INSTITUTE OF INFORMATION AND COMMUNICATION TECHNOLOGIES
BULGARIAN ACADEMY OF SCIENCES
CYBERNETICS AND INFORMATION TECHNOLOGIES • Volume **25**, No 4
Sofia • 2025 Print ISSN: 1311-9702; Online ISSN: 1314-4081
DOI: 10.2478/cait-2025-0038

ALEX: Automated Low-Light Enhancement eXpert for Intelligent Security Systems Using Vision Transformer

Alam Rahmatulloh[1,4], *Erna Haerani*[2], *Rohmat Gunawan*[1], *Eryan Ahmad Firdaus*[3], *Ghatan Fauzi Nugraha*[4]

[1]Departemen of Informatics, Faculty of Engineering, Siliwangi University, Indonesia
[2]Department of Information Systems, Faculty of Engineering, Siliwangi University, Indonesia
[3]Department of Informatics Engineering, University of Pertahanan, Indonesia
[4]Forensic and Security (FAST), Research Group, Siliwangi University, Indonesia
E-mails: alam@unsil.ac.id erna@unsil.ac.id rohmatgunawan@unsil.ac.id eryan.ahmad.firdaus@unhan.ac.id ghatan.fauzi.nurgraha@unj.ac.id

***Abstract**: Closed-Circuit TeleVision* (*CCTV*) *performance in low-light conditions often results in poor image quality. This study introduces Automated Low-Light Enhancement eXpert* (*ALEX*), *a new architecture that combines ViTRA with SwinIR to improve image clarity. ALEX utilizes Relative Lighten Cross-Attention* (*RLCA*) *and Relative Position Encoding* (*RPE*) *in the HVI color space to enhance light intensity and color, followed by SwinIR for depth restoration and resolution enhancement. Evaluation on benchmark datasets like LOLv1, LOLv2, and SID shows that ALEX outperforms existing methods like HVI-CIDNet and ViTRA, yielding sharper, more natural results based on PSNR, SSIM, and other metrics. Real-world CCTV tests demonstrate that ALEX improves image quality, even with dimmed or downscaled images. While the integration of SwinIR increases complexity and inference time, ALEX proves to be an effective low-light enhancement solution, offering significant potential for intelligent surveillance systems and future real-time applications on resource-constrained devices.*

***Keywords**: ALEX, Closed-circuit television, Low-light image enhancement, Relative lighten cross-attention, SwinIR.*

1. Introduction

Security quality assurance in an area can be achieved through surveillance, one of which is Closed-Circuit TeleVision (CCTV) [1]. Currently, both agencies and the public widely use CCTV as a security tool to monitor crime, which can occur at any time [2, 3]. However, unfavorable environmental conditions, such as insufficient lighting in these areas, pose a challenge for conventional CCTV [4]. Therefore, to overcome this, an integrated system with surveillance cameras that can handle low-light conditions is needed.

Several studies have used various deep learning-based models to address this issue. Ai and Kwon [5] integrated an Attention mechanism with a U-Net [6], adapted to improve image quality in critical areas, especially in dark conditions. This model is designed to improve the quality and increase the intensity of light without damaging the pixels in the image. The results of this study provide Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM), and Multiscale Structural Similarity (MS-SSIM) values of 21.20, 0.51, and 0.88, respectively, on the See In the Dark (SID) dataset. Research conducted by Qu et al. [7] proposed a new method called Double Domain Guided Network (DDNet), which is used to improve image quality in low-light conditions. DDNet uses two main modules, namely the Coarse Enhancement Module (CEM) Module which focuses on improving image color, and the Gradient Enhancement Module (GEM) Module which uses Laplacian of Gaussian (LoG) to enhance important image edge features. This model was tested on various datasets such as LOLv1, DICM, LIME, MEF, and TMDIED, which showed superior performance in improving image quality compared to other models. Bhandari et al. [8] developed a system to enhance images from nighttime surveillance cameras using infrared cameras. Images captured in low-light conditions were enhanced using enhancement techniques such as Histogram Equalization (HE), Adaptive Histogram Equalization (AHE), and Contrast-Limited Adaptive Histogram Equalization (CLAHE). After enhancement, the images were analyzed using Convolutional Neural Networks (CNNs) trained using transfer learning for object recognition. Experimental results showed that CLAHE improved object classification accuracy from 58% to 85%.

While some studies have shown improvements in the light quality of images captured by surveillance cameras, the models used are still limited to improving light intensity alone and provide minimal post-processing improvements. Image enhancement manipulates and adjusts images to make them more suitable for analysis [9]. Therefore, as an alternative solution to this problem, surveillance cameras can be directly integrated with Low-Light Image Enhancement (LLIE) models to achieve optimal light quality enhancement results. Current LLIE models have shown significant progress in improving image light quality [10]. Several LLIE models using GANs, such as EnlightenGAN [11], Wang et al.'s research [12], and Lee et al.'s research [13], have demonstrated very satisfactory light enhancement performance. This not only improves light quality but also sharpens image quality. Although it produces better image quality due to the competitive process between two networks, namely the generator and discriminator, this does not rule out the fact that GANs often experience instability and mode collapse issues [10, 14]. To address this, Transformer comes with greater use of self-attention for a more stable training process [15]. Furthermore, Transformer is more efficient with smaller datasets and can produce smoother image enhancement with finer details in LLIE tasks [10]. One transformer-based model that has a balance between improved light quality and training stability is the HVI-CIDNet (Yan et al., [16]). This model uses the Horizontal/Vertical-Intensity (HVI) color space specifically designed to address color noise and brightness artifacts in low-light images. HVI uses a polarized Hue/Saturation (HS) map and a trainable intensity function to reduce red and black artifacts. In addition, Yan et al. [16]. Developed the Color and Intensity Decoupling Network (CIDNet) that separates color and intensity processing to maximize Image enhancement results. Experiments show that this approach outperforms current methods in terms of image quality, addresses existing issues more efficiently, and produces more natural color restoration in low-light images.

Table 1. Characteristic comparison between ALEX and existing LLIE methods

Model	Enhancement type	Attention mechanism	Color space	Integration module	Test on CCTV	Notable limitation
ElightenGAN [11]	LLIE	Self-Attention	RGB	×	×	Training instability
HVI-CIDNet [16]	LLIE	Lighten Cross-Attention (LCA)	HVI	×	×	Limited spatial recovery
ViTRA [17]	LLIE	Relative Lighten Cross-Attention (RLCA)	HVI	×	×	No resolution enhancement
ALEX (our model)	Hybrid (LLIE + SR)	Relative Lighten Cross-Attention (RLCA)	HVI	SwinIR [17] integration	✓	Higher inference complexity

Based on previous research, a system integrated with CCTV cameras or surveillance cameras is needed that can enhance light intensity without destroying image detail. To meet this need, we propose developing an LLIE model combined with an image Super-Resolution (SR) model called ALEX (Automated Low-Light Enhancement eXpert). ALEX uses the LLIE model from our previous study, ViTRA [17], which is a further development of HVI-CIDNet [16] with the addition of a Relative Position Encoding (RPE) module [18] to the Lighten Cross-Encoding (LCA) module. Furthermore, to sharpen and improve image resolution, we added a SR imager process using SwinIR [19] after successfully enhancing the image's light quality. This combination is expected to create a model that can enhance light quality while improving the resolution of images captured by surveillance cameras.

Therefore, this study aims to enhance image visibility and clarity in low-light conditions by developing a hybrid model that combines illumination enhancement and spatial restoration. The main scientific contributions of this paper are as follows:

(1) We propose a new architecture, called ALEX, which integrates the ViTRA low-light enhancement framework with the SwinIR image restoration module to improve brightness and spatial detail simultaneously.

(2) We introduce a novel attention mechanism, namely Relative Lighten Cross-Attention (RLCA), that embeds Relative Position Encoding (RPE) within the Lighten Cross-Attention process to capture complex spatial relationships in the HVI colour space.

(3) We conduct extensive experiments on multiple benchmark datasets (LOLv1, LOLv2, SICE, SID, and LOL-Blur) as well as real-world CCTV data, demonstrating that ALEX achieves superior performance compared to state-of-the-art models such as HVI-CIDNet and ViTRA.

(4) We present an ablation and complexity analysis to investigate the effect of SwinIR integration on performance, computational cost, and inference time.

(5) confirming its robustness under dimmed and degraded visual conditions, thus proving its practical potential for real-world CCTV systems.

To further clarify the novelty and distinct features of the proposed model compared to existing low-light enhancement methods, Table 1 presents a characteristic comparison. The comparison highlights the architectural and methodological differences between ALEX and previous approaches.

2. Methods

In surveillance camera systems or CCTV, images captured in low-light conditions often suffer from degradation in the form of extreme darkness, high noise, and low contrast. Although methods based on Generative Adversarial Networks (GANs) have demonstrated realistic results, they are prone to training instability, visual artifacts, and high computational costs [20]. Transformer-based architectures, on the other hand, offer superior capabilities in modeling long-term spatial dependencies, but often neglect the relative positional relationships between pixels, which are crucial in Low-Light Image Enhancement (LLIE) tasks [10, 14].

To address these limitations, this study proposes ALEX, a novel architecture that extends the ViTRA framework [17] by integrating it with SwinIR [19], creating an architecture that not only enhances light quality but also reconstructs and improves image resolution. The ALEX architecture is shown in Fig. 1.

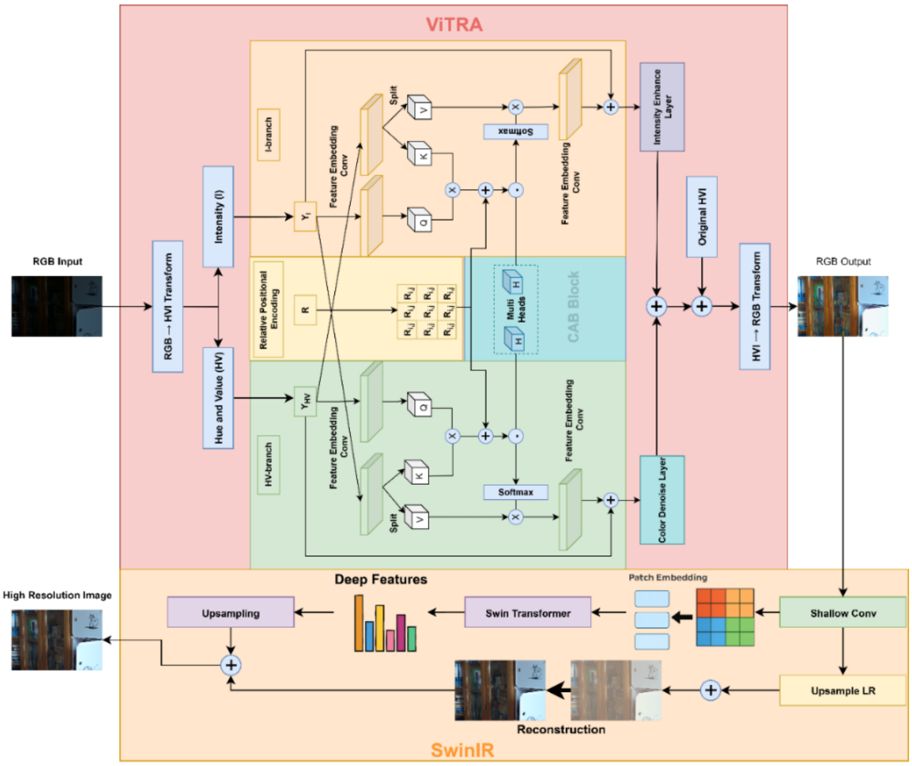

Fig. 1. The proposed ALEX architecture, drawn by the authors based on our previous ViTRA model [17] and the SwinIR framework [19]

In general, ALEX is an expansion of the ViTRA architecture built on HVI-CIDNet [16, 17], which consists of two parallel branches, namely the HV-branch (processing the Hue and Value components (color information) and the I-branch (processing the Intensity component (brightness). These two branches interact through the Lighten Cross-Attention (LCA) module [16], which is extended by

ViTRA by adding Relative Positional Encoding (RPE) [18], forming Relative Lighten Cross-Attention (RLCA) [17], which explicitly models the spatial relationship between pixels. When an image with low light intensity enters the ALEX architecture, the input image $I \in \mathbb{R}^{H \times W \times 3}$ will be converted to the HVI color space.

(1) $$\mathrm{HVI} = \mathrm{RGB} \rightarrow \mathrm{HVI}(I).$$

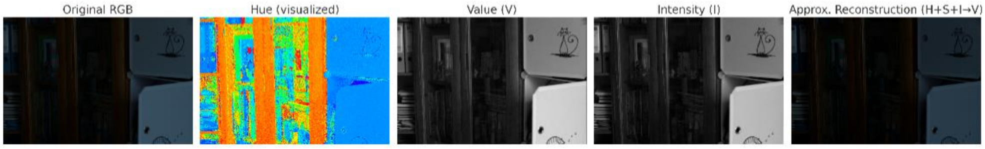

Fig. 2. Visualization of the RGB-to-HVI color space conversion showing the Hue (H), Value (V), Intensity (I), and reconstructed outputs

The RGB-to-HVI color space transformation, as depicted in Fig. 2, serves to isolate chromatic and luminance information, thereby enabling the model to process hue, value, and intensity components independently. After the conversion, the HVI component in the image is separated into $Y_{\mathrm{HV}} \in \mathbb{R}^{H \times W \times 2}$ (Hue-Value component) and $Y_{\mathrm{I}} \in \mathbb{R}^{H \times W \times 1}$ (Intensity component). These two branches will pass through the convolution layer to perform initial feature extraction using

(2) $$Q = W_Q . Y,\ K = W_K . Y,\ V = W_V . Y,$$

where W_Q, W_K, and W_V are the convolution weights to generate Query (Q), Key (K), and Value (V). Since RLCA requires RPE in its attention calculation, the relative position matrix needs to be initialized first by modeling the relative distance between pixels in the spatial grid. For example, for two pixels at positions (i, j) and (k, l), the relative distance vector is according to the next equation,

(3) $$\mathrm{Index}_{\mathrm{rel}} = (i - k + W - 1) \times (2w - 1) + (j - l + W - 1).$$

The relative position matrix $R \in \mathbb{R}^{(2W-1) \times (2W-1) \times H}$ (where W is the window size, H is the number of attention heads) stores the embedding of each relative offset. Each pair (Q, K) has a relative position index, which is used to extract spatial bias via equation

(4) $$R_{\mathrm{bias}} = R[\mathrm{Index}_{\mathrm{rel}}],$$

where R_{bias} is the relative spatial bias used to calculate the attention value on both parallel branches. The attention between branches can be calculated by entering the relative spatial bias using equation

(5) $$A = \mathrm{Softmax}\left(\frac{Q \otimes K}{\alpha H} + R_{\mathrm{bias}}\right),$$

where A is the attention value of RLCA and αH is a scaling factor based on the number of heads for numerical stability. By obtaining the value of A, the output of RLCA can be found using the next equation,

(6) $$\hat{Y} = W(V \odot A + Y),$$

where W is a linear weight matrix and $\odot$ is an element-wise multiplication. By using this RLCA module, the model can recognize complex spatial patterns such as edges, textures, and non-uniform dark areas, thereby improving the enhancement accuracy. After exiting the RLCA, the process continues by following the steps of HVI-CIDNet [16] to produce an image with improved light intensity quality.

After the initial enhancement process by the ViTRA module is completed, the resulting image (in RGB color space) does not immediately become the final output. Instead, the image enters the transformer-based spatial restoration module, SwinIR, to perform deep feature extraction and restore fine details such as texture, edges, and local structures that may still be degraded due to low light conditions or high noise. This flow is a key part of the ALEX architecture that combines the power of HVI color transformation with a Windows-based transformer model. The resulting image from ViTRA (which has been enhanced in intensity and color corrected) is used as the main input to the SwinIR module. Mathematically, it is formed by equation

(7) $$I_{\text{ViTRA}} \in \mathbb{R}^{H \times W \times 3},$$

where I_{ViTRA} is an image that has been enhanced in brightness with RGB channels, and $H \times W$ is the dimension of the image resulting from the enhancement by ViTRA. Although the color and brightness have been improved by ViTRA, it can still have lost fine details or local noise. So, after I_{ViTRA} is generated, the process will continue into the part of SwinIR, namely Shallow Convolution for initial feature extraction. This aims to convert the image to the initial feature representation through the equation

(8) $$F_0 = \text{Conv}_{3\times3}(I_{\text{ViTRA}}),$$

F_0 is a shallow convolution layer with $F_0 \in \mathbb{R}^{H \times W \times C_{0,}}$ and C_0 is the number of initial feature channels. This layer aims to extract basic features such as edge, texture, and color gradient efficiently. After obtaining the F_0 feature, the feature is then divided into small patches of size $P \times P$ to be used as a feature vector through linear projection in the next equation,

(9) $$E_p = \text{Linear}(P_p), \quad p = 1, 2, \dots, N,$$

where E_p is the embedding vector with $E_p \in \mathbb{R}^{C_e}$, P_p is the patch to p, size $P \times P \times C_0$, N is the total number of patches obtained from $N = \frac{HW}{P^2}$, and C_e is the embedding dimension. So that from this process produces a sequence of tokens $X = [E_1, E_2, E_3, \dots, E_N] \in \mathbb{R}^{N \times C_\text{e}}$. This spatial image form can be processed by the transformer in the Swin Transformer block. The core module of SwinIR is the Swin Transformer, which consists of several transformer blocks with Window-based Multi-head Self-Attention (W-MSA) and Shifted Window (SW-MSA) [21]. When entering the swin transformer block, the input will be processed first through W-MSA in a small window of size $w \times w$ using equation

(10) $$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^\text{T}}{\sqrt{d_k}} + B\right)V,$$

where, d_k is the key dimension and B is the relative bias of the position. After that, to connect the information between windows, the windows are stacked (shifted) in the next iteration using the SW-MSA module through the equation

(11) $$X_{\text{shifted}} = \text{Shift}(X).$$

After the shift, W-MSA is performed again to strengthen the interaction between windows. The process continues with hierarchical processing through downsampling and upsampling using patch merging in the next equation,

(12) $$X_{\text{next}} = \text{PatchMerge}(X) = \text{Conv}_{2\times2}(\text{Concat}(X_{\text{even}}, X_{\text{odd}}).$$

PatchMerge combines two neighboring patches into one, reduces the resolution by 2×, and increases the channel by 2×. This process repeats until it reaches the deepest feature level, followed by upsampling to restore the image resolution. Finally, SwinIR will perform reconstruction through upsampling using transpose convolution to produce a feature output with $F_{\mathrm{deep}} \in \mathbb{R}^{H\times W\times C}$. After that, Global Residual Learning is performed to sum the residuals with the upsampled version of the original image (to preserve the global structure) using the equation

$$\hat{I}_{\mathrm{HR}} = \mathrm{Conv}_{3\times3}\big(F_{\mathrm{deep}}\big) + \mathrm{Upsample}(I_{\mathrm{ViTRA}}), \tag{13}$$

where $\hat{I}_{\mathrm{HR}}$ is a significantly enhanced image, both in brightness, color, and spatial detail. This result is the final output of the entire ALEX architecture series. So, it can be concluded that after the ViTRA image is output in RGB color space, the image is entered into the SwinIR module for a deep spatial restoration process. Through shallow convolution, patch embedding, and Swin Transformer blocks, the model can extract global and local features efficiently. The upsampling and reconstruction processes then produce an image with optimal fine detail. With this integration, ALEX can produce images that are not only bright and natural but also clear and artifact-free, making it an ideal solution for intelligent security systems in low-light conditions.

Algorithm 1 summarizes the complete ALEX process to provide a clearer understanding of the technical implementation. It describes the step-by-step pipeline from input preprocessing in the HVI color space to enhancement and spatial restoration using SwinIR.

Algorithm 1. ALEX: Automated Low-Light Enhancement eXpert

```
Input: Low-light RGB image I_low
Output: Enhanced RGB image I_enhanced
Step 1. Convert I_low from RGB color space to HVI color space:
      (H, V, I) = RGB_to_HVI(I_low)
Step 2. Split the HVI image into two branches:
      HV_branch = (H, V)
      I_branch = I
Step 3. Perform initial feature extraction using convolution:
      F_HV = Conv(HV_branch)
      F_I = Conv(I_branch)
Step 4. Initialize Relative Position Encoding (RPE) matrices based on spatial distance
      RPE = f_relative_position(i, j, k, l)
Step 5. Apply Relative Lighten Cross-Attention (RLCA):
      Q_HV, K_I, V_I = Linear(F_HV, F_I)
      Attention = Softmax((Q_HV · K_I^T) / sqrt(d_k) + RPE)
      F_RLCA = Attention ⊙ V_I
Step 6. Fuse the HV and I branches via Lighten Cross-Decoding:
      F_fused = Fuse(F_RLCA, F_HV)
Step 7. Reconstruct the enhanced image in HVI space:
      I_HVI_enhanced = Decoder(F_fused)
Step 8. Convert I_HVI_enhanced back to RGB color space:
      I_RGB_enhanced = HVI_to_RGB(I_HVI_enhanced)
Step 9. Perform spatial restoration and resolution
enhancement using SwinIR:
      F_swin = ShallowConv(I_RGB_enhanced)
      Tokens = PatchEmbed(F_swin)
      F_swin = SwinTransformer(Tokens)
      I_enhanced = Reconstruction(F_swin)
Return: I_enhanced
```

3. Result and discussion

Experiments were conducted by training the ALEX model using datasets frequently used for LLIE tasks such as LOLv1 [22, 23], LOLv2 [24], SICE [25], dan Sony Total-Dark (STD) [26]. Selain itu, kami menggunakan dataset unpaired seperti DICM [27], LIME [28], MEF [29], NPE [30], and VV [31] to test the generalization of our developed model. Details regarding the datasets used can be seen in Table 2.

Table 2. Datasets summary

Dataset	Type	Pair/ Images	Resolution	Degrada-tion Type	Usage
LOLv1	Paired	500 pairs	400×600 to 600×900	Low-light, noise	Train/ Test
LOLv2	Paired	1000 pairs	Varied	Low-light, noise (real + synthetic)	Train/ Test
LOL-Blur	Paired	12000 pairs	Varied	Low-light + motion blur	Train/ Test
SICE	Paired	589 pairs	Up to 4000×3000	Multi-exposure	Train/ Test
SID	Paired	5094 images	4240×2832	Extreme low-light	Train/ Test
DICM	Unpaired	69 images	Varied	Low-light	Qualitative test
LIME	Unpaired	10 images	Varied	Low-light	Qualitative test
MEF	Unpaired	17 images	Varied	Low-light	Qualitative test
NPE	Unpaired	84 images	Varied	Low-light	Qualitative test
VV	Unpaired	24 images	Varied	High contrast low-light	Qualitative test

Training was carried out by following the settings in ViTRA [17], especially for determining the hyperparameters, and was performed using a single NVIDIA RTX 4060 GPU. The training used a batch size of 8, an image crop size of 256×256 pixels, and ran for 100 epochs, starting from epoch 0 with model checkpoints saved after each epoch. The AdamW optimizer was applied with an initial learning rate of 1×10^{-5} and a cosine annealing restart scheduler to ensure stable convergence. Data loading utilized 16 CPU threads in GPU mode. Several data augmentation techniques, including random rotation, flipping, and gamma adjustment, were used to improve generalization. The loss function combined multiple weighted components to balance illumination, structural, and perceptual accuracy, with final weights of $w_{\mathrm{HVI}} = 1.60$, $w_{\mathrm{E}} = 21.78$, $\mathrm{w}_{\mathrm{tv}} = 0.05$, and $w_{\mathrm{L1}} = w_D = w_P = 1 \times 10^{-7}$. These configurations followed the ViTRA [17] training strategy and provided stable optimization with consistent enhancement quality across datasets.

We used several evaluation methods to assess the performance of the proposed model. To measure image quality in the context of compression and image processing, we employed the Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) [32]. Following prior studies in low-light image enhancement [33-35], PSNR and SSIM are adopted as the primary evaluation metrics, as they jointly capture pixel-level fidelity and structural consistency, which are the most critical aspects for assessing enhancement quality. To further evaluate perceptual quality in a way that aligns more closely with human visual perception, we used the Learned Perceptual Image Patch Similarity (LPIPS) metric [36]. In addition, to assess image quality without requiring a reference, we applied two no-reference metrics: the Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) [37] and Natural Image Quality Evaluator (NIQE) [37]. The first evaluation was carried out by testing the ALEX model trained using the LOLv1, LOLv2-real, and LOLv2-synthetic datasets on various models. The results of the evaluation are shown in Tables 2 and 3.

Table 3. Quantitative comparison on LOLV1, LOLV2-Real, and LOLV2-Syn datasets without gamma-corrected (normal). Red color indicates the best value, blue color indicates the second-best value, and green color indicates the third-best value

Model	LOLv1			LOLv2-Real			LOLv2-Syn		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
ZeroDCF [35]	14.861	0.559	-	16.059	0.580	-	17.712	0.815	-
3DLUT [38]	14.350	0.445	-	17.590	0.721	-	18.040	0.800	-
DRBN [39]	16.290	0.617	-	20.290	0.831	-	23.220	0.927	-
RUAS [40]	16.405	0.500	-	15.326	0.488	-	13.765	0.638	-
EnlightenGAN [11]	17.480	0.651	0.322	18.230	0.617	0.309	16.570	0.734	0.220
Restomer [41]	22.365	0.816	-	18.693	0.834	-	21.413	0.830	-
LEDNet [42]	20.627	0.823	0.118	19.938	0.827	0.120	23.709	0.914	0.061
SNR-Aware [43]	24.610	0.842	-	21.480	0.849	-	24.140	0.928	-
PairLIE [44]	19.510	0.736	-	19.885	0.778	-	-	-	-
LLFlow [45]	21.149	0.854	0.119	17.433	0.831	0.176	24.807	0.919	0.067
LLFormer [46]	23.649	0.816	0.175	20.056	0.792	0.211	24.038	0.909	0.066
RetinexFormer [47]	25.153	0.846	0.131	22.794	0.840	0.171	25.670	0.930	0.059
TreEnhance [48]	21.960	0.810	-	-	-	-	-	-	-
IGDFormer [49]	24.11	0.821	-	22.73	0.833	-	25.33	0.937	-
FFTFormer [50]	24.345	0.844	0.0998	-	-	-	-	-	-
SwinLight GAN [51]	22.206	0.846	0.0840	-	-	-	-	-	-
Reis [52]	27.820	0.865	0.1390	-	-	-	-	-	-
DEANet++ [53]	22.542	0.850	-	-	-	-	-	-	-
ILR-Net [54]	23.762	0.865	0.1583	26.825	0.778	0.252	-	-	-
Dark2Light [55]	25.040	0.850	-	21.740	0.846	-	-	-	-
DDR [56]	19.820	0.778	0.232	-	-	-	-	-	-
HVI-CIDNet-wP [57]	23.809	0.857	0.0856	24.111	0.868	0.108	25.129	0.939	0.0450
HVI-CIDNet-oP [57]	23.500	0.870	0.1053	23.427	0.862	0.108	25.705	0.942	0.0471
ViTRA [17]	24.773	0.860	0.0896	23.866	0.870	0.102	25.716	0.946	0.0446
ALEX (Our Model)	**26.810**	**0.868**	**0.0835**	**27.016**	**0.905**	**0.111**	**25.993**	**0.970**	**0.0415**

Table 4 shows a comparison of various LLIE models on the LOLv1, LOLv2-real, and LOLv2-syn datasets without gamma correction. The ALEX model can provide performance that can outperform other models, especially on the LOLv2 dataset for PSNR and SSIM values. The same pattern is also shown in testing using the same dataset, but with additional gamma correction in Table 5. Although it has several evaluation values that cannot outperform other models, ALEX consistently shows real stability by always exceeding the test values of its baseline model (ViTRA [17]). Fig. 3 shows a qualitative comparison of image enhancement results on the LOLv1 and LOLv2 datasets.

Table 4. Quantitative comparison on LOLV1, LOLV2-Real, and LOLV2-Syn datasets with gamma-corrected (GT Mean)

Model	LOLv1			LOLv2-Real			LOLv2-Syn		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
ZeroDCF [35]	21.880	0.640	-	19.771	0.671	-	21.463	0.848	-
3DLUT [38]	21.350	0.585	-	20.190	0.745	-	22.173	0.854	-
DRBN [39]	19.550	0.746	-	-	-	-	-	-	-
RUAS [40]	18.654	0.518	-	19.061	0.510	-	16.584	0.719	-
EnlightenGAN [11]	20.003	0.691	0.317	-	-	0.301	-	-	0.213
Restomer [41]	26.682	0.853	-	26.116	0.853	-	25.428	0.859	-
LEDNet [42]	25.470	0.846	0.113	27.814	0.870	0.114	27.367	0.928	0.056
SNR-Aware [43]	26.716	0.851	-	27.209	0.871	-	27.787	0.941	-
PairLIE [44]	23.526	0.755	-	24.025	0.803	-	-	-	-
LLFlow [45]	25.190	0.930	0.110	25.421	0.877	0.158	27.961	0.930	0.063
LLFormer [46]	25.758	0.823	0.167	26.197	0.819	0.209	28.006	0.927	0.061
RetinexFormer [47]	27.140	0.850	0.129	27.694	0.856	0.166	28.992	0.939	0.056
HVI-CIDNet-wP [16]	27.715	0.876	0.079	28.134	0.892	0.101	29.367	0.950	0.040
HVI-CIDNet-oP [16]	28.141	0.889	0.099	27.762	0.881	0.101	29.566	0.950	0.044
ViTRA [17]	28.189	0.893	0.100	28.392	0.894	0.109	29.655	0.950	0.044
ALEX (Our Model)	**28.293**	**0.904**	**0.093**	**28.671**	**0.898**	**0.103**	**29.748**	**0.959**	**0.040**

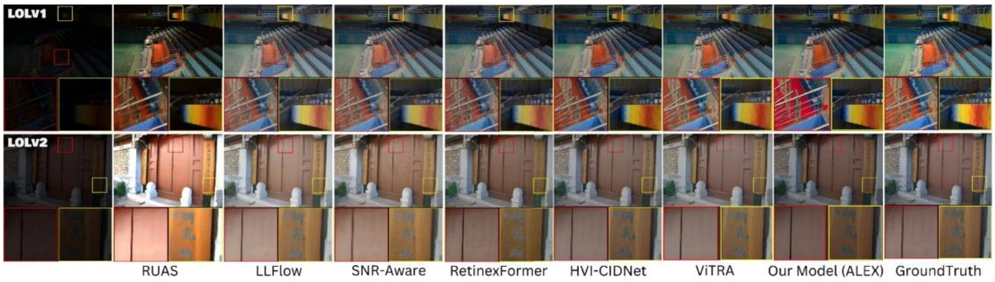

Fig. 3. Qualitative comparison between ALEX and other LLIE models on the LOL v1 and LOL v2 datasets (Zoom in to see the differences more clearly)

ALEX's significant improvements are highly detailed, sharpening even small details in the image. Despite extreme magnification, ALEX consistently maintains image quality, providing clear images even at higher magnifications. Next, we tested the SICE, SID, and LOL-Blur datasets to assess ALEX's ability to handle images with extremely low brightness and low quality. The results are shown in Table 5.

Table 5 shows that ALEX has superior test results on the LOL-Blur dataset. This indicates that our model is very good at improving images in both dark and blurry conditions. However, there is a degradation in PSNR and SSIM values on the SICE dataset compared to our previous model (ViTRA [17]), caused by over-enhancement, resulting in images with details and textures that are slightly far from the ground truth (see Fig. 3 and Fig. 4 for the difference). Furthermore, although ALEX is not the best model in testing on the SID dataset, especially in PSNR and SSIM values, our model provides a definite improvement compared to the ViTRA [17]. and HVI-CIDNet [16] models, which are the baselines of ALEX. In addition, ALEX consistently reduces the LPIPS value of the ViTRA [17] model for testing in Table 4 by an average of 4.67%. Thus, ALEX can produce image enhancements that are consistent with human perception compared to ViTRA [17]. To clarify the existing comparison, Fig. 3 and Fig. 4 shows a qualitative comparison of several LLIE models on the LOL-Blur dataset, and Fig. 5 on the SICE-Grad, SICE-Mix, and SID datasets.

Table 5. Quantitative comparison on SICE, SID, and LOL-Blur datasets. Red indicates the best value, blue indicates the second-best value, and green indicates the third-best value

Model	SICE-Grad			SICE-Mix			SID			LOL-Blur		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
Zero DCE [35]	12.475	0.644	0.334	12.428	0.633	0.382	14.087	0.090	0.813	17.680	0.542	0.422
RUAS [40]	8.628	0.494	0.499	8.684	0.493	0.525	12.622	0.081	0.920	-	-	-
LLFlow [45]	12.737	0.617	0.388	12.737	0.617	0.388	16.226	0.367	0.619	-	-	-
LEDNet [42]	12.551	0.576	0.383	12.668	0.579	0.412	20.830	0.648	0.471	25.271	0.859	0.141
Retinex Former [47]	-	-	-	-	-	-	-	-	-	22.904	0.824	0.236
IGD Former [49]	-	-	-	-	-	-	23.96	0.687	-	-	-	-
MIMO [58]	-	-	-	-	-	-	-	-	-	24.410	0.835	0.183
Swin Light GAN [51]	17.695	0.748	0.140	-	-	-	-	-	-	-	-	-
HVI-CIDNet [16]	13.446	0.648	0.318	13.425	0.636	0.362	22.904	0.676	0.411	26.572	0.890	0.120
ViTRA [17]	13.636	0.638	0.366	13.600	0.627	0.405	22.901	0.677	0.410	26.675	0.893	0.116
ALEX (Our Model)	**13.528**	**0.626**	**0.357**	**13.494**	**0.616**	**0.399**	**22.935**	**0.689**	**0.392**	**26.749**	**0.897**	**0.104**

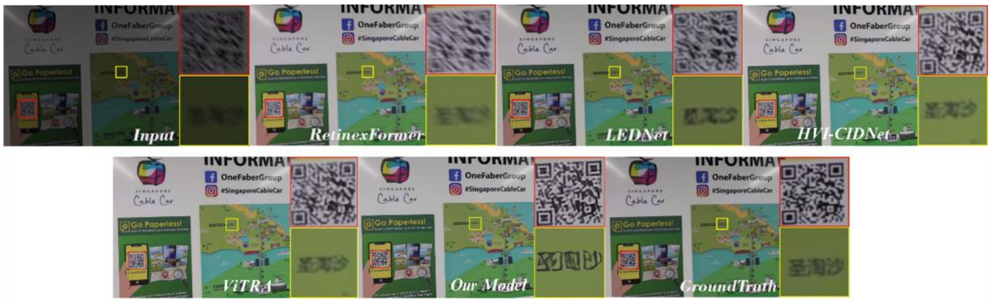

Fig. 4. Qualitative comparison on the LOL-Blur dataset (Zoom in to see the differences more clearly)

Furthermore, to test the generalization effectiveness of ALEX, we conducted tests on unpaired datasets (Fig. 6) such as DICM [27], LIME [28], MEF [29], NPE [30], and VV [31]. We used the Blind/Reference less Image Spatial Quality Evaluator (BRISQUE) [37] and Natural Image Quality Evaluator (NIQE) [59] metric calculation methods to measure the perceptual quality of images without ground-truth. Table 6 shows the test results on the unpaired dataset.

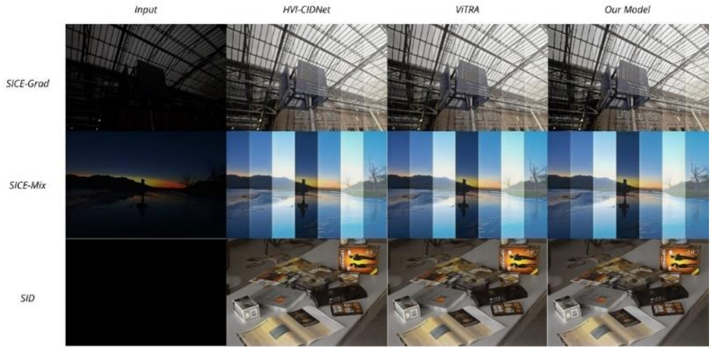

Fig. 5. Qualitative comparison of the SICE-Grad, SICE-Mix, and SID datasets
(Zoom in to see the differences more clearly)

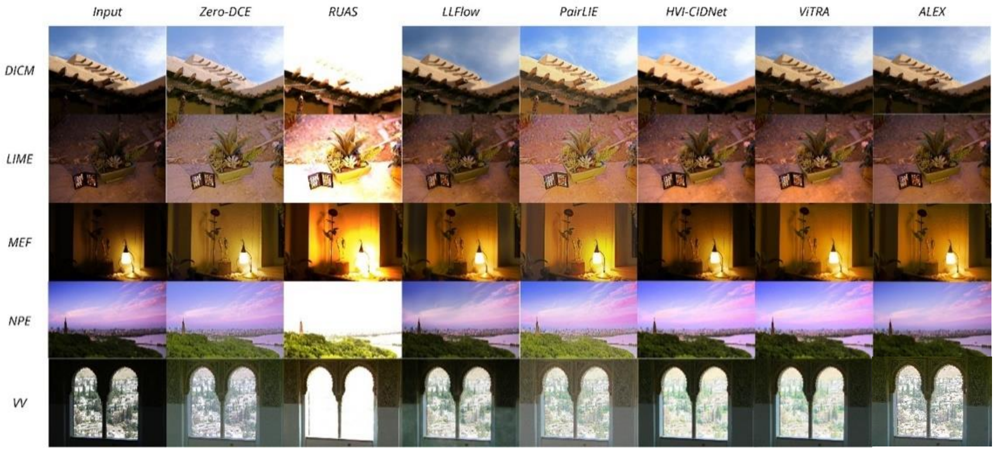

Fig. 6. Qualitative comparison on an unpaired dataset (Zoom in to see the differences more clearly)

Table 6. Quantitative comparison on unpaired datasets (DCIM, LIME, MEF, NPE, AND VV) based on BRISQUE↓ and NIQE↓ metrics

Model	DICM		LIME		MEF		NPE		VV	
	BRISQUE	NIQE	BRISQUE	NIQE	BRISQUE	NIQE	BRISQUE	NIQE	BRISQUE	NIQE
ZeroDCE [35]	27.56	4.58	20.44	5.82	17.32	4.93	20.72	4.53	34.66	4.81
RUAS [40]	38.75	5.21	27.59	4.26	23.68	3.83	47.85	5.53	38.37	4.29
LLFlow [45]	26.36	4.06	27.06	4.59	30.27	4.70	28.86	4.67	31.67	4.04
SNR-Aware [43]	37.35	4.71	39.22	5.74	31.28	4.18	26.65	4.32	78.72	9.87
PairLIE [44]	33.31	4.03	25.23	4.58	27.53	4.06	28.27	4.18	39.13	3.57
FFTFormer [50]	-	-	-	3.83	-	3.81	-	3.72	-	-
SwinLightGAN [51]	28.88	5.27	30.31	5.40	32.04	5.11	29.88	5.29	-	-
Reis [52]	29.40	4.87	-	-	-	-	-	-	-	-
DEANet++ [53]	-	-	-	3.57	-	3.32		3.07	-	-
ILR-Net [54]	20.33	1.73	13.00	2.29	29.73	2.82	24.03	2.68	-	-
HVI-CIDNet [16]	21.47	3.79	16.25	4.13	13.77	3.56	18.92	3.74	30.63	3.21
ViTRA	22.79	3.61	14.71	4.37	14.64	3.32	18.43	3.92	26.81	3.16
ALEX (Our Method)	22.04	2.62	14.29	4.20	13.81	2.93	18.02	3.68	26.39	3.02

The test results on the unpaired dataset in Table 6 show that ALEX generally improves image quality well, although it does not yet provide the best results on the DICM, LIME, and MEF datasets. However, as in previous tests, ALEX consistently provides better image quality improvements compared to the baseline ViTRA model [17]. This indicates that the combination of ViTRA [17] and SwinIR [19] in the ALEX model provides positive improvements compared to the baseline model.

Visually, ALEX can provide a significant difference compared to other models by showing sharper texture details and better resolution. A visualization of this comparison is shown in Fig. 6.

To measure the performance improvement of the developed model, we conducted an ablation study by comparing ALEX with the baseline model on data taken directly from CCTV recordings (the data can be seen at the link **https://s.id/ALEX-dataset**). We conducted tests with two schemes, namely CCTV images without any changes and CCTV images that were dimmed and reduced in resolution. The value measurements in this test used the BRISQUE and NIQE metrics because the data we used was unpaired. The quantitative results of this ablation study are shown in Table 7.

Table 7. Ablation study

Model	CCTV images without any changes		Dimmed CCTV images	
	BRISQUE↓	NIQE↓	BRISQUE↓	NIQE↓
HVI-CIDNet [16]	28.12	4.55	47.15	5.03
ViTRA [17]	27.85	4.08	46.15	5.30
ViTRA [17] + **SwinIR (ALEX)**	**24.42**	**3.87**	**20.17**	**3.98**

The ablation study results show that ALEX is superior compared to the baseline models HVI-CIDNet [16] and ViTRA [17] on original CCTV data. On the original CCTV image data, ALEX obtained the lowest values for the BRISQUE↓ and NIQE↓ metrics with scores of 24.42 and 3.87, respectively. Meanwhile, on the dimmed and reduced-resolution CCTV image data, ALEX obtained BRISQUE↓ and NIQE↓ measurements with scores of 20.17 and 3.98, respectively. These results demonstrate the superiority of ALEX in handling original data originating from CCTV, especially on dimmed and reduced-resolution images. In addition to quantitatively being superior to the baseline model, ALEX qualitatively provides better visuals, as evidenced by Fig. 7.

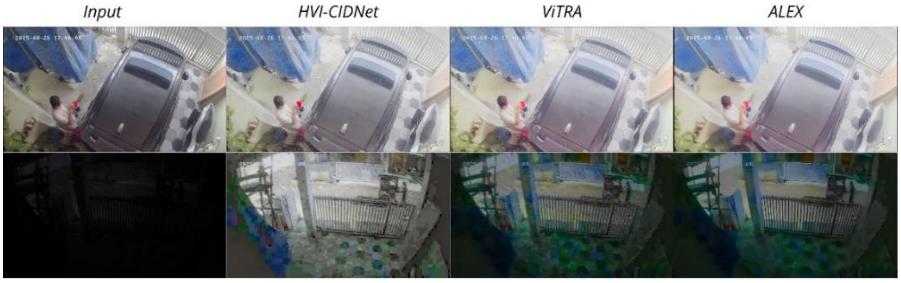

Fig. 7. Qualitative comparison for the ablation study (Zoom in to see the differences more clearly)

However, because the development of ALEX is a modification of the addition of modules to ViTRA [17], this causes the complexity of the model to increase, especially the speed of model inference. This is evidenced by Table 8, which shows the results of the comparison of inference speed between HVI-CIDNet [16], ViTRA [17], and ALEX on several datasets used in this study. In addition, this increase in model complexity also affects the Floating-Point Operations (FLOPs) of the ALEX model, which are larger compared to the two baseline models. The comparison of FLOPs between ALEX and the baseline model can be seen in Fig. 8.

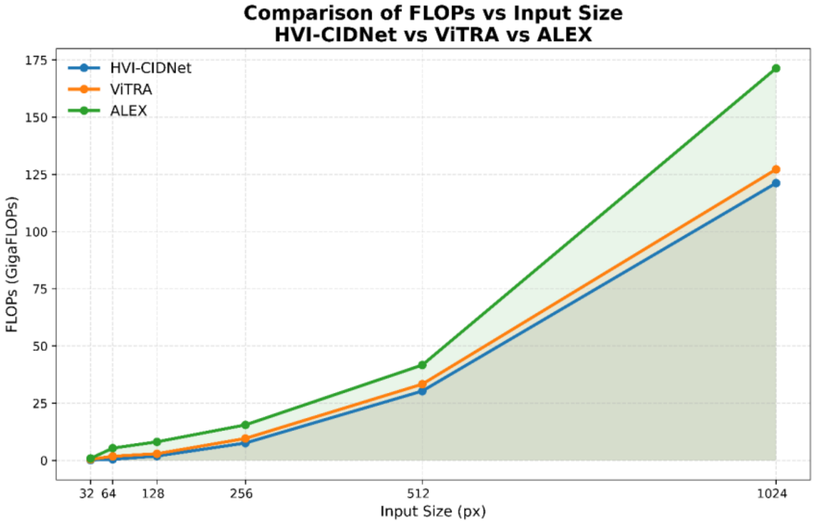

Fig. 8. FLOPs comparison between HVI-CIDNet, ViTRA, and ALEX at image input sizes of 32, 64, 128, 256, 512, and 1024

Table 8. Comparison of image inference speed (second/image) on several datasets between the baseline model (HVI-CIDNet) AND ViTRA

Dataset	HVI-CIDNet	ViTRA	ALEX
LOLv1	0.47 s per image	0.47 s per image	1.04 s per image
LOLv2-Real	0.24 s per image	0.26 s per image	0.42 s per image
LOLv2-Syn	0.14 s per image	0.17 s per image	0.35 s per image
DCIM	0.22 s per image	0.22 s per image	0.31 s per image
LIME	0.6 s per image	0.6 s per image	0.73 s per image
MEF	0.35 s per image	0.35 s per image	0.59 s per image
NPE	0.63 s per image	0.63 s per image	0.8 s per image
VV	3.38 s per image	5.12 s per image	10.84 s per image

Although ALEX has superior performance compared to the baseline model, this is inversely proportional to the existing model complexity. The addition of the SwinIR module [19] to ALEX is the main cause of the increased complexity of the architecture we developed. Therefore, our future research will address this by improving the ALEX model performance while reducing model complexity to accelerate model inference when used. After comparing the inference time and FLOPs in Table 8 and Fig. 8, it can be observed that adding the SwinIR [19] module increases the computational load of ALEX compared to the baseline models. To further clarify the computational cost of the proposed method, a theoretical analysis of its time complexity is presented below.

The overall computational complexity of ALEX depends mainly on two modules: the Relative Lighten Cross-Attention (RLCA) module inherited from ViTRA [17] and the SwinIR-based spatial restoration module [19]. For an input feature map of size $H \times W$ with channel dimension C, the RLCA module performs multi-head cross-attention between the HV and I branches [17]. The complexity of this process can be expressed as the equation

(14) $$O(\mathrm{HW}.C^2 + \frac{(\mathrm{HW})^2.C}{H_\alpha}),$$

where H_α denotes the number of attention heads. The first term corresponds to linear projections (for query, key, and value), while the second term arises from pairwise attention computation across all spatial positions. The SwinIR module applies window-based self-attention with a local window size M, effectively reducing the quadratic dependency on image size to a local region [19]. Thus, its complexity is approximately using the next equation

(15) $$O(\mathrm{HW}.C^2 + M^2.C),$$

which scales linearly with the image resolution when M is fixed. Therefore, the total computational complexity of ALEX can be summarized as equation

(16) $$O(\mathrm{HW}.C^2 + \frac{(\mathrm{HW})^2.C}{H_\alpha} + M^2.C),$$

when using local attention windows, the dominant term grows nearly linearly with the number of pixels, i.e., $O(\mathrm{HW})$. However, the constant factor is higher due to the multi-stage feature extraction and upsampling operations in SwinIR. This theoretical behavior aligns well with the empirical findings shown in Fig. 8, confirming that ALEX provides superior enhancement quality at the cost of moderately increased computational complexity.

4. Challenges and limitations

Although ALEX demonstrates promising results in controlled experiments and provides significant improvement in low-light image enhancement for surveillance systems, several challenges arise when applied to real-world CCTV environments. These challenges are primarily related to illumination dynamics, hardware constraints, environmental variability, and model generalization. The detailed discussion is presented below.

1) Challenges in real-world scenarios

(a) Illumination instability and dynamic lighting conditions

In practical surveillance scenarios, lighting conditions vary dramatically due to factors such as vehicle headlights, passing shadows, flickering neon signs, or sudden exposure to bright light sources. These abrupt changes can cause the model to produce inconsistent enhancement results, including overexposure in bright regions and underexposure in darker areas. Although ALEX's RLCA module can adaptively manage local illumination, it is still limited when illumination changes rapidly over time or across large spatial regions in the same frame.

(b) Noise, artifacts, and sensor degradation

Commercial CCTV cameras (especially low-cost models) often introduce significant compression artifacts and sensor noise due to limited hardware quality or bandwidth restrictions. In such cases, the captured frames contain high-frequency distortions that are difficult to distinguish from texture details. While ALEX improves color and brightness fidelity, it can unintentionally amplify these artifacts, leading to unnatural textures or "halo" effects, particularly around edges or reflective surfaces.

(c) Environmental disturbances and adverse weather

CCTV footage in outdoor environments is frequently affected by environmental conditions such as rain, fog, smoke, or haze, as well as camera motion or vibration. These factors alter image visibility in ways not fully represented in the training datasets. Since ALEX is primarily optimized for static low-light degradation, its performance under mixed degradation (e.g., low-light with fog or motion blur) may decline. Extending the model to handle such compound conditions requires further adaptation through multi-domain or multimodal training strategies.

(d) Real-time processing constraints

A major practical challenge for real-world surveillance applications is real-time operation. The integration of SwinIR in ALEX enhances spatial detail but also increases computational complexity, resulting in higher inference latency. On high-end GPUs, this is manageable; however, many CCTV systems rely on edge devices or embedded processors (e.g., NVIDIA Jetson, ARM-based units) with limited computational resources. Deploying ALEX on such systems requires model optimization techniques such as pruning, quantization, or lightweight transformer variants.

(e) Dataset limitations and domain gap

Although ALEX was trained and tested on several benchmark datasets (LOLv1, LOLv2, SICE, SID, and unpaired sets such as LIME and DICM), these datasets only partially represent the true complexity of real-world surveillance data. Real CCTV footage encompasses a broader range of lighting types, camera angles, and noise characteristics, resulting in a domain gap between benchmark performance and field deployment. Bridging this gap remains a challenge and motivates future collection of large-scale, real-world, low-light CCTV datasets.

(f) Temporal consistency in video enhancement

Currently, ALEX processes each frame independently. As a result, consecutive frames in a video sequence may exhibit slight color or brightness fluctuations, known as temporal inconsistency or flickering. In real-time surveillance systems, such inconsistencies can affect object tracking, motion analysis, and downstream recognition models. A temporal-aware variant of ALEX, possibly incorporating recurrent attention or 3D transformers, would be necessary to address this issue.

2) Limitations of the proposed method

Despite its strong performance, ALEX still presents several methodological and practical limitations that should be acknowledged:

(a) Computational complexity

The integration of the SwinIR transformer considerably increases the number of parameters and FLOPs, resulting in slower inference speed compared to ViTRA and HVI-CIDNet. This makes direct deployment on low-power hardware complex without further model optimization.

(b) Over-enhancement in extreme darkness

In scenes with extremely low illumination, ALEX may slightly exaggerate brightness or contrast, producing “washed-out” regions with reduced texture fidelity.

This occurs because the network prioritizes global brightness restoration over local texture preservation.

(c) Limited robustness to mixed degradations

ALEX performs optimally under low-light and mild noise conditions but is less effective when degradation includes motion blur, fog, or raindrops. These mixed distortions require joint restoration strategies that are outside the scope of the current architecture.

(d) Lack of temporal modeling

As mentioned, the current framework operates in a frame-by-frame manner. The absence of temporal consistency modeling leads to perceptual instability when enhancing continuous video streams.

(e) Dependency on pretrained models and high-quality datasets

The success of ALEX largely depends on the pretraining quality of ViTRA and SwinIR modules, as well as the fidelity of the datasets. Variations in sensor type or compression level can cause the model's learned representations to deviate from actual surveillance conditions.

(f) Energy consumption and hardware cost

The high computational cost not only impacts latency but also power consumption, which is a critical factor for continuous 24/7 surveillance deployment. Optimizing the model for efficiency is necessary for sustainability in large-scale systems.

In summary, while ALEX demonstrates promising results in controlled experiments and provides significant improvement in low-light image enhancement for surveillance systems, it faces several challenges when applied to real-world CCTV environments. These challenges, including illumination dynamics, hardware constraints, environmental variability, model generalization, and the methodological and practical limitations of ALEX, provide a clear direction for future research and improvements in image enhancement technology.

5. Conclusion

In this study, we propose ALEX (Automated Low-Light Enhancement eXpert) as a new model for image enhancement in CCTV-based security systems under low-light conditions. ALEX is built on the ViTRA architecture ViTRA [17] that integrates Relative Lighten Cross-Attention (RLCA) with Relative Position Encoding (RPE) [15] to improve light intensity and image color. Furthermore, the results from ViTRA are processed with SwinIR to perform spatial restoration and resolution enhancement. With this combination, ALEX can produce images that are not only brighter but also have sharper and more natural spatial details. Experimental results show that ALEX consistently outperforms baseline models such as HVI-CIDNet [16] and ViTRA [17] on various benchmark datasets, including LOLv1, LOLv2, SID, SICE, and LOL-Blur. Quantitative evaluation using PSNR, SSIM, LPIPS, BRISQUE, and NIQE metrics demonstrates that ALEX significantly improves image quality, both in terms of ground-truth agreement and human visual perception. Tests on unpaired datasets

(DICM, LIME, MEF, NPE, and VV) also demonstrate ALEX's generalizability under varying real-world lighting conditions.

Furthermore, direct testing on original CCTV images, both unmodified and dimmed and reduced in resolution, demonstrates ALEX's superior performance compared to the baseline. This model not only delivers better quantitative results but also exhibits clearer visual quality and greater detail in the objects within the images. This confirms ALEX's potential for use in real-world applications in intelligent surveillance systems. However, this study also identified limitations in model complexity. The addition of the SwinIR module [19] increases the number of FLOPs and decreases inference speed compared to the baseline. This can be a bottleneck when ALEX is applied to real-time systems or devices with limited computing resources. Therefore, this challenge needs to be addressed in future model development.

For future research, several potential development directions exist. First, focus on optimizing architectural efficiency so that ALEX can still provide high-quality enhancements with lower computational overhead, enabling it to be used on edge devices or real-time systems. Second, integrating ALEX with object detection and recognition models will enhance its practical value in surveillance systems, as it not only produces clearer images but also directly supports event analysis or suspicious activity detection. Third, exploring model compression methods such as pruning, quantization, or distillation could be a solution to accelerate inference without significantly losing image quality. Furthermore, ALEX development could focus on robustness to more extreme environmental conditions, such as combinations of low lighting with rain, fog, or blur caused by camera movement. Finally, the use of more diverse and complex real-world CCTV datasets will further validate ALEX's performance and ensure the model's widespread applicability in real-world surveillance scenarios.

References

1. Piza, E. L., B. C. Welsh, D. P. Farrington, A. L. Thomas. CCTV Surveillance for Crime Prevention: A 40-Year Systematic Review with Meta-Analysis. – Criminology and Public Policy, Vol. **18**, 2019, No 1, pp. 135-159.
2. Denise Cuevas, Q. P., J. P. Carlo Corachea, E. B. Escabel, M. A. Bautista. Lou Effectiveness of CCTV Cameras Installation in Crime Prevention. – College of Criminology Research Journal, Vol. **7**, 2016.
3. Sheela, A. J., S. Balaji, B. Balaji, U. Hemanth Kumar. A Survey on Crime Detection using CCTV Systems. – In: Proc. of 3rd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA'23), 2023, pp. 254-261.
4. Ghari, B., A. Tourani, A. Shahbahrami, G. Gaydadjiev. Pedestrian Detection in Low-Light Conditions: A Comprehensive Survey. – Image and Vision Computing, Vol. **148**, 2024, 105106.
5. Ai, S., J. Kwon. Extreme Low-Light Image Enhancement for Surveillance Cameras Using Attention U-Net. – Sensors (Switzerland), Vol. **20**, 2020, No 2.
6. Ronneberger, O., P. Fischer, T. Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. – In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Springer Verlag, 2015, pp. 234-241.

7. Qu, J., R. W. Liu, Y. Gao, Y. Guo, F. Zhu, F.-Y. Wang. Double Domain Guided Real-Time Low-Light Image Enhancement for Ultra-High-Definition Transportation Surveillance. – IEEE Transactions on Intelligent Transportation Systems, Vol. **25**, 2024, No 8, pp. 9550-9562.
8. Bhandari, A., A. Kafle, P. Dhakal, P. R. Joshi, D. B. Kshatri. Image Enhancement and Object Recognition for Night Vision Traffic Surveillance. – In: G. Ranganathan, X. Fernando, S. F., E. A. Y., Eds. Soft Computing for Security Applications. Springer Singapore, Singapore, 2022, pp. 733-748.
9. Werdiningsih, I., I. Puspitasari, R. Hendradi. Recognizing Daily Activities of Children with Autism Spectrum Disorder Using Convolutional Neural Network Based on Image Enhancement. – Cybernetics and Information Technologies, Vol. **25**, 2025, No 1, pp. 78-96.
10. Jingchun, Z., G. Eg Su, M. Shahrizal Sunar. Low-Light Image Enhancement: A Comprehensive Review on Methods, Datasets, and Evaluation Metrics. – Journal of King Saud University – Computer and Information Sciences, 2024.
11. Jiang, Y., X. Gong, D. Liu, Y. Cheng, C. Fang, X. Shen, J. Yang, P. Zhou, Z. Wang. EnlightenGAN: Deep Light Enhancement Without Paired Supervision. – IEEE Transactions on Image Processing, Vol. **30**, 2021, pp. 2340-2349.
12. Wang, L., L. Zhao, T. Zhong, C. Wu. Low-Light Image Enhancement Using Generative Adversarial Networks. – Scientific Reports, Vol. **14**, 2024, No 1.
13. Lee, M. H., Y. H. Go, S. H. Lee, S. H. Lee. Low-Light Image Enhancement Using CycleGAN-Based Near-Infrared Image Generation and Fusion. – Mathematics, Vol. **12**, 2024, No 24.
14. Tian, Z., P. Qu, J. Li, Y. Sun, G. Li, Z. Liang, W. Zhang. A Survey of Deep Learning-Based Low-Light Image Enhancement. – Sensors, Vol. **23**, 2023, No 18, pp. 1-22.
15. Vaswani, A., N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, I. Polosukhin. Attention is All You Need. – Advances in Neural Information Processing Systems, Vol. **2017-Decem**, 2017, No Nips, pp. 5999-6009.
16. Yan, Q., Y. Feng, C. Zhang, G. Pang, K. Shi, P. Wu, W. Dong, J. Sun, Y. Zhang. HVI: A New Color Space for Low-Light Image Enhancement. – In: Proc. of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR'25), IEEE, 2025, pp. 5678-5687.
17. Darmawan, I., A. Rahmatulloh, R. Gunawan, R. Wahjoe Witjaksono, G. Fauzi Nugraha. ViTRA: Vision Transformer with Relative Position Embedding Attention for Low-Light Image Quality Improvement. – IEEE Access, Vol. **13**, 2025, pp. 160588-160601.
18. Shaw, P., J. Uszkoreit, A. Vaswani. Self-Attention with Relative Position Representations. – In: Proc. of NAACL HLT 2018 – 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies – Proceedings of the Conference, Vol. **2**, 2018, pp. 464-468.
19. Liang, J., J. Cao, G. Sun, K. Zhang, L. Van Gool, R. Timofte. SwinIR: Image Restoration Using Swin Transformer. – In: Proc. of IEEE/CVF International Conference on Computer Vision Workshops (ICCVW'21), IEEE, 2021, pp. 1833-1844.
20. Chen, Z., K. Pawar, M. Ekanayake, C. Pain, S. Zhong, G. F. Egan. Deep Learning for Image Enhancement and Correction in Magnetic Resonance Imaging – State-of-the-Art and Challenges. – Journal of Digital Imaging, 2023, pp. 204-230.
21. Liu, Z., Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, B. Guo. Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows. – In: Proc. of IEEE/CVF International Conference on Computer Vision (ICCV'21), IEEE, 2021, pp. 9992-10002.
22. Wei, C., W. Wang, W. Yang, J. Liu. Deep Retinex Decomposition for Low-Light Enhancement. – In: Proc. of British Machine Vision Conference (BMVC'18). Vol. **2019**. 2018, 61772043.
23. Bogdanova, V. Image Enhancement Using Retinex Algorithms and Epitomic Representation. – Cybernetics and Information Technologies, Vol. **10**, 2010, No 3, pp. 20-30.
24. Yang, W., W. Wang, H. Huang, S. Wang, J. Liu. Sparse Gradient Regularized Deep Retinex Network for Robust Low-Light Image Enhancement. – IEEE Transactions on Image Processing, Vol. **30**, 2021, pp. 2072-2086.

25. Cai, J., S. Gu, L. Zhang. Learning a Deep Single Image Contrast Enhancer from Multi-Exposure Images. – IEEE Transactions on Image Processing, Vol. **27**, 2018, No 4, pp. 2049-2062.

26. Chen, C., Q. Chen, J. Xu, V. Koltun. Learning to See in the Dark. – In: Proc. of IEEE/CVF Conference on Computer Vision and Pattern Recognition, IEEE, 2018, pp. 3291-3300.

27. Lee, C., C. Lee, C.-S. Kim. Contrast Enhancement Based on Layered Difference Representation of 2D Histograms. – IEEE Transactions on Image Processing, Vol. **22**, 2013, No 12, pp. 5372-5384.

28. Guo, X., Y. Li, H. Ling. LIME: Low-Light Image Enhancement via Illumination Map Estimation. – IEEE Transactions on Image Processing, Vol. **26**, 2017, No 2, pp. 982-993.

29. Ma, K., K. Zeng, Z. Wang. Perceptual Quality Assessment for Multi-Exposure Image Fusion. – IEEE Transactions on Image Processing, Vol. **24**, 2015, No 11, pp. 3345-3356.

30. Wang, S., J. Zheng, H. M. Hu, B. Li. Naturalness Preserved Enhancement Algorithm for Non-Uniform Illumination Images. – IEEE Transactions on Image Processing, Vol. **22**, 2013, No 9, pp. 3538-3548.

31. Vonikakis, V., R. Kouskouridas, A. Gasteratos. On the Evaluation of Illumination Compensation Algorithms. – Multimedia Tools and Applications, Vol. **77**, 2018, No 8, pp. 9211-9231.

32. Wang, Z., A. C. Bovik, H. R. Sheikh, E. P. Simoncelli. Image Quality Assessment: from Error Visibility to Structural Similarity. – IEEE Transactions on Image Processing, Vol. **13**, 2004, No 4, pp. 600-612.

33. Liu, R., L. Ma, J. Zhang, X. Fan, Z. Luo. Retinex-Inspired Unrolling with Cooperative Prior Architecture Search for Low-Light Image Enhancement. – In: Proc. of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR'21), IEEE, 2021, pp. 10556-10565.

34. Zhang, Y., J. Zhang, X. Guo. Kindling the Darkness. – In: Proc. of 27th ACM International Conference on Multimedia, ACM, New York, NY, USA, 2019, pp. 1632-1640.

35. Guo, C., C. Li, J. Guo, C. C. Loy, J. Hou, S. Kwong, R. Cong. Zero-Reference Deep Curve Estimation for Low-Light Image Enhancement. – In: Proc. of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR'20), IEEE, 2020, pp. 1777-1786.

36. Zhang, R., P. Isola, A. A. Efros, E. Shechtman, O. Wang. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. – In: Proc. of IEEE/CVF Conference on Computer Vision and Pattern Recognition, IEEE, 2018, pp. 586-595.

37. Mittal, A., A. K. Moorthy, A. C. Bovik. No-Reference Image Quality Assessment in the Spatial Domain. – IEEE Transactions on Image Processing, Vol. **21**, 2012, No 12, pp. 4695-4708.

38. Zeng, H., J. Cai, L. Li, Z. Cao, L. Zhang. Learning Image-Adaptive 3D Lookup Tables for High Performance Photo Enhancement in Real-Time. – IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. **2020**, pp. 2058-2073.

39. Yang, W., S. Wang, Y. Fang, Y. Wang, J. Liu. From Fidelity to Perceptual Quality: A Semi-Supervised Approach for Low-Light Image Enhancement. – In: Proc. of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR'20), IEEE, 2020, pp. 3060-3069.

40. Liu, R., L. Ma, J. Zhang, X. Fan, Z. Luo. Retinex-Inspired Unrolling with Cooperative Prior Architecture Search for Low-Light Image Enhancement. – In: Proc. of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR'21), IEEE, 2021, pp. 10556-10565.

41. Zamir, S. W., A. Arora, S. Khan, M. Hayat, F. S. Khan, M.-H. Yang. Restormer: Efficient Transformer for High-Resolution Image Restoration. – In: Proc. of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR'22), IEEE, 2022, pp. 5718-5729.

42. Zhou, S., C. Li, C. Change Loy. LEDNet: Joint Low-Light Enhancement and Deblurring in the Dark. – In: Lecture Notes in Computer Science Computer Vision. Springer, Nature, Switzerland, 2022, 573-589.

43. Xu, X., R. Wang, C.-W. Fu, J. Jia. SNR-Aware Low-light Image Enhancement. – In: Proc. of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR'22), IEEE, 2022, pp. 17693-17703.

44. Fu, Z., Y. Yang, X. Tu, Y. Huang, X. Ding, K.-K. Ma. Learning a Simple Low-Light Image Enhancer from Paired Low-Light Instances. – In: Proc. of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR'23), IEEE, 2023, pp. 22252-22261.
45. Wang, Y., R. Wan, W. Yang, H. Li, L.-P. Chau, A. Kot. Low-Light Image Enhancement with Normalizing Flow. – In: Proc. of AAAI Conference on Artificial Intelligence, Vol. **36**, 2022, No 3, pp. 2604-2612.
46. Wang, T., K. Zhang, T. Shen, W. Luo, B. Stenger, T. Lu. Ultra-High-Definition Low-Light Image Enhancement: A Benchmark and Transformer-Based Method. – Proceedings of the AAAI Conference on Artificial Intelligence, Vol. **37**, 2023, No 3, pp. 2654-2662.
47. Cai, Y., H. Bian, J. Lin, H. Wang, R. Timofte, Y. Zhang. Retinexformer: One-Stage Retinex-Based Transformer for Low-light Image Enhancement. – In: Proc. of IEEE International Conference on Computer Vision, 2023, pp. 12470-12479.
48. Cotogni, M., C. Cusano. TreEnhance: A Tree Search Method for Low-Light Image Enhancement. – Pattern Recognition, Vol. **136**, 2023, 109249.
49. Wen, Y., P. Xu, Z. Li, W. Xu. (ATO) An Illumination-Guided Dual Attention Vision Transformer for Low-Light Image Enhancement. – Pattern Recognition, Vol. **158**, 2025.
50. Pei, X., Y. Huang, W. Su, F. Zhu, Q. Liu. FFTFormer: A Spatial-Frequency Noise Aware CNN-Transformer for Low Light Image Enhancement. – Knowledge-Based Systems, Vol. **314**, 2025.
51. He, M., R. Wang, M. Zhang, F. Lv, Y. Wang, F. Zhou, X. Bian. SwinLightGAN: A Study of Low-Light Image Enhancement Algorithms Using Depth Residuals and Transformer Techniques. – Scientific Reports, Vol. **15**, 2025, No 1.
52. Reis, M. J. C. S. Low-Light Image Enhancement Using Deep Learning: A Lightweight Network with Synthetic and Benchmark Dataset Evaluation. – Applied Sciences (Switzerland), Vol. **15**, 2025, No 11.
53. Jiang, Y., J. Zhu, L. Li, H. Ma. A Joint Network for Low-Light Image Enhancement Based on Retinex. – Cognitive Computation, Vol. **16**, 2024, No 6, pp. 3241-3259.
54. Yin, M., J. Yang. ILR-Net: Low-Light Image Enhancement Network Based on the Combination of Iterative Learning Mechanism and Retinex Theory. – PLoS ONE, Vol. **20**, 2025, No 2.
55. Li, R.-K., M.-H. Li, S.-Q. Chen, Y.-T. Chen, Z.-H. Xu. Dark2Light: Multi-Stage Progressive Learning Model for Low-Light Image Enhancement. – Optics Express, Vol. **31**, 2023, No 26, 42887.
56. Zhang, W., H. Zhang, X. Liu, X. Guo, X. Wang, S. Li. Unsupervised Low-Light Image Enhancement Based on Explicit Denoising and Knowledge Distillation. – Computers, Materials and Continua, Vol. **82**, 2025, No 2, pp. 2537-2554.
57. Yan, Q., Y. Feng, C. Zhang, P. Wang, P. Wu, W. Dong, J. Sun, Y. Zhang. You Only Need One Color Space: An Efficient Network for Low-light Image Enhancement. – arXiv Preprint arXiv 2402.05809, Vol. **2024**, pp. 1-11.
58. Cho, S.-J., S.-W. Ji, J.-P. Hong, S.-W. Jung, S.-J. Ko. Rethinking Coarse-to-Fine Approach in Single Image Deblurring. – In: Proc. of IEEE/CVF International Conference on Computer Vision (ICCV'21), IEEE, 2021, pp. 4621-4630.
59. Mittal, A., R. Soundararajan, A. C. Bovik. Making a "Completely Blind" Image Quality Analyzer. – IEEE Signal Processing Letters, Vol. **20**, 2013, No 3, pp. 209-212.

Received: 06.09.2025, First revision: 23.10.2025, Accepted: 28.10.2025