

채널 어텐션을 이용한 AHDR 모델의 성능 평가

윤석준 이근택 조남익

서울대학교 뉴미디어통신공동연구소

ysj624637@ispl.snu.ac.kr

Performance Evaluation of AHDR Model using Channel Attention

Youn, Seok Jun Lee, Keuntek Cho, Nam Ik

Seoul National University Institute of New Media and Communication

요약

본 논문에서는 기존 AHDRNet에 channel attention 기법을 적용했을 때 성능에 어떠한 변화가 있는지를 평가하였다. 기존 모델의 병합 망에 존재하는 DRDB(Dilated Residual Dense Block) 사이, 그리고 DRDB 내의 확장된 합성곱 레이어(dilated convolutional layer) 뒤에 또다른 합성곱 레이어를 추가하는 방식으로 channel attention 기법을 적용하였다. 데이터셋은 Kalantari의 데이터셋을 사용하였으며, PSNR(Peak Signal-to-Noise Ratio)로 비교해본 결과 기존의 AHDRNet의 PSNR은 42.1656이며, 제안된 모델의 PSNR은 42.8135로 더 높아진 것을 확인하였다.

1. 서론

영상처리에서 해상도를 높이기 위해 많은 연구가 진행되었으며, 특히, 심층 신경망(DNN, Deep Neural Network)이 발전함에 따라 영상처리에 있어서 많은 성과가 있었다. AHDRNet(Attention-guided Network for Ghost-Free High Dynamic Range Imaging)은 차세대 고품질 이미지 기술로 주목을 받고 있는 HDRI에 spatial attention을 적용시켜 더욱 높은 성능을 보여주었다[1].

한편, spatial attention 뿐 아니라 channel attention 기술의 적용 역시 다양한 분야에서 성능향상을 보여주었다. RCAN(Residual Channel Attention Networks)은 이미지 슈퍼 해상도(Image Super-Resolution) 분야에 channel attention을 적용시켜 더 좋은 성능을 보여주었다.[2].

본 논문에서는 AHDRNet에 channel attention 기법을 적용한 새로운 방식의 성능을 실험적으로 평가하였다. 기존 모델의 병합 망에 존재하는 DRDB(Dilated Residual Dense Block) 사이, 그리고 DRDB 내의 확장된 합성곱 레이어 뒤에 또다른 합성곱 레이어를 추가하는 방식으로 channel attention 기법을 적용하였다. 데이터셋은 Kalantari의 데이터셋을 사용하였으며, HDRI 성능을 다양한 톤 매핑 기법을 기반으로 한 PSNR(Peak Signal-to-Noise Ratio)로 비교해본 결과 기존의 모델과 어떠한 성능의 차이가 있는지를 세 종류의 PSNR을 사용하여 평가하였다[3].

2. AHDR과 Channel Attention 기법

인간의 시각 기관과 카메라의 visual system의 가장 큰 차이점은 인지할 수 있는 휘도의 최대 범위가 다르다는 것이다. 이를 동적 범위(DR, dynamic range)라 하며, 이미지에서 가장 밝은 부분과 가장 어두운 부분의 비를 나타낸다. 카

메라의 visual system은 상대적으로 인간의 시각 기관에 비해 dynamic range가 좁으며, 이 차이로 인해 우리가 카메라로 찍힌 사진을 보았을 때 어두운 그림자 부분이 뭉개지거나 밝은 영역이 지나치게 하얗게 날아가는 것을 볼 수 있다. 이렇듯 카메라가 갖고 있는 저동적 범위를 고동적 범위로 끌어 올려주는 기술을 HDRI(High Dynamic Range Imaging)이라 한다.

HDRI 분야가 연구됨에 따라 사람들은 더 나은 퍼포먼스를 얻을 수 있는 방법을 모색하기 시작했다. 그 중 눈여겨 볼 만한 기술로 어텐션 모듈을 도입한 합성곱 블록(CBAM, Convolutional Block Attention Module)을 들 수 있다[4]. CBAM은 BAM(Bottleneck Attention Module)의 후속 연구로, 기존의 BAM이 channel 및 spatial attention을 하나의 3D 합성곱 피쳐로 더해서 구현을 하는 형태에 변형을 주어서 더 향상된 성능을 보여주었다. 이 BAM과 CBAM에서 주요한 개념으로 앞서 언급된 channel attention과 spatial attention을 뽑을 수 있다[5].

Channel attention은 피쳐(feature) 간의 채널 관계(inter-channel relationship)를 통해 생성되는데, 각 피쳐 맵에서는 어떠한 피쳐가 검출되었는지에 가장 중점을 둔다. 기본적으로 효율적인 계산을 위해 spatial dimension을 압축했으며, GAP(Global Average Pooling)만을 사용하던 기존 BAM의 방식과 달리 GMP(Global Max Pooling) 또한 추가로 사용하여 병렬로 구성이 가능케 하였다. Channel attention map은 식 (1)처럼 표현할 수 있으며, 이 때

$$M_c(F) = \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))) \quad (1) \\ = \sigma(W_1(W_0(F_{avg}^c)) + W_1(W_0(F_{max}^c)))$$

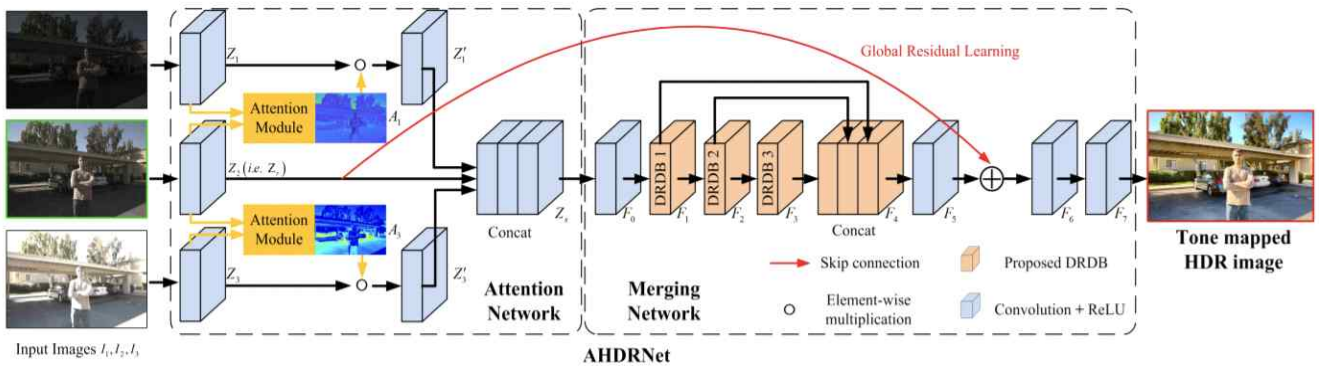


그림 1 AHDRNet의 전체적인 구조. 피쳐의 추출을 위한 어텐션 망(attention network)과 고동적 범위 이미지의 예측을 위한 병합 망(merging network)으로 이루어져 있다[1].

$M_c(F)$ 는 channel attention map, MLP 는 multi-layer perceptron, W_0, W_1 은 MLP 의 weight, σ 는 sigmoid function, 그리고 F_{avg}^c 와 F_{max}^c 는 average pooling과 max pooling으로 생성된 피쳐이다.

Spatial attention은 피쳐 간의 공간 관계(inter-spatial relationship)를 통해 생성되며, average pooling과 max pooling을 채널 축으로 적용시켜 효과적으로 공간 정보를 강조한다.

CBAM에서는 이러한 각 attention module의 특징 외에도 배치와 순서에도 신경을 썼다. 모듈을 병렬로 구성하는 것보다 직렬(sequential)로 구성하는 것이, 그리고 spatial attention보다 channel attention이 앞선 순서로 배치되어 있는 것을 최종 구조로 채택하였다. Attention module을 여러 합성곱 신경망 모델들에 적용했을 때 더 좋은 퍼포먼스를 보여주는 것을 확인할 수 있었다.

이러한 attention module을 HDR에 적용한 모델로 AHDRNet을 찾아볼 수 있다[1]. HDR에서 쉽게 찾아볼 수 있는 고질적인 문제인 유령 효과(ghosting effect)를 해결하기 위해 AHDRNet은 attention module과 DRDB(Dilated Residual Dense Block)을 사용했으며, 전체 구조는 위의 그림 1과 같다.

AHDRNet은 크게 어텐션 망(attention network)와 병합 망(network)으로 구성되어 있다. 어텐션 망에서 서로 다른 노출량(EV, Exposure Value)과 휘도를 가진 세 장의 이미지가 입력으로 들어오면 spatial attention module을 먼저 거치게 된다. 중간 휘도를 가진 기준 이미지(reference image)를 기반으로 새로운 피쳐 맵들을 뽑아내면 이를 하나의 묶음으로 만든 후 병합 망으로 내보낸다.

병합 망의 구조는 그림 2와 같으며, 기본적으로 [6]에 사용된 residual dense network를 사용한다. 폐색(occlusion)으로 인해 저동적 범위 이미지에서 충분히 얻지 못한 정보를 끌어내기 위해, RDB(Residual Dense Block)에 추가로 확장된 합성곱 레이어(2-dilated convolutional layer)를 적용시켰다[7]. 이러한 구조를 DRDB(Dilated

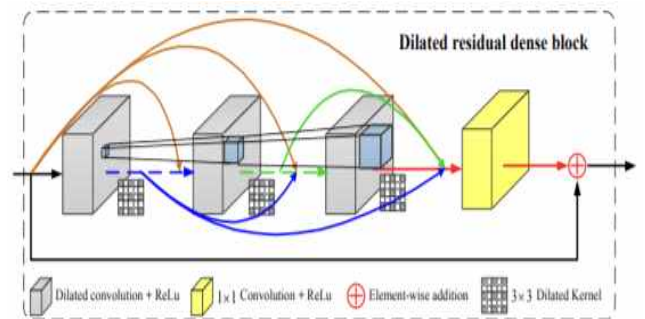


그림 2 DRDB의 전체적인 구조[1].

Residual Dense Block)이라 하는데, DRDB는 6개의 합성곱 레이어로 이루어져 있는데, [8]의 dense block과 달리 한 block의 입력과 출력 사이에 local residual skip-connection을 적용시켰다. 또한 DRDB를 사용한 결과 각 블록에서의 수용 영역(receptive field)을 효율적으로 넓힐 수 있었다. 병합 망으로 들어온 피쳐 맵 묶음은 폐색(occlusion)으로 인해 저동적 범위 이미지에서 충분히 얻지 못한 정보를 이 과정에서 더 끌어낼 수 있다. 또한 AHDRNet은 추가로 global residual learning을 적용시켜 결과로 나올 이미지가 기준 이미지의 특징을 잘 살릴 수 있도록 해주었다.

3. 실험

앞서 언급된 AHDRNet이 spatial attention에 집중하고 있다는 점에 착안하여, 본 논문에서 attention 모델의 성능 개선을 위해 2단계의 변화를 적용하였다. 먼저 병합 망의 DRDB 레이어들 사이에 합성곱 블록(convolutional block)을 추가하였고, 그 후 DRDB 구조 내에 존재하는 확장된 합성곱 레이어들 사이 사이에 추가적인 합성곱 블록을 연결하여 channel attention을 적용하고 성능 평가를 하였다.

실험에 사용한 Kalantari의 데이터셋은 74개의 트레이닝 데이터셋과 15개의 테스트 데이터셋으로 구성되어 있으며, 각 set은 노출량이 서로 다른 세 장의 저동적 범위 이미지를

포함하고 있다[3].

HDR을 통해 얻어낸 고동적 범위의 이미지는 최종적으로 출력될 때 톤매핑된(tone-mapped) 이미지로 나오게 되는데, 기존 AHDRNet 및 제안된 모델 모두 μ -law를 기반으로 톤매핑하였다. 톤매핑된 이미지 $T(H)$ 는 식 (2)와 같이 표현할 수 있다.

$$T(H) = \frac{\log(1 + \mu H)}{\log(1 + \mu)} \quad (2)$$

이 때 H는 고동적 범위의 이미지이며, μ 는 압축량을 결정짓는 매개 변수로 두 모델 모두 $\mu=5000$ 이다. 이렇게 얻어진 톤매핑된 이미지를 토대로 기존의 모델과 제안된 모델의 성능 평가 비교를 위해 PSNR(Peak Signal-to-Noise Ratio) metric을 사용하였다.

AHDRNet의 저자(Yan, Qingsen, et al)가 공개한 training code가 따로 존재하지 않았기 때문에, 본 실험은 논문에 언급된 네트워크 설정과 학습 매개 변수 설정을 참고하여 동일한 환경에서 재현하여 실험하였다. 이 과정에서 주요 매개 변수들은 기존의 논문에 언급된 setting 값과 동일한 값을 사용하였다. 트레이닝 과정에서 Adam Optimizer를 사용하였으며, batch size는 8, 그리고 임의로 crop할 패치 사이즈는 256×256 로 고정시켰다. learning rate는 기존의 논문에서는 1×10^{-5} 로 되어 있으나, 본 실험에서는 빠른 학습을 위해 1×10^{-4} 로 통일시켰다. 기존 모델과 제안된 모델 모두 2000 epoch 만큼 트레이닝 시켰을 때 나온 모델을 기준으로 test PSNR 값을 측정해 본다.

4. 실험 결과

Kalantari의 테스트 데이터셋 안에 들어있는 총 15개의 ground truth 이미지를 제안된 모델의 결과 이미지와 각각 비교하여 PSNR을 측정하였다. 그 후 얻어낸 15개의 PSNR 값들의 평균을 통해 성능을 평가하였다. 그 결과 기존의 AHDRNet 모델의 PSNR은 42.1656였으며, 제안된 모델의 PSNR은 42.8135였다.

표 1 Kalantari의 데이터셋으로 실험했을 때 모델에 따른 average PSNR 값의 비교

model	AHDRNet	AHDRNet+CA
PSNR	42.1656	42.8135

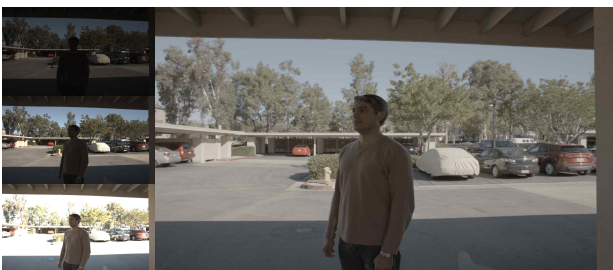


그림 3 HDR을 거친 후 입력과 출력 이미지의 비교



그림 4 HDR을 거친 후 입력과 출력 이미지의 비교

위의 그림 3, 그림 4에서 왼쪽에 배치한 세 개의 작은 이미지들은 입력으로 들어간 테스트 데이터셋으로, 위에서부터 차례로 낮은, 중간, 높은 노출량을 갖고 있다. 오른쪽에 배치한 큰 이미지는 입력으로 들어온 세 이미지를 제안된 모델에 통과시켰을 때 얻어낸 톤매핑된(tone-mapped) 이미지이다.

5. 결론

본 논문은 AHDRNet에 channel attention을 적용하여 성능 평가를 해보았다. AHDRNet의 병합 망에 존재하는 DRDB 레이어 뒤, 그리고 DRDB 내 확장된 합성곱 레이어들 사이에 합성곱 블록을 추가로 연결하여 channel attention을 적용시켰다. PSNR 비교해 보았을 때, 기존의 AHDRNet 모델에 비해 제안된 모델의 성능이 높아진 것을 확인했다.

감사의 글

이 논문은 2021년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원(2021R1A2C2007220)과 2021년도 BK21 플러스 사업에 의하여 지원을 받아 수행된 연구임.

참고문헌

- [1] Yan, Qingsen, et al. "Attention-guided network for ghost-free high dynamic range imaging." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019.
- [2] Zhang, Yulun, et al. "Image super-resolution using very deep residual channel attention networks." Proceedings of the European conference on computer vision (ECCV). 2018.
- [3] Kalantari, Nima Khademi, and Ravi Ramamoorthi. "Deep high dynamic range imaging of dynamic scenes." ACM Trans. Graph. 36.4 (2017): 144-1.
- [4] Woo, Sanghyun, et al. "Cbam: Convolutional block attention module." Proceedings of the European conference on computer vision (ECCV). 2018.
- [5] Park, Jongchan, et al. "Bam: Bottleneck attention module." arXiv preprint arXiv:1807.06514 (2018).

- [6] Zhang, Yulun, et al. "Residual dense network for image super-resolution." Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.
- [7] Yu, Fisher, and Vladlen Koltun. "Multi-scale context aggregation by dilated convolutions." arXiv preprint arXiv:1511.07122 (2015).
- [8] Huang, Gao, et al. "Densely connected convolutional networks." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.