# EE243 Final Project

**Adolfo Bugarin** **Anish Kale** **Aryan Ramachandra**
Computer Science and Engineering
University of California, Riverside
{abuga006, akale014, arama081}@ucr.edu

## Abstract

*Urban surveillance systems frequently operate in poor lighting, resulting in dark, low-contrast footage that degrades the performance of object detection models. Traditional low-light enhancement methods either require paired training data or introduce distortions, making them unsuitable for real-time deployment in smart cities. Using Zero-DCE to pre-process low-light surveillance footage and improve detection accuracy will provide reliable performance gains in nighttime urban monitoring scenarios without the need for ground truth references or heavy computation. In this project we 1) replicate the results using the provided implementation in the paper; 2) test the model on new datasets including images from low-light surveillance systems; 3) provide a thorough analysis.*

## 1. Introduction

In recent years, smart cities have increasingly relied on surveillance systems for public safety, traffic monitoring, and urban management. However, these systems often operate under challenging lighting conditions, particularly during nighttime, resulting in dark, low-contrast footage that significantly degrades the performance of downstream computer vision models such as object detectors and trackers. Enhancing such low-light images is therefore critical to maintaining reliable surveillance operations. Traditional image enhancement methods, including histogram equalization, Retinex-based models, and supervised deep learning approaches, have shown limited success in these scenarios. Many of these techniques either introduce unnatural distortions or rely heavily on paired training datasets containing both low-light and corresponding well-lit reference images — which are difficult, if not impossible, to collect at scale in real-world urban settings.

The Zero-Reference Deep Curve Estimation (Zero-DCE) model, introduced by Li et al. [1], addresses these limitations by proposing a fully unsupervised learning framework for low-light image enhancement. Instead of relying on paired datasets, Zero-DCE learns enhancement curves directly from unpaired low-light images using a carefully designed set of self-supervised loss functions. Its lightweight architecture allows for real-time processing, making it particularly well-suited for deployment in surveillance pipelines where both speed and quality are crucial.

In this project, we build upon the Zero-DCE framework with the goal of validating its suitability for urban surveillance applications. Specifically, we (1) replicate the results presented in the original Zero-DCE paper, (2) extend the evaluation to additional low-light surveillance datasets such as LOL-V2 Real and DarkFace, and (3) conduct both quantitative and qualitative analyses of the enhancement results to assess its practical performance under real-world conditions.

## 2. Related Works

Low-light image enhancement has been an active area of research, with approaches ranging from traditional image processing to advanced deep learning-based models.

Early methods such as Histogram Equalization (HE) [2] and Contrast Limited Adaptive Histogram Equalization (CLAHE) [3] attempted to redistribute pixel intensity values to improve brightness and contrast. However, these methods often fail to handle complex lighting variations and may introduce artifacts or noise amplification.

Retinex-based methods, inspired by human visual perception, decompose images into illumination and reflectance components to separately process lighting conditions. Representative works include Single-Scale Retinex (SSR) [4] and Multi-Scale Retinex (MSR) [5], though these approaches struggle with color distortion and lack of robustness in extreme low-light scenes.

With the rise of deep learning, supervised models such as EnlightenGAN [6] and Retinex-Net [7] have demonstrated significant improvements by learning complex mappings between paired low-light and normal-light images. However, the dependency on paired data limits their scalability and real-world applicability.

To overcome these challenges, unsupervised and zero-reference methods have emerged. The Zero-DCE model [1] proposes a curve estimation approach that learns enhancement mappings directly from low-light images using self-supervised loss terms. More recently, methods like KinD++ [8] have further explored decomposition-based frameworks while maintaining data efficiency.

# 3. Method and Experiments

## 3.1. Model Overview

In this project, we adopt the Zero-Reference Deep Curve Estimation (Zero-DCE) model, originally proposed by Li et al. [1], to address the challenge of low-light image enhancement for urban surveillance systems. The model directly predicts a pixel-wise parametric enhancement curve, avoiding the need for paired training data, which is often infeasible to collect in real-world scenarios such as street-level surveillance.

Unlike conventional deep image enhancement pipelines that directly predict an enhanced output image, Zero-DCE instead predicts curve parameters $\alpha$, which are applied to each pixel using the following quadratic enhancement function:

$$LE(I(x); \alpha) = I(x) + \alpha \cdot I(x) \cdot (1 - I(x))$$

where $I(x)$ denotes the pixel intensity normalized to $[0, 1]$. This formulation allows for controlled, iterative enhancement without introducing excessive distortions. In our implementation, the network predicts 24 curve parameters which are applied across 8 iterative enhancement steps.

The network architecture itself is a fully convolutional 7-layer design (DCE-Net), which is both lightweight and computationally efficient. This is particularly well-suited for real-time deployment in surveillance applications, where resource constraints and latency are significant considerations.

## 3.2. Loss Functions

Because no paired ground-truth images are available for supervision, the network relies on four carefully designed self-supervised loss functions that capture human-perceived image quality:

- **Exposure Control Loss:** Penalizes deviations from a desired mean brightness level across the output image.
- **Color Constancy Loss:** Promotes color balance by reducing discrepancies between RGB channels. This loss was omitted in grayscale evaluations where color channels are not present.
- **Spatial Consistency Loss:** Encourages local smoothness while preserving structure to avoid over-enhancement of specific regions.
- **Illumination Smoothness Loss:** Enforces smooth transitions across neighboring pixels in the predicted enhancement curves, suppressing noise amplification.

These losses jointly guide the network toward generating perceptually pleasing enhanced images that retain natural textures, suppress noise, and maintain balanced exposure.

## 3.3. Training Procedure

The model was adapted from the official PyTorch implementation provided by the authors of Zero-DCE [12], with GPU acceleration enabled. We made modifications to enable evaluation on additional datasets. The following training hyperparameters were preserved to match the original setup:

- Optimizer: Adam
- Learning rate: $1 \times 10^{-4}$
- Batch size: 8
- Image resolution: $512 \times 512$
- Epochs: Until convergence based on loss stabilization

By closely following the original training configuration, we aimed to faithfully reproduce the results while also evaluating the model's generalization performance on new low-light surveillance datasets.

## 3.4. Datasets and Evaluation Setup

To evaluate both replication accuracy and domain generalization, we trained and tested the model under multiple dataset conditions:

- **LOL Dataset [1]:** The original dataset used in Zero-DCE, consisting of paired low-light and normal-light images. Only low-light images were used during training to preserve the unsupervised nature of the model.
- **LOL-V2 Real Dataset:** A newer version containing a more diverse set of real-world low-light scenes, offering a more challenging testbed for generalization.
- **DarkFace Dataset [9]:** A dedicated surveillance dataset with low-light facial images captured under urban nighttime conditions. This dataset closely reflects the kinds of practical surveillance scenarios that motivated this project.

## 3.5. Experiment Design

We structured our experiments into three distinct phases:

(a) Original (Low-Light) (b) LOL-Trained Result (c) Original Paper Dataset Result (d) DarkFace-Trained Result

Figure 1. Qualitative enhancement results on a sample low-light surveillance image using models trained on different datasets.

- **Replication Experiment:** The model was trained using the LOL dataset and tested on the test split of the dataset used in the original Zero-DCE paper. This served as a sanity check to validate whether our PyTorch reimplementation could replicate the original reported performance.

- **Cross-Dataset Generalization Experiment:** The model trained on LOL was directly applied to the DarkFace dataset, simulating a real-world deployment scenario where the model encounters unseen surveillance imagery captured under different lighting and scene conditions.

- **In-Domain Performance Experiment:** The model was both trained and tested on the LOL dataset itself to establish an upper-bound reference for enhancement performance.

Each of these experiments was evaluated both qualitatively, by visually inspecting enhancement quality (see Figure 1), and quantitatively, by calculating the Peak Signal-to-Noise Ratio (PSNR) on enhanced outputs.

### 3.6. Evaluation Metrics

We primarily relied on PSNR as a quantitative measure to compare image reconstruction quality across datasets. Higher PSNR values typically indicate better reconstruction fidelity with less noise and fewer enhancement artifacts.

- For in-domain LOL evaluation, the model achieved an average PSNR of 27.94 dB.

- For replication on the dataset in the original paper, we obtained an average PSNR of 27.89 dB.

- For generalization to the DarkFace dataset, we observed an average PSNR of 27.87 dB.

The small variance across these results suggests strong generalization capability of Zero-DCE even when applied to surveillance datasets it was not explicitly trained on.

### 3.7. Limitations of Grayscale Inputs

While we experimented with grayscale inputs for Zero-DCE, this adaptation presents several challenges. The original Zero-DCE framework is specifically designed for color image enhancement using pixel-wise curve estimation in RGB space. Converting the input to grayscale removes all chromatic information, fundamentally altering the learning task. As a result, the model can no longer perform color correction or color-based noise suppression, reducing its capacity to learn meaningful enhancement beyond contrast and brightness.

Additionally, Zero-DCE's pretrained weights and network architecture are tailored for three-channel RGB input. Switching to a single-channel input invalidates these weights and eliminates the benefits of transfer learning. Loss functions such as the color constancy loss $L_{\text{color}}$ also become irrelevant, and must be removed or replaced to avoid computational issues or ineffective gradient updates.

Moreover, there is no widely accepted grayscale-specific low-light dataset, and synthetic grayscale degradation (e.g., darkening or noise) does not accurately reflect the complexity of real-world low-light conditions. Evaluation of perceptual quality also becomes more difficult, as single-channel PSNR or SSIM may not fully capture the subjective improvements seen in enhanced images.

Unless there is a specific grayscale-driven application—such as thermal, medical, or document imagery—we recommend retaining RGB inputs to fully utilize the design and strengths of the Zero-DCE architecture.

## 4. Discussion and Conclusion

Our results demonstrate that Zero-DCE is an effective and practical solution for enhancing low-light surveillance imagery. The model achieved consistent PSNR scores across both the LOL dataset and unseen datasets like DarkFace, confirming its strong generalization capabilities. Visual inspection also revealed noticeable improvements in contrast and detail (Figure 1b), which are essential for downstream tasks such as object or face detection in nighttime surveillance.

A major advantage of Zero-DCE is its zero-reference, self-supervised training framework, which removes the need for paired data—a major bottleneck in real-world applications. Its lightweight architecture further supports real-time deployment in resource-constrained settings.

Although we experimented with grayscale variants of the model, we found them to be suboptimal. Converting RGB images to grayscale not only removes critical color information but also changes the input-output relationship that the model is designed to learn, making the training objective less meaningful. Furthermore, loss functions that rely on color information—such as the color constancy loss used in the original Zero-DCE paper—become irrelevant, and pretrained weights tuned for three-channel input cannot be reused. In addition, there is a lack of large, publicly available grayscale-only low-light datasets, making it infeasible to train and evaluate such models effectively in a real-world context.

While performance was generally strong, extreme low-light scenes with heavy noise may still challenge the model (Figure 1d). Also, PSNR alone may not fully capture perceptual quality, suggesting future work should incorporate metrics like SSIM or LPIPS, and evaluate impact on downstream vision tasks.

In summary, Zero-DCE offers a scalable, efficient enhancement solution for smart city surveillance. Future directions include tighter integration with detection pipelines and exploring adaptive or scene-aware enhancement strategies.

## References


[1] C. Li, C. Guo, J. Han, C. Y. Loy, and M. Yang, "Zero-reference deep curve estimation for low-light image enhancement," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 1780–1789. 1, 2

[2] S. M. Pizer, E. P. Amburn, J. D. Austin, et al., "Adaptive histogram equalization and its variations," *Computer Vision, Graphics, and Image Processing*, vol. 39, no. 3, pp. 355–368, 1987. 1

[3] K. Zuiderveld, "Contrast limited adaptive histogram equalization," in *Graphics Gems IV*, Academic Press, 1994, pp. 474–485. 1

[4] D. J. Jobson, Z. Rahman, and G. A. Woodell, "A multiscale retinex for bridging the gap between color images and the human observation of scenes," *IEEE Transactions on Image Processing*, vol. 6, no. 7, pp. 965–976, 1997. 1

[5] Z. Rahman, D. J. Jobson, and G. A. Woodell, "Retinex processing for automatic image enhancement," *Journal of Electronic Imaging*, vol. 13, no. 1, pp. 100–110, 1996. 1

[6] Y. Jiang, X. Gong, D. Li, et al., "EnlightenGAN: Deep light enhancement without paired supervision," *IEEE Transactions on Image Processing*, vol. 30, pp. 2340–2349, 2021. 1

[7] C. Wei, W. Wang, W. Yang, and J. Liu, "Deep retinex decomposition for low-light enhancement," in *Proc. British Machine Vision Conference (BMVC)*, 2018. 1

[8] Y. Zhang, J. Zhang, and X. Guo, "Kind++: A deeper and stronger network for low-light image enhancement," *IEEE Transactions on Image Processing*, vol. 30, pp. 5110–5123, 2021. 2

[9] S. Rakshit, "Dark Face Dataset," Available: https://www.kaggle.com/datasets/soumikrakshit/dark-face-dataset, 2019. [Accessed: Jun. 8, 2025]. 2

[10] C. S. Chan, "Exclusively Dark Image Dataset (Ex-Dark)," Available: https://github.com/cs-chan/Exclusively-Dark-Image-Dataset, 2018. [Accessed: Jun. 8, 2025].

[11] Y. Wang, "CVPR 2020 Semi-Low-Light Dataset," Available: https://github.com/flyywh/CVPR-2020-Semi-Low-Light, 2020. [Accessed: Jun. 8, 2025].

[12] C. Li, "Zero-DCE: Zero-Reference Deep Curve Estimation for Low-Light Image Enhancement," GitHub repository, https://github.com/Li-Chongyi/Zero-DCE, 2020. 2