

다중 채널 입력 Convolution residual neural networks 기반의 초해상화 기법

염광영, 김문철
한국과학기술원 전기 및 전자 공학과
yky5464@kaist.ac.kr, mkim@ee.kaist.ac.kr

Super-resolution based on multi-channel input convolutional residual neural network

Gwang-Young Youm, Munchurl Kim
Korea Advanced Institute of Science and Technology School of Electrical Engineering

요약

최근 Convolutional neural networks(CNN) 기반의 초해상화 기법인 Super-Resolution Convolutional Neural Networks (SRCNN) 이 좋은 PSNR 성능을 발휘하는 것으로 보고되었다 [1]. 하지만 많은 제안 방법들이 고주파 성분을 복원하는데 한계를 드러내는 것처럼, SRCNN 도 고주파 성분 복원에 한계점을 지니고 있다. 또한 SRCNN 의 네트워크 층을 깊게 만들면 좋은 PSNR 성능을 발휘하는 것으로 널리 알려져 있지만, 네트워크의 층을 깊게 하는 것은 네트워크 파라미터 학습을 어렵게 하는 경향이 있다. 네트워크의 층을 깊게 할 경우, gradient 값이 아래(역방향) 층으로 갈수록 발산하거나 0 으로 수렴하여, 네트워크 파라미터 학습이 제대로 되지 않는 현상이 발생하기 때문이다. 따라서 본 논문에서는 네트워크 층을 깊게 하는 대신에, 입력을 다중 채널로 구성하여, 네트워크에 고주파 성분에 관한 추가적인 정보를 주는 방법을 제안하였다. 많은 초해상화 기법들이 고주파 성분의 복원 능력이 부족하다는 점에 착안하여, 우리는 네트워크가 고주파 성분에 관한 많은 정보를 필요로 한다는 것을 가정하였다. 따라서 우리는 네트워크의 입력을 고주파 성분이 여러 가지 강도로 입력되도록 저해상도 입력 영상들을 구성하였다. 또한 잔차신호 네트워크(residual networks)를 도입하여, 네트워크 파라미터를 학습할 때 고주파 성분의 복원에 집중할 수 있도록 하였다. 본 논문의 효율성을 검증하기 위하여 set5 데이터와 set14 데이터에 관하여 실험을 진행하였고, SRCNN 과 비교하여 set5 데이터에서는 2, 3, 4 배에 관하여 각각 평균 0.29, 0.35, 0.17dB 의 PSNR 성능 향상이 있었으며, set14 데이터에서는 3 배의 관하여 평균 0.20dB 의 PSNR 성능 향상이 있었다.

1. 서론

초해상화는 저해상도 이미지를 고해상도 이미지로 변환하는 기술을 의미한다. 초해상화는 유일한 해가 존재하지 않는 역문제(ill-posed inverse problem)로 알려져 있다 [1], [7]. 다시 말해, 하나의 저해상도 이미지에는 여러가지 고해상도 이미지가 존재한다는 것이다. 여러가지의 고해상도 이미지 중 최적의 고해상도 이미지를 찾는 것이 초해상화 기술의 본질이다. 최근의 많은 초해상화 기법들은 예시를 통하여 저해상도 이미지와 고해상도 이미지 간의 관계를 학습하고, 이 관계에 기반하여 저해상도 이미지를 고해상도 이미지로 변환한다 [1], [7].

딥 러닝 기술 역시 훈련 영상을 이용하여 입력과 출력의 관계를 학습하는 방법으로 최근 여러 영상처리 및 컴퓨터 비전 문제에 적용되어 괄목할만한 성능을 보여주었다. 초해상화 문제에서는 Dong 이 Convolution neural networks (CNN)을 이용한 Super-Resolution Convolutional Neural Networks (SRCNN)을 제안하였다 [1]. SRCNN 은 3 개의 Convolutional layer 로 구성되어 있다.

딥 러닝에서, 층을 깊게 할수록 더 좋은 성능을 얻을 수 있다 [2], [3]. 하지만 층을 깊게 할 경우, gradient 가 아래 층으로

갈수록 발산하거나 0 으로 수렴하여, 학습이 제대로 되지 않는 현상이 발생한다. 따라서 다양한 고주파 성분을 내포하고 있는 입력을 다중 채널로 구성하여 네트워크를 통한 입력과 출력 관계를 학습함으로써 보다 좋은 성능을 얻을 수 있도록 하였다.

대부분의 초해상화 기법들은 전처리과정에 bicubic 보간법을 이용하여 저해상도 이미지를 고해상도 이미지와 같은 해상도로 만드는데, 이 과정에서 저주파성분은 잘 보존되나 고주파 성분을 잃어버리게 된다 [1], [7]. 이 때문에, 대부분의 초해상화 기법들은 영상의 저주파영역은 잘 복원하지만 고주파영역은 잘 복원해내지 못하게 된다.

본 논문에서는 다중 채널 입력을 고주파 영역이 여러 강도로 강화된 저해상도 이미지로 구성하여, SRCNN 을 통해 초해상화 과정을 학습하는 다중입력 SRCNN 을 제안한다. 다중입력 SRCNN 에서는 네트워크가 보다 풍부한 고주파 영역 관한 정보를 학습과정에서 이용할 수 있도록 하였다. 또한 다중입력 SRCNN 이 다양한 보간법의 장점을 활용 할 수 있도록, 전처리 과정에서 다양한 보간법으로 해상도를 증대 시켰다[8]. 이를 통하여 본 논문에서 제안하는 다중입력 SRCNN 이 고주파영역까지 잘 표현할 수 있는 특징을 추출할 수 있고, 이는 향상된 고주파 영역 복원을 가능하게 한다.

또한 저해상도 이미지가 고해상도 이미지의 저주파영역에

관한 성분을 가지고 있다고 가정하여, 다중입력 잔차신호 SRCNN 을 이용하여 고해상도 영상의 고주파 성분을 학습하는 데에 초점을 맞추도록 하였다.

2. 제안하는 방법

본 논문에서는 SRCNN 을 기반으로 한 초해상화 방법을 제안하였다. 그림 1 은 제안된 다중 입력 잔차신호 SRCNN 방법의 구조를 보여준다. SRCNN 의 입력이 단 채널인 반면, 제안된 방법은 입력이 다중 채널이다. 우리는 18 개의 채널을 입력으로 두었고, 이들은 다양하게 고주파영역이 강화되고 다양한 방법으로 보간된 저해상도 이미지들로 구성되어 있다. 첫번째 층은 3D-convolution 을 통하여 패치의 m_1 개의 특징 값들을 추출한다. 두번째 층은 m_1 개의 특징 값들을 고해상도 이미지를 위한 m_2 개의 특징 값들로 비선형 매핑을 수행한다. 세번째 층은 m_2 개의 특징 값들로 고해상도 이미지를 복원한다.

또한 잔차신호 네트워크(residual networks)를 도입하여 네트워크가 고주파 성분을 집중적으로 학습하도록 하였다. 고해상도 이미지와 저해상도 이미지 차이는 고주파 성분을 의미하므로, 네트워크의 출력이 고해상도 영상과 저해상도 영상의 차이가 되게 하였고, 네트워크 출력이 저해상도 영상을 더하여 고해상도 이미지를 추론하였다.

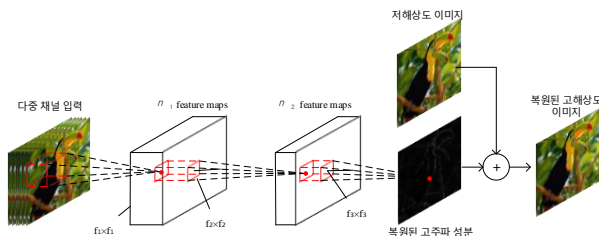


그림 1. 제안하는 다중입력 잔차신호 SRCNN 의 구조

SRCNN 과 PSNR 성능의 공정한 비교를 위하여 SRCNN 과 마찬가지로, $f_1 = 9, f_2 = 1, f_3 = 5$ 로 두었고 $m_1 = 64, m_2 = 32$ 로 두었다. 또한 SRCNN 과 마찬가지로, activation 함수로 Rectified linear unit (ReLU)을 사용하였다 [4].

다중 채널 입력을 구성하기 위하여, 전처리과정에서 저해상도 영상의 고주파 영역을 여러 강도로 강화시켰다. 그후, 여러 가지 강도로 고주파 영역이 강화된 저해상도 영상들을 다양한 보간법을 이용하여 고해상도 영상의 해상도와 같게 만들고 이들로 다중 채널 입력을 구성하였다. 고주파 영역을 강화하는 것은 MATLAB 의 “imsharpen” 함수를 이용하였고, “Amount” 파라미터 값을 조절하여 강도를 조절하였다 [6]. “Amount” 값을 0 부터 2 까지 0.4 간격들의 값 즉 0, 0.4, 0.8.....2 들의 값을 사용하였다. 여기서 0 의 “Amount” 값은 고주파 영역의 강화가 전혀 이루어지지 않았음을 의미한다. 이를 통하여 제안된 다중 입력 잔차신호 SRCNN 은 고주파 성분의 보다 많은 정보를 활용할 수 있게 되었다. 또한 보간법에는 bicubic, bilinear, nearest neighbor 의 보간법을 사용하였다. Bilinear 는 저주파 영역을 잘 보존하는 경향이 있고, nearest neighbor 보간 법은 고주파 영역을 잘 보존하는 경향이 있다 [8]. Bicubic 보간법은 이 두 보간법의 중간적인 성질을 보인다. 전처리 과정에서 여러 방법의 보간법을 사용하여 networks 는 다양한 보간법의 장점들을 활용할 수 있게 되었다.

3. Networks 학습

실험을 위하여, 다른 초해상화 기법들의 실험과 마찬가지로, 고해상도 이미지를 bicubic 보간법으로 줄인 것을 저해상도 영상이라고 가정하였다[1]. 또한 다른 초해상화 기법들과 마찬가지로 Y 채널에 관하여 실험을 진행하였다. 뿐만 아니라, 제안된 다중 입력 잔차신호 SRCNN 의 파라미터가 평균자승오차를 최소화 시키는 관점에서 학습되도록 하였다. SRCNN 과 마찬가지로, 91 영상 데이터 세트를 학습 데이터로 사용하였다 [1]. 하지만, 다중 채널 입력에 따라 학습해야 할 네트워크 파라미터 수가 증가하였고, 이로 인해 필요한 학습 데이터의 양이 증가 되었다. 따라서 각각의 학습데이터를 회전시키면서 data augmentation 을 수행하였다. 공정한 성능 비교를 위하여 SRCNN 에도 같은 data augmentation 을 수행하였다. SRCNN 과 마찬가지로 첫 번째와 두 번째 층의 learning rate 를 10^{-4} 로 두고 마지막 층의 learning rate 를 10^{-5} 로 두었다. 또 weight decay parameter 를 0.0005, momentum parameter 를 0.9 로 두었다. Mini-batch gradient descent 역전파 방법으로 학습을 진행하였으며, “MatConvnet” toolbox 가 학습에 사용되었다 [5].

표 1. Set5 에 관한 제안된 방법과 SRCNN 의 성능 비교

Set5	Scales	Bicubic	SRCNN	제안하는 방법
Baby	2	37.07	38.30	38.37
bird	2	36.81	40.69	41.12
butterfly	2	27.43	32.29	32.93
head	2	34.86	35.61	35.66
woman	2	32.14	35.10	35.36
Average	2	33.66	36.40	36.69

Baby	3	33.91	34.98	35.04
bird	3	32.58	35.12	35.56
butterfly	3	24.04	27.68	28.30
head	3	32.88	33.51	33.61
woman	3	28.56	30.97	31.48
Averages	3	30.39	32.45	32.80

Baby	4	31.78	33.03	32.95
bird	4	30.18	32.12	32.46
butterfly	4	22.10	25.08	25.40
head	4	31.59	32.18	32.24
woman	4	26.46	28.54	28.77
Average	4	28.42	30.19	30.36

표 2. Set14 에 관한 제안된 방법과 SRCNN 의 성능 비교

Set14	Scale	Bicubic	SRCNN	MC-SRCNN
Baboon	3	23.21	23.61	23.64
barbara	3	26.25	26.55	26.55
bridge	3	24.40	25.14	25.23
coastguard	3	26.55	27.21	27.26
comic	3	23.12	24.27	24.42
face	3	32.82	33.52	33.62
flowers	3	27.23	28.86	29.14
foreman	3	31.18	34.47	34.86
lenna	3	31.68	33.34	33.52
man	3	27.01	28.16	28.29
monarch	3	29.43	32.37	32.87

pepper	3	32.39	34.37	34.66
ppt3	3	23.71	25.99	26.37
zebra	3	26.63	28.93	29.07
Averages	3	27.54	29.05	29.25

4. 실험 결과

제안된 방법의 효과를 알아보기 위하여, set5 와 Set14 에 관하여 실험을 진행하였다[1]. 객관적인 성능의 지표로 PSNR 을 도입하였다. 표 1 과 표 2 는 Bicubic, SRCNN 과 제안된 다중입력 잔차신호 SRCNN 에 대해 각각 set5 에서와 set14 에서의 PSNR 성능 비교이다. 두개의 표에서 확인 할 수 있듯이, 제안된 방법은 SRCNN 보다 set5 데이터에서는 2, 3, 4 배에 관하여 각각 평균 0.29, 0.35, 0.17dB 의 PSNR 성능 향상이 있었으며, set14 데이터에서는 3 배의 관하여 평균 0.20dB 의 PSNR 성능 향상이 있다. 이는 제안된 방법이 residual learning 을 통하여 networks 가 고주파 영역을 학습하는데 집중할 수 있고, 다양한 보간법과 고주파 성분 강화로 인하여 보다 풍부한 정보를 가지고 있어, 이를 토대로 고해상도 이미지를 잘 복원할 수 있는 특징 추출을 통해 성능 향상이 이루어졌다고 판단된다.

5. 결론

본 논문에서는 SRCNN 기반의 다중 채널 입력을 이용하는 초해상화 기법을 제안하였다. 대부분의 초해상화 기법들이 고주파영역을 잘 복원하지 못한다는 사실을 바탕으로, 본 논문에서는 네트워크가 고주파 영역에 관한 정보가 더 필요하다는 가정을 하였다. 이러한 가정을 바탕으로, 다중 채널 입력을 고주파 영역이 강화된 이미지들로 구성하였고, 이는 네트워크가 고주파 영역을 보다 더 잘 복원할 수 있는 특징 값을 추출하도록 하였다. 또한 저해상도 영상이 고해상도 영상의 저주파 성분을 포함하고 있다는 가정을 하에, 잔차신호 네트워크를 이용하여 네트워크가 학습과정에서 고주파영역의 복원에 집중하도록 하였다. 그 결과, 제안된 방법은 SRCNN 보다 우수한 성능을 보여주었다.

감사의 글

본 논문 연구는 연구재단 중견연구자사업 핵심연구(개인) 과제(과제번호: 2014R1A2A2A01006642)로 수행되었음.

참고문헌

- [1] C. Dong, C. C. Loy, K. He and X. Tang, "Learning a deep convolutional network for image super-resolution," *ECCV 2014, Part IV, LNCS 8692*, Springer, pp. 184-199, 2014.
- [2] A. Krizhevsky, S. Ilya S, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, pp. 1097-1105, 2012.
- [3] Y. Sun, X. Wang, and X. Tang, "Deep convolutional network cascade for facial point detection," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 3476-3483, 2013.

- [4] V. Nair, and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," *Int. Conf. Machine Learning*, pp.807-814, 2010.
- [5] A. Vedaldi and L. Karel. "MatConvNet: Convolutional neural networks for matlab," *Proceedings of the 23rd Annual ACM Conference on Multimedia Conference*, pp. 689-692, 2015.
- [6] R.C Gonzales, R.E Woods and S.L Eddins, *Digital image processing using MATLAB*, Pearson Prentice Hall, 2004.
- [7] J. S. Choi and M. C. Kim, "Super-Interpolation with Edge-Orientation-Based Mapping Kernels for Low Complex 2x Up-Scaling," *IEEE Transactions on Image Processing*, vol. 25, issue 1, pp. 469-483, Jan. 2016.
- [8] D. Han, "Comparison of commonly used image interpolation methods," *ICCSEE, Hangzhou, China*, pp. 1556-1559, 2013.