An Integrated Deep Learning Architecture for Illumination-Aware Object Detection

Mostafa Habibi, Mehrdad Nourani

Predictive Analytics & Technologies Laboratory
Department of Electrical & Computer Engineering
The University of Texas at Dallas
Richardson, Texas 75080, USA

Outline

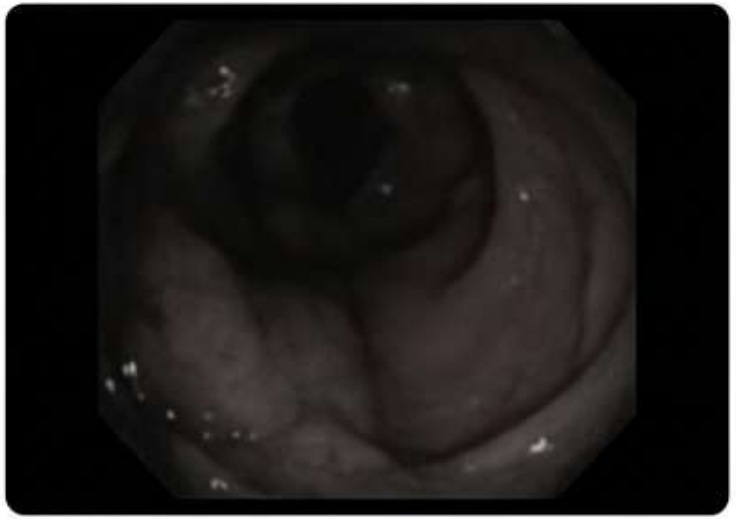
- Motivation
- Low Light Image Enhancement (LLIE)
- Prior Works
- Proposed Methodology
  - Customized Illumination Enhancer
  - Customized Object Detector
  - Three-Step Training Strategy
- Experimental Results
- Conclusion

Motivation

- Low-light images
  - Poor visibility, Low contrast, and High noise
- Computer vision models become less accurate and less reliable

Autonomous Driving: Missed pedestrians or cyclists due to low contrast.

Surveillance: Security breaches and false alarms from noisy feeds.



Medical Imaging: Diagnostic errors in endoscopy automated polyp identification.

Low-Light Image Enhancement (LLIE)

- LLIE is a preprocessing step that
  - Restores brightness
  - Improves contrast
  - Reduces noise
  - Preserves textures and colors

- LLIE improves downstream tasks
  - Recovering lost features
  - Producing stronger image

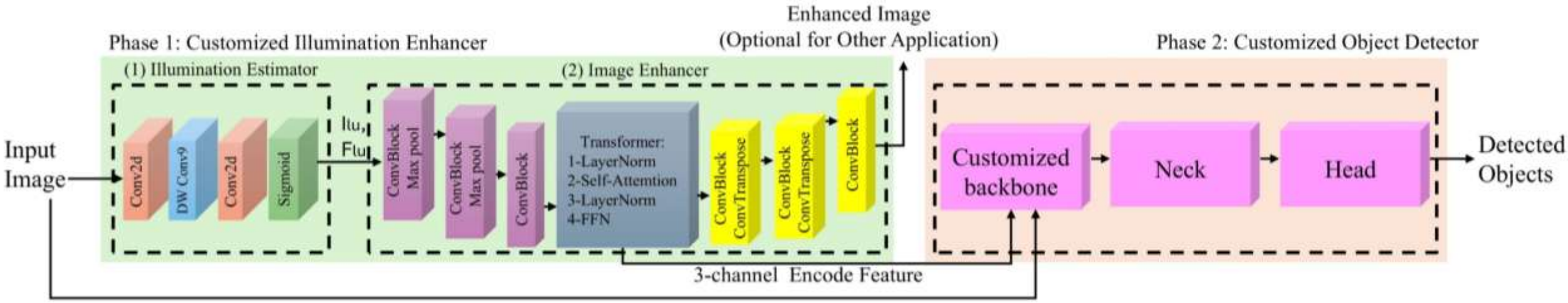
Prior Work on LLIE

Strategy	Main Idea	Representative Methods	Strength	Limitations
Traditional	Handcrafted enhancement (contrast/illumination adjustment)	CLAHE, MMBEBHE, SSR, MSR, LIME	Simple, fast, no training	Noise amplification Halo artifacts Weak generalization
Machine Learning	Learn enhancement parameters from data	Color estimation, sparse representation, Gaussian process, Training Enhance (RL + MCTS)	More adaptive than traditional methods	Scalability issues Struggles in extreme low-light nonlinear degradation
Deep Learning LLIE	Learn low→normal mapping + illumination-aware restoration	LLNet, MBLLEN, MIRNet, RetinexNet, KinD, URetinex-Net, RetinexFormer, LLFormer	Strong performance, Better feature learning, Global context (Transformers)	High compute cost Weak dark-region features Noise/color distortion Video instability
End-to-End Low-Light Detection	Integrate enhancement inside detector	IA-YOLO, Dark-YOLO, NLE-YOLO	Joint optimization for detection in low light	Compute overhead Robustness issues in very dark/mixed-light scenes

- A robust, real-time system that integrates low-light detection with stable feature mapping is still highly in demand.

Proposed Methodology

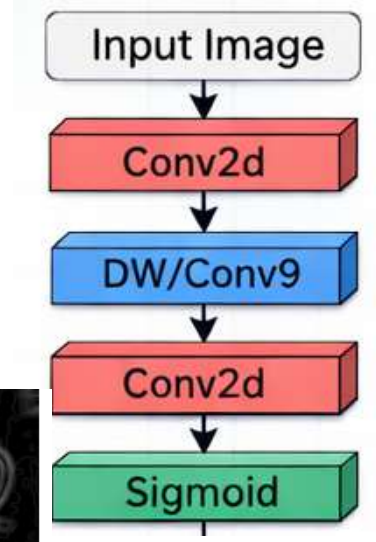
- We propose an Illumination-Aware Detection framework that performs
  - Phase 1: Low-light image enhancement
  - Phase 2: Object detection



Phase1: Customized Illumination Enhancer (Retinex)

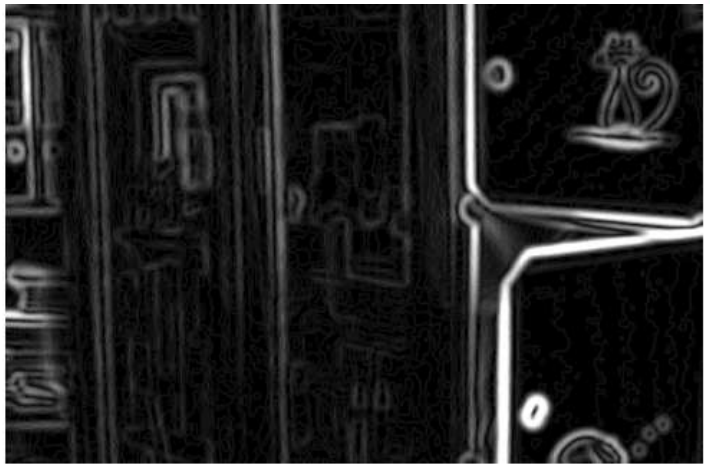
(1) Illumination Estimator

- Predicts a per-pixel illumination map $\hat{L}$ and Extracts high-level illumination features F_{lu}.
  - Light-adjusted Image ($I_{lu} = I \odot \hat{L}$): Normalizes dark regions using the predicted map.
  - Illumination Feature Map (F_{lu}): Encodes global exposure and local variations.



Original

per-pixel Illumination Map $\hat{L}$



Illumination Feature Map F_{lu}

Phase1: Customized Illumination Enhancer (U-Net Transformer)

(2) Image Enhancer

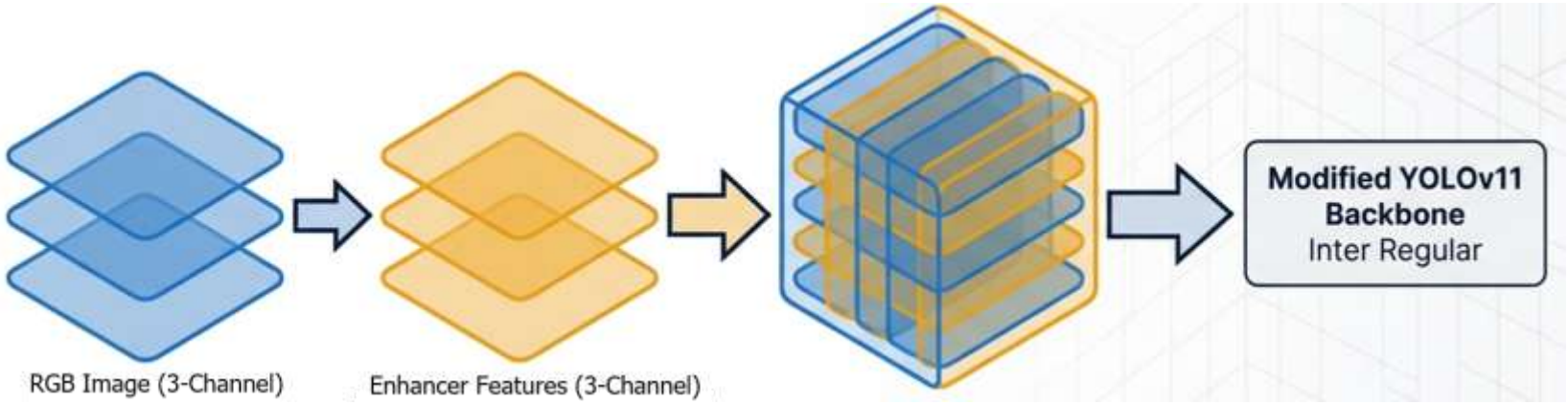
- Combines U-Net skip connections with a Transformer.
- Generates the final enhanced image $\hat{L}$ by aligning recovered structure with corrected lighting.
- Training loss function

$$L_{enh} = \| \hat{I} - I_{gt} \| + \lambda_1 \| \nabla \hat{L} - \nabla L_{gt} \|$$

Forces the enhanced image $\hat{I}$ to be close to the ground-truth normal-light image I_{gt}.	λ_1 controls the tradeoff between "match the target image" and "keep illumination smooth."	Forces the illumination map to be smooth and realistic.

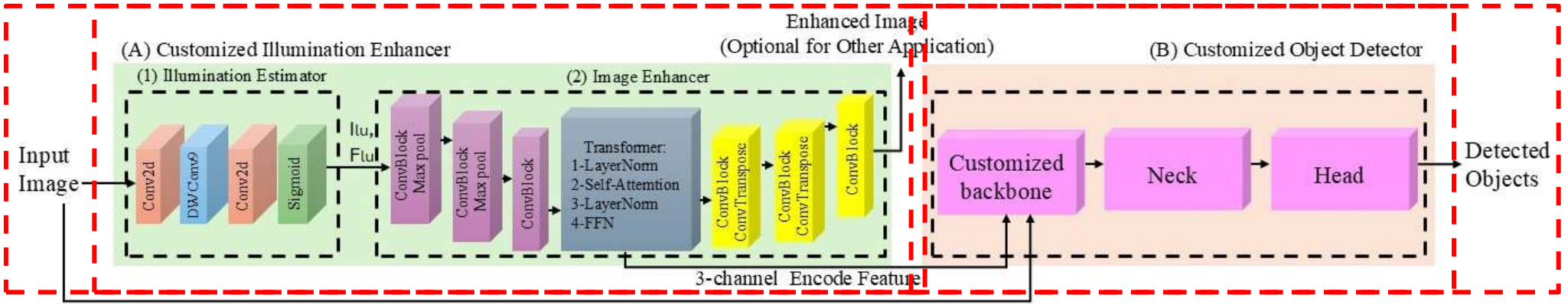
Phase 2: Customized Object Detector (Based on YOLOv11)

- YOLOv11 architecture: Backbone → Neck → Detection Head
- Modified Backbone: 6 channels instead of 3



- Detector learns from
  - **Appearance + Lighting cues**
- Neck + Head kept the same to preserve
  - Efficiency and Multi-scale feature fusion

Three-Step Learning Strategy



Step 3

- **COCO dataset (~330K images, 80 categories)**
- **Enhancer networks are trained only as frozen feature extractors** (no weight updates)
- **Customized YOLO detector learns object detection**
- Output
  - 3-channel illumination-aware encode feature map from the transformer module
  - Enhanced image produced (useful for monitoring)

Two Examples of Enhancement & Detection

Original Image

Our Methodology

Our Enhancer

Experimental Results – Enhancer Module

- Evaluation on the public LOL (Low-Light) dataset
  - A paired low-light benchmark with 500 image pairs
  - Used 10-fold cross-validation on the training set for performance evaluation

- Metrics:
  - Model Parameters (Memory Efficiency)
  - PSNR (Peak Signal-to-Noise Ratio)
  - SSIM (Structural Similarity)

Method	Parameters (M)	PSNR (dB)	SSIM
RetinexNet	0.83	16.77	0.56
KinD	4.10	20.87	0.80
RetinexFormer	8.90	23.60	0.86
Our Enhancer	**1.18**	**22.41**	**0.84**

Experimental Results – Joint Training

- Evaluated on the public ExDARK dataset
  - 7,363 real low-light images collected across 10 lighting conditions with object annotations for detection
- On joint training, our platform learns
  - Illumination-invariant features from real low-light data
  - More accurate detection in challenging lighting

Method	Parameters (M)	mAP (%)	Key Innovation
IA-YOLO	7.2	63.8	Image-adaptive dual-branch fusion
NLE-YOLO	7.5	68.5	YOLOv5 with fusion enhancement
Dark-YOLO	6.8	71.3	Multi-attention integration
Proposed	**3.78**	**69.1**	**Illumination-aware joint training**

Experimental Results – Latency and Computational Cost

- Evaluation of end-to-end pipeline across 1,000 images with batch size = 1

Metric	Value
Avg. Latency (ms/img)	123.25
Avg. Enhancement (ms/img)	63.41
Avg. Detection (ms/img)	59.84
Latency Min (ms)	21.26
Latency Std (ms)	172.09
Latency Max (ms)	577.5

Metric	Value
Enhancer FLOPs (G)	1.974
YOLO FLOPs (G)	2.912
Total FLOPs (G)	4.886
Peak GPU Memory (GB)	2.24
Avg. FPS (derived)	8.11

Experimental Results – Ablation Study

- We compared three configurations on the ExDark test set.
  - **Baseline**: Standard YOLOv11 with official weights (no low-light training)
  - **No Enhancer**: YOLOv11 trained on low-light data but with the 3-channel illumination features disabled
  - **Proposed (Full):** The complete pipeline using fused RGB and illumination-aware features

Configuration	F1 Score (%)	Key Takeaway
YOLOv11 (Pretrained)	7.52	Too strict, fails to detect objects in Low-light (3.91% Recall)
YOLOv11 (Fine-tuned, No Enhancer)	57.32	Better, but still misses 60% of objects (40.18% Recall)
Proposed (Full Pipeline)	**83.54**	**Balance Precision (85.62%) and Recall (81.55%)**

Conclusions

- Highlights
  - End-to-end illumination-aware detection framework that integrates image enhancement + object detection in a real-time CNN-based architecture
  - Strong results on LOL dataset, COCO, and ExDark with **competitive accuracy** and **few parameters** suitable for **real-time detection** in low light environments
- Limitations
  - Has not been fully tested on harder real-world image problems (noisy sensor, motion blur, defocus blur, resolution loss, overexposure/high glare)
- Future work
  - Extend to condition-adaptive enhancement and perform cross-dataset evaluation.

Thanks for Your Attention!

Mostafa Habibi, Mehrdad Nourani
Department of ECE
The University of Texas at Dallas
{mostafa.habibidehsheikhi, nourani}@utdallas.edu