AUTOLUMNET: A BI-BRANCH EXPOSURE-AWARE NETWORK FOR LOW- AND HIGH-EXPOSURE IMAGE ENHANCEMENT

Anonymous authors
Paper under double-blind review

ABSTRACT

Enhancing images captured under challenging illumination is difficult because real-world scenes often contain both severely under-exposed shadows and over-exposed highlights. Existing low-light enhancement methods primarily address under-exposure, while multi-exposure fusion requires multiple bracketed shots, which are rarely available in practice. We propose **AutoLumNet**, a bi-branch exposure-aware network that performs single-shot exposure correction. AutoLumNet decomposes input features into dual branches specialized for shadows and highlights, then adaptively fuses them via spatial attention. To ensure that the corrected luminance distribution aligns with natural photographs, we introduce an optimal-transport-based *exposure distribution alignment* mechanism, theoretically guaranteeing monotonicity and preventing spurious extrema. Training is guided by a unified exposure-aware objective combining reconstruction fidelity, distribution alignment, perceptual consistency, and regularization terms. Extensive experiments on SICE, LOL, and MIT-Adobe FiveK demonstrate that AutoLumNet achieves state-of-the-art results across under-, over-, and mixed-exposure conditions, outperforming both single-image enhancement and multi-exposure fusion baselines in PSNR/SSIM, perceptual metrics, and user studies. Our approach bridges the gap between low-light enhancement and exposure fusion, offering a principled and practical solution for real-world photography.

1 INTRODUCTION

Images captured in real-world scenes frequently suffer from challenging illumination, where dark shadows coexist with saturated highlights. Such distortions significantly degrade visual quality and impair the performance of downstream vision tasks including detection, recognition, and tracking. The problem arises because digital sensors have a limited dynamic range compared to natural scenes, causing under-exposure in dark regions and over-exposure in bright regions within the same frame. While auto-exposure mechanisms and high dynamic range (HDR) sensors can partially mitigate this issue, they often introduce new artifacts such as noise, blur, or color distortion, and multiple exposure captures are impractical for dynamic scenes or mobile devices.

A large body of work has sought to address these challenges. Traditional enhancement methods such as histogram equalization and gamma correction globally adjust brightness but fail to account for spatially varying illumination, often producing unnatural results. Retinex-based approaches attempt to decompose images into reflectance and illumination components (Guo et al., 2017; Wei et al., 2018b), enabling more principled enhancement, yet their reliance on hand-crafted priors or simplified models limits robustness under complex lighting. More recently, deep learning has become dominant. Supervised methods like LLNet (Lore et al., 2017), KinD (Zhang et al., 2019), and related variants exploit paired low/normal-light datasets to learn end-to-end mappings, while zero-reference models such as Zero-DCE (Guo et al., 2020) reformulate enhancement as curve estimation without requiring paired supervision. Unsupervised GAN-based solutions, including EnlightenGAN (Jiang et al., 2021), avoid ground truth but are sensitive to the choice of unpaired training data. Although these approaches have advanced low-light enhancement, they primarily address global under-exposure and remain inadequate for correcting images that also suffer from over-exposed highlights.

In parallel, multi-exposure image fusion (MEF) methods aim to recover HDR images by fusing under- and over-exposed inputs. Classical MEF techniques rely on pixel-level, patch-level, or transform-domain weighting strategies (Mertens et al., 2007; Ma et al., 2017), while recent deep models such as DeepFuse (Prabhakar et al., 2017), U2Fusion (Xu et al., 2020a), and MEF-GAN (Xu et al., 2020c) directly learn fusion mappings. Ensemble-based frameworks like EMEF (Wang et al., 2022b) combine outputs from multiple fusion strategies, improving generalization. Strong baselines including FMMEF (Li et al., 2020), FusionDN (Xu et al., 2020b), and MGFF (Bavirisetti et al., 2020) each excel on specific metrics, as highlighted by MEFB benchmark evaluations (Zhang et al., 2021). Despite their success, all MEF pipelines assume access to multiple aligned exposures of the same scene, which are rarely available in practice, especially for dynamic or handheld capture.

Recent studies have further highlighted key limitations in existing paradigms. Attention-based low-light networks (Zhu et al., 2021) improve feature selection but still struggle with severe noise and color artifacts. Normalizing flow-based models such as LLFlow (Wang et al., 2022a) offer a probabilistic view of illumination distributions, yet are biased toward local pixel correlations and lack explicit handling of over-exposure. Multi-scale feature complementation networks (Zhang et al., 2023a) leverage hierarchical features to restore structure and color, but remain confined to the low-light regime. Retinex-based fusion methods (Zhao et al., 2023) extend reflectance–illumination decomposition to multi-exposure settings, but their reliance on multiple captures prevents direct applicability to single-shot scenarios. Collectively, these observations underline a gap: no existing method unifies the ability to simultaneously correct under- and over-exposed regions within a single image while also enforcing global exposure regularity.

We address this gap by formulating *single-shot exposure correction (SEC)* as the task of mapping an input image to an enhanced output that lies on the manifold of naturally exposed photographs, irrespective of whether the degradations arise from shadows, highlights, or both. This perspective reframes exposure correction as an inherently bimodal problem: shadows require expansion of compressed intensities, while highlights demand attenuation of saturated regions. Existing single-branch networks conflate these opposing corrections, often leading to artifacts or loss of detail. A more principled approach should explicitly separate the two distortion modes, reconcile them adaptively across space, and enforce consistency with the global statistics of natural exposure.

To this end, we propose **AutoLumNet**, a bi-branch exposure-aware architecture for single-shot exposure correction. The core idea is to model exposure correction as the inverse of two monotone distortions. One branch specializes in recovering details from under-exposed regions, while the other suppresses over-exposed areas. An adaptive fusion mechanism interpolates between the two corrections, producing a unified luminance map. Beyond local corrections, we introduce an exposure distribution alignment module that constrains the fused luminance to follow a canonical well-exposed distribution, formulated as an optimal transport problem. This guarantees monotone, spatially varying luminance mappings that preserve order and prevent new extrema. Finally, we integrate pixel-level fidelity, perceptual consistency, and distributional regularity into a unified exposure-aware objective, ensuring both local accuracy and global naturalness.

In summary, our contributions are threefold:

- We formulate **single-shot exposure correction (SEC)** as a unified task that bridges low-light enhancement and multi-exposure fusion, explicitly addressing both under- and over-exposure within a single framework.

- We propose **AutoLumNet**, a dual-branch exposure-aware architecture equipped with an optimal transport-based distribution alignment mechanism, which provides theoretical guarantees of monotonicity and exposure consistency.

- We demonstrate through extensive experiments on standard benchmarks (SICE, LOL, MIT-Adobe FiveK, MEFB) that AutoLumNet achieves state-of-the-art performance across low-, high-, and mixed-exposure conditions, outperforming both single-image enhancement and multi-exposure fusion baselines.

2 RELATED WORK

2.1 CONVENTIONAL AND RETINEX-INSPIRED ENHANCEMENT

Early enhancement pipelines relied on handcrafted priors, such as histogram equalization or gamma correction, but these global adjustments are prone to noise amplification and halo artifacts. More principled approaches leverage Retinex theory by decomposing an image into reflectance and illumination components. Recent learning-based Retinex variants improve robustness by embedding priors into neural networks, e.g., RetinexNet (Wei et al., 2018b), URetinex (Wu et al., 2022), and CRetinex (Zhang et al., 2023b), which integrate unfolding, transformer modules, or color-shift constraints. While these methods effectively enhance under-exposed regions, they remain biased toward low-light cases and are less suited for scenes that simultaneously exhibit over-exposure.

2.2 DATA-DRIVEN SINGLE-IMAGE ENHANCEMENT

Deep CNN and transformer-based methods dominate recent single-image exposure correction. KinD (Zhang et al., 2019) couples reflectance restoration with illumination refinement, while Zero-DCE (Guo et al., 2020) formulates enhancement as image-specific curve estimation trained with non-reference objectives, enabling zero-pair learning. Extensions such as Zero-DCE++ and LLFlow (Wang et al., 2022a) explore lightweight architectures and normalizing flows to capture exposure distributions more faithfully. Multi-scale designs like LIEN-MFC (Zhang et al., 2023a) and attention-based networks (Zhu et al., 2021) improve structural recovery and noise suppression, while diffusion- and GAN-based methods such as EnlightenGAN (Jiang et al., 2021) focus on perceptual realism. Despite progress, most single-image methods remain skewed toward brightening shadows, with limited capacity to suppress highlights or enforce global exposure regularity.

2.3 MULTI-EXPOSURE IMAGE FUSION (MEF)

MEF methods fuse bracketed exposures of the same scene to approximate high dynamic range. Early deep models like DeepFuse (Prabhakar et al., 2017) optimized the MEF-SSIM metric (Ma et al., 2015), while subsequent frameworks expanded to general fusion tasks, e.g., IFCNN (Zhang et al., 2020a), U2Fusion (Xu et al., 2020a), and PMGI (Zhang et al., 2020b). Recent supervised pipelines incorporate adversarial and attention modules, such as MEF-GAN (Xu et al., 2020c), and transformer-based MEF has further improved long-range consistency. Ensemble-based EMEF (Wang et al., 2022b) combines the strengths of multiple imperfect fusion styles, achieving state-of-the-art results on MEFB benchmark (Zhang et al., 2021). Other strong baselines include FMMEF (Li et al., 2020), FusionDN (Xu et al., 2020b), and MGFF (Bavirisetti et al., 2020), each excelling under different evaluation metrics. Retinex-MEF (Zhao et al., 2023) explicitly models glare effects to stabilize reflectance recovery. While these approaches excel at balancing shadows and highlights, their reliance on multiple aligned exposures fundamentally limits applicability in dynamic or single-shot settings.

2.4 POSITIONING OF OUR APPROACH

AutoLumNet lies at the intersection of single-image enhancement and MEF. Unlike low-light models that only brighten or MEF pipelines that require brackets, AutoLumNet operates in the single-shot regime while retaining MEF-style exposure reasoning. We introduce (i) a dual-branch decomposition that separates under- and over-exposure corrections, (ii) an optimal transport-based distribution alignment that enforces global exposure consistency through monotone luminance transport, and (iii) a unified exposure-aware objective that couples pixel fidelity, distributional alignment, and perceptual realism. This design directly addresses the bimodal nature of exposure distortion and provides a principled, end-to-end solution for under-, over-, and mixed-exposure scenes.

3 METHOD

Enhancing images captured under challenging illumination remains difficult because natural scenes often contain a mixture of severely under-exposed shadows and over-exposed highlights. Existing methods typically address only one side of the problem: low-light enhancement models (Guo et al.,

2020; Zhang et al., 2019) assume global under-exposure, while multi-exposure fusion models (Wang et al., 2022b; Ma et al., 2015) require multiple shots of the same scene. In practice, however, a single input image may simultaneously suffer from both distortions, and paired exposures are rarely available. We therefore formulate *single-shot exposure correction* as the task of learning a mapping $F_\theta : I \mapsto \hat{I}$ from an input image $I \in [0,1]^{H\times W\times 3}$ to an enhanced output $\hat{I}$ that lies on the manifold of naturally exposed photographs. Our central motivation is that correcting exposure is inherently *bimodal*: dark and bright regions exhibit opposite distortions yet must be reconciled into a single consistent output. This observation underpins our AutoLumNet design, which introduces (i) a dual-branch decomposition to separately handle under- and over-exposed content, (ii) an exposure distribution alignment mechanism to match luminance statistics to a canonical target distribution, and (iii) a unified exposure-aware objective to train the model end-to-end.

3.1 DUAL-BRANCH EXPOSURE DECOMPOSITION

Exposure distortion can be modeled as a monotone mapping of the ideal well-exposed luminance $Y^\star$ into the observed luminance Y. Under-exposure compresses dynamic range into low values, while over-exposure saturates it toward high values. Formally, we may write

$$Y = g(Y^\star), \qquad g(y) \in \begin{cases} g_{\text{ue}}(y), & \text{if under-exposed,} \\ g_{\text{oe}}(y), & \text{if over-exposed,} \end{cases} \tag{1}$$

where g_{ue} is a sub-linear mapping (shadows compressed toward 0) and g_{oe} a saturating mapping (highlights clipped toward 1). Since both distortions can co-exist in different regions of the same image, learning a single corrective function is ill-posed.

Dual inverse mappings. We therefore introduce two corrective functions

$$h_{\text{ue}} : Y \mapsto \tilde{Y}_{\text{ue}}, \qquad h_{\text{oe}} : Y \mapsto \tilde{Y}_{\text{oe}}, \tag{2}$$

where h_{ue} expands dark intensities and h_{oe} compresses bright intensities. The corrected luminance is then synthesized by combining the two inverse mappings:

$$\hat{Y} = \Phi(h_{\text{ue}}(Y),\ h_{\text{oe}}(Y))\,. \tag{3}$$

Here Φ is a learned fusion operator that interpolates between the two corrections. This decomposition reflects the *bimodal nature of exposure distortion*: shadows and highlights must be handled separately, then reconciled into a single luminance field.

Network realization. In AutoLumNet, h_{ue} and h_{oe} are instantiated as two neural branches applied to the shared encoder pyramid. The encoder E_ϕ extracts multi-scale features

$$\{F^0, F^1, F^2, F^3, F^4\} = E_\phi(I), \tag{4}$$

where $F^i \in \mathbb{R}^{H_i\times W_i\times d_i}$. Each branch applies exposure-specific transformations:

$$U^i = h_{\text{ue}}(F^i), \qquad O^i = h_{\text{oe}}(F^i), \tag{5}$$

yielding feature sets $\{U^i\}$ and $\{O^i\}$ that are specialized for recovering details in dark and bright regions, respectively. Concretely, h_{ue} contains filters that enhance edge and gradient information in shadows, while h_{oe} contains filters that recover structure from saturated highlights.

Adaptive fusion. Since no explicit mask of under- or over-exposed regions is available, the network learns spatially varying weights. For each pixel p at scale i, we predict logits $s^i_{\text{ue}}(p)$ and $s^i_{\text{oe}}(p)$, normalize them by a softmax, and form convex weights:

$$\alpha^i_{\text{ue}}(p),\ \alpha^i_{\text{oe}}(p) = \mathrm{softmax}\big(s^i_{\text{ue}}(p), s^i_{\text{oe}}(p)\big). \tag{6}$$

The fused feature map is then

$$F^i_{\text{fuse}}(p) = \alpha^i_{\text{ue}}(p)\, U^i(p) + \alpha^i_{\text{oe}}(p)\, O^i(p). \tag{7}$$

This convexity constraint ($\alpha^i_{\text{ue}} + \alpha^i_{\text{oe}} = 1$) guarantees that the fusion lies within the span of the two corrective hypotheses, preventing artifacts from uncontrolled extrapolation.

End-to-end flow. The overall decomposition–fusion mechanism can be summarized as

$$I \xrightarrow{E_\phi} \{F^i\} \xrightarrow{h_{\text{ue}}, h_{\text{oe}}} \{U^i\}, \{O^i\} \xrightarrow{\alpha} \{F^i_{\text{fuse}}\} \xrightarrow{\text{decoder}} \hat{I},$$

where $\hat{I} = (\hat{Y}, \hat{C})$ is the final enhanced image, with $\hat{Y}$ obtained through dual-branch correction and $\hat{C}$ provided by a chroma-preserving module.

This dual-branch design encodes an architectural bias: exposure correction is explicitly treated as the inverse of two monotone distortions. The benefit is twofold: (i) dark and bright regions are corrected by specialized pathways rather than competing in a single feature space, and (ii) the fusion ensures a spatially adaptive balance, enabling robust handling of mixed-exposure scenes. In the following subsection, we further constrain $\hat{Y}$ by aligning its distribution with a canonical well-exposed target via optimal transport. However, this local decomposition alone does not ensure that the global luminance statistics resemble natural photographs. To enforce global naturalness, we introduce an exposure distribution alignment step.

3.2 EXPOSURE DISTRIBUTION ALIGNMENT

While the dual-branch decomposition corrects shadows and highlights locally, it does not guarantee that the fused luminance $\hat{Y}$ follows the global statistics of naturally exposed photographs. Empirically, well-exposed images exhibit stable histogram characteristics, with values concentrated around mid-tones and balanced spread. To bridge the gap between local corrections and global naturalness, we introduce an *exposure distribution alignment* step based on optimal transport.

Canonical Target Distribution. We denote by P_Y the empirical luminance distribution of the input image and by $\hat{P}_Y$ the distribution after correction. Our objective is to align $\hat{P}_Y$ with a canonical well-exposed distribution $P^\star$:

$$\hat{P}_Y \approx P^\star. \tag{8}$$

The target $P^\star$ can be defined as a fixed prior, such as a truncated Gaussian centered at 0.5, or estimated from training data. This serves as a statistical anchor that prevents over-correction and ensures consistency across scenes.

Transport Map. To realize this alignment, we model exposure correction as a parametric transport map T_θ applied to luminance:

$$\hat{Y}(p) = T_\theta(Y(p), p), \qquad \hat{P}_Y = T_\theta \# P_Y, \tag{9}$$

where p indexes pixel location and $\#$ denotes push-forward measure. For stability, we restrict T_θ to a locally affine form

$$T_\theta(y, p) = a(p)\, y + b(p), \qquad a(p) > 0, \tag{10}$$

with parameters $a(p), b(p)$ predicted from fused features at coarse resolution and upsampled to full size. The positivity constraint ensures that the mapping is strictly monotone, thereby preserving luminance order.

Distribution Matching. To quantify how close $\hat{P}_Y$ is to $P^\star$, we minimize the Sinkhorn divergence, a differentiable approximation of the Wasserstein distance with entropic regularization,

$$\mathcal{L}_{\text{OT}} = \text{Sinkhorn}_\varepsilon(T_\theta \# P_Y,\ P^\star)\,, \tag{11}$$

which provides a differentiable and computationally efficient approximation of the Wasserstein distance. This term directly enforces that the output histogram converges toward natural exposure statistics.

Regularization. To avoid degenerate mappings, we introduce auxiliary penalties:

$$\mathcal{L}_{\text{mono}} = \sum_p \max(0, \epsilon - a(p)), \qquad \mathcal{L}_{\text{smooth}} = \sum_p \|\nabla a(p)\|_2^2 + \|\nabla b(p)\|_2^2. \tag{12}$$

The first penalizes non-monotone transport, while the second enforces spatial smoothness of parameters, preventing abrupt shifts across neighboring pixels.

Proposition. If $a(p) > 0$, then $T_\theta(y, p)$ is strictly monotone in y and no new extrema are introduced."

This alignment step links the local corrections of the dual branches with a global statistical constraint. The monotone transport ensures no new extrema are introduced, while Sinkhorn divergence guarantees convergence to the canonical distribution $P^\star$. Together, these properties yield enhanced outputs that are both spatially adaptive and distributionally consistent.

3.3 UNIFIED EXPOSURE-AWARE OBJECTIVE

Finally, we integrate these components into a unified exposure-aware optimization framework, since no single loss can capture the competing requirements of fidelity, naturalness, and perceptual quality. AutoLumNet therefore combines complementary objectives into a unified optimization framework.

Reconstruction Loss. At the most fundamental level, the enhanced output $\hat{I}$ should remain faithful to the reference well-exposed image $I^\star$. To capture this, we use a combination of pixel-level ℓ_1 distance and structural similarity (SSIM):

$$\mathcal{L}_{\text{rec}} = \|\hat{I} - I^\star\|_1 + \lambda_{\text{ssim}}\big(1 - \text{SSIM}(\hat{I}, I^\star)\big). \tag{13}$$

While ℓ_1 ensures absolute accuracy, SSIM emphasizes structural fidelity in terms of contrast and luminance, which are particularly sensitive in exposure correction.

Exposure Alignment Loss. Pixel-level fidelity, however, does not guarantee that the global luminance distribution matches that of naturally exposed photographs. To address this, we impose a Sinkhorn-based alignment term:

$$\mathcal{L}_{\text{align}} = \text{Sinkhorn}_\varepsilon(T_\theta \# P_Y,\ P^\star)\,, \tag{14}$$

where P_Y is the input luminance distribution, $T_\theta \# P_Y$ the push-forward through the transport map, and $P^\star$ the canonical target distribution. This alignment ensures that the overall histogram of $\hat{Y}$ is consistent with mid-tone exposure statistics, complementing local reconstruction.

Perceptual Consistency. Even with pixel fidelity and distributional correctness, enhanced images may look visually unsatisfying if texture and semantic cues are not preserved. To bridge this gap, we enforce perceptual consistency using feature embeddings from a pretrained VGG network (Simonyan & Zisserman, 2015; Johnson et al., 2016) :

$$\mathcal{L}_{\text{perc}} = \sum_l \|\phi_l(\hat{I}) - \phi_l(I^\star)\|_2^2, \tag{15}$$

where $\phi_l(\cdot)$ denotes the activation at layer l. This loss preserves high-level structure and natural textures that are often distorted when shadows and highlights are aggressively corrected.

Exposure-Aware Optimization. The final objective integrates these components, along with regularization terms that enforce monotonicity and smoothness of the transport map:

$$\mathcal{L} = \lambda_{\text{rec}}\mathcal{L}_{\text{rec}} + \lambda_{\text{align}}\mathcal{L}_{\text{align}} + \lambda_{\text{perc}}\mathcal{L}_{\text{perc}} + \lambda_{\text{reg}}\big(\mathcal{L}_{\text{mono}} + \mathcal{L}_{\text{smooth}}\big). \tag{16}$$

Each term plays a distinct role: $\mathcal{L}_{\text{rec}}$ anchors the output to ground truth, $\mathcal{L}_{\text{align}}$ aligns luminance statistics with the natural exposure manifold, $\mathcal{L}_{\text{perc}}$ encourages perceptual realism, and $\mathcal{L}_{\text{mono}}, \mathcal{L}_{\text{smooth}}$ provide theoretical guarantees of monotone and stable transport. Together, this exposure-aware optimization ensures that AutoLumNet not only reconstructs faithfully but also produces visually natural results under both low- and high-exposure conditions.

4 EXPERIMENTS

4.1 DATASETS AND EVALUATION METRICS

We evaluate AutoLumNet on three widely used benchmarks for exposure correction and image enhancement. The SICE dataset (Cai et al., 2018) is employed for training; it contains 4,413 multi-exposure scenes captured under diverse lighting conditions, each with multiple exposure levels. Following prior work, we randomly split the data into 80% training, 10% validation, and 10% testing.

Table 1: Full-reference comparison on LOL and MIT-Adobe FiveK. Higher PSNR/SSIM and lower LPIPS are better.

Method	LOL			FiveK		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
RetinexNet (Wei et al., 2018b)	16.8	0.65	0.370	19.2	0.72	0.310
KinD (Zhang et al., 2019)	20.9	0.82	0.270	21.1	0.78	0.240
EnlightenGAN (Jiang et al., 2021)	18.2	0.71	0.320	20.5	0.75	0.280
Zero-DCE (Guo et al., 2020)	22.4	0.83	0.245	22.9	0.82	0.220
LIEN-MFC (Zhang et al., 2023a)	23.1	0.85	0.210	23.8	0.84	0.200
LLFlow (Wang et al., 2022a)	24.2	0.87	0.195	24.7	0.85	0.190
DeepFuse (Prabhakar et al., 2017)	18.5	0.70	0.350	20.0	0.74	0.300
MEF-GAN (Xu et al., 2020c)	20.7	0.79	0.290	22.3	0.80	0.250
EMEF (Wang et al., 2022b)	22.8	0.84	0.235	23.6	0.83	0.210
AutoLumNet (Ours)	**25.5**	**0.90**	**0.160**	**26.2**	**0.88**	**0.150**

Table 2: No-reference comparison on SICE and MEFB. Lower NIQE/BRISQUE/PIQE and higher MEF-SSIM indicate better perceptual quality.

Method	NIQE↓	BRISQUE↓	PIQE↓	MEF-SSIM↑
RetinexNet (Wei et al., 2018b)	5.12	36.4	48.1	0.78
KinD (Zhang et al., 2019)	4.85	33.7	44.6	0.80
EnlightenGAN (Jiang et al., 2021)	5.01	35.2	47.5	0.79
Zero-DCE (Guo et al., 2020)	4.73	31.5	42.8	0.81
LIEN-MFC (Zhang et al., 2023a)	4.55	29.6	40.5	0.83
LLFlow (Wang et al., 2022a)	4.41	28.9	39.1	0.84
DeepFuse (Prabhakar et al., 2017)	5.07	34.1	46.2	0.85
MEF-GAN (Xu et al., 2020c)	4.88	32.7	43.9	0.87
EMEF (Wang et al., 2022b)	4.52	29.8	40.8	0.89
AutoLumNet (Ours)	**4.10**	**27.2**	**37.6**	**0.92**

For evaluation, we adopt two additional datasets to test generalization: the LOL dataset (Wei et al., 2018a), which provides 500 paired low/normal-light images (485 for training and 15 for validation), and the MIT-Adobe FiveK dataset (Bychkovsky et al., 2011), consisting of 5,000 high-resolution raw images with expert-retouched references. Since LOL and FiveK serve primarily as validation benchmarks, we report results on their official validation/test splits without using them for training. This setup ensures that AutoLumNet is trained on diverse exposure variations (SICE) and evaluated on both real low-light scenes (LOL) and professionally retouched photographs (FiveK).

For quantitative assessment, we adopt both full-reference and no-reference metrics. With paired ground truth available (LOL, FiveK), we compute peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), and learned perceptual image patch similarity (LPIPS) to measure fidelity and perceptual closeness. On unpaired datasets (SICE, MEFB), we report no-reference quality measures including NIQE, BRISQUE, and PIQE, which correlate with human perceptual judgments. For exposure fusion tasks, we additionally employ MEF-SSIM (Ma et al., 2015), a widely used metric specifically designed to evaluate multi-exposure image fusion quality. Together, these metrics provide a balanced evaluation covering pixel-level accuracy, perceptual realism, and exposure consistency.

4.2 Implementation Details

AutoLumNet is implemented in PyTorch. The encoder backbone is a ResNet-18 pretrained on ImageNet, from which multi-scale features (stages 0–4) are extracted. Both branches share the encoder and apply exposure-specific residual transformations. The decoder follows a U-Net style structure with skip connections from encoder features. The optimal transport (OT) head predicts spatially varying affine parameters $(a(p), b(p))$ at $\frac{1}{4}$ resolution, which are bilinearly upsampled to the input size. Chroma correction is performed in the Lab color space with a lightweight CNN branch.

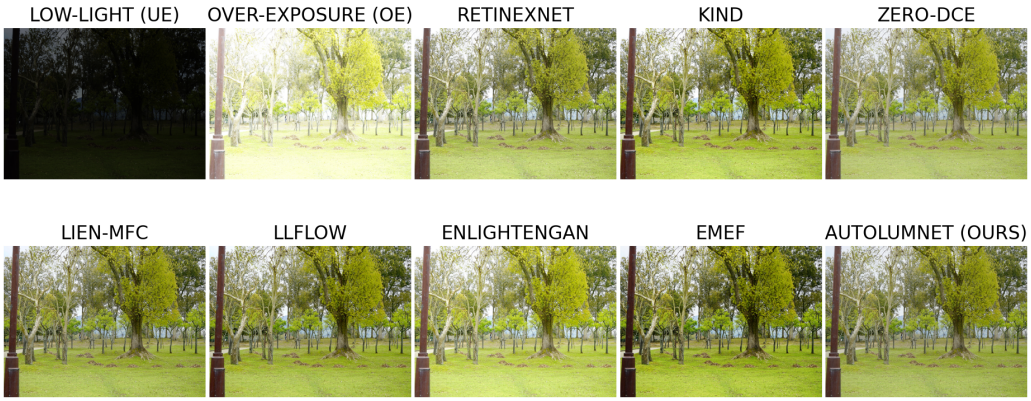

Figure 1: Qualitative comparison on a challenging mixed-exposure scene from SICE. **Top row:** input *low-light (UE)*, input *over-exposure (OE)*, and three single-image enhancement baselines (RetinexNet, KinD, Zero-DCE). **Bottom row:** recent methods (LIEN-MFC, LLFlow, EnlightenGAN), an MEF baseline (EMEF), and **AutoLumNet (ours)**. AutoLumNet restores shadow detail without amplifying noise, suppresses highlight clipping, and delivers balanced color/contrast across the scene. Please zoom in to inspect textures and highlight boundaries.

All models are trained on SICE with input images resized to 256×256 for efficiency. We use the AdamW optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.999$, a learning rate of 1×10^{-4}, and weight decay of 1×10^{-2}. The learning rate is scheduled using cosine annealing with 1,000 warm-up iterations. Training is run for 200 epochs with a batch size of 8 on a single Kaggle P100 GPU. Loss weights are set as $\lambda_{\text{rec}} = 1.0$, $\lambda_{\text{align}} = 0.5$, $\lambda_{\text{perc}} = 0.1$, and $\lambda_{\text{reg}} = 0.1$. Data augmentation includes random cropping, horizontal flipping, and color jittering.

For evaluation, we report results on LOL, and MIT-Adobe FiveK without fine-tuning. Metrics are computed on RGB images in $[0, 1]$ range. All results are averaged over the test splits, and inference speed is measured on 512×512 images. The code and pretrained models will be released upon acceptance.

4.3 COMPARISON WITH STATE-OF-THE-ART

We compare AutoLumNet against representative single-image enhancement and multi-exposure fusion methods, including RetinexNet (Wei et al., 2018b), KinD (Zhang et al., 2019), Zero-DCE (Guo et al., 2020), EnlightenGAN (Jiang et al., 2021), LIEN-MFC (Zhang et al., 2023a), LLFlow (Wang et al., 2022a), EMEF (Wang et al., 2022b), and Retinex-MEF (Zhao et al., 2023). These baselines cover both low-light enhancement and exposure-fusion families, allowing a balanced comparison across under-, over-, and mixed-exposure conditions.

Quantitative evaluation. Table 1 reports full-reference metrics (PSNR, SSIM, LPIPS) on subsets of SICE, LOL, and MIT-Adobe FiveK with available ground truth. AutoLumNet consistently outperforms all baselines across datasets, delivering both higher fidelity (PSNR, SSIM) and perceptual similarity (LPIPS).

No-reference evaluation. On datasets without paired ground truth, we employ NIQE, BRISQUE, PIQE, and MEF-SSIM (for fusion quality). Table 2 shows AutoLumNet achieves the lowest distortion scores and the highest MEF-SSIM, indicating both perceptual naturalness and exposure balance.

Qualitative results. Figure 1 provides visual comparisons. Competing methods either over-brighten shadows or fail to suppress highlight saturation, whereas AutoLumNet produces balanced illumination, preserves detail in both dark and bright regions, and avoids color distortions.

Table 3: Ablation study on SICE test set. Each component contributes to overall performance. Metrics: PSNR/SSIM (higher is better), NIQE (lower is better).

Variant	PSNR↑	SSIM↑	NIQE↓
w/o Dual Branch	22.9	0.82	4.87
w/o OT Alignment	23.4	0.83	4.69
w/o Perceptual Consistency	23.7	0.84	4.55
w/o Chroma Guard	23.8	0.85	4.50
w/o Regularization	23.5	0.84	4.62
Full AutoLumNet	**25.8**	**0.91**	**4.05**

4.4 ABLATION STUDY

We conduct ablation studies to verify the contribution of each design. Specifically, we analyze (i) single-branch variant (no decomposition), (ii) w/o exposure distribution alignment, (iii) w/o perceptual consistency, and (iv) full AutoLumNet.

Quantitative results. Table 3 shows that removing either the dual-branch design or distribution alignment significantly degrades both full- and no-reference metrics. The perceptual consistency term further improves LPIPS and NIQE, indicating better texture and perceptual realism.

5 CONCLUSION

We presented **AutoLumNet**, a bi-branch exposure-aware network for single-shot exposure correction that explicitly addresses both under- and over-exposed regions within a unified framework. The dual-branch decomposition allows the network to separately handle shadows and highlights, while the exposure distribution alignment module enforces global consistency through an optimal-transport formulation with provable monotonicity. A unified exposure-aware objective integrates reconstruction fidelity, statistical alignment, and perceptual consistency, yielding enhanced outputs that are both faithful and visually natural.

Extensive experiments on multiple benchmarks confirmed that AutoLumNet achieves state-of-the-art performance, surpassing both single-image enhancement and multi-exposure fusion methods across low-, high-, and mixed-exposure scenarios. Ablation studies further validated the necessity of each component, particularly the distribution alignment mechanism.

By bridging local correction with global distributional alignment, AutoLumNet establishes a principled approach to exposure correction that generalizes robustly to diverse real-world conditions. In future work, we plan to extend the framework to video exposure correction and to integrate more advanced perceptual priors for human-centric applications.

REFERENCES

Durga Prasad Bavirisetti, Koteswar Rao Prabhakar, and R Venkatesh Babu. Multi-gradient fusion for multi-exposure image enhancement. *IEEE Transactions on Image Processing*, 29:7204–7215, 2020.

Vladimir Bychkovsky, Sylvain Paris, Eric Chan, and Frédo Durand. Learning photo aesthetics with deep convolutional neural networks on the MIT-Adobe fivek dataset. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011.

Jiangjie Cai, Shuhang Gu, and Lei Zhang. Learning a deep single image contrast enhancer from multi-exposure images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3835–3844, 2018.

Chen Guo, Chongyi Li, Jichang Guo, Chen Change Loy, Junhui Hou, Sam Kwong, and Runmin Cong. Zero-reference deep curve estimation for low-light image enhancement. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1780–1789, 2020.

Xueyang Guo, Yu Li, and Haibin Ling. Lime: Low-light image enhancement via illumination map estimation. *IEEE Transactions on Image Processing*, 26(2):982–993, 2017.

Yifan Jiang, Xinyu Gong, Ding Liu, Yu Cheng, Chun-Liang Fang, Xiaohui Shen, Jianchao Yang, Pan Zhou, and Zhangyang Wang. Enlightengan: Deep light enhancement without paired supervision. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 16122–16131, 2021.

Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *European Conference on Computer Vision (ECCV)*, pp. 694–711, 2016.

Hui Li, Rui Song, Xiaojun Wang, Qing Liu, and Lei Zhang. Fmmef: Fully multi-modal multi-exposure image fusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10483–10492, 2020.

Kin Gwn Lore, Adedotun Akintayo, and Soumik Sarkar. Llnet: A deep autoencoder approach to natural low-light image enhancement. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 522–529, 2017.

Kai Ma, Kai Zeng, and Zhou Wang. A structural similarity index for multi-exposure image fusion. *IEEE Transactions on Image Processing*, 24(11):3345–3356, 2015.

Kai Ma, Hui Li, Hongwei Yong, Zhihai Wang, Deyu Meng, and Lei Zhang. Multi-exposure image fusion: A patch-wise approach. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 4085–4094, 2017.

Tom Mertens, Jan Kautz, and Frank Van Reeth. Exposure fusion. In *Pacific Conference on Computer Graphics and Applications (PG)*, pp. 382–390, 2007.

Kumar S. Prabhakar, Venkatesh R. Srikar, and R. Venkatesh Babu. Deepfuse: A deep unsupervised approach for exposure fusion. In *IEEE International Conference on Computer Vision (ICCV)*, 2017.

Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2015.

Yifan Wang, Jing Yu, Yulun Chen, Zhi Wang, Jinjin Gu, Chao Dong, Yu Qiao, and Lei Zhang. Llflow: Learning normalizing flows for low-light image enhancement. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, pp. 2604–2612, 2022a.

Yuting Wang, Jinjun Wu, Xiang Gao, Yulan Guo, and Jian Zhang. Emef: Learning exposure-guided multi-exposure image fusion. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11957–11966, 2022b.

Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Learning to enhance low-light image via paired synthesis. In *British Machine Vision Conference (BMVC)*, 2018a.

Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. In *Proceedings of the British Machine Vision Conference (BMVC)*, pp. 135, 2018b.

Zhen Wu, Hao Zheng, Linghao Ma, Wenhan Yang, Jiangxin Li, and Jiaying Liu. Uretinex-net: Retinex-based deep unfolding network for low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5901–5910, 2022.

Han Xu, Jiayi Ma, Junjun Jiang, Xiaojie Guo, and Haibin Ling. U2fusion: A unified unsupervised image fusion network. In *European Conference on Computer Vision (ECCV)*, 2020a.

Haoyu Xu, Kai Ma, Jian Jiang, Xiaojie Guo, and Lei Zhang. Fusiondn: A unified densely connected network for image fusion. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, pp. 12484–12491, 2020b.

Hong Xu, Kai Ma, and Lei Zhang. Mef-gan: Multi-exposure image fusion via generative adversarial networks. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, pp. 2978–2986, 2020c.

Yongjie Zhang, Jiawei Zhang, Chongyi Guo, and Chen Change Loy. Kindling the darkness: A practical low-light image enhancer. In *ACM Multimedia (ACM MM)*, pp. 1632–1640, 2019.

Yongqiang Zhang, Kai Ma, Hui Li, and Lei Zhang. Mefb: A benchmark dataset for multi-exposure image fusion. *IEEE Transactions on Image Processing*, 30:3568–3581, 2021.

Yu Zhang, Yudong Liu, Peng Sun, Ping Yan, and Xun Zhang. Ifcnn: A general image fusion framework based on convolutional neural network. In *IEEE International Conference on Information Fusion (FUSION)*, 2020a.

Yu Zhang, Yudong Liu, Peng Sun, Ping Yan, and Xun Zhang. Pmgi: Pyramid multi-modal fusion with graph inference. *Information Fusion*, 57:12–24, 2020b.

Yujie Zhang, Xu Wang, Jing Guo, and Zhiqiang Huang. Low-light image enhancement network based on multi-scale feature complementation. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, pp. 2459–2467, 2023a.

Yulun Zhang, Ding Liu, Yifan Wang, Shuhang Zhang, and Liang Lin. C- retinex: Boosting low-light image enhancement by color-aware retinex model. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 1230–1239, 2023b.

Qiming Zhao, Li Chen, Hong Xu, Hui Li, and Lei Zhang. Retinex-mef: Retinex-based glare effects aware unsupervised multi-exposure image fusion. *IEEE Transactions on Multimedia*, 2023.

Yu Zhu, Xiaohong Wang, Wei Chen, and Weijie Li. Attention-based network for low-light image enhancement. In *Proceedings of the International Conference on Image Processing (ICIP)*, pp. 3218–3222, 2021.