

가우시안 어둠 분포 학습을 통한 잠재 확산 기반 저조도 이미지 밝기 복원 기법에 관한 연구

송순용
한국전자통신연구원
soony@etri.re.kr

A Study on Gaussian Darkness Modeling in Latent Diffusion for Low-Light Image Enhancement

Soonyong Song
Electronics and Telecommunications Research Institute (ETRI)

요 약

본 논문은 잠재 확산 기반의 저조도 개선 기법인 LD-Gauss 를 제안하였다. 저조도 이미지를 안정적으로 밝게 만들기 위해, 잠재 공간에서 어두움을 가우시안 분포로 근사하고 그 분포의 평균과 분산을 조건부로 추정하는 뒤, 확산 모델의 역과정에서 이를 이용해 원래의 광량을 복원하는 것이 핵심 아이디어이다. 제안한 모델은 사전학습된 Latent Diffusion 의 pretrained weight 를 freeze 하고, 저조도 조건으로부터 잠재 복원량을 예측하는 추정 네트워크와 잠재 U-Net 만 학습한다. 실험 결과, PSNR 은 9.66dB 에서 16.45dB 로 약 6.79dB 향상되었고, SSIM 은 0.351 에서 0.478 로 증가하였다. 반면 LPIPS 는 0.328 에서 0.369 로 열화되어 밝기가 올라가면서 세부 질감과 자연스러움의 손실이 발생됨을 시사하였다. 시각적인 평가 역시 전반적인 노출 회복에는 성공하였지만, 부분적 디테일 붕괴와 과도한 밝기 보정으로 인한 인공적인 느낌이 관찰되었다. 이 결과는 PSNR 과 SSIM 지표가 픽셀의 일치도를 선호하는 반면, LPIPS 가 사람의 시각과 상관성이 높다는 기존의 결과와 일치한다.

I. 서 론

저조도 영상은 촬영 단계에서 밝기가 부족하고 센서 잡음이 상대적으로 커지기 때문에, 단순히 톤을 조정하는 것 만으로는 충분한 밝기 회복과 디테일 유지를 동시에 달성하기 어렵다. 특히, 노이즈 특성은 광 신호의 세기에 따라 달라지기 때문에 푸아송 및 가우시안 잡음 모델을 통해 물리적 현실성을 반영한 저조도 합성이 필요하다. [1] 전통적으로, 저조도 이미지 개선 분야에서 GAN 과 같은 딥러닝 기법이 널리 사용되었다. 하지만, 실제 어두운 환경에서는 학습 이미지의 부족이나 과도 보정, 질감 손실 같은 한계가 존재하였다. 최근에는 Diffusion 기반 방법들이 이미지 품질 회복과 구조 보존 측면에서 유망한 결과를 보여주고 있다. Latent Diffusion 과 같은 확산 모델은 고해상도 이미지 합성 및 복원 작업에서 성공적으로 활용되는 모델로, VAE 를 활용하여 잠재 공간에서 확산 과정을 수행함으로써 픽셀 공간 기반보다 계산 효율성과 시각 품질 면에서 모두 우수한 결과를 보여준다. [2]

본 연구에서는 저조도에서의 어둠을 잠재 공간에서 가우시안 분포로 모델링하고, 이를 확산 복원 과정에 통합하는 새로운 접근법인 LD-Gauss 기법을 제안한다. 이 방법은 사전학습 가중치를 고정한 상태에서, 밝기

손실 복원량(평균 및 분산)을 조건적으로 추정하는 중간 네트워크와 잠재 U-Net 기반 확산 역과정을 학습 함으로써, 잠재공간에서의 통계적인 분포 근사화를 통해 노출 부족 문제와 잡음으로 무너진 구조를 동시에 복원하는 것을 목표로 한다.

II. 본론

본 연구의 목표는 잠재 공간에서 노출 손실의 확률 구조를 모델링하여 저조도 입력을 원래 노출로 복원하는 것이다. 구체적으로는 세 단계로 요약할 수 있다. 첫째, 원본 이미지를 sRGB 에서 선형 광도로 바꾼 뒤, 노출 스톱 감소, 푸아송 잡음 추가, 가우시안 잡음 추가, sRGB 역변환을 거쳐 저조도 데이터를 합성한다. 둘째, 사전학습된 Latent Diffusion 의 사전학습 가중치를 그대로 고정하고, 고정된 scaling factor 규칙에 따라 입력과 저조도의 잠재 특성을 얻는다. 셋째, 저조도 영상에서 얻은 밝기, 엷지, 주저파 조명 신호와 추정 잡음 세기 맵을 조건으로 삼아 잠재 복원량의 평균과 분산을 예측하는 네트워크와 확산 역과정의 노이즈 예측 U-Net 을 함께 학습한다. 확산 손실과 함께 평균 및 분산이 실제 잠재 차이와 들어맞도록 Gaussian NLL 을 적용하고, Latent Diffusion 복호화 후의 사이클 일치 손실을 보조로 합산하였다. 이런 구성은 물리 합성 단계에서의 푸아송-가우시안 가정과, Latent Diffusion

비용 효율성을 동시에 달성한다. 데이터셋은 STL-10 을 사용하였다. 이 데이터셋은 96x96 해상도의 컬러 이미지로 구성되고, 지도 학습에 쓸 수 있는 라벨 수는 적지만 비지도 학습을 위한 대량의 비라벨 데이터를 포함한다. 본 연구에서는 회귀성 복원 과제이므로 라벨 정보는 사용하지 않았다. 저해상도이긴 하지만, 잠재 해상도가 8 배 축소되는 Latent Diffusion 구성을 시험하기에는 충분한 해상도를 갖는다. 정량 평가에서는 PSNR, SSIM, LPIPS 세 가지 지표를 사용하였다. PSNR 은 평균 제곱 오차 기반이기 때문에 밝기나 색이 일치할수록 값이 커지고, SSIM 은 구조적 유사성을 통해 대비, 명암, 구조의 보존을 측정한다. LPIPS 는 사전학습 신경망의 특성 공간에서 거리로 유사도를 측정한 지표로 사람의 시각과 더 높은 상관을 보이는 것으로 알려져 있다. 이 세 가지 지표를 통해 복원된 이미지의 품질을 균형적으로 평가할 수 있다. [3] 실험 결과, 저조도 입력 대비 복원 전과 복원 후를 비교하였을 때, PSNR 은 9.66dB 에서 16.45dB 로 6.79dB 상승했고, SSIM 은 0.351dptj 0.478 로 0.127 증가하여 약 36% 향상이 있음을 확인하였다. 반면 LPIPS 는 0.328 에서 0.369 로 0.041 증가하여 특성의 거리 측면에서 손실이 있음을 확인하였다. 이러한 결과는 노출과 전반적인 구조 회복에는 성공하였지만, 자연스러운 질감은 충분히 복원되지 못하였음을 의미한다. 그림 1 에는 학습된 모델로 복원한 결과를 나타내었다. 시각적으로 전체 밝기는 target 이미지와 비슷하게 올라가고 전반적인 실루엣은 유사하게 나타났지만, 세부적인 특징인 무늬, 경계면 등에서 세부 디테일이 무너지는 것으로 판단되며, 일부 장면에서는 과다 보정으로 인한 인공적인 결과가 출력되었다. LPIPS 성능의 열화는 이러한 결과와 일치한다. 이러한 결과는 세 가지 원인에 의한 것으로 분석할 수 있다. 첫째, 잠재 영역에서 해상도의 축소 때문으로, Latent Diffusion 기본 구성처럼 축소 후 복원 절차를 진행하면 고주파 표현 능력의 저하가 발생한다. 둘째, 본 모델은 노출과 잡음의 통계적 복원에 초점을 둔다. 밝기를 올리고 노이즈를 정규화하는 방향의 손실이 강하면 텍스처를 복원하기 위해 필요한 미세 특징이 희생될 수 있다. 셋째, 학습 데이터셋의 해상도가 낮은 것과 이미지의 다양성 부족이 문제일 수 있다. 실사용에 가까운 질감을 확보하기 위해서는 큰 해상도로 다양한 장면들의 학습이 필요하다. 이번 결과에서 PSNR 과 SSIM 이 유의미하게 좋아진 것은 제안한 방법론은 합리적으로 선택된 것을 의미한다. 제안한 방법의 절차가 밝기와 구조를 복원하기에는 효과적인 것을 보여준다. 반면 LPIPS 가 열화된 사실은 텍스처 손실 향이 더 필요한 것을 시사한다.

III. 결론

본 논문은 잠재 공간에서 어둠을 가우시안으로 근사해 밝기를 복원하는 LD-Gauss 를 제안하였다. 제안 방법은 PSNR 상승과 SSIM 개선으로 밝기와 구조 회복이 가능함을 보였고, 동시에 LPIPS 의 열화와 시각적 평가에서의 디테일 붕괴와 과보정이라는 한계를 드러냈다. 그럼에도 본 결과는 잠재공간에서 노출 손실의 확률적인 분포를 학습하여 밝기를 복원하려는 접근은 유효함을 보여주며, 고해상도 데이터를 활용한 학습, 톤 제어, 저손실 VAE 채택 등의 방법을 적용하여 단점을 보완해 나간다면 실제 응용으로 이어질 가능성이 충분할 것으로 판단된다.



그림 1. 저조도 이미지 밝기 복원 결과: 입력 이미지 (좌), 복원 결과 (중), 목표 이미지(우)

ACKNOWLEDGMENT

본 연구는 한국전자통신연구원 연구운영비지원사업의 일환으로 수행되었음. [25ZR1100, 자율적으로 연결·제어·진화하는 초연결 지능화 기술 연구]

참 고 문 헌

- [1] Rombach, Robin, et al. "High-resolution image synthesis with latent diffusion models." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022.
- [2] Jiang, Hai, et al. "Lightendiffusion: Unsupervised low-light image enhancement with latent-retinex diffusion models." European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2024.
- [3] Zhang, Richard, et al. "The unreasonable effectiveness of deep features as a perceptual metric." Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.