

Super-Resolution을 위한 Deconvolution 적용 고속 컨볼루션 뉴럴 네트워크

이동현^{*}, 이호성^{**}, 이규중^{***}, 이혁재^{****}

Fast Very Deep Convolutional Neural Network with Deconvolution for Super-Resolution

Donghyeon Lee^{*}, Ho Seong Lee^{**}, Kyujoong Lee^{***}, Hyuk-Jae Lee^{****}

ABSTRACT

In super-resolution, various methods with Convolutional Neural Network(CNN) have recently been proposed. CNN based methods provide much higher image quality than conventional methods. Especially, VDSR outperforms other CNN based methods in terms of image quality. However, it requires a high computational complexity which prevents real-time processing. In this paper, the method to apply a deconvolution layer to VDSR is proposed to reduce computational complexity. Compared to original VDSR, the proposed method achieves the 4.46 times speed-up and its degradation in image quality is less than -0.1 dB which is negligible.

Key words: Single Image Super Resolution, Convolutional Neural Network, Deep Learning, Complexity Reduction, Memory Reduction

1. 서 론

단일 영상 초해상도(SISR, Single Image Super-Resolution) 기술은 저해상도 영상(Low resolution image)에 필터 기반의 연산을 적용하여 좋은 화질의 높은 해상도의 영상(Super resolution image)을 출력하는 기술을 의미한다[1]. SISR은 의료 영상[2], 보안 영상[3], TV[4]등 고해상도 영상이 필요한 분야에 적용되는 기술로 기존부터 많은 연구들이 진행 되어 왔다. 기존 SISR의 방법으로는 Bicubic 보간법과 같

이 픽셀 단위의 주변 유사도를 활용하는 방법, 에러를 예측함으로써 보간법을 개선한 자가 열화 보상 기법[18], 패치의 부분적인 공간을 보간하는 주변 데이터 내재화(Neighbor embedding)[5] 그리고 사전 기반(dictionary based) 방법[6]등이 있으며 최근 기계학습을 이용한 랜덤 포레스트(Random Forests)[7]를 사용한 방법도 있다.

기존 방법들의 공통된 문제점들은 단일 영상 초해상도 영상처리 출력 영상에 블러 현상이 있으며, 특히 고주파수 영역의 픽셀 값의 오류가 발생하여 시각

※ Corresponding Author : Kyujoong Lee, Address: (31460) 70, Sunmoon-ro 221 beon-gil, Tangjeong-myeon, Asan-si, Chungcheongnam-do, Korea, TEL : +82-41-530-2271, FAX : +82-41-530-2933, E-mail : kyujoonglee@sunmoon.ac.kr

Receipt date : Sep. 22, 2017, Revision date : Oct. 31, 2017
Approval date : Nov. 3, 2017

^{*} Inter-university Semiconductor Research Center (ISRC), Department of Electrical Engineering and Computer Science, Seoul National University (E-mail : dhlee@capp.snu.ac.kr)

^{**} Inter-university Semiconductor Research Center (ISRC), Department of Electrical Engineering and Computer Science, Seoul National University (E-mail : hslee@capp.snu.ac.kr)

^{***} Dept. of Electronic Eng., School of Engineering, Sun Moon University

^{****} Inter-university Semiconductor Research Center (ISRC), Department of Electrical Engineering and Computer Science, Seoul National University (E-mail : hjlee@capp.snu.ac.kr)

적인 결점이 존재하는 경우가 많다[8]. C. Dong은 컨볼루션 뉴럴 네트워크(CNN, Convolutional Neural Network)를 단일 영상 초해상도영상처리에 처음으로 적용한 SRCNN(Super-Resolution Convolutional Neural Network) [9]을 제안하여, 기존의 방법들 대비 월등히 높은 화질의 고해상도 영상(I^{SR})을 얻을 수 있음을 제시하였다. SRCNN은 컨볼루션 뉴럴 네트워크를 기존 방법들 보다 높은 고화질 영상을 출력하나, 저해상도 영상을 Bicubic 보간법을 통해 고해상도 영상으로 변환하는 전처리 과정이 필요하다. 이러한 전처리 과정은 컨볼루션 뉴럴 네트워크 이외에도 추가로 연산을 요구한다. 더불어 전처리과정 결과인 고해상도 영상에 대해서 컨볼루션 뉴럴 네트워크 연산을 수행하므로 컨볼루션 뉴럴 네트워크의 연산량 또한 영상의 면적에 비례하여 증가한다. 이러한 SRCNN의 연산량을 줄이기 위해 FSRCNN(Fast Super-Resolution Convolutional Neural Network) [10], ESPCN(Efficient Sub-pixel Convolutional Neural network)[2]과 같은 방법이 제안되었다. FSRCNN과 ESPCN은 SRCNN의 컨볼루션 뉴럴 네트워크의 마지막 레이어를 디컨볼루션 레이어로 변경함으로써, 저해상도 영상을 컨볼루션 뉴럴 네트워크에 입력하여 고해상도 영상을 출력한다. 이 방법으로 컨볼루션 뉴럴 네트워크 전체 연산량을 감소하여 동작 시간을 향상 시켰으며, 고해상도 출력 영상의 화질 저하 또한 최소화 하였다. FSRCNN, ESPCN은 기존의 SRCNN의 구조를 기반으로 제안되었으며, SRCNN의 구조를 벗어난 단일 영상 초해상도 영상 처리용 컨볼루션 뉴럴 네트워크인 VDSR(Very Deep convolutional networks for Super-Resolution) [11]이 제안되었다. VDSR 네트워크 구조는 Oxford 대학교의 Visual Geometry Group에서 제안한 영상 인식용 컨볼루션 뉴럴 네트워크인 VGG(Visual Geometry Group) [12] 구조를 기반으로 하여, 3x3 필터 크기의 20개 연속된 컨볼루션 레이어로 구성된다. 기존 SRCNN이 3개의 레이어를 사용하는 데에 반해, VDSR은 20개의 레이어를 사용하기 때문에 논문의 저자가 그 특징을 표현하기 위해 “Very Deep”이란 용어를 사용하였다. VDSR은 기존 SRCNN과 달리 잔차 신호를 학습함으로써, 영상의 화질이 매우 향상되었다. 또한, 잔차 신호는 0에 가까운 값이 많기 때문에 처리 속도가 빨라졌다. 그러나, VDSR은 20개

의 연속된 고해상도 영상 크기의 컨볼루션 레이어를 통과하는 구조로 여전히 실시간 동작에는 한계가 있다. 따라서, 본 논문은 VDSR의 연산량을 줄이기 위해, ESPCN 방법이 SRCNN에 적용한 디컨볼루션 레이어를 적용하는 방법과 개선된 실험 결과를 제시하고자 한다.

본 논문은 단일 영상 초해상도용 CNN 구조인 VDSR에 ESPCN에 적용된 디컨볼루션 레이어를 적용한 단일 영상 초해상도용 CNN구조를 제안한다. 디컨볼루션 레이어를 적용함으로써 VDSR은 저해상도 영상을 Bicubic 보간법으로 전처리과정을 거쳐 고해상도 크기의 영상을 사용하나 제안하는 CNN구조는 저해상도 영상을 전처리과정 없이 단일 영상 초해상도 연산을 수행한다. 제안하는 CNN구조는 디컨볼루션 레이어의 적용으로 VDSR과 비교하여 영상 확대 비율 x2 기준으로 연산량을 4배 감소 하였다. 이 연산량 감소는 제안하는 CNN 구조가 VDSR과 비교하여 테스트 연산을 4.46배 빠르게 동작하게 하며 평균 PSNR(Peak Signal-to-Noise Ratio)이 31.22 dB로 VDSR과 비교하여 PSNR저하가 0.1dB보다 작아 효율적인 단일 영상 초해상도용 CNN 구조임을 보인다.

본 논문은 총 다섯 개의 장으로 이루어지며 2장부터의 구성은 다음과 같다. 2장에서는 컨볼루션 뉴럴 네트워크 기반 super resolution 관련 연구인 SRCNN, FSRCNN, ESPCN 그리고 VDSR과 같이 본 연구의 기초가 되는 알고리즘들에 대해 설명한다. 3장에서는 기존의 VDSR 네트워크 구조에 디컨볼루션 네트워크를 적용한 방법 및 제안하는 네트워크를 학습하기 위한 방법을 기술한다. 4장에서는 실험 결과 및 고찰을 통해 제안한 네트워크 구조의 우수성을 실험 결과를 토대로 증명하며, 5장에서 본 논문의 결론을 기술한다.

2. 컨볼루션 뉴럴 네트워크 기반 super resolution 관련 연구

본 장은 기존의 단일 영상 초해상도 영상처리용 컨볼루션 뉴럴 네트워크인 SRCNN 및 FSRCNN, ESPCN, VDSR 대해 설명한다. SRCNN은 3 개의 레이어로 구성 된다. 첫 번째 레이어는 패치 추출 및 표현(Patch extraction and representation)을 수행하

며, 두 번째 레이어는 비선형 연결(Non-linear mapping), 마지막 레이어는 재건(reconstruction)을 수행한다. SRCNN은 컨볼루션 뉴럴 네트워크의 입력 영상으로 저해상도의 영상(Low resolution image, I^{LR})을 Bicubic 보간법을 이용하여 확대한 영상(High resolution image, I^{HR})을 사용한다. 따라서 컨볼루션 뉴럴 네트워크 연산 이전 preprocessing 연산 과정이 필요하며 고해상도 영상의 크기만큼 컨볼루션 뉴럴 네트워크의 연산을 수행하므로 연산량 및 메모리 사용량이 크다. 많은 연산량 및 메모리 사용량을 해결하기 위해 FSRCNN과 ESPCN 컨볼루션 뉴럴 네트워크 구조가 제안되었다. FSRCNN은 SRCNN의 두 번째 레이어인 비선형 연결 레이어를 4개의 작은 컨볼루션 레이어들로 나누었으며, SRCNN의 마지막 레이어인 재건 컨볼루션 레이어를 디컨볼루션 레이어로 바꾸었다. Table 1에서 보이는 것처럼, FSRCNN은 SRCNN 대비 파라미터 개수를 4.59배 줄이고, 입력 영상 메모리 사용량을 영상 확대 비율의 제곱만큼($1/\text{scale factor}(=s)^2$)로 감소시켰다. ESPCN은 SRCNN의 구조를 유지한 채 마지막 재건 레이어만 디컨볼루션 레이어로 교체하여 SRCNN 대비 파라미터 개수를 2.53배 줄이고, 입력 영상 메모리 사용량을 $1/s^2$ 로 감소시켰다. 동일한 신호라면 컨볼루션 뉴럴 네트워크의 파라미터 개수는 연산량과 비례하기 때문에, 파라미터 개수의 감소는 연산량의 감소를 의미한다.

VDSR은 VGG 컨볼루션 뉴럴 네트워크 구조를 단일 영상 초해상도 영상 처리에 적용하였다. VDSR은 기존 컨볼루션 네트워크 방법들의 단점인 작은 컨텍스트 크기로 인한 출력 영상 화질의 한계에 주목하였다. VDSR은 3×3 크기의 컨볼루션 필터만으로 구성하였으며 20개의 레이어를 사용한다. VDSR구조는

41×41 의 큰 컨텍스트 윈도우를 사용하고(SRCNN, FSRCNN, ESPCN의 각각 컨텍스트 윈도우 크기는 17×17 과 21×21 , 9×9 이다.), 잔차 신호를 학습하여, 기존의 컨볼루션 뉴럴 네트워크 기반의 단일 영상 초해상도 영상 처리 방법들의 화질 성능을 뛰어넘는 결과를 보인다. VDSR은 잔차 신호를 대상으로 하기 때문에, 대부분의 신호가 0에 가까워서, SRCNN, FSRCNN, ESPCN에 비해 파라미터의 수가 많아도, 처리 속도가 향상 되었다. 그러나, VDSR은 20개의 연속된 고해상도 영상 크기의 컨볼루션 레이어의 구조이기 때문에 실시간 처리를 위해서는 연산량 감소가 필요하다. 본 논문은 VDSR의 동작 시간 향상을 위해 SRCNN의 동작 시간을 향상시킨 ESPCN의 디컨볼루션 레이어를 적용하여 컨볼루션 네트워크 전체 연산량을 감소하는 방법을 제안하고자 한다.

3. 제안한 컨볼루션 뉴럴 네트워크 기반 super resolution 방법

3.1 제안한 컨볼루션 뉴럴 네트워크 구조

본 논문은 컨볼루션 뉴럴 네트워크 기반으로 super resolution 영상(I^{SR} , Super resolution image)을 출력하는 방법을 제안한다. 제안하는 방법은 기존의 VDSR구조를 기반으로 super resolution 화질을 유지하면서 연산량 및 입력 영상 메모리 사용량이 모두 $1/s^2$ 만큼 감소한 컨볼루션 네트워크 구조를 제안한다. FSRCNN과 ESPCN은 기존의 컨볼루션 네트워크의 연산량을 줄이기 위한 방법으로 파라미터의 수를 줄이는 방법을 사용했다. 그러나, VDSR 구조는 컨볼루션 필터의 최소 단위인 3×3 크기의 필터를 사용하므로 필터 크기를 줄이는 방법은 불가능하므로,

Table 1. Comparison of the structure and the number of parameters of CNNs for SISR. The convolution layer is represented by CONV(R, M, N), and R, M, and N denote the size of convolution filter, the number of filters, and the number of channels of input data, respectively

	SRCNN-Ex	FSRCNN	ESPCN	VDSR
First part	Conv(9, 64, 1)	Conv(5, 56, 1)	Conv(5, 64, 1)	Conv(3, 64, 1)
Mid part	Conv(5, 32, 64)	Conv(1, 12, 56), 4*Conv(3, 12, 12), Conv(1, 56, 12)	Conv(3, 32, 64)	18*Conv(3, 64, 64)
Last part	Conv(5, 1, 32)	DeConv(9, 1, 56)	Conv(3, s^2 , 32)	Conv(3, 1, 64)
Parameters	57,184	12,464	21,184 ($s=2$)	664,704
Input image	I^{HR}	I^{LR}	I^{LR}	I^{HR}

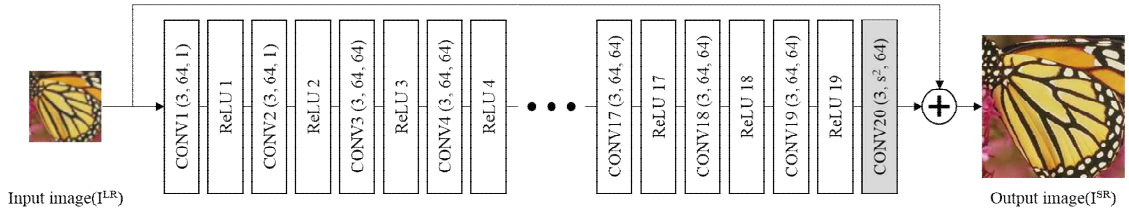


Fig. 1. Proposed CNN for super resolution. The CONV20 operates as a deconvolution layer(gray coloured layer).

디컨볼루션 레이어를 적용하여 저해상도 영상(I^{LR})을 입력받아 고해상도 영상(I^{SR})을 출력하여 실제 연산량 및 입력 영상 메모리 크기를 감소시켰다. Fig. 1은 제안한 super resolution 컨볼루션 뉴럴 네트워크의 구조를 보인다.

제안하는 CNN 구조는 CONV1부터 CONV19 까지 기존의 VDSR과 동일하며 마지막 컨볼루션 레이어인 CONV20는 디컨볼루션 레이어로 여러 개의 출력 채널을 생성한다. CONV20의 필터 개수(M)는 super resolution 영상의 영상 확대 비율의 제곱(= s^2)이다. 입력 영상과 CONV20 출력 결과 각 채널 데이터를 더하는 방법으로 고해상도 super resolution 영상을 출력한다. 제안하는 CNN 구조의 활성화 함수(Activation function)은 Rectified Linear Unit(ReLU)를 사용한다.

3.2 컨볼루션 뉴럴 네트워크의 연산량 및 메모리 감소 기법

기존의 CNN 구조인 VDSR은 Bicubic 보간법으로 전처리 과정을 거쳐 컨볼루션 뉴럴 네트워크에 고해상도 크기의 영상(I^{HR})을 입력하여 화질을 개선한 고해상도 영상을 출력한다. 반면 제안하는 컨볼루션 뉴럴 네트워크 구조는 전처리 과정 없이 저해상도(I^{LR})

의 영상을 입력하여 고해상도의 super resolution 영상(I^{SR})을 출력한다. 고해상도 크기의 영상(I^{HR})을 컨볼루션 네트워크에 입력하는 기존의 방법들은 전처리의 추가 연산 및 컨볼루션 레이어마다 고해상도의 영상에 대해서 연산을 수행해야 한다. 제안하는 컨볼루션 네트워크 구조는 기존 CNN 구조의 전처리 연산의 제거와 네트워크 내 연산량 및 메모리 사용을 감소하기 위해 마지막 컨볼루션 레이어인 CONV20을 디컨볼루션 레이어[8]로 사용하였다. Fig. 2는 디컨볼루션 레이어를 적용하여 저해상도 영상으로부터 고해상도 super resolution 영상을 생성하는 방법을 보인다.

제안하는 컨볼루션 뉴럴 네트워크는 저해상도 입력 영상 I^{LR} 과 CONV20 연산 결과 출력되는 s^2 개 채널의 잔차 신호를 더하여 고해상도 영상 I^{SR} 을 출력한다. I^{SR} 을 생성하기 위해 CONV20 결과 각 채널의 잔차 신호는 입력 영상과 픽셀 단위로 더한 데이터를 시계방향으로 배치한다. 식 (1)은 I^{SR} 을 생성하기 위해 디컨볼루션 레이어 결과 s^2 개의 채널로 이루어진 tensor(T)를 I^{SR} 에 배치하는 방법($PL_{SR}(T)_{x,y}$, Pixel Location in Super Resolution)을 보인다[8].

$$PL_{SR}(T)_{x,y} = T_{\text{floor}(x/r), \text{floor}(y/r), r \cdot \text{mod}(y,r) + \text{mod}(x,r)} \quad (1)$$

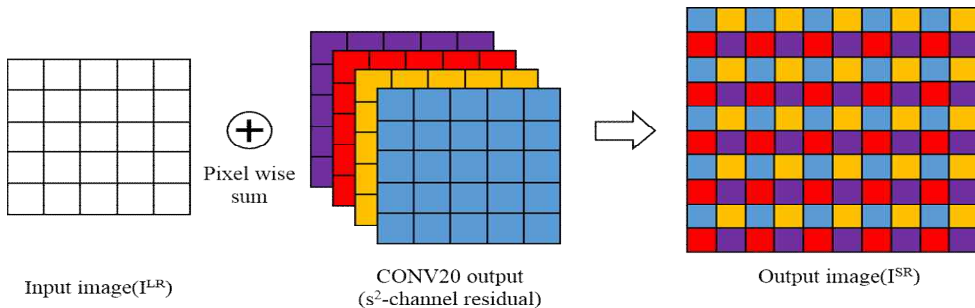


Fig. 2. Example of deconvolution layer($s=2$). The deconvolution layer is adopted to VDSR and this layer reduces computational costs and CNN's internal memory.

디컨볼루션 레이어를 적용한 제안하는 구조와 VDSR의 연산량을 영상 확대 비율이 2인 경우를 예로 들어 비교하면 다음과 같다. 제안하는 구조는 W(입력 영상 가로)*H(입력 영상 세로) 크기의 입력 영상을 전처리 과정 없이 그대로 사용하는 반면 VDSR은 전처리 과정에서 입력 영상을 가로, 세로 방향으로 2배씩 확대한 2W*2H 크기의 I^{HR} 영상에 대해 단일 영상 초해상도 연산을 수행 한다. 따라서 VDSR은 제안하는 구조와 비교하여 컨볼루션 뉴럴 네트워크에 입력할 영상을 저장하는 입력 영상 메모리 크기가 4배 크다. 제안하는 구조와 VDSR은 CONV1부터 CONV19까지 모두 3x3 크기의 64개 필터로 구성된다. 제안하는 구조는 입력 영상에 대해 필터 연산을 수행하나 VDSR은 I^{HR} 영상에 대해 필터 연산을 수행하므로 필터 연산 횟수가 4배 차이가 존재한다. 제안하는 구조의 컨볼루션 뉴럴 네트워크와 VDSR의 CONV20 연산을 비교하면 다음과 같다. 제안하는 구조의 컨볼루션 뉴럴 네트워크는 W*H 크기 64 채널 데이터를 W*H 크기 4 채널 데이터로 변환 한다. VDSR은 2W*2H 크기 64 채널 데이터를 2W*2H 크기 1 채널 데이터로 변환하므로 제안하는 구조와 VDSR의 CONV20 연산량은 동일하다.

따라서, 제안하는 방법은 디컨볼루션 레이어를 활용하여 네트워크를 구성하는 각 레이어당 연산량을 디컨볼루션 레이어를 제외하고 모두 1/s²로 감소하였다. 디컨볼루션 레이어의 연산량은 VDSR의 CONV20과 동일하다. 더불어 입력 영상을 I^{LR}을 사용함으로써 입력 영상 메모리 크기 또한 1/s²로 감소하였다.

3.2 초기 학습률 설정

본 장은 제안하는 컨볼루션 뉴럴 네트워크의 초기 학습률 (Initial Learning Rate) 설정에 대하여 기술한다. VDSR 구조는 초기 학습률 (Initial Learning

Rate)의 최적 값으로 0.1을 사용 하였다. 그러나, 제안하는 구조는 기존의 VDSR의 마지막 컨볼루션 레이어를 디컨볼루션 레이어로 사용하도록 변경했기 때문에, 제안하는 구조에 맞는 새로운 초기 학습률을 설정 해야 한다. 본 논문에서는 제안하는 구조를 학습하기 위한 초기 학습률을 0.5로 정하였으며, 이는 초기 학습률을 일정 범위 안에서 변경하며 실험적으로 정하였다. Table 2는 초기 학습률을 변경하며 측정 한 단일 영상 초해상도 영상처리 성능을 비교한 결과이다. 초기 학습률을 0.1부터 0.75까지 변경하여 측정 하였으며, 초기 학습률이 0.5인 경우 단일 영상 초해상도 출력 영상의 화질이 가장 좋음을 확인할 수 있다.

4. 실험 결과 및 고찰

4.1 실험 환경 구성

제안하는 super resolution 용 컨볼루션 뉴럴 네트워크의 실험 환경은 다음과 같다. 본 연구의 학습과 성능 측정을 위한 영상들은 VDSR과 동일하다. 학습용 데이터는 Berkeley Segment Dataset[13]의 291장의 영상을 사용 하였으며, 성능을 측정하기 위한 데이터는 ‘Set5’[14], ‘Set14’[15], ‘Urban100’[16], ‘B100’[6]을 사용하였다. 동작 속도를 측정하기 위한 환경으로는 Nvidia GTX 1080TI GPU를 사용 했으며 GPU 메모리를 30%로 제한하였다, 컨볼루션 뉴럴 네트워크의 구성 및 시뮬레이션은 TensorFlow를 사용하여 실험하였다.

제안하는 컨볼루션 뉴럴 네트워크의 초기 학습률은 실험적으로 정한 0.5로 설정하였다. 패치 크기는 네트워크의 컨텍스트와 동일하게 입력 영상 기준으로 41x41으로 설정 하였다. 영상 확대 비율이 2인 경우 출력 영상 기준 패치 크기는 82x82로 서로 겹치지 않게 생성하여, 기존 VDSR보다 학습 데이터의 패치

Table 2. Selecting initial learning rate for the proposed CNN

Learning Rate	Set 5		Set 14		Average	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
0.1	36.60	0.9531	32.44	0.9064	33.53	0.9187
0.25	37.1	0.9568	32.86	0.9106	33.98	0.9228
0.5	37.26	0.9577	32.97	0.9115	34.10	0.9237
0.75	37.16	0.9567	32.93	0.9105	34.04	0.9227

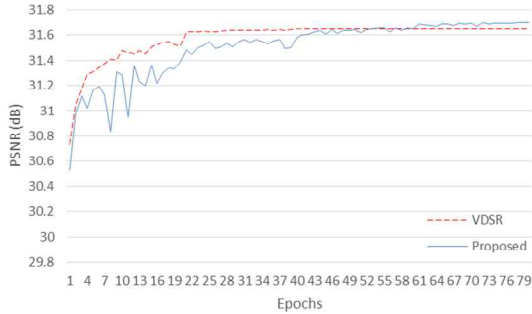


Fig. 3. The average PSNR test results of “B100” dataset in scale factor 2.

수는 1/4로 작다. 영상 확대 비율 3, 4에 대해서는 패치를 겹치게 하여 영상 확대 비율 2인 경우와 학습용 데이터 수를 동일하게 하여 네트워크를 학습 하였다. 컨볼루션 뉴럴 네트워크는 80 에폭 학습하였으며 20 에폭마다 학습률을 1/10으로 감소하였다. VDSR은 영상 확대 비율 2와 3, 4를 한 번에 학습 하며 각 에폭 마다 10 분의 시간이 소요된다. 제안하는 컨볼루션 뉴럴 네트워크 구조는 영상 확대 비율 2, 3, 4에 대해 에폭 마다 각각 12분, 38분, 60분의 시간이 소요 된다. 이는 영상 확대 비율의 제곱에 비례하여 학습 데이터의 패치 크기가 증가하기 때문이다.

4.2 실험 결과 비교 및 분석

본 장은 제안하는 컨볼루션 네트워크 구조를 이용한 단일 영상 초해상도 영상처리 결과를 정량적으로 비교한다. 제안하는 구조는 TensorFlow를 사용하여 구현하였으며, 동일한 실험 환경에서의 비교를 위해 기존의 VDSR도 TensorFlow[17]를 사용한 결과와 비교한다. Super resolution 영상의 화질을 비교하기 위해 PSNR과 SSIM을 사용하였다. Fig. 3은 제안하는 컨볼루션 뉴럴 네트워크와 VDSR의 각 에폭마다 B100 영상들에 대한 평균 PSNR 결과를 보인다. B100 영상들 평균 PSNR은 전체 영상 평균 PSNR과 가장 유사하여 각 에폭마다 단일 영상 초해상도 영상 처리 성능을 보이기에 적합하다. Fig. 3의 VDSR 실험 결과는 직접 컨볼루션 뉴럴 네트워크를 학습 한 결과로 Table 3의 공개된 결과[17]와 차이가 있다. 그러나 두 컨볼루션 뉴럴 네트워크 구조가 에폭이 증가할수록 평균 PSNR 값이 수렴하는 것을 확인할 수 있다. Table 3은 실험 결과를 제시하고 있다. 영상 확대 비율이 3인 경우는 기존의 VDSR 대비 PSNR과 SSIM이 각각 0.02dB, 0.0002 상승 하였으며, 영상 확대 비율이 2인 경우는 PSNR과 SSIM이 각각 0.07dB, 0.0015 하락 하였으며, 영상 확대 비율이 4인

Table 3. Result on various test dataset images. PSNR(dB) and SSIM are used to measure image quality. And processing time(sec) of each of CNN is measured.

Test dataset	Upscaling factor(=s)	VDSR(TensorFlow)[11, 17]			Proposed		
		PSNR	SSIM	time	PSNR	SSIM	time
Set5	x2	37.24	0.9583	0.47	37.26	0.9577	0.76
Set14		32.80	0.9118	0.61	32.97	0.9115	0.37
B100		31.73	0.8949	0.07	31.68	0.8939	0.02
Urban100		30.35	0.9096	1.05	30.21	0.9072	0.33
Average		31.29	0.9041	0.56	31.22	0.9026	0.20
Set5	x3	33.37	0.9189	0.46	33.55	0.9191	0.20
Set14		29.67	0.8316	0.61	29.78	0.8313	0.38
B100		28.72	0.7968	0.07	28.73	0.7963	0.01
Urban100		26.82	0.8204	1.05	26.84	0.8215	0.15
Average		28.02	0.8126	0.56	28.04	0.8128	0.10
Set5	x4	31.09	0.8791	0.47	31.00	0.8721	0.20
Set14		27.87	0.7666	0.61	27.87	0.7619	0.36
B100		27.19	0.7248	0.07	27.13	0.7208	0.01
Urban100		24.96	0.7448	1.04	24.81	0.7376	0.09
Average		26.30	0.7401	0.56	26.21	0.7346	0.07

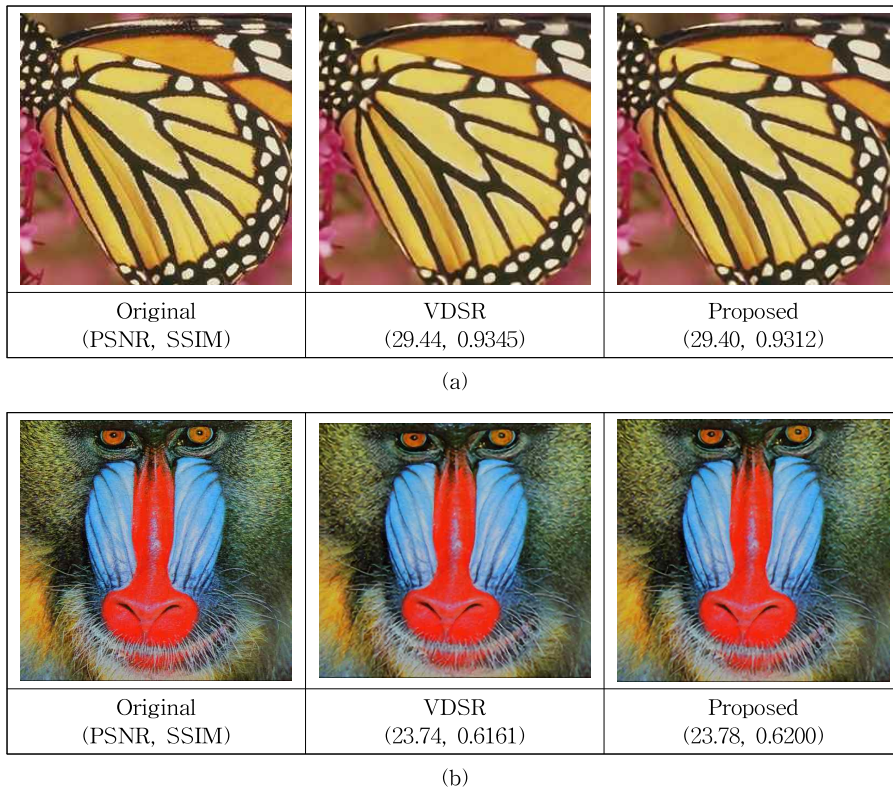


Fig. 4. Super-resolution result with scale factor x3. (a) Butterfly(Set5); (b) Baboon(Set14).

경우 또한 PSNR과 SSIM이 각각 0.09dB, 0.0055 하락 하였다. 영상 확대 비율이 2, 4에서 PSNR 및 SSIM의 하락 이유는 적은 양의 학습 데이터로 추정 된다. ESPCN은 디컨볼루션 레이어를 적용한 네트워크를 학습하기 위해 ImageNet내 50,000장의 영상으로 학습 하였다. 반면, 본 논문은 VDSR과 직접적인 비교를 위해 291장의 영상을 학습 하였다. 제안하는 컨볼루션 뉴럴 네트워크는 기존 VDSR 대비 파라미터가 $1/s^2$ 배 감소하였으며, 실제 CNN 동작 속도는 4.46배 향상되었다. Fig. 4는 제안하는 컨볼루션 네트워크와 VDSR의 단일 영상 초해상도 처리 결과 이미지를 제시하고 있다. 주관적 화질에서도 제안한 컨볼루션 뉴럴 네트워크가 기존 VDSR과 동등함을 확인 할 수 있다.

5. 결 론

본 논문은 컨볼루션 네트워크 기반 단일 영상 초해상도 영상 처리 방법 중 화질 성능이 높은 VDSR

구조의 연산량 및 입력 영상 메모리 사용량을 감소하는 방법을 제안 하였다. VDSR은 단일 영상 초해상도용 컨볼루션 네트워크 중 고해상도 영상의 화질이 매우 뛰어나지만 연산량이 작지 않아, 실시간 동작에 한계가 있다. VDSR은 컨볼루션 필터 중 가장 작은 크기의 3x3필터를 사용하므로 한 개의 큰 레이어를 작은 여러 개의 레이어로 나누어 파라미터 수를 감소하는 기존의 방법 적용이 불가능하다. 따라서 VDSR의 전체 연산량을 감소하기 위한 방법으로 ESPCN의 디컨볼루션 레이어 방법을 적용 하였다. ESPCN은 SRCNN 구조의 마지막 레이어를 디컨볼루션 레이어로 변경함으로써, SRCNN 대비 연산량을 감소한 방법이다. 본 논문은 VDSR에 디컨볼루션 레이어를 적용하여 전체 연산을 제거 하였으며 전체 연산량 및 입력 영상 메모리 크기를 감소하였다. 또한, 기존 VDSR과 비교하여 출력 영상의 화질을 유지함으로써 제안하는 방법이 기존 VDSR 대비 효율적인 컨볼루션 뉴럴 네트워크 구조임을 보였다.

REFERENCE

- [1] D. Glasner, S. Bagon, and M. Irani, "Super-resolution from a Single Image," *Proceeding of IEEE 12th International Conference on Computer Vision*, pp. 349-356, 2009.
- [2] S. Peled and Y. Yeshurun, "Superresolution in MRI: Application to Human White Matter Fiber Tract Visualization By Diffusion Tensor Imaging," *Magnetic Resonance in Medicine*, Vol. 45, No. 1, pp. 29-35, 2001.
- [3] L. Zhang, H. Zhang, H. Shen, and P. Li, "A Super-resolution Reconstruction Algorithm for Surveillance Images," *Signal Processing*, Vol. 90, No. 3, pp. 848-859, 2010.
- [4] T. Goto, T. Fukuoka, F. Nagashima, S. Hirano, and M. Sakurai, "Super-resolution System for 4K-HDTV", *Proceeding of IEEE 22nd International Conference on Pattern Recognition*, pp. 4453-4458, 2014.
- [5] H. Chang, D.Y. Yeung, and Y. Xiong, "Super-resolution through Neighbor Embedding," *Proceeding of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 1, pp. 275-282, 2014.
- [6] R. Timofte, V. De Smet, and L. Van Gool, "A+: Adjusted Anchored Neighborhood Regression for Fast Super-resolution," *Proceeding of Asian Conference on Computer Vision*, pp. 111-126, 2014.
- [7] S. Schuler, C. Leistner, and H. Bischof, "Fast and Accurate Image Upscaling with Super-resolution Forests," *Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3791-3799, 2015.
- [8] W. Shi, J. Caballero, F. Huszar, J. Totz, A.P. Aitken, and R. Bishop, "Real-time Single Image and Video Super-resolution Using an Efficient Sub-pixel Convolutional Neural Network," *Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1874-1883, 2016.
- [9] C. Dong, C. C. Roy, K. He, and X. Tang, "Image Super-resolution Using Deep Convolutional Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 38, No. 2, pp. 295-307, 2016.
- [10] C. Dong, C.C. Roy, and X. Tang, "Accelerating the Super-resolution Convolutional Neural Network," *Proceeding of European Conference on Computer Vision, Springer International Publishing*, pp. 391-407, 2016.
- [11] J. Kim, J.K. Lee, and K.M. Lee, "Accurate Image Super-resolution Using Very Deep Convolutional Networks," *Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1646-1654, 2016.
- [12] K. Simonyan and A. Zisserman, "Very Deep Convolutional networks for Large-scale Image Recognition," *Proceeding of International Conference on Learning Representations*, pp. 40-53, 2015.
- [13] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A Database of Human Segmented Natural Images and its Application to Evaluating Segmentation Algorithms and Measuring Ecological Statistics," *Proceedings of the IEEE International Conference on Computer Vision*, pp. 416-423, 2001.
- [14] M. Bevilacqua, A. Roumy, C. Guillemot and M.L.A. Morel, "Low-complexity Single-image Super-resolution Based on Nonnegative Neighbor Embedding," *Proceeding of the 23rd British Machine Vision Conference*, pp. 1-10, 2012.
- [15] R. Zeyde, M. Elad, and M. Protter, "On Single Image Scale-up Using Sparse-representations," *Proceeding of International Conference on Curves and Surfaces*, pp. 711-730, 2010.
- [16] JB. Huang, A. Singh, and N. Ahuja, "Single Image Super-resolution from Transformed Self-exemplars," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5197-5206, 2015.
- [17] A Tensorflow Implementation of "Accurate

Image Super-resolution Using Very Deep Convolution Networks,” <https://github.com/Jongchan/tensorflow-vdsr> (Accessed Sep., 21, 2016)

[18] W. Kim, K. Moon, and J. Kim, “Image Enhancement Using Improved Self Degradation Restoration Method,” *Journal of Korea Multimedia Society*, Vol. 16, No. 10, pp. 1180-1188, 2013.



이 규 중

2002년 서울대학교, 전자공학과 학사
2008년 University of Southern California, 전자공학과 석사
2013년 서울대학교, 전기컴퓨터공학부 박사

2013년~2017년 삼성전자 S.LSI 책임 연구원
2017년~현재 선문대학교 전자공학과 교수
관심분야: 딥러닝, 영상 처리, 영상 압축, 멀티미디어 SOC 설계



이 동 현

2012년 서강대학교, 전자공학과 학사
2012년~현재 서울대학교 전기정보공학부 석박사통합과정
관심분야: 영상 압축, 영상 인식, 영상 처리



이 혁 재

1987년 서울대학교, 전자공학과 학사
1989년 서울대학교 전자공학과 석사
1996년 Purdue University 전기컴퓨터공학 박사

1998년~2001년 Intel Corporation senior component design engineer
2001년~현재 서울대학교 전기정보공학부 교수
관심분야: 컴퓨터 구조 및 멀티미디어SoC 설계



이 호 성

2016년 서울대학교, 전기정보공학부 학사
2016년~현재 서울대학교, 전기정보공학부 석사과정
관심분야: 영상 인식, 영상 처리, Deep learning