야간 CCTV 영상의 사람 탐지 성능 향상에 관한 연구

김윤지 [1], 윤여찬 [2], 백민경 [3], 이용우 [4], 이우진 [5*]
[1] 동국대학교 컴퓨터 · AI 학과 석사과정
[2] 동국대학교 지리학과 석사과정
[3, 4] 동국대학교 컴퓨터공학전공 학사과정
[5] 동국대학교 컴퓨터 · AI 학과 교수

yunji03516@gmail.com, 2021112646@dgu.ac.kr, bmgbmk0423@gmail.com,
yongu3823@gmail.com, wj926@dgu.ac.kr

A Study on Enhancing Person Detection in Low-Light CCTV Image

Yun-Ji Kim[1], Yeo-Chan Yoon[2], Ming-Gyeong Baek[3], Yong-Woo Lee[4], Woo-Jin Lee[5]
[1, 5]Dept. of AI Convergence, Dongguk University-Seoul
[2]Dept. of Geography, Dongguk University-Seoul
[3, 4]Dept. of Computer Engineering, Dongguk University-Seoul

요 약

본 연구는 저조도 CCTV 환경의 사람 탐지 성능 저하 문제를 해결하고자, 기존의 시각적 개선 중심의 저조도 이미지 개선(LLIE) 기술의 한계를 분석한다. 이를 위해 YOLO 탐지 성능을 기준으로 최적의 LLIE 모델을 선정하고 YOLO 와 결합하는 파이프라인을 제안한다. LLVIP 데이터셋 기반의 비교 실험 결과, 객체의 구조를 가장 잘 보존하는 HVI-CIDNet 이 mAP 를 가장 효과적으로 향상시킴을 확인하였다. 이는 사람의 시각적 만족도보다 탐지 모델 친화적 개선이 최종 성능에 더 중요하다는 것을 보여준다. 따라서 본 연구는 향후 LLIE 기술이 후속 태스크와의 시너지를 고려해야 한다는 방향성을 제시한다.

1. 서론

오늘날 CCTV 영상은 범죄 예방 및 돌발 상황 발생 시 신속한 조치를 위해 중요한 기여를 하고 있다. 특히 사람 탐지(person detection)는 공공안전 분야에서 핵심적인 과제로, 다양한 상황에서 신뢰성 있는 성능을 요구한다. 하지만 실제 CCTV 환경은 조도가 낮고, 센서의 한계로 인한 노이즈가 존재하며, 해상도 역시 충분하지 않은 경우가 많다. 이러한 한계로 인해 CCTV 영상에서 기존의 사람 탐지 기법은 최적의 성능을 발휘하기 어렵다.

이러한 문제를 해결하기 위해, 저조도 이미지 개선(Low-Light Image Enhancement, LLIE) 기술을 탐지 모델의 전처리 단계로 활용하는 enhance-then-detect 접근 방식이 널리 연구되어 왔다. 하지만 대부분의 LLIE 기술은 주로 사람의 시각적 만족도를 높이는 데 최적화되어 있으며, 이는 탐지 모델의 기계적 인식과는 반드시 일치하지 않는다. 그 결과, LLIE 로 전처리된 입력은 탐지 성능을 일관되게 개선하지 못하거나, 불필요한 인공적 패턴을 생성하여 오히려 객체 탐지 성능을 저하시키는 원인이 될 수 있다. 이러한 한계는 사람이 보기 좋은 이미지와 탐지 모델이 이해하기 좋은 이미지 간의 근본적 괴리에서 비롯된다.

본 연구에서는 이러한 한계를 분석하고, 어떠한 LLIE 기술이 객체 탐지 성능에 실질적으로 기여하는가? 라는 질문에 답하고자 한다. 이를 위해 EnlightenGAN [1], Zero-DCE [2] 등 각기 다른 접근 방식으로 설계된 최신 LLIE 모델들을 대표적인 탐지 모델인 YOLO(You Only Look Once)[3]와 결합하여, 저조도 환경에서의 사람 탐지 성능을 체계적으로 비교 및 분석한다.

나아가 본 연구는 단순히 최적의 조합을 찾는 것에 그치지 않고, LLIE 모델의 결과물이 갖는 밝기, 대비, 노이즈와 같은 다양한 시각적 속성이 탐지 성능에 미치는 영향을 심층적으로 분석한다. 이를 통해 향후

기계 지각에 최적화된 이미지 개선 기술의 발전 방향을 제시하고, 공공 안전 분야에서 실질적인 가이드라인을 제공하는 것을 목표로 한다.

2. 관련 연구

CCTV 영상에서의 신뢰성 있는 사람 탐지는 열악한 조명 조건으로 인한 기술적 난제에 직면해 있다. 이를 해결하기 위해 다양한 저조도 이미지 개선(Low-Light Image Enhancement, LLIE) 기법이 제안되어 왔다. 초기 접근은 Retinex 이론[4][5]에 기반하여 조명 성분과 반사 성분을 분리함으로써 밝기와 대비를 복원하는 방식에 집중하였다. 최근에는 딥러닝의 발전과 함께 데이터로부터 직접 개선 전략을 학습하는 모델들이 등장하며 성능 향상을 주도하고 있다.

본 연구에서 고려하는 대표적인 LLIE 기법들은 각각 서로 다른 접근방식을 통해 문제를 해결한다. 예를 들어 Zero-DCE[2]는 비지도 학습 기반으로 저조도 보정을 수행한다. 이는 레퍼런스 데이터 없이도 조명 맵을 추정하여 이미지를 개선하는 방법이다. 반면 EnlightenGAN[1]는 생성적 적대 신경망(GAN)[6]을 활용하여 시각적으로 인상적인 저조도 개선 결과를 생성한다. 이 외에도 HVI-CIDNet[7]는 Retinex 이론을 기반으로 구축한 신경망 모델을 통해 이미지의 색상과 디테일을 효과적으로 복원한다. 그러나 이러한 기법들의 공통적인 목표는 인간 시각에서의 품질 향상에 최적화되어 있다는 점이다.

LLIE 기법을 객체 탐지 모델의 전처리기로 활용하는 enhance-then-detect 전략은 직관적 해결책으로 널리 주목받아 왔다. 그러나 여러 연구들은 LLIE 가 탐지 성능을 항상 보장하지 않으며, 오히려 노이즈 증폭이나 GAN 기반 기법의 인공적 패턴으로 인해 성능 저하가 발생할 수 있음을 보고한다[8]. 이는 곧 사람이 보기 좋은 이미지와 기계가 보기 좋은 이미지간 괴리가 있음을 보여준다.

이러한 문제를 완화하기 위해 end-to-end 방식으로 개선과 탐지를 공동 학습하는 시도도 제안된 바 있다[9][10]. 그러나 기존 연구들은 주로 특정 LLIE 기법을 적용했을 때의 성능 변화만 보고하거나, 새로운 복합 구조를 설계하는 데 초점을 맞추는 경향이 있었다. 또한 이들은 LLIE 기법의 특성(밝기, 대비, 노이즈 등)이 탐지 성능에 어떠한 영향을 미치는지에 대한 체계적인 비교, 분석이 부족하다.

따라서 본 연구는 이러한 연구 격차를 해소하고자, 서로 다른 접근 방식을 지닌 최신 LLIE 모델들을 대표적인 탐지 모델(YOLO)[3]과 결합하여, 저조도 환경에서의 성능 변화를 정량적으로 분석하고 그 원인을 규명하는 데 목적이 있다.

3. 제안 모델

본 연구는 저조도 CCTV 환경에서 발생하는 사람 탐지율 저하 문제에 대응하기 위해 순차적인 2 단계 방법론을 제시한다. 먼저, 여러 저조도 개선 모델을 후보군으로 설정하고, 각각의 후보 모델을 전처리 단계로 적용하여 YOLO[3] fine-tuning 한다. 이후, 모든 조합에 대해 최종 성능을 비교 및 분석함으로써, 각 LLIE 기법이 탐지 성능에 미치는 영향을 체계적으로 평가한다.

3.1. 최적의 저조도 개선 모델 선정

기존의 저조도 이미지 개선(LLIE) 연구는 주로 인간의 시각적 품질 향상에 초점을 맞추어 왔기 때문에, 개선된 이미지가 실제 객체 탐지 성능을 직접적으로 향상시킨다는 보장은 없다.

따라서 본 연구에서는 YOLO 탐지 성능에 미치는 효과를 정량적으로 분석하기 위해, 대표적인 LLIE 모델인 EnlightenGAN [1], Zero-DCE [2], HVI-CIDNet[7]을 후보군으로 설정하였다. 각 모델은 공식적으로 배포된 사전학습 가중치를 사용하였으며, HVI-CIDNet 의 경우, LOLv1/wo_perc.pth 가중치를 채택하였다. 이때 내부 파라미터인 alpha 값과 gamma 값은 기본 설정값인 1 로 설정하였다. 이 가중치는 실제 저조도 데이터셋(LOLv1)으로 학습되어 현실적인 노이즈 및 왜곡 패턴 처리에 강점을 가진다. 또한 탐지 성능에 불필요한 인공적 질감(artifact)을 유발할 수 있는 Perceptual Loss 를 학습 과정에서 제외하여(wo_perc), 객체의 본질적인 구조적 정보를 최대한 보존할 수 있다고 장점을 가진다.

실험은 원본 저조도 데이터셋과 각 LLIE 모델로 개선된 세 가지 데이터셋을 대상으로 동일한 YOLO 아키텍처를 독립적으로 fine-tuning 하는 방식으로 수행된다. 이후 모든 조합에 대해 성능 지표(mAP)를 비교하고, 이를 통해 LLIE 모델이 탐지 성능에 미치는 영향을 체계적으로 평가한다.

3.2. 객체 탐지 모델: YOLO 기반 파인튜닝

본 연구의 객체 탐지기로는 YOLO(You Only Look Once) 계열 모델을 채택하였다. YOLO 는 실시간에 가까운 처리 속도와 높은 정확도를 갖춘 1-stage 탐지기로, 신속한 대응이 요구되는 CCTV 환경에 합하다.

본 연구에서는 YOLO 모델을 저조도 환경에 특화되도록 fine-tuning 하였다. 특히 탐지 대상을 사람(person)으로 한정한 단일 클래스 학습(single-class training)을 수행함으로써, 모델이 특정 객체의 특징을 집중적으로 학습하도록 유도하였다. 이를 통해

저조도 CCTV 환경에서의 사람 탐지 성능을 효과적으로 향상시키고자 한다.

4. 실험 및 결과

4.1. 실험 데이터셋 (Experimental Dataset)

본 연구의 실험에는 LLVIP (Visible-Infrared Paired Dataset for Low-light Vision)[11] 데이터셋을 활용하였다. LLVIP 는 실제 야간 환경과 유사한 저조도 조건에서 촬영된 고품질 RGB 이미지를 제공하는 공개 데이터셋으로, CCTV 환경에서의 사람 탐지 성능을 검증하기에 적합하다.

저조도 조건에 대한 모델의 강건한 성능을 집중적으로 평가하기 위해, LLVIP 훈련 데이터 12,025 장 전체의 밝기(0=완전 암부, 1=완전 광부)를 분석했고, 그중 가장 어두운 하위 3,000 장의 이미지를 선별하여 이번 실험을 위한 데이터셋으로 최종 구축하였다.

4.2. 비교 실험 구성 및 학습 설정

저조도 이미지 개선(LLIE) 기법이 YOLO[3] 기반 탐지 모델의 성능에 미치는 영향을 정량적으로 비교 분석하기 위해, 다음과 같이 네 가지 조건의 입력 데이터셋을 구성하여 실험을 진행하였다.

1. 원본 저조도 이미지
2. EnlightenGAN[1]으로 개선된 이미지
3. Zero-DCE[2]로 개선된 이미지
4. HVI-CIDNet[7]으로 개선된 이미지

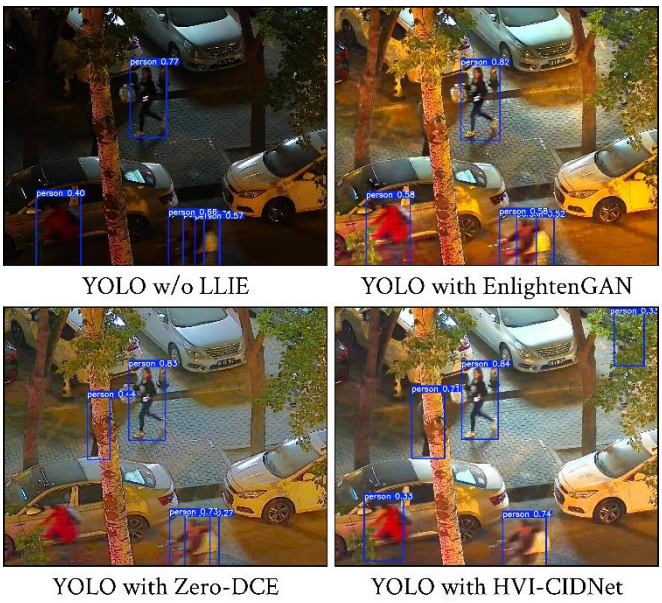
공정한 비교를 위해, 각 데이터셋에 대해 YOLO 탐지 모델을 동일한 조건 하에 독립적으로 미세 조정(fine-tuning)하였다. 모든 모델은 동일한 사전 학습 가중치(yolo11n.pt)에서 초기화되었으며, 100 epoch 동안 학습을 진행하였다. 이미지 입력 크기는 640x640 으로 고정하였다.

최적화 설정은 초기 학습률(learning rate) 0.003, warmup_epochs 는 3, weight_decay 는 0.0007 로 설정하였다. 저조도 영상의 특성을 고려하여 HSV 색 공간 증강(h=0.02, s=0.3, v=0.3)과 함께 Mosaic(0.1) 및 Affine 변환 (degrees=2, translate=0.1, scale=0.2)을 모든 실험에 일관되게 적용하였다. 이러한 통제된 실험 설계를 통해, 각 LLIE 전처리 기법이 최종 사람 탐지 성능(AP)에 미치는 영향을 정밀하게 비교 분석하였다.

<표 1> 저조도 개선(LLIE) 기법에 따른 YOLO 탐지 성능 비교

Method	Precision	Recall	AP50	AP50-95
YOLO w/o LLIE	0.918	0.813	0.884	0.475
YOLO with EnlightenGAN	0.940	0.779	0.879	0.480
YOLO with Zero-DCE	0.916	0.824	0.824	0.480
YOLO with HVI-CIDNet	0.940	0.812	0.899	0.481

위의 <표 1>은 저조도 개선 모델별 YOLO 성능을 Precision, Recall, AP50, AP50-95 지표로 정리한 결과이다. 표에서 확인할 수 있듯이, 적용된 저조도 개선 모델에 따라 최종 탐지 성능의 차이를 확인할 수 있다.. HVI-CIDNet 기반 YOLO 모델이 원본 저조도 이미지와 비교했을 때 AP50-95 점수는 0.475 에서 0.481 로, AP50 점수는 0.884 에서 0.899 로 상승하여 모든 지표에서 가장 안정적이고 균형 잡힌 성능 향상을 보였다. 반면 Zero-DCE 는 재현율(Recall)을 소폭 향상시켰으나 다른 주요 지표에서는 유의미한 개선은 관찰되지 않았다. EnlightenGAN 의 경우 높은 정밀도(Precision)를 달성한 반면 재현율은 오히려 감소하는 상충 관계를 보여 탐지 성능이 불안정함을 확인할 수 있었다.



(그림 1) 저조도 개선(LLIE) 기법에 따른 YOLO 탐지 결과.

또한 (그림 1)에서 볼 수 있듯이, HVI-CIDNet 을 적용한 YOLO 모델은 다른 모델들이 탐지에 실패하는 영역의 객체까지 일관되게 포착하여 강건한 성능을 입증하였다.

5. 결론

본 연구는 어떤 저조도 개선(LLIE) 기술이 객체 탐지 성능에 실질적으로 기여하는가? 라는 질문에서 출발하였다. EnlightenGAN[1], Zero-DCE[2], HVI-CIDNet[7] 각기 다른 방식으로 설계된 모델들을 YOLO 와 결합하여 체계적으로 분석한 결과, 객체의 구조적 정보를 가장 충실히 보존하는 HVI-CIDNet 이 탐지 성능을 가장 효과적으로 향상시키는 모델임을 입증했다.

이러한 성능 차이는 각 모델의 최적화 목표가 근본적으로 다르기 때문으로 해석된다. 시각적 만족도를 높이는 데 집중하는 생성형 모델(EnlightenGAN)은 때때로 탐지에 필수적인 객체의 미세한 질감을 변형시켜 오히려 재현율(Recall) 하락을 유발했다. Zero-DCE 와 같이 단순한 명암 조정(lightness adjustment) 방식을 사용하는 모델은 재현율(Recall)을 소폭 향상시키는 데 그쳤으며, 다른 주요 지표에서는 유의미한 전체 성능 개선이 관찰되지 않아 제한적인 효과를 보였다. 이를 통해 사람의 눈에 자연스러운 이미지와 기계 지각에 유리한 이미지는 다를 수 있다는 핵심적인 시사점을 확인했다. 반면 HVI-CIDNet 은 인공적인 질감 생성을 최소화하고 객체의 원본 윤곽선을 보존함으로써, YOLO[3]가 객체를 판단하는 데 사용하는 핵심 특징(feature)들을 효과적으로 유지했다.

결론적으로, 본 연구는 단순히 보기 좋은 이미지를 만드는 것을 넘어 탐지 모델 친화적인 개선 방식이 최종 성능에 결정적인 영향을 미친다는 것을 명확히 보여준다. 이는 향후 저조도 개선 기술이 시각적 품질 향상에만 머무르지 않고, 후속으로 이어질 컴퓨터 비전 태스크(downstream task)와의 시너지를 고려하는 방향으로 발전해야 함을 제안한다. 또한 공공 안전 분야의 실무자들에게는 LLIE 모델을 시스템에 도입할 때, 시각적 데모만으로 성능을 판단할 것이 아니라 최종 목표인 탐지 성능을 기준으로 모델을 직접 평가하고 선정해야 한다는 실질적인 가이드라인을 제공한다.

감사의 글

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학 ICT 연구센터육성지원사업(IITP-2025-RS-2020-II201789)과 인공지능융합혁신인재양성사업(IITP-2025-RS-2023-00254592), 한국연구재단의 지원(No. RS-2025-00556289)으로 수행되었음.

참고문헌

[1] Jiang, Yifan, et al. “Enlightengan: Deep light enhancement without paired supervision.” IEEE Transactions on Image Processing, vol. 30, pp. 2340-2349, 2021.

[2] Guo, Chunle, et al. “Zero-reference deep curve estimation for low-light image enhancement.” Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020.

[3] Redmon, Joseph, et al. “You only look once: Unified, real-time object detection.” Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[4] Land, Edwin H., and John J. McCann. “Lightness and retinex theory.” Journal of the Optical society of America 61.1, 1-11, 1971.

[5] Jobson, Daniel J., Zia-ur Rahman, and Glenn A. Woodell. “A multiscale retinex for bridging the gap between color images and the human observation of scenes.” IEEE Transactions on Image processing 6.7, 965-976, 1997.

[6] Goodfellow, Ian J., et al. “Generative adversarial nets.” Advances in neural information processing systems 27 2014.

[7] Yan, Qingsen, et al. “Hvi: A new color space for low-light image enhancement.” Proceedings of the Computer Vision and Pattern Recognition Conference. 2025.

[8] Hsu, Ling-Yuan. “AI-assisted deepfake detection using adaptive blind image watermarking.” Journal of Visual Communication and Image Representation 100, 104094, 2024.

[9] Guo, Haifeng, Tong Lu, and Yirui Wu. “Dynamic low-light image enhancement for object detection via end-to-end training.” 2020 25th International Conference on Pattern Recognition (ICPR). IEEE, 2021.

[10] Yin, Xiangchen, et al. “Pe-yolo: Pyramid enhancement network for dark object detection.” International conference on artificial neural networks. Cham: Springer Nature Switzerland, 2023.

[11] Jia, Xinyu, et al. “LLVIP: A visible-infrared paired dataset for low-light vision.” Proceedings of the IEEE/CVF international conference on computer vision. 2021.