

딥러닝 기반 문화재 영상에 대한 4 배 및 8 배 초해상화

*손채연, *김수예, **김주영, *김문철

*한국과학기술원 전기 및 전자 공학부, **한국전자통신연구원
thscodus97@kaist.ac.kr, sooyekim@kaist.ac.kr, kimjy1113@etri.re.kr,
mkimee@kaist.ac.kr

Deep Learning based x4 and x8 Super-Resolution for Cultural Property Images

*Chaeyeon Son, *Soo Ye Kim, **Juyoung Kim, *Munchurl Kim

*Korea Advanced Institute of Science and Technology Dep. of Electrical Engineering,
**Electronics and Telecommunications Research Institute

요 약

문화재 영상 데이터는 방대한 양으로 인해 고해상도로 모두 저장하기 어렵거나 시간이 지나 상대적으로 화질이 낮은 영상들이 다수 존재하기에 초해상화가 필요한 상황이 많다. 따라서 본 논문에서 처음으로 문화재 영상에 특화된 4 배 및 8 배 딥러닝 기반 초해상화 방식을 제안한다. 문화재 영상 데이터는 배경이 단조롭고 물체가 영상 중간에 위치한다는 특징이 있어 이를 고려해 중간 부분에서만 패치를 추출하는 방식을 적용하여 의미 있는 패치로 학습이 되도록 한다. 또 자연 영상 데이터 셋인 DIV2K 를 사용해 학습하는 방식과 직접 구성된 문화재 데이터 셋을 이용해 학습하는 방식, 그 둘을 적절히 함께 사용하여 학습하는 전이 학습 방법까지 세 가지로 학습하여 초해상화의 성능을 향상시키는 방법을 제안한다. 그 결과, 쌍삼차 보간법(Bicubic interpolation)보다 4 배 초해상화에서는 약 1.25dB, 8 배 초해상화에서는 약 1.26dB 의 성능 개선을 확인하였으며, 단순 DIV2K 로 학습한 방식보다는 4 배에서는 0.06dB, 8 배에서는 0.17dB 의 성능 개선을 확인하였다.

1. 서론

초해상화는 저해상도 영상을 고해상도의 영상으로 변화시키는 영상 처리 기술이다. 손실되었거나 전송, 보관 등의 이유로 저해상도로 저장된 이미지들을 선명한 고화질의 영상으로 볼 수 있게 하여 다양한 분야에서 활용될 수 있다. 예전의 쌍선형 보간법(Bilinear interpolation), 쌍삼차 보간법(Bicubic interpolation) 등의 간단한 보간법에서 시작해서, 머신 러닝을 이용하는 형태로 발전해왔다.

딥러닝 이전에 SelfExSR [1], SI [2] 등의 비딥러닝 머신러닝 방식이 대표적이었다. SI 는 적은 양의 데이터로 저해상도에서

고해상도 영상으로 선형 매핑을 학습하게 하는 방식인데, 초해상화 연구 분야에서는 많은 양의 데이터를 쉽게 확보할 수 있기 때문에 많은 양의 데이터를 활용하는 딥러닝 방식보다 성능이 떨어질 수 밖에 없다. 딥러닝은 합성곱 신경망으로 비선형 매핑을 통해 다양한 특징을 학습할 수 있는 반면 비딥러닝 방식은 선형 매핑을 배우기 때문에 한계가 존재한다. 또한 패치 사이즈 등의 하이퍼 파라미터가 약간만 변하더라도 성능이 크게 변화한다는 점에서 분명한 한계가 있다.

딥러닝을 이용하는 초해상화 방식은 SRCNN [3]을 시발점으로 VDSR [4], ESPCN [5], SRResnet [6], EDSR [7] 등 많은 기술들이 소개되어왔다. 초해상화에 적합한 형태로 모델이

빠르게 발전하며 꾸준한 성능의 향상을 보여왔는데, 특히 RCAN [8]은 크게 복잡하지 않은 구조를 가지면서도 벤치마크 테스트에서 꾸준히 높은 성능을 보이기에 자연 영상에서도 우수한 모델이 문화재 영상에서도 우수한 결과를 보여줄 것이라 기대해 본 연구는 해당 논문의 모델 구조를 수정해서 연구를 진행하였다.

기존의 초해상화 연구는 다양하게 이루어졌으나 대부분 특정되지 않은 일반적인 자연 영상에서만 이루어졌으며, 이제까지 문화재 영상 데이터에서는 검증되지 않았다. 문화재 영상이 대체로 단색 배경에서 영상 중간 부분에 문화재가 위치한다는 특징을 갖는다는 점에서 자연 영상과는 분명한 다른 특성이 있는 만큼 그에 특화된 방식이 필요하다. 본 논문은 딥러닝을 이용해 문화재 영상을 4 배와 8 배로 초해상화하기 위한 방법을 제안한다. RCAN 에서 착안한 모델 구조를 사용하면서, 문화재 영상의 특성을 이용하는 패치(patch) 추출 방식과 문화재 영상 데이터 셋을 다양한 방법으로 활용하는 딥러닝 학습 방식을 제안한다. 본 연구는 문화재 영상을 목적으로 하는 최초의 초해상화 연구로써 앞으로 문화재 관련 연구에서 도움이 될 것이라 기대한다.

2. 문화재 영상 데이터 셋 활용

2.1 문화재 영상 데이터 셋 구성

문화재 영상의 분포가 특정 문화재에 치우쳐지지 않도록 목재, 금속, 가죽, 종이, 자기, 합성 소재 등의 재질 별로 다양하게 데이터를 e-뮤지엄에서 수집하여 총 215,200 장이 고르게 분포되도록 구성하고 쌍삼차 보간법(Bicubic interpolation)으로 문제 상황에 맞는 4 배, 8 배 축소시켜서 학습 데이터로 사용하였다.

또한 결과 테스트를 위하여 문화재의 다양한 재질과 유형을 대표하는 20 개의 영상을 뽑았다.

2.2 문화재 영상 데이터 전처리 단계에서 패치 추출 방식

Figure.1. 같은 문화재 영상들을 살펴보면 대체로 단색의 배경에 문화재가 중앙에 위치한 것을 확인할 수 있다. 이미지 전체를 학습한다면 상관 없겠지만 영상 데이터가 크기 때문에 학습을 위해서는 작은 사이즈로 패치(patch)를 잘라서 학습을 하게 된다. 하지만 문화재 영상의 이러한 특성으로 인해 패치를 뽑을 때 이미지 전체에서 무작위로 뽑을 경우 문화재가 포함되어 있지 않고 단색의 배경만 있는 등의 의미 없는 패치가 뽑혀 오히려 학습에 방해될 수 있다. 이를 방지하기 위해 본 논문에서는 문화재 데이터에 특화된 패치 추출 방식을 제안한다.

문화재 영상의 중간 부분인 가로와 세로 각각 1/4 지점부터



Figure. 1. 문화재 영상 예시

3/4 지점까지 중에서 패치를 추출할 수 있도록 하지만 가운데 부분만 학습 시키기에는 문화재가 중간 부분을 벗어나 위치하는 영상도 존재할 수 있고, 문화재와 배경 사이의 경계나 배경도 분명히 학습해야 한다. 따라서 중간 부분에서 패치 뽑는 확률을 70%, 전체에서 패치를 뽑는 확률을 30%로 하여 패치를 뽑고 배치(batch)를 얻어 학습시킨다.

3. 학습 방법

3.1. 모델 구조

Figure. 2 에서 볼 수 있듯 본 논문에서는 [8]에서 사용된 잔차 속의 잔차 네트워크(Residual in Residual Network) 구조를 참고하여 학습을 진행했다. 논문에서 보다는 축소된 형태인 2 개의 ResGroup 으로 네트워크를 구성했는데, Figure. 3 에서 볼 수 있듯 각 그룹은 10 개의 ResBlock 을 가지며, 각 ResBlock 은 [7]에서 참고한 형태이다. 입력 이미지를 ResGroup 을 지나서

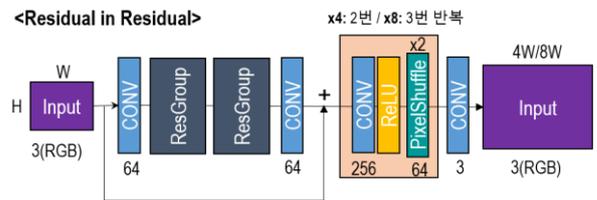


Figure. 2. 본 논문의 모델 구조(Residual in Residual Network)

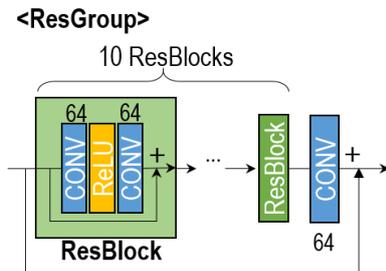


Figure. 3. 본 논문의 모델 구조에 사용된 ResGroup 내부 구조

나온 결과에 더한 후, 콘볼루션, ReLU 활성화 함수, 픽셀 셔플 [5]로 구성된 2 배 확대 모듈을 사용해 원하는 사이즈의 이미지로 키운다. 이 때 4 배 확대 네트워크의 경우 위 모듈을 2 번, 8 배 확대 네트워크의 경우 3 번 반복한다.

3.2. 문화재 데이터 활용 학습 방식

3.2.1. DIV2K 영상으로 학습

NTIRE2017 에서 주어진 DIV2K [9] 데이터 셋은 2K 화질의 영상들로 학습 용 800 장, 테스트 용 100 장으로 구성되어 있다. 어느 한 쪽에 특정되지 않은 일반적인 자연 영상으로 구성되어 있어 초해상화에 필요한 일반적인 이미지의 특성을 쉽게 배울 수 있다.

3.2.2. 문화재 영상으로 학습

앞서 2.1 에서 설명한 문화재 데이터 셋만 가지고 학습을 하며, 2.2 에서의 패치 추출방식을 이용해 학습을 한다. 일반적인 자연 영상보다는 문화재 영상의 특징만 익히도록 한다.

3.2.3. DIV2K 로 학습 후 문화재 영상으로 학습(전이 학습)

3.2.1 에서 학습된 웨이트를 전이 학습하여 문화재 영상에서 추가로 학습한다. 모델이 자연 영상의 특징을 익힌 상태로 2.1 의 문화재 영상 데이터 셋을 이용해 추가 학습을 하여, 자연 이미지에서 초해상화를 위한 유의미한 특징을 뽑아낼 수 있는 필터를 가진 상태로 문화재 영상 데이터로 학습한다.

4. 실험결과

4.1. 실험 환경 설정

모든 실험에서 L1 손실(loss)을 사용하고, 아담 최적화 기법(Adam optimizer)으로 최적화를 진행하고, 초기 학습률은 $1e-4$ 로 설정하였다. 배치 크기는 8 이며, 32×32 로 잘려진 패치를 사용한다.

한가지 데이터 셋으로만 학습하는 경우 65K 번의 반복(iteration)까지 학습하는 중에 가장 낮은 유효 손실(valid loss)을 갖는 웨이트로 테스트를 진행하였고, 전이 학습을 하는 경우에는 DIV2K 로 학습해서 나온 가장 우수한 웨이트로 문화재 영상 데이터에서 10K 번의 반복만큼 학습을 추가로 진행하여 역시 가장 낮은 유효 손실을 갖는 웨이트로 테스트를 진행하였다.

4.2. 실험 결과

문화재 테스트 영상 20 장에 대하여 쌍삼차 보간법과 각 실험에서의 PSNR 과 SSIM 수치를 구하여 성능을 비교하였다.

	4 배		사용 iteration
Metric	PSNR (dB)	SSIM	
Bicubic	32.19	0.8566	
Ours (DIV2K)	33.38	0.8722	64K
Ours(문화재)	33.29	0.8709	56K
Ours(전이)	33.44	0.8729	64K+800

Table. 1. 초해상화 방법에 따른 성능 비교(4 배)

	8 배		사용 iteration
Metric	PSNR (dB)	SSIM	
Bicubic	29.13	0.7959	
Ours (DIV2K)	30.22	0.8112	65K
Ours(문화재)	30.19	0.8094	51K
Ours(전이)	30.39	0.8117	65K+3100

Table. 2. 초해상화 방법에 따른 성능 비교(8 배)

문화재 데이터만 사용한 경우보다는 다양한 영상으로 구성된 DIV2K 로 학습한 경우가 성능이 더 좋았으며, 전이 학습을 통해 자연 영상으로 학습 후 문화재 영상으로 추가 학습으로 미세 조정된 실험이 4 배, 8 배 모두에서 가장 좋은 결과를 보였다.





Figure. 4. 위부터 아래로 Bicubic, Ours(DIV2K), Ours(전이), Ours(문화재) 모델의 4 배 초해상화한 결과 영상



Figure. 5. 왼쪽 위부터 시계방향으로 Bicubic, Ours(DIV2K), Ours(전이), Ours(문화재) 모델의 8 배 초해상화한 결과 영상

딥러닝 학습 모델의 결과 영상은 모두 쌍삼차 보간법의 결과보다 수치적으로도 육안으로도 좋은 결과를 보인다. 특히 결과 영상에 대한 평균 PSNR 수치는 4 배 초해상화에서 최소 1.1 dB, 8 배 초해상화에서는 최소 1.09 dB 로 크게 개선된 성능을 볼 수 있다. 딥러닝 방식들 중에는 수치적으로는 전이학습을 통한 학습 모델의 결과 영상이 가장 우수하지만 육안으로는 미세한 차이만을 확인할 수 있다.

5. 결론

문화재 영상 데이터는 배경이 단조롭고 물체가 영상 중간에 위치한다는 특징이 있어 이를 고려해 특화하여 네트워크 학습을 할 필요가 있다. 이를 위해 문화재 데이터 셋을 새로 구성하고, 중앙 부분에서만 패치를 추출하는 방식을 적용하였으며, 자연 영상 데이터 셋인 DIV2K 과 문화재 데이터를 적절히 함께 사용하여 학습하는 방식을 제안하였다. 그 결과, 쌍삼차 보간법보다 4 배 초해상화에서는 약 1.25dB, 8 배 초해상화에서는 약 1.26dB 가 증가하고, 단순 DIV2K 학습 방식보다는 4 배에서는 0.06dB, 8 배에서는 0.17dB 가 증가한 성능을 보였다.

Acknowledgements

본 연구는 문화체육관광부 및 한국콘텐츠진흥원의 2020 년도 문화기술연구개발 지원사업으로 수행되었음. [R2020040045, 실감형 문화유산 체험을 위한 애셋 기반 지능형 큐레이션 및 서비스 운영 기술 개발]

참고문헌

- [1] J. B. Huang, A. Singh, and N. Ahuja, "Single image super-resolution from transformed self-exemplars", In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 5197-5206, 2015.
- [2] J. Choi and M. Kim, "Super-Interpolation With Edge-Orientation-Based Mapping Kernels for Low Complex 2× Upscaling." IEEE Transactions on Image Processing, vol. 25, no. 1, pp. 469-483, January 2016, doi: 10.1109/TIP.2015.2507402.
- [3] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution", In European Conference on Computer Vision (ECCV), pages 184-199. Springer, 2014.
- [4] J. Kim, J. K. Lee and K. M. Lee, "Accurate image super-resolution Using very deep convolutional networks," Proc. of IEEE Conference on Computer Vision and Pattern Recognition, pp. 1646-1654, June 2016.

- [5] W. Shi, J. Caballero, F. Huszar, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang, "Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network", In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 1874-1883, 2016.
- [6] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, et al., "Photo-realistic single image super-resolution using a generative adversarial network", arXiv preprint arXiv:1609.04802, 2016.
- [7] B. Lim, S. Son, H. Kim, S. Nah and K.M. Lee, "Enhanced Deep Residual Networks for Single Image Super-Resolution", In CVPRW, 2017.
- [8] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong and Y. Fu, "Image Super-Resolution Using Very Deep Residual Channel Attention Networks", In ECCV, 2018.
- [9] E. Agustsson and R. Timofte, "Ntire 2017 challenge on single image super-resolution: Dataset and study," Proc. of IEEE Conference on Computer Vision and Pattern Recognition Workshop, pp. 1122-1131, July 2017.