특집논문 (Special Paper)
방송공학회논문지 제29권 제3호, 2024년 5월 (JBE Vol.29, No.3, May 2024)
https://doi.org/10.5909/JBE.2024.29.3.235
ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

Transformer 기반의 HDR 영상 복원 알고리즘을 통한 효과적인 Ghost Artifact 제거

박 예 인[a)], 강 석 주[a)‡]

Effective Ghost Artifact Removal through Transformer-based HDR Image Reconstruction

Ye-In Park[a)] and Suk-Ju Kang[a)‡]

요 약

High dynamic range (HDR) imaging의 주요 문제는 다중 노출의 low dynamic range (LDR) 이미지들을 병합할 때 발생하는 ghost artifact를 제거하는 것이다. 주어진 다중 노출 이미지들 간의 움직임이 심한 영역에서는 필요한 정보가 로컬 영역에 존재하지 않기 때문에 local 특징 추출에 특화된 convolutional neural network (CNN) 기반 방법으로는 ghost artifact가 불가피하다. 이 문제를 해결하기 위해 우리는 local connectivity 특성을 지닌 CNN과 global dependency 획득이 가능한 transformer를 결합시킨 새로운 HDR 이미지 복원 네트워크를 제안한다. 제안 방법은 CNN을 통해 다중 노출의 LDR 이미지들에서 고차원의 feature들을 추출하고, 추출된 feature들을 중간 노출 LDR 기준으로 낮은 노출 LDR과는 Low-transformer에, 높은 노출 LDR과는 High-transformer에 입력시켜 global 정보를 캡처한다. 본 논문의 실험 결과를 통해 제안 방법이 기존 방법들과 비교하여 우수한 성능을 보임을 입증한다.

Abstract

The primary issue in high dynamic range (HDR) imaging is the removal of ghost artifacts that occur when merging multi-exposure low dynamic range (LDR) images. In areas with significant motion between the given multi-exposure images, necessary information is not present in the local region, making ghost artifacts inevitable for convolutional neural network (CNN)-based methods specialized in local feature extraction. To address this issue, we propose a novel HDR image reconstruction network that combines CNNs with transformers capable of acquiring global dependency while retaining local connectivity characteristics. Our method extracts high-dimensional features from multi-exposure LDR images using CNNs and then feeds these features into Low-transformers for low-exposure LDR images and High-transformers for high-exposure LDR images, based on a medium exposure LDR image, to capture global information. Through experiments presented in this paper, we demonstrate the superior performance of the proposed method compared to the existing methods.

Keyword : High dynamic range imaging, Ghost artifact removal, Multi-exposure low dynamic range images, Transformer, Attention mechanism

Ⅰ. 서 론

High dynamic range (HDR) 이미지는 10비트 또는 12비트로 구성되어 real world의 광범위한 밝기 범위를 보다 잘 포착할 수 있어, 영상에서 풍부한 정보와 색상을 제공한다. 그러나 대부분의 디지털 카메라와 기존 센서들은 8비트의 low dynamic range (LDR) 이미지만 캡처할 수 있어, 밝은 영역이 과도하게 밝아지거나 어두운 영역이 깊게 어두워져 데이터 손실이 크다는 문제가 발생한다. 따라서 LDR 이미지에서 손실된 데이터를 복구하고, 넓은 조도 범위를 제공할 수 있는 HDR 이미지를 생성하기 위한 딥러닝 기반의 연구들이 진행되고 있다. 일반적인 HDR 재구성 방법인 다중 노출 이미지 융합은 노출 값이 다른 LDR 이미지들을 정렬한 후 HDR 이미지로 병합한다[1,2]. 그러나 정렬된 LDR 이미지를 획득하는 것은 카메라나 물체의 움직임으로 인해 생성되는 HDR 이미지에 ghost artifact가 발생한다. 이 문제를 해결하기 위해 optical-flow 기반 방법[3]이 제안되었지만 ghost artifact는 여전히 발생한다. Deep neural network 기반 방법 중 convolutional neural network (CNN)를 사용하는 방법은 HDR 이미지 재구성에 유망한 것으로 보이지만 local connectivity 특성으로 인해 local feature 추출에 중점을 두기 때문에 ghost artifact 제거에 어려움이 있다. 이러한 한계를 극복하기 위해 transformer 구조를 통해 이미지를 패치 단위로 분할하고 패치 간의 attention 가중치를 계산하여 global 정보 추출이 가능하다[4]. 따라서 본 논문에서는 움직임이 있는 영역과 움직임이 없는 영역에 대한 feature를 효과적으로 추출하기 위해서 local feature 추출에 특화된 CNN과 global feature 추출에 강력한 transformer를 결합한 새로운 HDR 이미지 복원 네트워크를 제안한다. 제안 방법을 통해 기존 방법들 대비 ghost artifact의 상당 부분을 제거하여 보다 풍부한 정보와 색상을 갖는 고화질의 HDR 이미지를 생성할 수 있다. 본 논문은 제안 방법의 네트워크 구조에 대한 상세한 설명과 기존 방법 대비 높은 정성적 및 정량적 성능 평가 위주로 구성되어 있다.

a) 서강대학교 전자공학과(Dept. of Electronic Engineering, Sogang University)
‡ Corresponding Author : 강석주(Suk-Ju Kang)
E-mail: sjkang@sogang.ac.kr
Tel: +82-2-705-8466
ORCID: https://orcid.org/0000-0002-4809-956X
※ This research was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2020M3H4A1A02084899), and by the MSIT(Ministry of Science and ICT), Korea, under the ITRC(Information Technology Research Center) support program(IITP-2024-RS-2023-00260091) supervised by the IITP(Institute for Information & Communications Technology Planning & Evaluation).
· Manuscript March 26, 2024; Revised April 17, 2024; Accepted April 18, 2024.

Ⅱ. 제안 방법

본 논문에서는 움직임이 있는 다중 노출 LDR 이미지들로 HDR 이미지 생성 시 발생하는 ghost artifact를 완화하기 위한 새로운 CNN-transformer 기반의 HDR 이미지 복원 네트워크를 제안한다. 네트워크에 LDR 이미지들을 입력하기에 앞서 여러 관련 연구들[5,6,7]에서 사용된 것처럼 우리는 감마 보정을 통해 LDR 이미지를 해당되는 HDR 표현으로 변환한다. LDR 이미지와 매핑된 HDR 이미지를 입력으로 함께 사용하고 픽셀 값은 모두 [0,1]로 정규화한다. 그림 1과 같이 전체 네트워크는 크게 feature 추출 네트워크, 병렬적인 transformer, fusion 네트워크로 구성된다. 다음 하위 섹션들에서 각 서브 네트워크들에 대하여 설명하겠다.

1. Feature 추출 네트워크

Feature 추출 네트워크에서는 세 개의 다중 노출 LDR 이미지들을 각각 CNN 및 channel attention을 통해 각 이미지에 해당되는 고차원 feature를 추출한다. CNN에 있어서는 하나의 CNN 레이어와 하나의 MultiConv 모듈로 구성이 되는데, MultiConv는 3개의 서로 다른 커널 사이즈를 갖는 CNN 레이어들을 병렬적으로 구성하고, 이들을 concatenation 연산을 통해 연결한 후에 하나의 CNN 레이어를 거친 뒤 MultiConv의 입력 feature와 더해 최종 feature를 출력시키는 구조이다. 이때, 병렬적으로 구성되는 3개의 CNN 레이어의 커널 사이즈는 1, 3, 5이며, 입력 채널과 출력 채널 수는 모두 64로 동일하다. 다중 스케일을 갖는 MultiConv를 통해 상세한 local feature 추출이 가능해진다. 작은 커널 사이즈를 통한 세밀한 텍스처부터, 큰 커널 사이즈를 통한 넓은 영역의 구조적 패턴까지 다양한 정보를 효

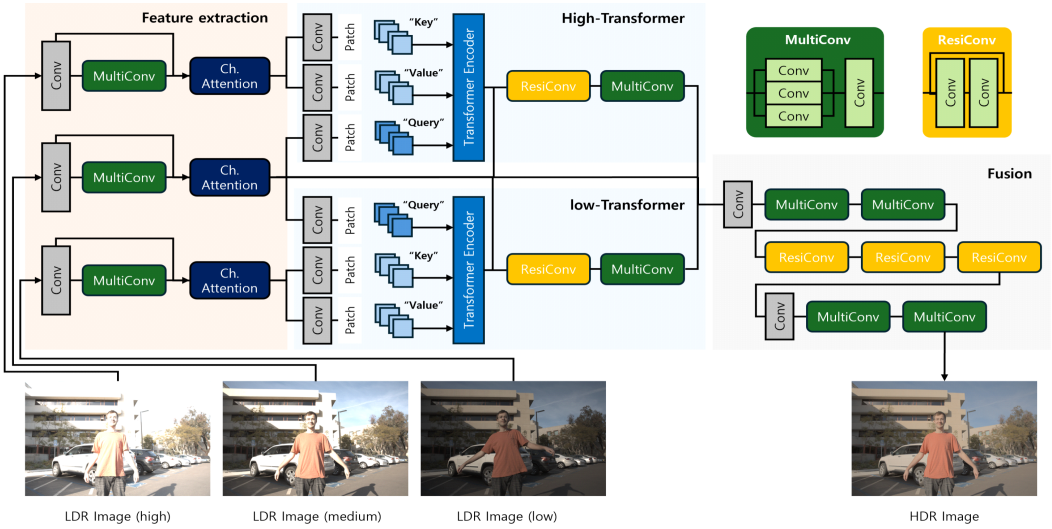

그림 1. 제안 방법의 전체 네트워크 구조
Fig. 1. Overall architecture of the proposed method

과적으로 학습할 수 있다. 각 스케일에서 추출된 feature들을 결합함으로써 보다 포괄적인 정보를 활용할 수 있어 고해상도의 디테일을 유지하면서도 전체적인 이미지의 대비와 색상의 다이나믹을 향상시켜 최종적으로 복원될 HDR 이미지에 만족스러운 효과를 기여할 수 있다. 이렇게 획득된 feature가 처음 CNN 레이어의 feature와 결합되고 결합된 feature가 channel attention[8]에 입력된다. Channel attention은 average pooling 및 fully-connected layer를 통해 추출된 feature에 sigmoid function을 적용하고 그 가중치를 곱 연산하는 과정을 담는다. 이를 통해, 각 이미지를 대표하는 유익한 local feature로의 선택이 가능하다.

2. 병렬적인 transformer 구조

Feature 추출 네트워크에서 출력된 각 LDR 이미지에 대한 feature들이 transformer에 입력된다. transformer는 병렬적으로 두 개의 네트워크가 동일한 구조로 구성되는데, Low-transformer는 낮은 노출 LDR과 참조 이미지인 중간 노출 LDR의 feature를 입력으로 받고, High-transformer는 높은 노출 LDR과 중간 노출 LDR의 feature를 입력으로 받는다. 각 transformer 구조에서 query는 참조 이미지인 중간 노출 LDR의 feature를 입력한다. 낮은 노출 LDR과 높은 노출 LDR feature는 각 transformer 구조에서 key와 value로 입력되며 그들에 대한 모든 패치들이 탐색되어 움직임이 있는 영역에서 필요한 정보들이 attention 가중치를 통해 획득될 수 있도록 하기 때문에 global 정보 추출이 가능하다. 이러한 방법으로 추출된 feature는 feature 추출 네트워크에서 출력된 중간 노출 LDR 이미지의 feature와 결합되어 하나의 ResiConv 모듈과 하나의 MultiConv 모듈을 통과한다. ResiConv는 두 개의 연속적인 CNN 레이어로 구성되며, 이전 레이어의 feature가 뒤에서 연결되는 residual block 형태이다. ResiConv에 입력되는 feature와 출력되는 feature의 channel 수는 모두 128로 동일하다. 이렇게 residual block의 구성을 통해 입력에 가까운 feature를 네트워크의 심층 레이어에 직접적으로 전달하게 함으로써 심층 네트워크에서 발생할 수 있는 gradient vanishing 문제를 완화할 수 있고, 안정적인 학습이 가능해진다. 또한, 이전 레이어에서 학습된 feature 정보가 손실 없이 전달될 수 있기 때문에 중요한 정보를 보존하는데 효과적인 방법이다. Low-transformer와 High-transformer 각각에서 출력된 feature와 feature 추출 네트워크에서 출력된 중간 노출 LDR 이미지의 feature가 결합되어 fusion 네트워크에 입력된다.

3. Fusion 네트워크

Fusion 네트워크는 여러 개의 MultiConv 모듈과 Resi-

Conv 모듈의 시퀀스로 구성된다. 추출된 feature들을 결합하고 의미있는 고차원의 feature를 추출하여 최종적으로 예측된 HDR 이미지가 생성된다. HDR 이미지를 복원하는데 사용되는 loss function은 다음과 같다.

$$L(H,\hat{H}) = \| T(H) - T(\hat{H}) \|_2 \quad (1)$$

H와 $\hat{H}$는 각각 ground truth (GT) HDR 이미지와 복원된 HDR 이미지를 의미하고, $T(\cdots)$와 $\| \cdots \|_2$는 각각 μ-law을 사용한 tone mapping과 l_2 norm을 의미한다.

Ⅲ. 실험 결과

1. 데이터셋

우리는 제안 방법의 성능을 검증하기 위해 Kalantari 데이터셋[5]을 사용하였다. 이 데이터셋은 74개의 training 샘플과 15개의 test 샘플로 구성되어 있다. 각 샘플에는 {-2, 0, +2} 또는 {-3, 0, +3}의 노출 편향이 서로 다른 3개의 정렬되지 않은 이미지들이 포함되어 있다. 학습에 있어 전체 이미지를 무작위로 자른 256×256 크기의 패치를 사용하였고, 무작위 회전 및 뒤집기를 적용하여 데이터 증강을 통해 학습 샘플을 다양화하였다.

2. 정성적 성능 평가

그림 2와 그림 3을 통해 제안 방법과 관련 최신 방법들[5,9,10]의 결과 이미지를 시각적으로 확인할 수 있다. 입력되는 LDR 이미지들과 부분적으로 확대한 LDR 패치들을 통해 알 수 있듯이 움직임의 크기가 상당히 크다. 그림 2에

그림 2. Kalantari 데이터셋에 대한 제안 방법과 기존 방법들의 정성적 결과 비교
Fig. 2. Comparison of qualitative results of the proposed method and the existing methods on the Kalantari dataset

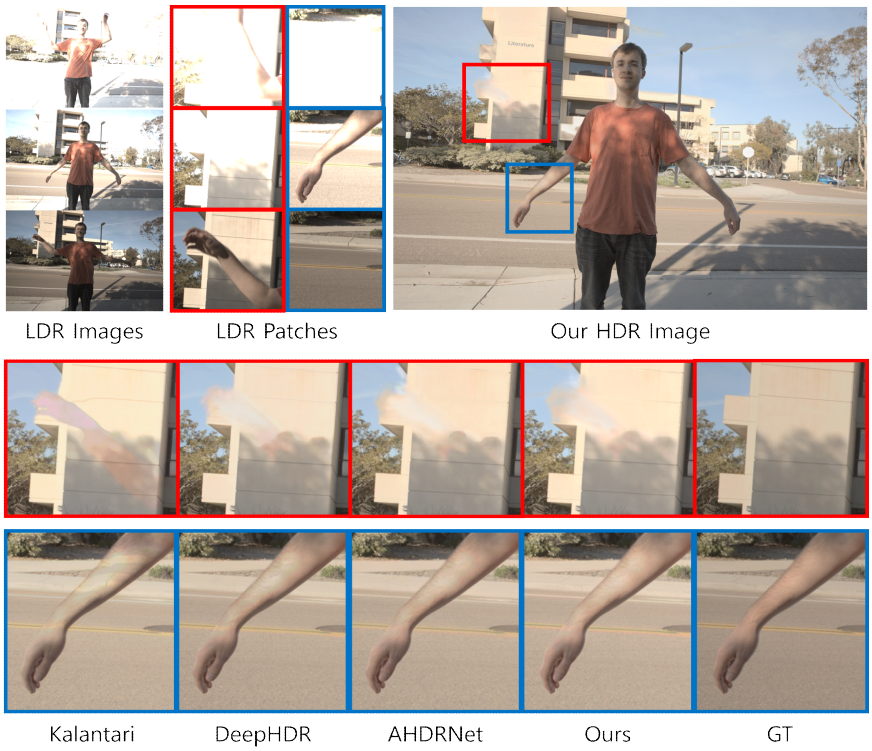

그림 3. Kalantari 데이터셋에 대한 제안 방법과 기존 방법들의 정성적 결과 비교
Fig. 3. Comparison of qualitative results of the proposed method and the existing methods on the Kalantari dataset

서의 중간 노출 LDR에 대한 빨간 박스와 파란 박스 패치를 보면, 사람의 팔이 포화되어 낮은 노출 LDR에서 세부 정보를 가져와야 하고, 큰 움직임으로 인해 해당되는 세부 사항을 global 영역에 걸쳐 획득되어야만 한다. 비교 방법들은 global 영역에 걸쳐 필요한 정보들을 탐색 및 획득하는 과정이 없기 때문에 ghost artifact가 큰 것을 확인할 수 있다. 이에 반해, 제안 방법은 ghost artifact를 효과적으로 제거하여 시각적으로 만족스러운 결과를 얻는 것이 가능하다. 그림 3에서의 중간 노출 LDR에 대한 빨간 박스 패치를 보면, 건물이 포화되어 낮은 노출 LDR에서 세부 정보를 가져와야 한다. 그러나, 낮은 노출 LDR의 패치에서는 건물 앞에 사람의 팔이 등장하여 건물이 크게 가려져 복원에 있어 가장 어려운 영역이다. 해당 영역에 대해서도 제안 방법이 비교 방법들 대비 ghost artifact를 효과적으로 줄인 것을 확인할 수 있다. 그림 3에서의 중간 노출 LDR에 대한 파란 박스 패치를 보면, 사람의 팔이 부분적으로 포화되어 있어 낮은 노출 LDR에서 해당 영역에 대한 세부 정보를 가져와야 한다. 그러나 낮은 노출 LDR에는 사람의 팔이 존재하지 않기 때문에 global 영역에 걸쳐 필요한 정보만 선택적으로 가져와야 하지만 비교 방법들은 그러한 과정이 없기 때문에 사람의 팔에 뒷 배경에 대한 텍스처가 혼재되어 나타나는 것을 확인할 수 있다. 이에 반해, 제안 방법은 사람의 팔 영역에 필요한 정보를 global 영역에 걸쳐 획득하기 때문에 ground truth와 가장 유사한 결과 이미지를 보이는 것으로 확인된다.

3. 정량적 성능 평가

우리는 제안 방법을 두가지 평가 척도를 기반으로 평가하였다. μ-law를 사용한 tone mapping 이미지에서 peak-signal-to-noise ratio (PSNR)과 structure similarity (SSIM), 즉 PSNR-μ와 SSIM-μ를 측정하여 결과 HDR 이미지의 성

표 1. Kalantari 데이터셋에 대한 제안 방법과 기존 방법들의 정량적 결과 비교

Table 1. Comparison of quantitative results of the proposed method and the existing methods on the Kalantari dataset

Method	Kalantari[5]	DeepHDR[9]	AHDRNet[10]	Proposed
PSNR-μ	42.74	41.63	43.68	43.73
SSIM-μ	0.9877	0.9869	0.9902	0.9904

능을 평가하였다[11]. 표 1은 제안 방법이 비교 방법들과 비교하였을 때 두 가지 평가 척도에서 각각 43.73, 0.9904로 가장 높은 성능을 내는 것을 보여준다.

Ⅳ. 결 론

본 논문에서는 global dependency 획득이 가능한 transformer와 local connectivity 특성을 지닌 CNN을 결합시켜, 다중 노출 LDR 이미지들 간 움직임이 있는 영역과 움직임이 없는 영역을 모두 효과적으로 복원시켜 고화질의 HDR 이미지를 생성하는 새로운 HDR 이미지 복원 네트워크를 제안하였다. 실험 결과, 기존 방법들을 통해 생성된 HDR 이미지들의 경우 ghost artifact가 상당 부분 확인되는데 반해 제안 방법을 통해 생성된 HDR 이미지는 ghost artifact의 범위가 작고 우수한 표현력을 갖는 것을 입증하였다. 또한, 대표적인 평가 척도에 있어서도 높은 성능을 보임을 입증하였다. 그러나 복원이 필요한 포화 영역이 다른 노출 이미지에서 크게 가려진 경우에 대해서는 transformer 구조만으로는 한계가 분명하기 때문에 generative AI 기법의 추가 적용을 통한 성능 개선 연구가 필요할 것으로 사료된다. 본 논문에서 제안한 방법을 통해 HDR 이미지 생성 기술이 발전되기를 기대하며, 사진 및 비디오 산업, 의료 이미징, 자율 주행, 보안 시스템 등 다양한 분야에 혁신적인 결과가 나오기를 기대한다.

참 고 문 헌 (References)

[1] P. E. Debevec and J. Malik, “Recovering high dynamic range radiance maps from photographs,” in Seminal Graphics Papers: Pushing the Boundaries, Volume 2, pp. 643 – 652, 2023.
doi: https://doi.org/10.1145/258734.258884.

[2] M. Granados, B. Ajdin, M. Wand, C. Theobalt, H.-P. Seidel, and H. P. Lensch, “Optimal hdr reconstruction with linear digital cameras,” in 2010 IEEE computer society conference on computer vision and pattern recognition, pp. 215 – 222. IEEE, 2010.
doi: https://doi.org/10.1109/CVPR.2010.5540208.

[3] A. Dosovitskiy, P. Fischer, E. Ilg, P. Hausser, C. Hazirbas, V. Golkov, P. Van Der Smagt, D. Cremers, and T. Brox, “Flownet: Learning optical flow with convolutional networks,” in Proceedings of the IEEE international conference on computer vision, pp. 2758 – 2766, 2015.
doi: https://doi.org/10.1109/ICCV.2015.316.

[4] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, “Attention is all you need,” Advances in neural information processing systems, vol. 30, 2017.
doi: https://dl.acm.org/doi/10.5555/3295222.3295349.

[5] N. K. Kalantari, R. Ramamoorthi et al., “Deep high dynamic range imaging of dynamic scenes.” ACM Trans. Graph., vol. 36, no. 4, pp. 144 – 1, 2017.
doi: http://dx.doi.org/10.1145/3072959.3073609.

[6] Q. Yan, L. Zhang, Y. Liu, Y. Zhu, J. Sun, Q. Shi, and Y. Zhang, “Deep hdr imaging via a non-local network,” IEEE Transactions on Image Processing, vol. 29, pp. 4308 – 4322, 2020.
doi: https://doi.org/10.1109/TIP.2020.2971346.

[7] Z. Pu, P. Guo, M. S. Asif, and Z. Ma, “Robust high dynamic range (hdr) imaging with complex motion and parallax,” in Proceedings of the Asian Conference on Computer Vision, 2020.
doi: https://doi.org/10.1007/978-3-030-69532-3_9.

[8] Z. Liu, Y. Wang, B. Zeng, and S. Liu, “Ghost-free high dynamic range imaging with context-aware transformer,” in European Conference on Computer Vision, pp. 344 – 360. Springer, 2022.
doi: https://doi.org/10.1007/978-3-031-19800-7_20.

[9] S. Wu, J. Xu, Y.-W. Tai, and C.-K. Tang, “Deep high dynamic range imaging with large foreground motions,” in Proceedings of the European Conference on Computer Vision (ECCV), pp. 117 – 132, 2018.
doi: https://doi.org/10.1007/978-3-030-01216-8_8.

[10] Q. Yan, D. Gong, Q. Shi, A. v. d. Hengel, C. Shen, I. Reid, and Y. Zhang, “Attention-guided network for ghost-free high dynamic range imaging,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1751 – 1760, 2019.
doi: https://doi.org/10.1109/CVPR.2019.00185.

[11] A. Hore and D. Ziou, “Image quality metrics: Psnr vs. ssim,” in 2010 20th international conference on pattern recognition, pp. 2366 – 2369. IEEE, 2010.
doi: https://doi.org/10.1109/ICPR.2010.579.

저 자 소 개

박 예 인

- 2018년 : 동덕여자대학교 컴퓨터학과 졸업
- 2020년 ~ 현재 : 서강대학교 전자공학과 석박통합 재학
- ORCID : https://orcid.org/0000-0001-7713-5753
- 주관심분야 : 컴퓨터비전, 영상처리, 딥러닝, 시계열 데이터 분석

강 석 주

- 2006년 : 서강대학교 전자공학과 졸업
- 2011년 : 포항공과대학교 전자전기공학과졸업 (공학박사)
- 2011년 ~ 2012년 : LG Display 선임 연구원
- 2012년 ~ 2015년 : 동아대학교 전기공학과 조교수
- 2015년 ~ 2021년 : 서강대학교 전자공학 부교수
- 2021년 ~ 현재 : 서강대학교 전자공학 교수
- ORCID : https://orcid.org/0000-0002-4809-956X
- 주관심분야 : 멀티미디어 영상신호처리, 컴퓨터 비전, 딥러닝 시스템 설계