

지각적 손실 함수 및 차이 영상을 이용한 CNN 기반의 고해상도 복원

정준영, 박인규

인하대학교 정보통신공학과

{cheongjunyoung@gmail.com, pik@inha.ac.kr}

요약

본 논문은 CNN(convolutional neural network)을 이용하여 고해상도 영상을 복원하는 효과적인 학습방법을 제안한다. 제안하는 알고리즘은 RGB 영상이 아닌 저해상도 영상의 휘도 영상(luminance)과 고해상도 영상의 휘도 영상의 차이 영상(residual)을 학습하여 네트워크 모델의 수용성을 향상시키고 화소 기반의 손실함수 대신 미리 학습된 VGG-16 네트워크를 이용한 지각적 손실함수(perceptual loss function) 및 차이 영상(residual image)를 기반으로 고해상도 영상 복원을 위해 제안하는 CNN 를 학습한다.

1. 서론

고해상도 복원 문제(super-resolution)는 노이즈 제거(de-noising), 안개 및 비 제거(de-hazing, de-raining), 반사 영상 제거(reflection removal) 등의 저하된 영상으로부터 선명한 영상을 복원하는 ill-posed 문제 중 하나로 오랜 기간동안 영상처리 분야의 중요 문제로 자기 유사성(self-similarity), 디셔너리(dictionary) 등의 다양한 학습기반의 기법들이 연구되어 왔다.

특히 최근 각광받고 있는 CNN 기반의 학습기법은 컴퓨터비전분야의 고 수준 문제인 영상분류뿐만 아니라 ill-posed 로 모델링 되는 다양한 저 수준 문제에 대해서도 뛰어난 성능을 보이고 있다.

SRCNN[1]은 보간 기법을 통해 향상된 저해상도 영상과 고해상도 영상 간의 관계를 CNN 를 학습하여 고해상도 복원 문제에 대한 효율성을 보였다. VDSR[2]은 SRCNN[1] 보다 깊은 CNN 모델을 이용하였으며 특히 RGB 3 채널 영상이 아닌 고해상도 단일 휘도 영상과 저해상도 단일 휘도 영상의 차이 영상을 학습하여 CNN 의 수용성 향상을 통해 높은 고해상도 복원성능을 보였다. 나아가 Johnson 등 [3]은 저해상도 영상과 고해상도 영상 간의 화소 값 기반의 content 관계를 학습하는 대신 미리 학습된 CNN 모델의 특징을 기반으로 한 지각적 손실함수를 제안하여 인지적인 관점에서 고해상도 복원 성능의 향상을 이뤘다.

본 논문에서 제안하는 기법은 이러한 기존연구들의 이점을 기반으로 한 CNN 학습 기법으로 차이 영상 학습을 통한 CNN 모델의 수용성 향상과 고 수준 영상분류 문제를 위해 미리 학습된 VGG 네트워크의 특징 기반의 지각적 손실을 이용한 학습을 통해 인지적 관점에서 향상된 고해상도 영상 복원 문제를 해결하고자 한다.

2. 제안하는 기법

제안하는 CNN 을 이용한 고해상도 복원 기법은 기존 연구된 기법들의 효율성을 기반으로 네트워크 모델 설계 및 학습을 수행한다.

2.1 차이 영상(residual image)

제안하는 CNN 는 VDSR [2]와 동일한 구조로 그림 1 과 같이 3x3 convolutional filter 로 이뤄진 convolution layer 와 non-linear activation 함수인 relu layer 의 반복적인 구조의 형태로 구현하였다. 이와 같이 작은 수용영역(receptive field)인 3x3 convolutional filter 의 convolution layer 와 non-linear activation layer 의 연속된 구조는 넓은 수용영역을 적은 수의 변수(parameter)로 구현 할 수 있다는 이점이 있다.

학습 데이터로는 RGB 3 채널 영상을 입력 영상으로 전부 이용하는 대신 bi-cubic 보간 기법으로 향상된 저해상도 영상의 휘도 영상을 입력영상으로, 입력 영상과 고해상도 영상의 휘도 영상의 차이 영상을 출력 영상으로 학습하여 CNN 의 수용성을 향상시켰다.

2.2 지각적 손실함수(Perceptual loss function)

기존 많은 연구들[1,2]은 CNN 학습을 위해 손실함수로 저 수준의 손실 함수인 화소 기반의 Euclidean distance 를 이용하였다. 그러나 이러한 손실함수는 인지적인 관점에서의 손실을 표현하지 못해 낮은 수준의 손실을 갖는 결과 영상을 도출하여도 인지적 관점에서는 낮은 수준의 영상으로 판단 될 수

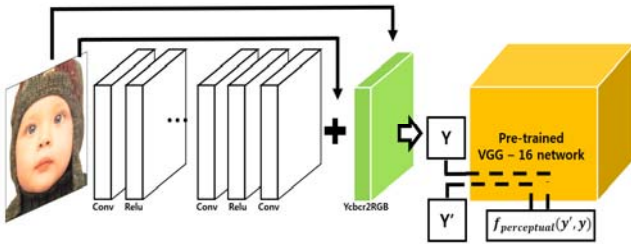


그림 1. 제안하는 CNN 학습 기법.

있다. 따라서 Johnson 등 [3]과 같이 미리 학습된 고수준의 특징을 검출하는 VGG-네트워크의 특징을 기반으로 한 다음과 같은 손실함수를 이용하여 학습을 수행한다.

$$f_{\text{perceptual}}^{\phi, j}(y', y) = \frac{1}{C_j H_j W_j} \|\phi_j(y') - \phi_j(y)\|^2 \quad (1)$$

$\phi_j(y')$, $\phi_j(y)$ 는 각각 원본영상과 제안하는 CNN에 대한 VGG 네트워크의 j 번째 layer의 출력 영상을 의미한다. 또한 smoothness 를 고려하기 위해 Charbonnier penalty 함수를 추가하였다.

$$f_{\text{smoothness}}(y) = \frac{1}{CHW} \sqrt{(\nabla y)^2 + \varepsilon} \quad (2)$$

2.3 CNN 학습방법

학습 데이터는 VDSR[2] 학습 데이터와 동일하게 이용하였으며 네트워크의 수용영역의 크기와 동일하게 41x41 크기의 작은 영상으로 나눠 학습을 수행하였다. CNN 학습 시, 높은 learning rate 을 이용하기 위해 gradient clip 기법을 적용하여 exploding 또는 vanishing gradient 문제를 해결하였으며 지각적 손실함수는 미리 학습된 VGG 네트워크의 relu_3 layer의 결과를 적용하였다.

3. 실험결과

본 논문에서 이용한 Test 영상은 Set5[4], Set14[5] Dataset 이며 영상의 가로, 세로 해상도를 1/4 로 저하시킨 후 다시 복원하는 실험을 수행하였다. 제안하는 CNN 기법의 성능을 확인하기 위하여 bi-cubic, VDSR[2] 기법의 성능을 비교 측정하였다.

실험결과, 각 화소 값에 대한 저 수준의 손실함수를 학습한 VDSR[2]의 경우 저 수준의 측정단위인 PSNR 및 SSIM (표 1)의 결과가 제안하는 기법에 비해 높게 측정되었음에도 불구하고 그림 2 에서 보는 바와 같이 인지적 관점에서 제안하는 기법에 비해 낮은 가시성을 보였다.

x4	Bi-cubic (PSNR / SSIM)	VDSR (PSNR / SSIM)	Proposed (PSNR / SSIM)
Set5[4]	28.42 / 0.81	31.35 / 0.88	29.19 / 0.83
Set14[5]	26.01 / 0.70	28.02 / 0.77	26.70 / 0.72

표 1. Test 영상에 대한 평균 PSNR 및 SSIM.

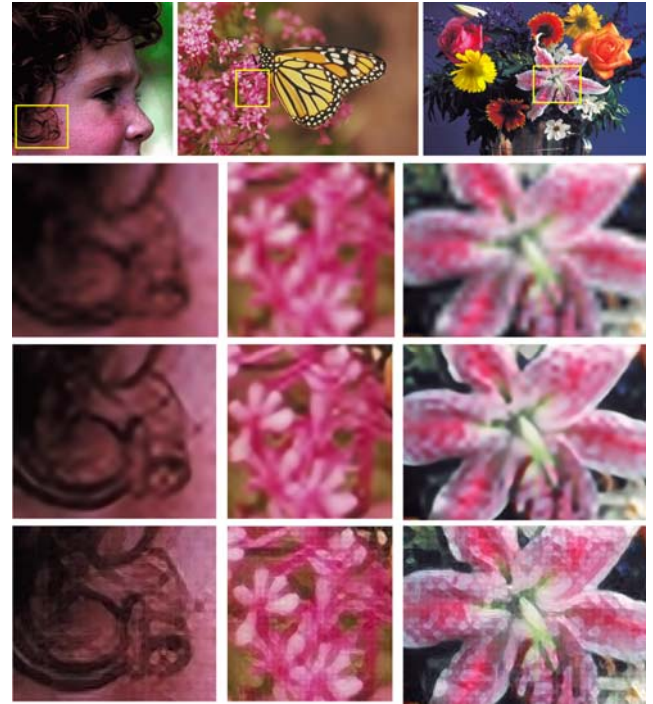


그림 2. Bi-cubic, VDSR[2], 제안하는 기법 비교결과 (1 행: 원본영상. 2 행: Bi-cubic. 3 행: VDSR[2]. 4 행: 제안하는 기법).

4. 결론

본 논문은 지각적 손실함수 및 차이 영상을 이용한 CNN 기반의 고해상도 복원 기법을 제안한다. 본 기법을 통해 CNN의 수용성을 향상, 인지적 관점에서 고해상도 영상을 복원 할 수 있으며 실험 결과를 통해 본 기법의 효용을 보였다.

5. 감사의 글

이 논문은 2016 년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2016R1A2B4014731).

6. 참고문헌

- [1] C. Dong *et al.*, "Image super-resolution using deep convolutional network," IEEE PAMI, vol. 38, no. 2, pp. 295-307, February 2016.
- [2] J. Kim *et al.*, "Accurate image super-resolution using very deep convolutional networks," Proc. of IEEE CVPR, June 2016.
- [3] J. Johnson *et al.*, "Perceptual loss for real-time style transfer and super-resolution," Proc. of ECCV, October 2016.
- [4] M. Bevilacqua *et al.*, "Low-complexity single-image super-resolution based on nonnegative neighbor embedding," Proc. of BMVC, September, 2012.
- [5] R. Zeyde *et al.*, "On single image scale-up using sparse-representations," Lecture Notes in Computer Science, vol. 6920, pp. 711-730, June 2012.