma}, \frac{\theta}{\gamma}]$

θ: predefined range
γ: current learning rate

방법론 | 핵심 아이디어

- Single CNN for multi-scale factor super-resolution
    - Size of input patch = Size of receptive fields (41 x 41)
    - Images are divided into sub-images with no overlap.
    - Mini-batch size is 64.
    - Sub-images from different scales can be in the same batch.

방법론 | 핵심 아이디어

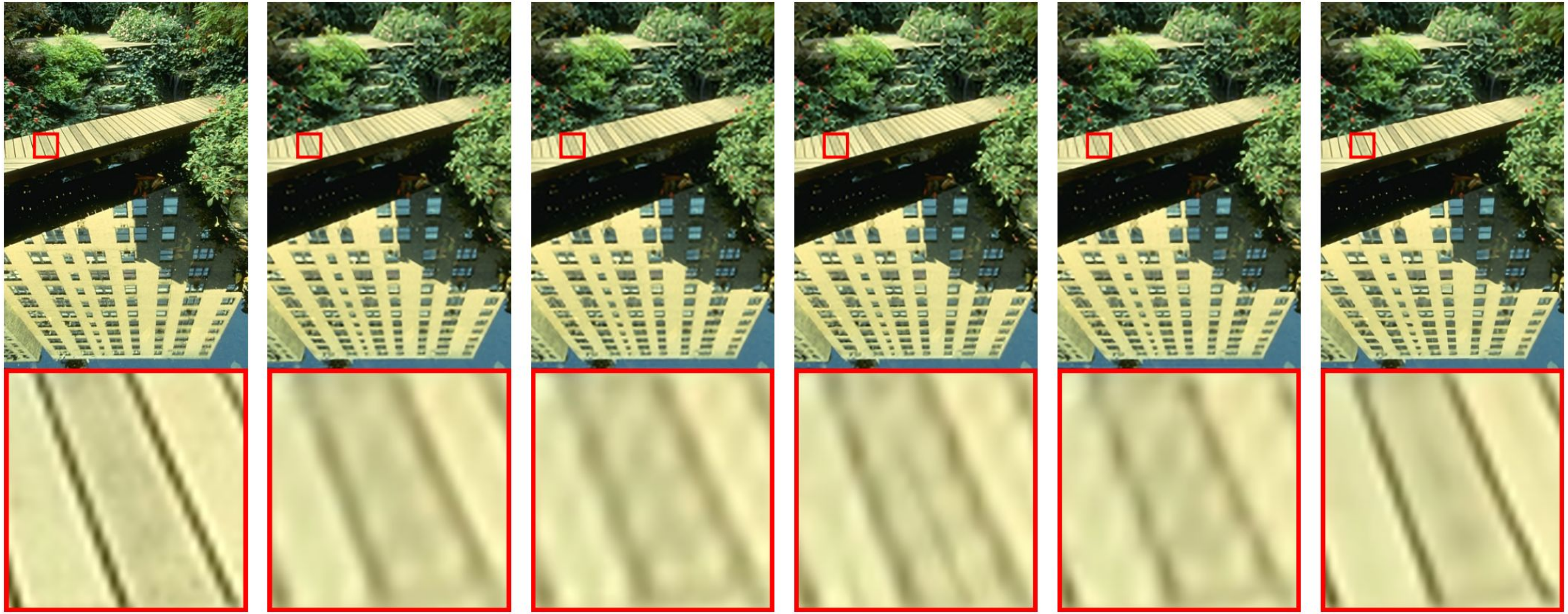
- Hyperparameters (Final model)
  - **depth** = 20
  - **batch_size** = 64
  - **momentum** = 0.9
  - **weight_decay** = 0.0001
  - **weight_initialization** = he_normal()
  - **activations** = ReLU
  - **epochs** = 80
  - **learning_rate** = 0.1
  - **learning_rate** was decreased by a factor 10 every 20 epochs.

주요결과 | 실험 결과

Dataset	Scale	Bicubic PSNR/SSIM/time	A+ [22] PSNR/SSIM/time	RFL [18] PSNR/SSIM/time	SelfEx [11] PSNR/SSIM/time	SRCNN [5] PSNR/SSIM/time	VDSR (Ours) PSNR/SSIM/time
Set5	×2	33.66/0.9299/0.00	36.54/0.9544/0.58	36.54/0.9537/0.63	36.49/0.9537/45.78	36.66/0.9542/2.19	37.53/0.9587/0.13
	×3	30.39/0.8682/0.00	32.58/0.9088/0.32	32.43/0.9057/0.49	32.58/0.9093/33.44	32.75/0.9090/2.23	33.66/0.9213/0.13
	×4	28.42/0.8104/0.00	30.28/0.8603/0.24	30.14/0.8548/0.38	30.31/0.8619/29.18	30.48/0.8628/2.19	31.35/0.8838/0.12
Set14	×2	30.24/0.8688/0.00	32.28/0.9056/0.86	32.26/0.9040/1.13	32.22/0.9034/105.00	32.42/0.9063/4.32	33.03/0.9124/0.25
	×3	27.55/0.7742/0.00	29.13/0.8188/0.56	29.05/0.8164/0.85	29.16/0.8196/74.69	29.28/0.8209/4.40	29.77/0.8314/0.26
	×4	26.00/0.7027/0.00	27.32/0.7491/0.38	27.24/0.7451/0.65	27.40/0.7518/65.08	27.49/0.7503/4.39	28.01/0.7674/0.25
B100	×2	29.56/0.8431/0.00	31.21/0.8863/0.59	31.16/0.8840/0.80	31.18/0.8855/60.09	31.36/0.8879/2.51	31.90/0.8960/0.16
	×3	27.21/0.7385/0.00	28.29/0.7835/0.33	28.22/0.7806/0.62	28.29/0.7840/40.01	28.41/0.7863/2.58	28.82/0.7976/0.21
	×4	25.96/0.6675/0.00	26.82/0.7087/0.26	26.75/0.7054/0.48	26.84/0.7106/35.87	26.90/0.7101/2.51	27.29/0.7251/0.21
Urban100	×2	26.88/0.8403/0.00	29.20/0.8938/2.96	29.11/0.8904/3.62	29.54/0.8967/663.98	29.50/0.8946/22.12	30.76/0.9140/0.98
	×3	24.46/0.7349/0.00	26.03/0.7973/1.67	25.86/0.7900/2.48	26.44/0.8088/473.60	26.24/0.7989/19.35	27.14/0.8279/1.08
	×4	23.14/0.6577/0.00	24.32/0.7183/1.21	24.19/0.7096/1.88	24.79/0.7374/394.40	24.52/0.7221/18.46	25.18/0.7524/1.06

Table 3: Average PSNR/SSIM for scale factor ×2, ×3 and ×4 on datasets Set5, Set14, B100 and Urban100. Red color indicates the best performance and blue color indicates the second best performance.

주요결과 | 실험 결과

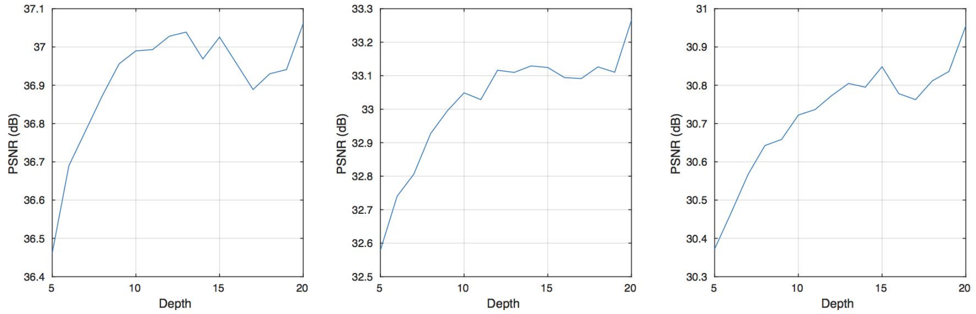


Ground Truth	A+ [22]	RFL [18]	SelfEx [11]	SRCNN [5]	VDSR (Ours)
(PSNR, SSIM)	(22.92, 0.7379)	(22.90, 0.7332)	(23.00, 0.7439)	(23.15, 0.7487)	(23.50, 0.7777)

Figure 6: Super-resolution results of "148026" (*B100*) with scale factor ×3. VDSR recovers sharp lines.

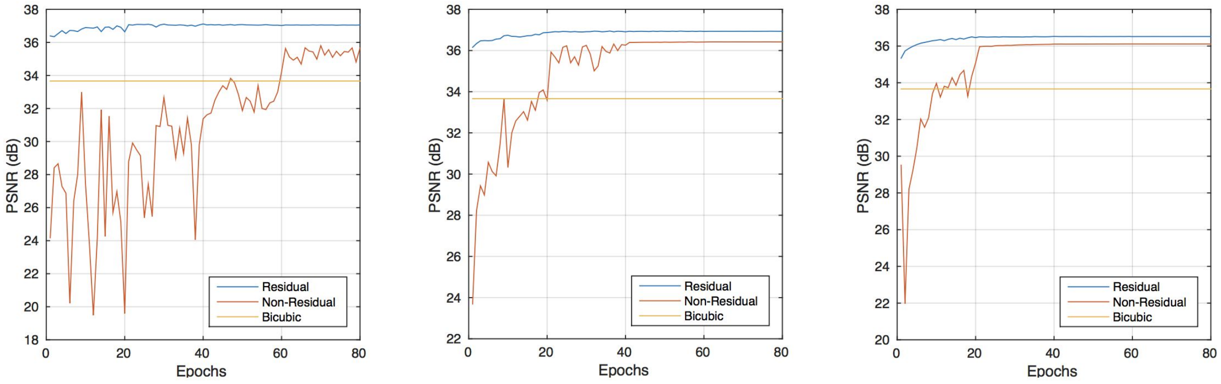
주요결과 | 실험 결과

-



(a) Test Scale Factor 2 **(b)** Test Scale Factor 3 **(c)** Test Scale Factor 4

Figure 3: Depth vs Performance



(a) Initial learning rate 0.1 **(b)** Initial learning rate 0.01 **(c)** Initial learning rate 0.001

리뷰의견 |

- 방법과 결과가 유효한가?
  - SRCNN을 20개의 레이어로 확장. 표현력 증가
- 효율적인 방법인가?
  - Residual learning과 adjustable gradient clipping을 통해 빠른 학습 달성
- 독창적인가?
  - Image restoration 문제에서 input-output의 유사성을 residual learning 적용
- 중요한 결과를 만들어냈는가?
  - 학습 속도 대폭 개선.
  - SRCNN과 비교하여 확연히 개선된 결과

부록 | 1. 구현 코드 (Caffe and Matlab)

Caffe and Matlab Implementation (Author):
http://mmlab.ie.cuhk.edu.hk/projects/SRCNN.html

부록 | 2. CNN 기반 관련 논문 리스트 (2014-2016)

- **Deep Networks for Image Super-Resolution with Sparse Prior** (ICCV2015),
  Zhaowen Wang et al.
- **Robust Single Image Super-Resolution via Deep Networks with Sparse Prior** (TIP2016),
  Ding Liu et al.
- **Accurate Image Super-Resolution Using Very Deep Convolutional Networks** (CVPR2016),
  Jiwon Kim et al.
- **Deeply-Recursive Convolutional Network for Image Super-Resolution** (CVPR2016),
  Jiwon Kim et al.
- **Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network** (CVPR2016),
  Wenzhe Shi et al.
- **Accelerating the Super-Resolution Convolutional Neural Network** (ECCV2016),
  Dong Chao et al.

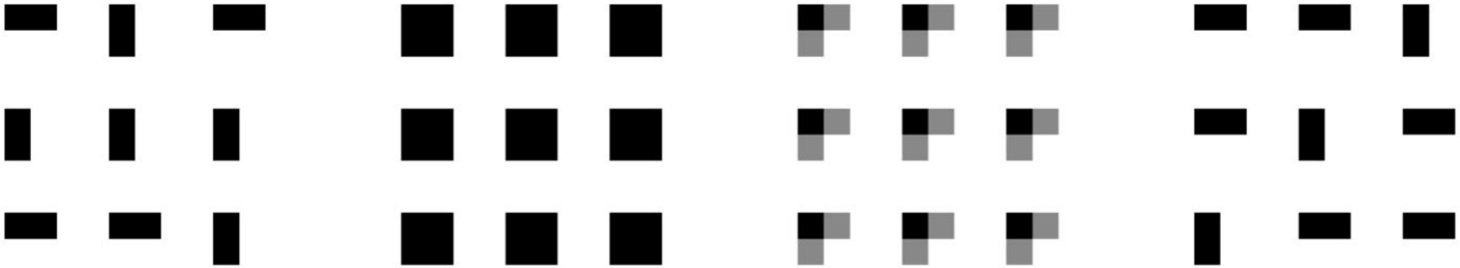
부록 | 2. GAN 기반 관련 논문 리스트 (2014-2016)

- **Perceptual Losses for Real-Time Style Transfer and Super-Resolution** (ECCV2016),
  Justin Johnson et al.
- **Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network**,
  Christian Ledig et al.
- **Amortised Map Inference for Image Super-Resolution**,
  Casper Kaae Sønderby et al.
- **EnhanceNet: Single Image Super-Resolution through Automated Texture Synthesis**,
  Mehdi S. M. Sajjadi et al.

부록 | 3. Challenge on Super-Resolution

부록 | Open Questions

- CNN 기반 모델은 GAN 기반 모델보다 무조건 성능이 나쁘다고 볼 수 있는가?



(I) I_{HR} (II) I_{LR} (III) I_{est} MSE (IV) I_{est} Adv.

부록 | Open Questions

- PSNR은 여전히 유효한 성능 지표인가?

감사합니다

fb.com/taegyun.jeon
github.com/tgjeon
taylor.taegyun.jeon@gmail.com

Paper reviewed by
Taegyun Jeon