

다중 노출 High Dynamic Range 이미징을 위한 경량화 네트워크

*이근택 **조남익

서울대학교 전기정보공학부, 뉴미디어통신공동연구소

*leekt000@ispl.snu.ac.kr **nicho@snu.ac.kr

Lightweight Network for Multi-exposure High Dynamic Range Imaging

*Lee, Keuntek **Cho, Nam Ik

Dept. of Electrical and Computer Eng., INMC, Seoul National University

요약

최근 영상 및 비디오 분야에 심층 신경망(DNN, Deep Neural Network)을 사용한 연구가 다양하게 진행됨에 따라 High Dynamic Range (HDR) 이미징 기술에서도 기존의 방법들 보다 우수한 성능을 보이는 심층 신경망 모델들이 등장하였다. 하지만, 심층 신경망을 사용한 방법은 큰 연산량과 많은 GPU 메모리를 사용한다는 문제점이 존재하며, 이는 심층 신경망 기반 기술들의 현실 적용 가능성에 제한이 되고 있다. 이에 본 논문에서는 제한된 연산량과 GPU 메모리 조건에서도 사용 가능한 다중 노출 HDR 경량화 심층 신경망을 제안한다. Kalantari Dataset에 대해 기존 HDR 모델들과의 성능 평가를 진행해 본 결과, PSNR- μ 와 PSNR-1 수치에서 GPU 메모리 사용량 대비 우수한 성능을 보임을 확인하였다.

1. 서론

다중 노출 HDR 이미징(Multi-exposure High Dynamic Range imaging)은 각각 다른 노출 정도로 촬영된 LDR(Low Dynamic Range) 이미지들로부터 하나의 HDR 이미지를 복원하는 기술이다. 특히, 다른 HDR 이미징 기술보다 다중 노출 HDR 이미징은 기술적으로 어려운 점들이 존재한다. 첫 번째는 입력으로 사용되는 LDR 이미지들의 공간 정보가 다르기 때문에 복원된 HDR 이미지에서 ghosting artifact와 같은 의도치 않은 문제점이 발생할 수 있다는 점이다. 두 번째는 각각의 LDR 이미지들에 존재하는 저조도 및 고조도 영역으로 인한 부족한 이미지 정보 때문에 복원된 HDR 이미지의 품질이 낮아질 수 있다는 점이다.

과거에는 LDR 이미지들 간의 optical-flow 추적을 통해, 미리 입력 LDR 이미지들을 사전 정렬하여 사용하는 방식을 통해 이와 같은 문제를 해결하고자 연구하였다[3, 4]. 최근에는 심층 신경망, 특히 합성곱 신경망(CNN, Convolutional Neural Network)을 통한 방법이 이미지 초해상, 이미지 잡음 제거 등의 다양한 이미지 복원 분야에서 주로 연구되어 왔으며, 기존의 방법들 보다 우수한 성능 결과를 보여주고 있다. 합성곱 신경망을 사용한 방법들에서도 위에서 언급한 다중 노출 HDR 이미징의 문제를 해결하기 위해, 입력 LDR 이미지들의 피쳐를 병합하는 방법에 대해 다양한 구조의 네트워크를 제안하였다[1, 2, 6, 7].

AHDRNet에서는 spatial-attention 모듈을 사용하여 입력 LDR 이미지들을 피쳐 공간에서 정렬한다[6]. 구체적으로, 입력 LDR 이미지들 중 ground-truth HDR 이미지와 동일한 공간 정보를 가지는 LDR 이미지를 reference로, spatial-attention 모듈을 통해 다른 LDR 이미지들의 피쳐들의 attention map을 학습한다. 이러한 방법을 통해

AHDRNet은 피쳐들의 공간 정보를 효율적으로 정렬하였고, 고품질의 HDR 이미지를 생성하는 결과를 보였다.

NHDRNet은 non-local block 구조를 차용하여 LDR 이미지들의 피쳐를 병합하는 방법으로 사용하였다[1]. Non-local module과 동시에 세 가지 다른 필터를 사용하여 다양하고 견고한 피쳐들을 추출하는 tripple-pass 모듈을 병렬적으로 구성하여, 기존 다중 노출 HDR 이미징 모델들보다 우수한 성능을 달성하였다.

이와 같은 심층 합성곱 신경망들의 발전에도 불구하고, 아직 GPU 메모리 사용량 및 무거운 연산량과 같은 문제는 존재한다. 더욱이, 다중 노출 HDR 이미징 기술의 경우에는 입력이 여러개의 LDR 이미지가기 때문에, 네트워크 크기 증가에 따른 연산량 및 메모리 사용량이 비교적 크게 증가하게 된다.

본 논문에서는 다중 노출 HDR 이미징의 특성을 고려하여, discrete wavelet transform 인코더와 inverse wavelet transform 디코더, selective kernel merging 모듈을 사용하는 효율적인 다중 노출 HDR 이미징 경량화 네트워크를 제안한다. 또한 기존 심층 합성곱 신경망을 사용한 방법들과 비교하여 상대적으로 적은 연산량과 GPU 메모리 사용량을 보이지만, 우수한 성능을 달성함을 실험으로 보인다.

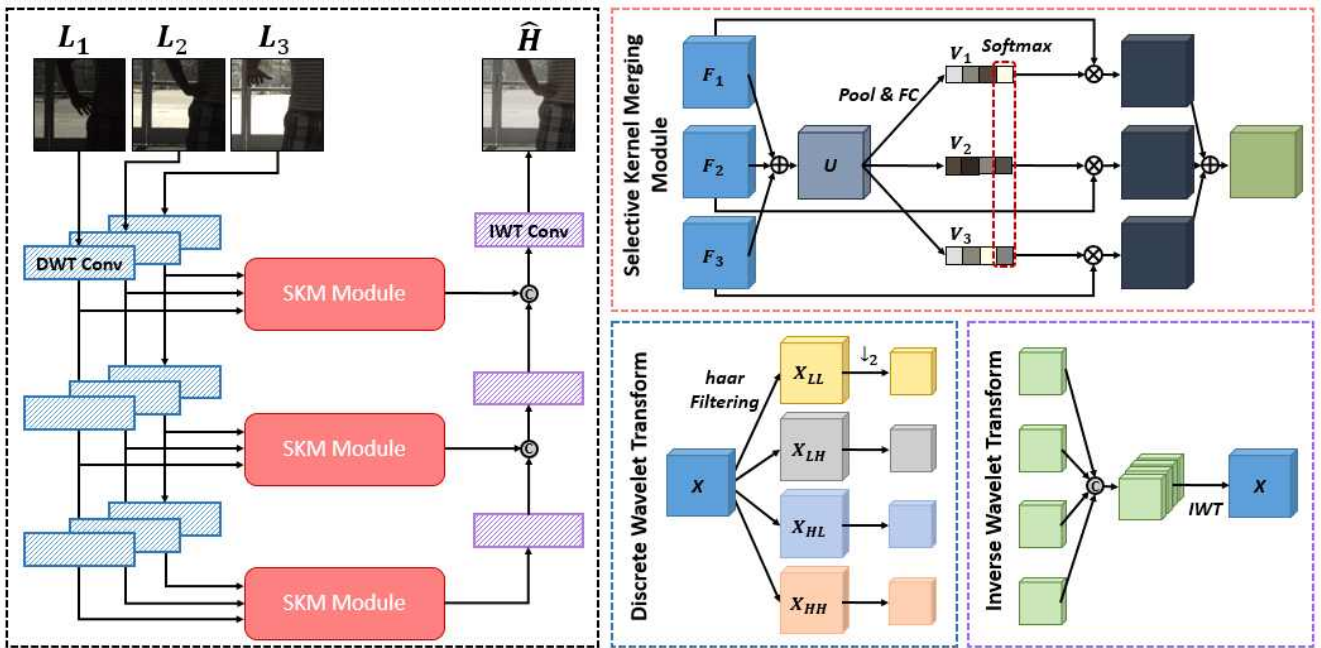


그림 1. 제안하는 경량화 다중 노출 HDR 이미징 네트워크 구조

2. 본론

2.1 네트워크 구조

제안하는 심층 합성곱 신경망은 기존에 HDR 이미징 기술에서 안정적인 성능을 보인 U-Net 구조를 사용하여 구성하였다[5]. 세부적인 사항으로는, 3-stage U-Net으로 구성하였으며 전체적인 네트워크의 구성은 아래의 그림 1과 같다. 총 3개의 LDR 이미지들 L_1, L_2, L_3 이 네트워크 입력으로 들어가고, 예측 HDR 이미지 \hat{H} 을 출력한다.

본 네트워크의 특징으로, 기존의 U-Net에서의 인코더는 피쳐 크기를 줄이기 위해 1보다 큰 stride값을 갖는 convolution 층으로 구성된 것과 다르게 DWT(Discrete Wavelet Transform)을 통해 피쳐 크기를 줄이고, 채널을 늘리는 방법을 사용하였다. 디코더도 마찬가지로 기존에서는 transposed convolution 층으로 구성된 것과 달리 IWT(Inverse Wavelet Transform)을 사용하여 피쳐들의 크기를 키우고, 채널을 줄이도록 설계하였다. 서론에서 언급하였듯이, 다중 노출 HDR 이미징에서는 입력이 다수의 LDR 이미지가기 때문에, 인코더 네트워크의 크기 증가에 따른 전체 네트워크의 메모리 사용량과 연산량이 상대적으로 큰 폭으로 증가하게 된다. 제안하는 wavelet을 사용한 인코더/디코더 구성은 필터의 값과 연산이 고정되어 있기 때문에, 단순히 convolution 층이나 transposed convolution으로 인코더/디코더를 구성하는 것 보다 적은 메모리를 사용한다. 본 논문에서는 DWT와 IWT에서 사용하는 필터로 haar 필터를 사용하였다. haar 필터의 구성은 수식 1과 같다.

$$f_{LL} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, f_{LH} = \begin{bmatrix} -1 & -1 \\ 1 & 1 \end{bmatrix}, f_{HL} = \begin{bmatrix} -1 & 1 \\ -1 & 1 \end{bmatrix}, f_{HH} = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$

수식 1. haar wavelet filter 구성

2.2 Selective Kernel Merging Module

SKNet에서는 selective kernel unit을 사용하여 분류 모델에서 우수한 성능을 달성함을 보였다[8]. 본 네트워크에서는 인코더를 통과한

LDR 이미지들의 피쳐를 효율적으로 병합하기 위해 SKM(Selective Kernel Merging) module을 사용한다.

그림 2에 표시되었듯이, SKM module로 입력되는 LDR 이미지들의 피쳐 F_1, F_2, F_3 은 모두 더해 피쳐 U 를 생성한다. 이후 순차적으로 average-pooling과 fully connected layer를 통과하여 weight 벡터 V_1, V_2, V_3 를 생성한다. 마지막으로 softmax function을 생성된 V_1, V_2, V_3 에 채널 방향으로 적용하여 입력으로 들어온 피쳐 F_1, F_2, F_3 에 곱하고, 이를 모두 더해 병합된 피쳐 M 을 생성한다. SKM module을 통해 각 입력 LDR 이미지 피쳐들에서 집중적으로 활용할 채널을 학습할 수 있으며, 더욱이 SKM module 내에서 유일하게 생성하는 1차원 weight vector는 그 다른 병합 모듈들에서 생성하는 피쳐들의 크기에 비해 상대적으로 작은 크기를 가지기 때문에 적은 GPU 메모리로 효율적인 피쳐 병합을 수행 할 수 있다.

3. 실험

3.1 학습 방법

학습과 평가를 위해 Kalantari의 데이터셋을 사용하였다[2]. Kalantari 데이터셋은 74개의 학습 샘플과 15개의 평가 샘플로 구성되어 있다. 학습에 사용되는 이미지들은 각각 256×256 크기의 패치로 crop하여 사용하였고, 과적합 현상을 방지하기 위해 상하 및 좌우 flip과 rotation을 적용하여 데이터 증강(augmentation)을 적용하였다.

학습을 위한 배치 크기는 8로 설정하였고, 손실함수는 L1 loss를 사용하였으며, Adam optimizer를 사용하여 학습을 진행하였다[9]. 학습률(learning rate)은 1e-4에서 시작하여 100,000번째 학습 진행마다 0.1만큼씩 곱하여 하향 조절하였다.

네트워크 입력으로는 3채널 RGB 이미지를 사용하는 대신, 이전 연구들에서 실험하였던 것과 동일하게 HDR 도메인으로 gamma correction된 이미지와 concatenate하여 6채널 데이터를 입력으로 사



LDR 이미지들

네트워크 예측 HDR 이미지

그림 2. 평가 데이터 샘플 중 입력 LDR 이미지들과 네트워크 예측 HDR 이미지

용하였다[1, 2, 6, 7].

3.2 실험 결과

성능 평가를 위해서, Kalantari 데이터셋의 15개 평가 샘플에 대하여 네트워크 결과를 AHDRNet[6], NHDRNet[1], HDR-GAN[7] 모델들과 비교해 보았다[2]. 평가 샘플은 1440×960 크기 설정으로 진행하였으며, tone-mapping 이후의 성능인 PSNR-μ와 HDR 이미지 간의 성능인 PSNR-I에 대하여 성능을 측정하였다. NHDRNet의 경우에는, 저자들의 코드가 공개되지 않았고, 논문에 GPU 메모리 사용량에 대한 내용이 없어, 성능을 제외한 부분은 생략하였다. 표 1의 성능 결과와 GPU 메모리 사용량을 확인해 보면, 심층 합성곱 신경망을 사용한 모델들이 전체적으로 매우 큰 GPU 메모리 사용량을 보이는 반면에, 본 논문의 방법은 절반가량의 GPU 메모리 사용으로 효율적인 수준의 성능을 보여준다는 점을 확인 할 수 있다.

표 1. Kalantari 평가 데이터셋에 대한 성능 평가 및 메모리 측정

	PSNR-I	PSNR-μ	GPU Memory
HDR-GAN[7]	41.57	43.92	8,833MB
AHDRNet[6]	41.14	43.63	13,512MB
NHDRNet[1]	-	42.41	N/A
Ours	41.14	43.39	4,529MB

4. 결론

최근 심층 네트워크의 발전에 따라 HDR 이미징 분야에서도 성능적인 향상이 빠르게 이루어졌지만, 아직까지 네트워크들이 사용하는 GPU 메모리와 연산량이 크다는 점이 문제로 존재하며, 네트워크를 현실에서 실제로 사용하기에는 아직 한계점으로 남아있다.

위와 같은 문제를 해결하기 위해, 본 논문에서는 다중 노출 HDR 이미지를 위한 경량화 네트워크를 제안하였다. 제안한 네트워크에서는 기존 U-Net 구조에서 DWT와 IWT를 인코더/디코더 부분에서 사용하여, 입력 LDR 이미지의 피처를 효율적으로 추출하였으며, SKM module을 통해 입력으로 들어오는 LDR 이미지들을 상대적으로 적은 연산량만을 사용해 효율적으로 병합하였다. 결과적으로, 이전의 심층 합성곱 신경망 모델들과의 성능 비교에서 GPU 메모리 사용 대비 효율적인 성능을 보임을 실험적으로 확인하였다.

본 논문의 연구와 더불어, 입력으로 사용되는 LDR 이미지들의 특징점이나 주목할 만한 부분을 효율적으로 추출 할 수 있는 방법에 대해 연구된다면, 보다 효율적으로 LDR 이미지 피처들을 병합 할 수 있으며, 나아가 기존 네트워크들 보다 경량화 된 네트워크를 설계할 수 있을 것으로 기대한다.

감사의 글

이 논문은 2021년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 [NO.2021-0-01343, 인공지능대학원지원(서울대학교)]

참고문헌

[1] Yan, Qingsen, et al. "Deep HDR imaging via a non-local network." IEEE Transactions on Image Processing 29 (2020): 4308-4322.
 [2] Kalantari, Nima Khademi, and Ravi Ramamoorthi. "Deep high dynamic range imaging of dynamic scenes." ACM Trans. Graph. 36.4 (2017): 144-1.

-
- [3] Kang, Sing Bing, et al. "High dynamic range video." *ACM Transactions on Graphics (TOG)* 22.3 (2003): 319-325.
- [4] Zimmer, Henning, Andrés Bruhn, and Joachim Weickert. "Freehand HDR imaging of moving scenes with simultaneous resolution enhancement." *Computer Graphics Forum*. Vol. 30. No. 2. Oxford, UK: Blackwell Publishing Ltd, 2011.
- [5] Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." *International Conference on Medical image computing and computer-assisted intervention*. Springer, Cham, 2015.
- [6] 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.
- [7] Niu, Yuzhen, et al. "HDR-GAN: HDR image reconstruction from multi-exposed ldr images with large motions." *IEEE Transactions on Image Processing* 30 (2021): 3885-3896.
- [8] Li, Xiang, et al. "Selective kernel networks." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019.
- [9] Kingma, Diederik P., and Jimmy Ba. "Adam: A method for stochastic optimization." *arXiv preprint arXiv:1412.6980* (2014).