

잡음제거 합성곱 신경망을 이용한 이미지 복원 방법

김선재[†], 이정호^{**}, 이석환^{***}, 전동산^{****}

Image Restoration Method using Denoising CNN

Seonjae Kim[†], Jeongho Lee^{**}, Suk-Hwan Lee^{***}, Dongsan Jun^{****}

ABSTRACT

Although image compression is one of the essential technologies to transmit image data on a variety of surveillance and mobile healthcare applications, it causes unnecessary compression artifacts such as blocking and ringing artifacts by the lossy compression in the limited network bandwidth. Recently, image restoration methods using convolutional neural network (CNN) show the significant improvement of image quality from the compressed images. In this paper, we propose Image Denoising Convolutional Neural Networks (IDCNN) to reduce the compression artifacts for the purpose of improving the performance of object classification. In order to evaluate the classification accuracy, we used the ImageNet test dataset consisting of 50,000 natural images and measured the classification performance in terms of Top-1 and Top-5 accuracy. Experimental results show that the proposed IDCNN can improve Top-1 and Top-5 accuracy as high as 2.46% and 2.42%, respectively.

Key words: Convolutional Neural Networks, Image Classification, Artifact Reduction, Image Denoising, Image Restoration

1. 서 론

최근 심층 신경망 기반 딥러닝 기법들은 이미지 분류와 객체 탐지와 같은 고수준 컴퓨터 비전(computer vision) 분야에서 높은 성능을 보이고 있으며 대중들에게 많은 관심을 받고 있다. 특히 이미지 분류는 컴퓨터 비전 분야의 주요 문제로 심층 신경망 기반 기술들이 개발된 이후 비약적으로 성능이 향상된 분야다. 이미지넷 대규모 시각 인식 대회(Image Net Large Scale Visual Recognition Challenge, ILSVRC)에서는 많은 연구자가 1,000개 클래스를 가

지는 이미지 데이터셋을 분류하는 심층 신경망을 개발하였고, Top-1과 Top-5 정답률을 이용해 제안된 심층 신경망 구조의 이미지 분류성능을 측정하였다 [1]. 이미지넷 대회에서 제안된 심층 신경망 구조는 대체로 여러 개의 합성곱 계층(convolutional layers)과 전 결합 계층(fully connected layers)들로 구성되며, 원본 이미지와 상응하는 정답 클래스로 구성된 이미지넷 데이터셋을 이용해 학습되었다. 이러한 심층 신경망 구조는 이미지 분류뿐만 아니라 객체 탐지/분할(object detection/segmentation), 얼굴 인식(face recognition), 이미지 초해상화(image super

* Corresponding Author : Dongsan Jun, Address: (49315) 37, Nakdong-daero 550, Saha-gu, Busan, Korea, TEL : +82-51-200-5823, FAX : +82-51-200-7783, E-mail : dsjun@dau.ac.kr

Receipt date : Dec. 21, 2021, Revision date : Jan. 11, 2022
Approval date : Jan. 12, 2022

[†] Department of Computer Engineering, Dong-A University (E-mail : sjkim@donga.ac.kr)

^{**} Department of Convergence IT Engineering, Kyungnam University (E-mail : ljhsorpion@paran.com)

^{***} Department of Computer Engineering, Dong-A University (E-mail : skylee@dau.ac.kr)

^{****} Department of Computer Engineering, Dong-A University

* Acknowledgement: This work was supported by the Technology Commercialization support program, through the Korea Innovation Foundation funded by the Ministry of Science and ICT.

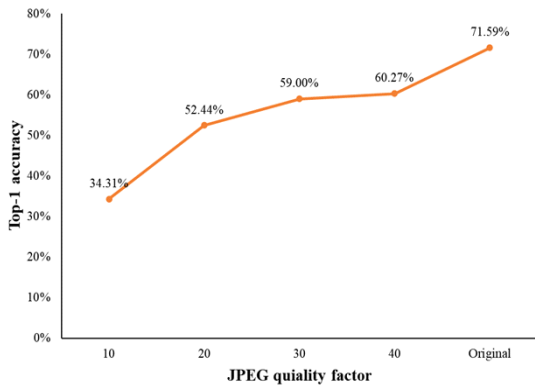


Fig. 1. Top-1 accuracy of VGG-16 model on ImageNet test dataset, which compressed with different quality factors by JPEG.

resolution, SR)와 같은 다른 분야의 컴퓨터 비전 문제에 대해 기반이 되는 구조로 사용되었으며 높은 성능을 보였다[2-8]. 하지만 원본 이미지를 입력받는 심층 신경망 기반 기술들은 사물인터넷(Internet of Things, IoT) 서비스나 휴대전화와 같은 저비트율(low bitrate) 환경에서 적용되기 어렵기 때문에 이미지 압축이 필요하게 된다. 일반적으로 이미지 압축은 고주파 성분을 제거하기 위한 블록 기반 부호화로 인해 블록킹 현상과 링잉 효과와 같은 원하지 않는 압축 손실이 발생한다. 이때 압축을 통해 양자화된 이미지는 원본 이미지 데이터만 학습한 이미지 분류 신경망의 성능을 떨어지게 하는 요인이 될 수 있다. Fig. 1은 ImageNet 테스트 데이터셋을 이미지 압축용 코덱인 JPEG(Joint Photographic Experts Group)[9]으로 원본 이미지를 압축한 이후 대표 이미지 분류 신경망 중 하나인 VGG16 모델에 입력했을 때 분류성능을 Top-1 정답률로 나타낸다. 이때 원본 이미지의 분류성능에 비해 압축된 이미지의 분류성능은 급격히 낮아지는 것을 확인할 수 있다. 예를 들어 원본 이미지에 대한 Top-1 정답률은 71.59%임에 반해 Quality Factor(QF) 10으로 압축된 이미지에 대한 분류성능은 34.31%로 원본 이미지에 비해 37.28%의 성능 저하가 일어남을 알 수 있다.

심층 신경망 기반 기술들은 이미지 잡음 제거

(image denoising)와 같은 저수준(low-level) 컴퓨터 비전 분야에서도 높은 잠재력을 보였다[10]. 일반적으로 이미지 잡음 제거 기법들은 JPEG 코덱 등으로 압축된 손실 이미지를 입력받고 원본 이미지에 가깝게 이미지를 개선하도록 학습한다. 대표적인 신경망 기반 잡음 제거 기법으로 2015년에 발표된 ARCNN [11]이 있으며, 이를 기반으로 많은 연구자들이 더 깊고 복잡한 신경망 구조를 제안해 객관적, 주관적 화질을 향상시켰다. 이때 이미지 잡음 제거 기법은 이미지 분류 신경망의 입력 이미지의 전처리로 사용함으로써 이미지 분류성능을 개선하는 데 사용될 수 있다. 본 논문에서는 Fig. 2와 같이 압축된 이미지의 잡음을 제거하기 위해 이미지 잡음 제거 합성곱 신경망(Image Denoising Convolutional Neural Networks, IDCNN)을 제안하며, 제안하는 이미지 잡음 제거 기법을 이미지 분류 신경망의 전처리로 사용했을 때 성능을 측정한다.

본 논문의 구성은 다음과 같다. 2장에서 심층 신경망 기반 이미지 분류 기법들과 이미지 잡음 제거 기법들에 대해 요약한 다음, 3장에서 제안하는 이미지 잡음 제거 기법을 설명한다. 그리고 4장에서는 실험 방법을 설명하고 제안 기법의 성능을 평가하며, 5장에서 결론을 맺는다.

2. 관련 연구

심층 신경망 기반 딥러닝 기법들은 각각 이미지 분류와 같은 고수준 컴퓨터 비전 문제와 이미지 개선과 같은 저수준 컴퓨터 비전 분야에 대해 높은 잠재력을 보여주었다. 본 절에서는 이미지 분류와 이미지 잡음 제거를 위한 신경망과 관련해 이전에 진행된 연구들을 설명한다.

2.1 심층 신경망 기반 이미지 분류

Krizhevsky는 5개의 합성곱 계층과 3개의 전결합 계층으로 구성된 AlexNet을 제안했다[12]. AlexNet은 2개의 GPU(Graphics Processing Units)를 이용한 병렬 처리를 위해 병렬 구조로 설계되었으며,

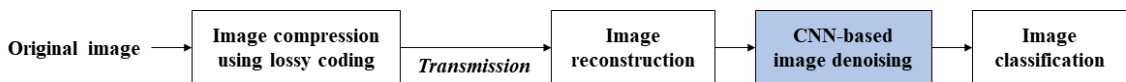


Fig. 2. Flowchart of the proposed methods for preprocessing of image classification networks.

ReLU(Rectified Linear Unit) 활성화 함수, 드롭아웃, 오버랩 풀링과 같은 학습 전략을 사용하였다. Simonyan는 대규모 시각 인식을 위해 합성곱 계층을 각각 11, 13, 16, 19개를 이용한 VGG(Visual Geometry Group) 신경망을 제안했다[13]. VGG는 신경망의 깊이와 이미지 분류성능의 상관관계를 효과적으로 조사했으며, 신경망의 깊이가 너무 깊어지게 될 때 발생하는 기울기 소실 문제에 대해서도 논의하였다. VGG 신경망이 소개된 이후 많은 연구들이 이미지 분류성능을 높이기 위해 효율적으로 더 깊은 심층 신경망을 설계하는 데 집중했다.

He는 잔차 학습(residual learning)을 적용한 ResNet(Residual Network)을 제안했으며, 제안하는 ResNet은 합성곱 계층 간에 적용되는 지역 스킵 연결(local skip connection)과 입력력 특성 맵 간에 적용되는 전역 스킵 연결(global skip connection)을 적용했다[14]. ResNet은 잔차 학습을 통해 100개 이상의 합성곱 계층을 기울기 소실 문제없이 효과적으로 학습하는 것을 성공적으로 보여주었다. Huang은 이전에 사용된 특성 맵을 계층적으로 연결해 재사용하기 위해 밀집 연결(dense connection)을 적용한 DenseNet을 제안했다[15]. DenseNet은 밀집 연결을 통해 더 유연하고 풍부한 특성들을 파라미터의 증가 없이 사용할 수 있었다. Xie는 ResNet과 대비해 파라미터의 수와 복잡도를 감소시키고 이미지 분류성능을 개선하기 위해 ResNeXt를 제안했다[16]. ResNeXt는 특성 맵을 여러 개의 경로로 나누어 합성곱 연산을 수행하고 다시 결합해주는 임베딩 기법을 이용해 파라미터의 수를 줄일 수 있었다.

Table 1은 앞서 설명한 이미지 분류를 위한 심층 신경망에 대해 50,000개의 이미지넷 테스트 원본 이미지 셋에 대해 분류를 수행한 결과를 나타낸다. 본문에서는 TorchVision 패키지에서 제공하는 미리 학습된 이미지 분류 신경망의 가중치를 이용하였다[17]. 이때 Top-1과 Top-5 정답률은 신경망 출력의 상위 1개 또는 상위 5개의 예측 결과 범위 내에 정답이 존재하는 경우의 비율이다.

2.2 심층 신경망 기반 압축 이미지 잡음 제거

딥러닝 기술의 발전으로 이미지 초해상화, 이미지 개선과 같은 저수준 컴퓨터 비전 분야 역시 확률 기반 기법들에서 심층 신경망 기반 기법들로 패러다임이 변화하였다. 이 분야에 대해 첫 번째 제안 방법으로 Dong은 3개의 합성곱 계층으로 구성된 SRCNN(Super Resolution CNN)를 제안하였다[5]. SRCNN은 보간 기법으로 생성된 고해상도 이미지를 입력으로 지정했으며, 입력과 원본 이미지 간 픽셀들의 매핑 함수를 학습시켰다. SRCNN이 발표된 이후 이전까지 이미지 분류에 사용된 기법들을 적용한 많은 이미지 개선 기법들이 소개되었다. SRCNN과 같은 신경망 구조는 압축으로 인한 잡음을 개선하는 데에도 사용될 수 있다. Dong은 JPEG 이미지 압축으로 생성되는 잡음을 제거하기 위해 SRCNN에서 1개의 합성곱 계층이 추가된 ARCNN(Artifact Reduction CNN)를 제안했다[11]. ARCNN은 JPEG 압축으로 생기는 블로킹, 링잉, 블러링 현상들을 제거해 객관적 화질뿐만 아니라 주관적 화질을 크게 개선했다. Dai은 비디오 압축 코덱인 고효율 비디오 코딩(High Efficiency Video Coding, HEVC)[18]의 I-frame 화질을 개선하기 위해 심층 신경망 기반 인루프 필터로써 VRCNN(Variable-filter-size Residue-learning CNN)을 제안했다[19]. VRCNN은 ARCNN과 비교해 이미지 개선 성능을 높였으며, 작은 크기의 필터를 이용해 파라미터의 개수도 줄일 수 있었다. Zhang은 17개의 합성곱 계층과 배치 정규화(batch normalization)와 잔차 학습을 적용한 DnCNN(Denoising CNN)을 제안했다[20]. DnCNN은 앞서 설명한 이미지 개선 신경망과 다르게 깊은 합성곱 계층들을 채택했으며, 배치 정규화를 통해 안정적인 학습을 진행할 수 있었다.

3. 제안한 방법

3.1 Image Denoising CNN(IDCNN) 구조

본 논문에서는 Fig. 3과 같이 압축 이미지의 품질

Table 1. The performance of the classification networks on the ImageNet test image dataset.

	AlexNet [12]	VGG16 [11]	ResNet101 [14]	DenseNet121 [15]	ResNeXt101 [16]
Top-1 accuracy (%)	56.55	71.59	77.37	74.43	79.31
Top-5 accuracy (%)	79.07	90.38	93.55	91.97	94.53

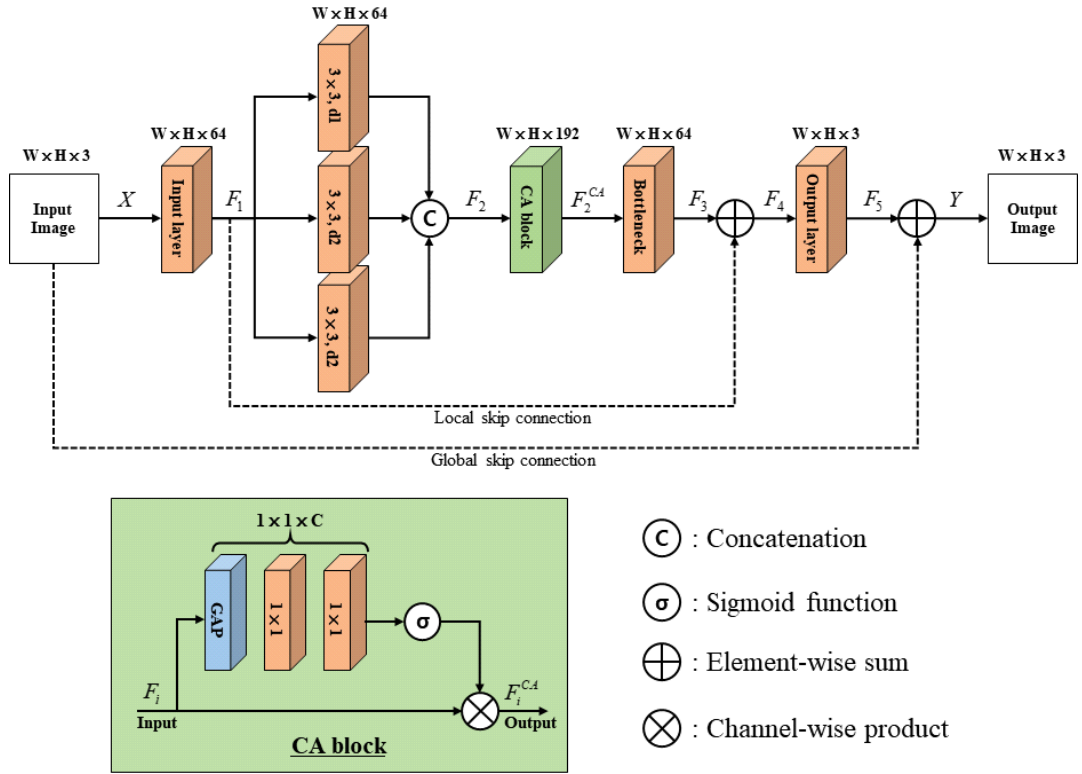


Fig. 3. Overall architecture of the proposed IDCNN.

개선 및 이미지 분류성능을 높이기 위해 Image Denoising CNN(IDCNN) 구조를 제안하고 있다. 제안된 IDCNN은 입력 계층, 특징 추출 및 취합 계층, 출력계층으로 구성된다. 이때 학습 시 빠른 수렴 속도와 잔차 학습을 구현하기 위해 2개의 스킵 연결이 존재하며 각각 입력 이미지를 출력 특징 맵에 더해주는 전역 스킵 연결과 IDCNN 내부 특징 맵과 연결된 지역 스킵 연결이 존재한다. 제안된 IDCNN에서 이

전 특징 맵 F_{i-1} 에 대해 합성곱 연산으로 생성되는 특징 맵 F_i 는 Eq. (1)과 같이 계산된다.

$$F_i = \delta_i(W_i * F_{i-1} + B_i) \tag{1}$$

여기서 δ_i , W_i , B_i , $*$ 는 각각 i 번째 계층의 Parametric Rectified Linear Unit(PReLU) 활성화 함수, 합성곱 계층의 가중치, 편향, 그리고 합성곱 연산을 나타낸다. 본 논문에서는 합성곱 연산을 통해 생성되

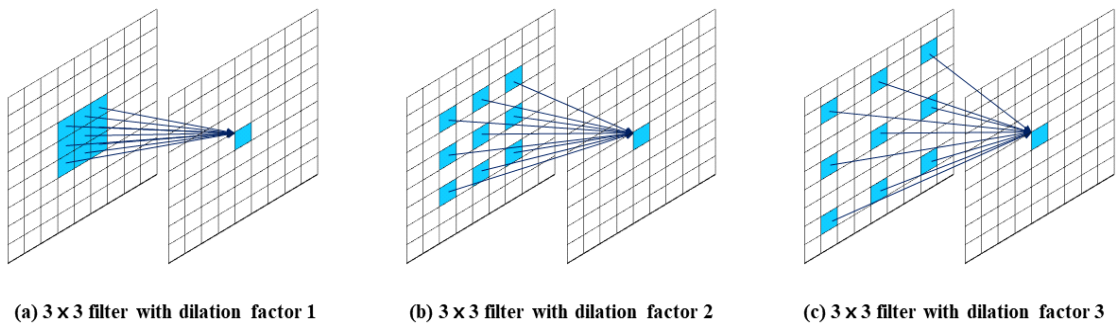


Fig. 4. Visualization of dilated convolution operation between three different dilation factors.

는 출력 특징 맵 F_i 의 채널 수는 64로 설정하였다.

제안된 IDCNN은 입력, 출력계층에서 3×3 크기의 필터를 사용하며, Bottleneck 계층은 1×1 크기의 필터를 사용한다. 특징 추출 계층에서는 3×3 크기의 필터를 사용하는 합성곱 계층이 존재하지만, 더 넓은 수용 영역(receptive field)을 사용하기 위해 서로 다른 Dilation Factor(DF)를 적용하였다[21]. Fig. 4는 서로 다른 DF에 대해 합성곱 연산이 적용되는 예시를 보여주며, 일반적인 합성곱 연산에 대해 추가적인 파라미터는 발생하지 않지만 넓은 수용 영역을 이용할 수 있음을 알 수 있다. 이때 DF에 따라 합성곱 연산에 사용되는 패딩의 범위도 변하게 되며, 필터의 크기 $Filter_w$ 와 DF가 주어졌을 때 패딩의 크기는 Eq. (2)와 같이 계산된다.

$$Padding\ Size = \lfloor (Filter_w \times DF - (DF - 1)) / 2 \rfloor \quad (2)$$

여기서 함수 $\lfloor \cdot \rfloor$ 는 정수 단위 내림 연산을 의미한다.

특징 취합 계층은 이전 계층에서 서로 다른 필터로 생성된 출력 특징 맵을 채널 차원으로 연결하며(concatenate), 각 채널 간 중요도를 판단하기 위해 Channel Attention(CA) 블록에 입력한다[22]. CA 블록으로 입력된 특징 맵 F_i 는 Global Average Pooling(GAP) 함수를 통해 $1 \times 1 \times C$ 형상으로 변환된 이후, 2개의 1×1 합성곱 계층과 시그모이드 활성화 함수를 통해 각 채널의 가중치를 결정하게 된다. 이후 생성된 가중치와 CA 블록의 입력 특징 맵의 채널 간 곱을 통해 Channel Attention이 적용된 특징 맵 F_i^{CA} 를 생성하게 된다. Bottleneck 계층은 이전 계층까지 사용된 특징 맵의 차원을 줄이기 위해 사용되며, 1×1 합성곱 연산을 통해 원래 채널의 개수로 변환한 특징 취합 계층의 출력 특징 맵은 지역 스킵 연결을 통해 F_4 를 생성하며, 출력계층에 입력된다. 출력계층은 이미지를 복원하기 위해 RGB 색상 요소에 해당하는 3개의 채널을 가지는 복원된 잔차 신호를 출력한다. 이후 Eq. (3)에 따라 전역 스킵 연결을 통해 최종적으로 복원된 이미지 Y 가 생성된다.

$$Y = X + F_5 \quad (3)$$

본 논문에서는 IDCNN의 입력 이미지 X 와 복원된 이미지 Y 각각에 대한 이미지 분류성능을 측정하고 그 결과를 제시한다.

3.2 IDCNN 학습

제안된 IDCNN은 압축된 이미지 X 를 입력받고 원본 이미지 O 에 가깝도록 복원 이미지 Y 를 출력하는 것을 학습의 목표로 한다. 따라서 IDCNN은 Eq. (4)와 같이 픽셀 간 차이를 나타내는 평균 제곱 오차(Mean Squared Error, MSE)를 손실 함수로 지정하고 하고, 이를 최소화하는 방향으로 신경망을 학습하게 된다.

$$L(\theta) = \frac{1}{N} \sum_{i=0}^{N-1} \|O_i - Y_i\|_2^2 \quad (3)$$

여기서 θ , N 는 각각 제안된 IDCNN의 파라미터, 배치의 크기를 나타낸다. Table 2는 제안된 IDCNN을 학습하기 위한 하이퍼 파라미터를 보여준다. 신경망이 학습되는 과정에서 128개 데이터로 구성된 배치 단위로 손실 함수인 MSE를 계산하고, Adam 최적화 기법[23]을 통해 신경망 파라미터 θ 를 업데이트 하게 된다. 이때 초기 학습률은 10^{-4} 로 지정하였으며, 학습이 진행됨에 따라 최적의 파라미터를 정교하게 탐색하기 위해 40 에폭 단위로 학습률을 10배 줄여주어 학습률 감소 기법을 적용하였다.

4. 실험 결과 및 고찰

4.1 Image Restoration

제안된 IDCNN을 학습하기 위해 자연 이미지 500장으로 구성된 BSDS500[24] 원본 데이터셋을 사용하였으며, 그중 400개의 이미지는 학습용, 100개의 이미지는 검증용 데이터셋으로 사용하였다. 다양한 종류의 압축을 임의로 생성하기 위해 원본 이미지 데이터셋은 이미지 압축 코덱인 JPEG을 이용해 각각 Quality Factor(QF) 10, 20, 30으로 압축을 진행하였다. 학습용 데이터셋의 전처리과정으로 $50 \times 50 \times 3$

Table 2. Hyper parameters used in the proposed IDCNN.

Hyper parameter	Options
Loss function	Mean squared error(MSE)
Optimizer	Adam[23]
Number of epochs	100
Batch size	128
Learning rate	10^{-4} to 10^{-6}
Activation function	PReLU, Sigmoid
Padding mode	Zero padding

Table 3. Experimental environment of the proposed IDCNN.

Settings	Options
Patch size	50×50×3
Number of patches	86,400
Operating system version	Ubuntu 16.04
Cuda version	10.1
Deep learning framework	Pytorch 1.4.0

Table 4. Average PSNR (dB) on validation dataset (100 images of BSDS500) where the best results of PSNR are shown in bold.

Quality Factor (QF)	JPEG [9]	ARCNN [11]	IDCNN (ours)
10	25.64	26.64	26.90
20	27.85	28.78	29.10
30	29.09	29.95	30.36
Average	27.53	28.46	28.79

크기의 패치들을 오버랩 되지 않게 추출하였으며, 불필요한 이미지의 경계 부분을 제거하였다. 결과적으로 본 논문에서는 86,400개의 이미지 패치 데이터를 이용해 제안된 IDCNN 학습을 진행하였다. Table 3은 상기 기술한 입력 데이터의 정보와 실험에 사용된 환경을 나타낸다.

제안된 신경망의 이미지 분류성능 향상 검증에 앞서 압축 이미지 품질 개선 성능을 확인해보았다. 이미지 품질 개선 성능을 검증하기 위해 BSDS500 검증용 데이터셋(100개 이미지)에서 원본 이미지와 제안 방법으로 복원된 이미지 간 최대 신호 대 잡음

비(Peak Signal-to-Noise Ratio, PSNR)를 계산했으며 Table 4에서 그 결과를 보여준다. 공정한 비교를 위해 JPEG[9]으로 압축된 이미지의 PSNR과 더불어 이전에 제안된 기법인 ARCNN[11]으로 복원된 이미지의 PSNR을 함께 측정하였다. 이때 제안된 IDCNN의 압축 이미지 복원 성능은 평균적으로 JPEG, ARCNN보다 각각 1.26dB, 0.33dB 높은 PSNR을 가짐을 확인할 수 있다.

4.2 Image Classification

제안된 IDCNN의 이미지 분류성능을 검증하기 위해 앞서 설명한 대표적인 이미지 분류 신경망 기법인 VGG16[13], ResNet101[14], DenseNet121[15] 구조를 이용하였으며, TorchVision[17]에서 제공하는 이미 학습된 모델 파라미터를 사용하였다. 이미지 분류는 1,000개의 클래스가 각각 50장으로 구성된 ImageNet 테스트용 데이터셋[1]을 사용하였으며, JPEG, ARCNN, 그리고 제안 방법인 IDCNN으로 복원된 이미지 데이터셋을 사전에 생성해 분류성능을 측정하였다. Table 5, 6은 각각의 기법으로 복원된 이미지에 대한 분류성능을 Top-1, Top-5 정확도를 이용해 나타낸다. Top-1 정확도의 경우 ResNet101 분류 모델을 제외한 모든 경우 IDCNN 복원 이미지가 가장 높은 성능을 보이며, 제안된 IDCNN은 3가지 분류 모델에 대해 JPEG, ARCNN 보다 각각 2.72%, 0.26% 높은 것을 확인할 수 있다. Top-5 정확도의 경우 제안된 IDCNN은 JPEG, ARCNN 보다 각각 2.63%, 0.20% 높은 평균 정확도를 가짐을 확인할 수 있다. 이에 따라 일반적으로 압축된 이미지를 분류 신경망에 그

Table 5. Top-1 accuracy on ImageNet test dataset where the best results of accuracy are shown in bold.

Classification Networks	QF	JPEG[9]	ARCNN[11]	IDCNN(ours)
VGG16[13]	10	34.31%	43.23%	43.45%
	20	52.44%	55.52%	55.39%
	30	59.00%	60.05%	60.47%
ResNet101[14]	10	55.53%	58.84%	58.53%
	20	66.89%	67.23%	67.21%
	30	69.89%	70.01%	70.40%
DenseNet121[15]	10	49.56%	54.06%	54.71%
	20	61.96%	62.70%	63.15%
	30	65.72%	65.84%	66.52%
Average	-	57.26%	59.72%	59.98%

Table 6. Top-5 accuracy on ImageNet test dataset where the best results of accuracy are shown in bold.

Classification Networks	QF	JPEG[9]	ARCNN[11]	IDCNN(ours)
VGG16[13]	10	58.87%	68.94%	69.05%
	20	77.40%	79.89%	79.60%
	30	82.33%	82.90%	83.52%
ResNet101[14]	10	78.61%	82.03%	81.82%
	20	87.70%	88.06%	88.03%
	30	89.62%	89.60%	89.80%
DenseNet121[15]	10	74.08%	78.39%	78.81%
	20	84.43%	85.06%	85.43%
	30	86.99%	86.98%	87.62%
Average	-	80.00%	82.43%	82.63%

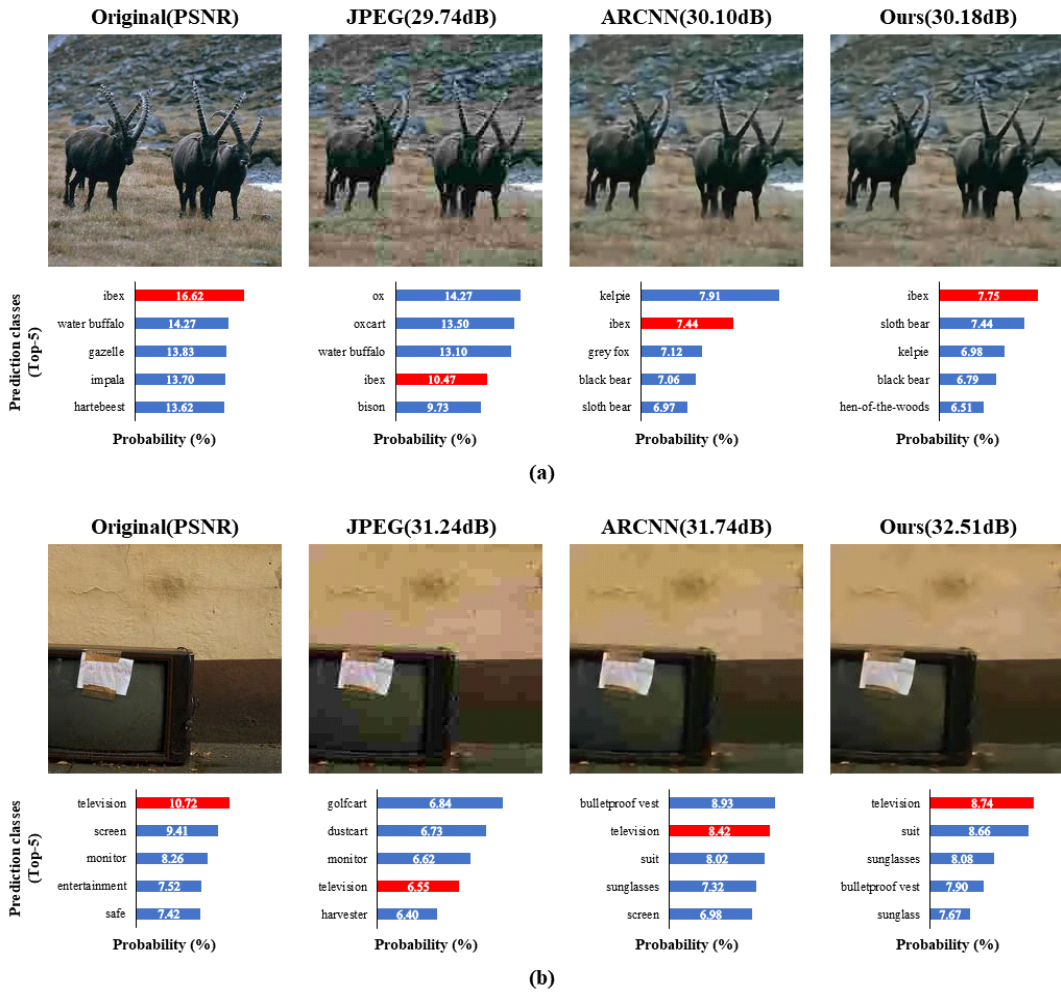


Fig. 5. Visual comparisons of reconstructed images and the Top-5 prediction classes on each image, where red bar indicate correct prediction class, (a) Results on '40252' image where 'ibex' is the correct class and (b) Results on '22274' image where 'television' is the correct class.

대로 입력하는 것 보다 전처리를 통해 개선된 이미지를 입력하는 것이 분류 신경망의 성능을 높이는 데 도움을 줄 수 있음을 알 수 있다.

Fig. 5는 ImageNet 테스트용 데이터셋에 대해 JPEG으로 압축된 이미지를 제안 방법으로 복원한 이미지의 시각화 예시를 나타내며, 이를 VGG16 분류 모델 입력하였을 때 예측된 상위 5개의 클래스를 함께 나타낸다. JPEG으로 압축된 이미지는 원본 이미지에 비해 블러킹 현상과 색차 성분의 변형이 발생하며, 특히 평탄한 배경 및 물체의 경계 부분이 변형되거나 소실됨을 확인할 수 있다. 제안된 IDCNN은 압축으로 발생하는 변형 및 손실을 제거하여 원본 이미지에 가장 가깝게 복원하는 것을 확인할 수 있다. Fig. 5의 예시에서 원본 이미지는 VGG16 분류 모델이 정답 클래스를 1순위로 예측하지만, JPEG으로 압축된 이미지는 정답 클래스의 예측 순위가 떨어지는 것을 확인할 수 있다. 또한 제안된 IDCNN은 이미지 복원을 통해 가장 높은 PSNR을 기록하며, 원본 이미지의 분류 예측과 동일하게 정답 클래스를 1순위로 예측함을 확인할 수 있다.

5. 결 론

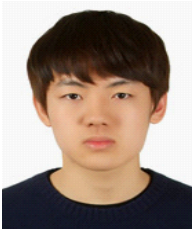
신경망 기반 이미지 처리는 최근 높은 성능을 보이며, 이 기술을 활용하기 위해 여러 분야에서 연구가 진행되고 있다. 하지만 모바일 기기 또는 사물인터넷과 같이 제한된 전송 환경에서는 이미지의 압축이 필요하게 된다. 이미지 압축은 이미지의 변형이나 손실을 일으키므로 원본 이미지로 학습된 신경망 모델은 압축된 이미지를 정상적으로 처리하지 못하는 문제가 존재한다. 본 논문에서는 합성곱 기반 이미지 잡음 제거 신경망 IDCNN을 제안하며, 제안 신경망을 압축된 이미지의 분류 전처리 단계에 사용해 이미지의 품질뿐만 아니라 이미지 분류성능을 개선하였다. 제안된 IDCNN은 BSDS500 검증용 데이터셋에 대해 JPEG 코덱으로 압축된 이미지와 이전 신경망 기반 이미지 복원 기법인 ARCNN 보다 각각 1.26dB, 0.33dB 높은 평균 PSNR을 가지며, 주관 품질 또한 가장 높은 것을 확인하였다. 이미지 분류성능을 검증하기 위해 본 논문에서는 50,000장의 이미지로 구성된 ImageNet 테스트용 데이터셋을 압축, 복원하고 대표적인 분류 신경망인 VGG, ResNet, DenseNet에 입력하여 Top-1, Top-5 정답률을 확인하였다. Top-

1 정답률의 경우, 제안된 IDCNN은 JPEG, ARCNN 보다 각각 2.72%, 0.26% 높은 평균 정확도를 가지며, Top-5 정답률은 각각 2.63%, 0.20% 높은 평균 정확도를 가지는 것을 확인하였다. 제안 방법은 이미지 분류 외에도 객체 탐지, 이미지 분할과 같은 고수준 컴퓨터 비전 문제에 대해서도 동일한 방법으로 적용 가능하며, 특히 원본 이미지로만 학습된 신경망 모델의 경우 재학습을 수행하지 않더라도 전처리만으로 성능 향상에 도움을 줄 수 있음을 기대할 수 있다.

REFERENCE

- [1] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, et al., "Imagenet Large Scale Visual Recognition Challenge," *International Journal of Computer Vision*, Vol. 115, No. 3, pp. 211-252, 2015.
- [2] N. Zhang, Y. Feng, and E.J. Lee, "Activity Object Detection Based on Improved Faster R-CNN," *Journal of Korea Multimedia Society*, Vol. 24, No. 3, pp. 416-422, 2021.
- [3] H.J. Park, Y.W. Lee, and B.G. Kim, "Efficient Tire Wear and Defect Detection Algorithm Based on Deep Learning," *Journal of Korea Multimedia Society*, Vol. 24, No. 8, pp. 1026-1034, 2021.
- [4] K.T. Kim, W.Y. You, and J.Y. Choi, "Attention Deep Neural Networks Learning based on Multiple Loss functions for Video Face Recognition," *Journal of Korea Multimedia Society*, Vol. 24, No. 10, pp. 1380-1390, 2021.
- [5] C. Dong, C.C. Loy, 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, 2015.
- [6] J. Kim, J.M. Lee, and K.M. Lee, "Accurate Image Super-resolution Using Very Deep Convolutional Networks," *Proceeding of IEEE conference on Computer Vision and Pattern Recognition*, pp. 1646-1654, 2016.
- [7] C. Leding, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta et al., "Photo-real-

- istic Single Image Super-resolution Using a Generative Adversarial Network,” *Proceeding of IEEE conference on Computer Vision and Pattern Recognition*, pp. 4681–4689, 2017.
- [8] T. Tong, G. Li, X. Liu, and Q. Gao, “Image Super-resolution Using Dense Skip Connections,” *Proceeding of IEEE International Conference on Computer Vision*, pp. 4799–4807, 2017.
- [9] G.K. Wallace, “The JPEG Still Picture Compression Standard,” *IEEE Transactions on Consumer Electronics*, Vol. 38, No. 1, pp. 18–34, 1992.
- [10] J. Liu, D. Liu, W. Yang, S. Xia, X. Zhang, and Y. Dai, “A Comprehensive Benchmark for Single Image Compression Artifact Reduction,” *IEEE Transactions on Image Processing*, Vol. 29, pp. 7845–7860, 2020.
- [11] C. Dong, Y. Deng, C.C. Loy, and X. Tang, “Compression Artifacts Reduction by a Deep Convolutional Network,” *Proceeding of IEEE International Conference on Computer Vision*, pp. 576–584, 2015.
- [12] A. Krizhevsky, I. Sutskever, and G.E. Hinton, “Imagenet Classification with Deep Convolutional Neural Networks,” *Advances in Neural Information Processing Systems*, Vol. 25, pp. 1097–1105, 2012.
- [13] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-scale Image Recognition,” *arXiv Preprint*, arXiv:1409.1556, 2014.
- [14] K. He, X. Zhang, S. Xiangyu, and J. Sun, “Deep Residual Learning for Image Recognition,” *Proceeding of IEEE conference on Computer Vision and Pattern Recognition*, pp. 770–778, 2016.
- [15] G. Haug, Z. Liu, L.V.D. Maaten, and K.Q. Weinberger, “Densely Connected Convolutional Networks,” *Proceeding of IEEE conference on Computer Vision and Pattern Recognition*, pp. 4700–4708, 2017.
- [16] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, “Aggregated Residual Transformations for Deep Neural Networks,” *Proceeding of IEEE conference on Computer Vision and Pattern Recognition*, pp. 1492–1500, 2017.
- [17] TorchVision Pre-trained Models, <https://pytorch.org/vision/stable/models.html> (accessed December, 09, 2021)
- [18] G.J. Sullivan, J.R. Ohm, W.J. Han, and T. Wiegand, “Overview of the High Efficiency Video Coding (HEVC) Standard,” *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 22, No. 12, pp. 1649–1668, 2012.
- [19] Y. Dai, D. Liu, and F. Wu, “A Convolutional Neural Network Approach for Post-processing in HEVC Intra Coding,” *Proceeding of International Conference on Multimedia Modeling*, pp. 28–39, 2017.
- [20] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, “Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising,” *IEEE Transactions on Image Processing*, Vol. 26, No. 7, pp. 3142–3155, 2017.
- [21] F. Yu and V. Koltun, “Multi-scale Context Aggregation by Dilated Convolutions,” *arXiv Preprint*, arXiv:1511.07122, 2015.
- [22] J. Hu, L. Shen, and G. Sun, “Squeeze-and-excitation Networks,” *Proceeding of IEEE conference on Computer Vision and Pattern Recognition*, pp. 7132–7141, 2018.
- [23] D.P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization,” *arXiv Preprint*, arXiv:1412.6980, 2014.
- [24] 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,” *Proceeding of IEEE International Conference on Computer Vision*, pp. 416–423, 2001.



김 선 재

2020년 8월 경남대학교 정보통신 공학과(공학사)

2021년 9월~현재 동아대학교 컴퓨터공학과(석사과정)

관심분야: 영상압축, 딥러닝 기반 영상처리



이 정 호

1994년 2월 건국대학교 컴퓨터 및 정보통신공학과 (공학석사)

2019년 9월~현재 경남대학교 융합IT공학과 (공학박사과정)

관심분야: 영상압축, 해양안전, 재난안전 등



이 석 환

1999년 경북대학교 전자공학과 (공학사)

2001년 경북대학교 전자공학과 (공학석사)

2004년 경북대학교 전자공학과 (공학박사)

2020년~현재 동아대학교 컴퓨터공학과 교수

2021년~현재 한국멀티미디어학회 논문지편집위원장

관심분야: 멀티미디어 콘텐츠보호, 디지털영상처리, 컴퓨터그래픽, 인공지능 및 기계학습



전 등 산

2002년 2월 부산대학교 전자컴퓨터공학부(공학사)

2004년 2월 KAIST 전기및전자공학과(공학석사)

2011년 2월 KAIST 전기및전자공학과(공학박사)

2021년~현재 동아대학교 컴퓨터공학과 조교수

관심분야: 영상압축, 지능형 영상처리, 머신러닝/딥러닝 등