

Article

Not peer-reviewed version

A Swin-Transformer-Based Network for Adaptive Backlight Optimization

[Jin Li](#), Rui Pu, [Junbang Jiang](#)^{*}, [Man Zhu](#)^{*}

Posted Date: 27 January 2026

doi: 10.20944/preprints202601.2127.v1

Keywords: Mini-LED; adaptive local dimming; partitioned backlight; deep learning



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

A Swin-Transformer-Based Network for Adaptive Backlight Optimization

Jin Li ¹, Rui Pu ¹, Junbang Jiang ^{1,*} and Man Zhu ^{2,*}

¹ National "111 Research Center" Microelectronics and Integrated Circuits, School of Science, Hubei University of Technology, Wuhan 430068, China

² Intelligent Transportation Systems Research Center, Wuhan University of Technology, Wuhan 430063, China

* Correspondence: jiangjunbang@hbut.edu.cn (J.J.); man.zhu.393@whut.edu.cn (M.Z.)

Abstract

To address luminance discontinuity, halo artifacts, and insufficient temporal stability commonly observed in Mini-LED local dimming systems, this paper proposes an adaptive backlight optimization network, termed SwinLightNet, based on a hierarchical attention mechanism. From a luminance modeling perspective, the proposed method exploits multi-scale feature correlations to achieve spatially smooth and content-adaptive backlight distributions, while incorporating temporal luminance constraints to enhance stability in video scenes. Experimental results demonstrate that SwinLightNet consistently outperforms conventional local dimming algorithms and representative learning-based methods in terms of PSNR, SSIM, and subjective visual quality, validating its effectiveness for Mini-LED backlight optimization.

Keywords: Mini-LED; adaptive local dimming; partitioned backlight; deep learning

1. Introduction

In recent years, advances in Mini-LED backlight modules have enabled display systems to surpass conventional liquid crystal displays (LCDs) in several key aspects, including higher contrast ratios, more precise local dimming control, wider color gamut, reduced light leakage and flicker, as well as improved energy efficiency [1–3]. Display technology serves as one of the most important means for information presentation and acquisition [4,5], and is deeply integrated into daily life through devices such as smartphones, tablets, televisions, projectors, and virtual/augmented reality systems. With the growing demand for enhanced visual quality, high dynamic range (HDR) displays have emerged as an effective solution. Compared with conventional images, HDR content provides a wider dynamic range and richer details, enabling more faithful reproduction of natural scenes. This typically requires high peak luminance, strong dark-level performance, accurate grayscale representation, wide color gamut, and high bit depth [6–8].

In Mini-LED-based display systems, image luminance is not solely determined by individual pixels but results from the joint modulation of the backlight and pixel transmittance. Since the spatial resolution of backlight dimming zones is significantly lower than that of display pixels, the backlight modulation process inherently suffers from spatial under-sampling. When images contain high-contrast structures, this mismatch often leads to halo artifacts and luminance leakage near region boundaries, thereby degrading overall visual consistency [9–11].

To alleviate these issues, existing studies have explored increasing the number of backlight zones or designing more sophisticated local dimming strategies. Song *et al.* [12] proposed a deep neural network-based pixel compensation method for quantum-dot backlit LCDs, which directly generates compensation images from the input to optimize zone-level luminance allocation. Chia *et al.* [13] employed convolutional neural networks (CNNs) for backlight luminance control, overcoming the limited generalization capability of hand-crafted features, particularly for HDR content. Zheng *et al.* [14] introduced a lightweight CNN framework for real-time HDR backlight

dimming with improved computational efficiency. Zhang *et al.* [15] applied deep CNNs to enhance the contrast of dual-modulation LCD systems while significantly reducing power consumption. Han *et al.* proposed a deep reinforcement learning framework based on decomposed graph scheduling, combining the advantages of double deep Q-networks (DDQN) to achieve both flexible policy learning and real-time responsiveness [16]. Salh *et al.* further integrated generative adversarial networks (GANs) with deep distributed Q-networks for intelligent scheduling in Internet of Things (IoT) environments, improving training stability and decision-making capability [17,18]. However, despite their effectiveness, CNN-based models remain constrained by local receptive fields when modeling long-range global dependencies, motivating the exploration of alternative architectures for more comprehensive spatio-temporal luminance modeling.

Unlike convolutional structures, Transformer architectures leverage self-attention mechanisms to model global relationships across the entire spatial domain, offering new opportunities for backlight optimization. Nevertheless, directly adopting standard Transformers often incurs high computational complexity, making it difficult to satisfy the real-time and stability requirements of display systems.

To address these challenges, this paper proposes a backlight optimization network termed SwinLightNet, based on a lightweight Swin Transformer architecture. By employing hierarchical shifted-window self-attention, the proposed network enables more accurate modeling of both global and local luminance relationships, resulting in smoother and more uniform backlight distributions. This leads to notable improvements in peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM), while effectively suppressing halo artifacts and enhancing overall visual quality. For video sequences, SwinLightNet further incorporates a temporal fusion module that performs dynamic backlight adjustment based on inter-frame luminance similarity, constraining abrupt zone-level luminance variations between consecutive frames. As a result, luminance flicker caused by block-wise transitions is significantly reduced, improving viewing comfort in dynamic scenes.

2. Adaptive Spatial-Contrast-Enhanced Local Dimming Method Based on Mini-LED Technology

2.1. Overall System Design

To achieve fine-grained backlight modulation for complex image content, this work develops an adaptive backlight optimization network, termed SwinLightNet, based on the Swin Transformer architecture. Driven by local luminance characteristics, the proposed network learns region-level backlight response relationships through multi-scale spatial modeling, enabling a unified formulation and performance enhancement over conventional statistical backlight strategies.

During the spatial modeling stage, the input image is mapped to a partition-level luminance representation, from which an initial backlight estimate is generated via a luminance-aware mapping. This estimation captures the energy distribution of dark and bright regions while incorporating an adaptive luminance adjustment mechanism to provide differentiated responses across varying intensity ranges, thereby enhancing image details without excessive amplification. Subsequently, a local neighborhood constraint is applied to refine the backlight estimation, enforcing spatial consistency and effectively suppressing luminance discontinuities at partition boundaries.

To address flicker artifacts caused by temporal variations of backlight in video sequences, SwinLightNet introduces an adaptive temporal fusion strategy based on inter-frame luminance statistics. This strategy jointly characterizes luminance variation trends between adjacent frames and employs dynamic regulation to constrain extreme changes, thereby improving temporal smoothness while preserving content responsiveness.

At the output stage, the optimized backlight distribution is mapped to pixel-level space to construct a continuous full-resolution backlight layer, which is then combined with the original image through luminance-guided fusion and compensation. Through this collaborative optimization

process, SwinLightNet consistently produces high-quality backlight distributions across diverse content scenarios, significantly enhancing image contrast and overall visual consistency.

Figure 1 illustrates the overall workflow of the SwinLightNet algorithm. By integrating multiple optimization strategies during backlight computation, the proposed framework preserves image details, improves display quality, enhances resource utilization efficiency, and reduces hardware power consumption.

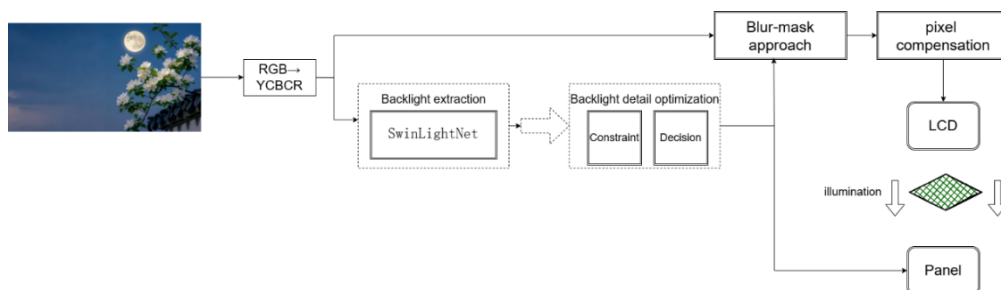


Figure 1. Flowchart of the SwinLightNet Algorithm.

2.2. Backlight Extraction

In LED display systems, accurate estimation of backlight distribution directly determines image luminance hierarchy and local contrast performance, making backlight modeling a critical component of the display optimization pipeline. As the foundational module of the SwinLightNet framework, the backlight extraction module aims to construct a spatially adaptive backlight representation from the input image, providing reliable input for subsequent constraint enforcement and temporal optimization.

This module performs luminance modeling on a block-level backlight control grid constrained by the physical partition structure of Mini-LED displays. To better align with the high sensitivity of the human visual system to luminance variations, the input RGB image is first converted into the YCrCb color space, and only the luminance component is retained for backlight computation, thereby reducing the influence of chromatic information on backlight estimation.

Subsequently, the luminance channel is divided into $r_b \times c_b$ non-overlapping subregions according to a predefined partition scheme, with each region corresponding to a physical LED backlight unit. Given an input image resolution of $H \times W$, the spatial dimensions of each backlight partition are defined as

$$h_b = \left\lfloor \frac{H}{r_b} \right\rfloor, w_b = \left\lfloor \frac{W}{c_b} \right\rfloor \quad (1)$$

Here, r_b and c_b denote the numbers of backlight partitions along the vertical and horizontal directions, respectively. Through this mapping, pixel-level luminance information is transformed into block-level representations consistent with the hardware structure, laying the foundation for constructing high-fidelity and spatially consistent backlight distributions.

2.2.1. Network Architecture

This paper proposes an end-to-end network for predicting brightness adjustment factors that combines convolutional feature extraction with the global modeling capabilities of the Swin Transformer in order to address the limitations of convolutional networks in capturing local details and global structural information in low-resolution local brightness adjustment tasks. The technique seeks to produce high-resolution factor distributions for backlight modification through upsampling and precisely restore spatially continuous and structurally consistent brightness dimming coefficient maps from low-resolution inputs. As shown in Figure 2, the complete network consists of three parts: an encoder, a Swin Transformer feature augmentation module, and a decoder.

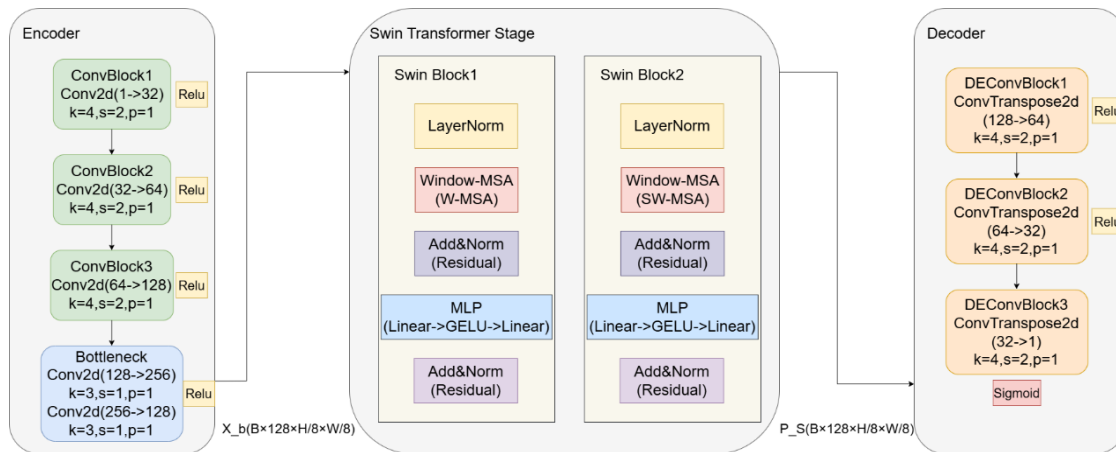


Figure 2. Network Architecture Diagram.

Three subsampling convolutional layers, each using a 4×4 convolution with stride 2 and padding 1, make up the encoder. In order to successfully extract multi-scale texture and structural features, this gradually lowers the input feature resolution to $H/2$, $H/4$, and $H/8$ while increasing the number of channels to 32, 64, and 128, respectively. All convolutional layers employ the ReLU activation function to enhance nonlinear modeling capability. The Bottleneck module then uses two consecutive 3×3 convolutions to further improve the latent representation.

This study offers a two-layer Swin Transformer module between the encoder and decoder to overcome the shortcomings of convolutions in representing global illumination trends and long-range dependency. Each Swin Block consists of an MLP (Linear \rightarrow GELU \rightarrow Linear), residual connections, windowed multi-head self-attention (W-MSA/SW-MSA), and LayerNorm. The network may compute effective attention inside local windows while improving global modeling capabilities through cross-window correlations by alternating between regular and shifted windows, yielding an enhanced feature representation P_s with superior edge preservation and global consistency. After processing through two Swin Block levels.

In the end, the decoder gradually reduces the number of feature channels from $128 \rightarrow 64 \rightarrow 32 \rightarrow 1$ by gradually restoring spatial resolution through three deconvolution modules. Lastly, it creates a brightness adjustment factor map adjusted to the $[0,1]$ range using a Sigmoid activation function. In order to create the final backlight adjustment layer, this map is further upsampled to full resolution using bilinear interpolation.

2.2.2. Loss Function

In this study, we developed a multi-loss joint optimization framework to guarantee that the projected luminance factor map achieves high accuracy, strong structural consistency, and spatial smoothness. The total loss function consists of three components: to improve the structural preservation capability of predictions in edge and texture regions, a structural similarity loss term

$$\mathcal{L}_{SSIM} = 1 - SSIM(X_d, X_{gt}) \quad (2)$$

is introduced. First, the L1 loss term $\mathcal{L}_{L1} = \|X_d - X_{gt}\|_1$ limits the pixel-level error between the predicted value X_d and the ground truth X_{gt} , guaranteeing overall luminance consistency. Lastly, a gradient-based smoothing constraint

$$\mathcal{L}_{smooth} = \|\nabla_x X_d\|_1 + \|\nabla_y X_d\|_1 \quad (3)$$

is used to suppress noise and lessen local oscillations, a gradient-based smoothing constraint is introduced to enforce spatial continuity of the brightness factor map. After incorporating these elements, the overall loss function is described as

$$L = \lambda_1 L_{L1} + \lambda_2 L_{SSIM} + \lambda_3 L_{smooth} \quad (4)$$

where λ_1 , λ_2 , and λ_3 are the weight parameters. This seeks to achieve more accurate and aesthetically pleasing brightness control effects by balancing the impact of various loss terms on the final model performance.

2.3. Backlight Constraint

The backlight constraint module is positioned between backlight estimation and temporal optimization within the SwinLightNet framework. Its primary purpose is to structurally regulate the initial backlight distribution without compromising content adaptivity, thereby enhancing spatial continuity and luminance stability. Unlike approaches that directly limit backlight amplitudes, this module constrains the relative relationships among backlight zones from the perspectives of local consistency and contrast preservation, effectively alleviating block artifacts and abrupt luminance transitions commonly observed in local dimming systems.

The module operates on an $r_b \times c_b$ backlight partition grid and consists of two consecutive stages: local luminance consistency enforcement and contrast-preserving adaptive adjustment. The output is a constrained backlight grid that serves as a stable input for subsequent temporal fusion.

2.3.1. Local Luminance Consistency Constraint

After initial backlight estimation, discontinuous luminance variations may arise between adjacent partitions due to content differences or estimation errors. To address this issue, a neighborhood-based local consistency constraint is introduced to structurally refine the luminance of each backlight partition.

Let $B_{i,j}^{raw}$ denote the initial backlight estimate. The local reference luminance is defined as a weighted average of neighboring partitions:

$$\bar{B}_{i,j} = \sum_{(m,n) \in \mathcal{N}_{i,j}} w_{m,n} B_{m,n}^{raw} \quad (5)$$

where $\mathcal{N}_{i,j}$ represents the neighborhood set of partition (i,j) , and the weights $w_{m,n}$ are normalized according to spatial distance. Based on this reference, the luminance of the current partition is adjusted as

$$B_{i,j}^{(0)} = B_{i,j}^{raw} + \alpha (\bar{B}_{i,j} - B_{i,j}^{raw}) \quad (6)$$

where $\alpha \in [0,1]$ is a tuning parameter that balances the degree of local smoothing and the original luminance response. By reducing abnormal inter-partition differences, this constraint effectively mitigates luminance discontinuities at block boundaries while preserving the overall luminance structure.

2.3.2. Contrast-Preserving Adaptive Adjustment

Following local consistency refinement, a contrast-preserving adaptive adjustment mechanism is applied to further enhance the stability of the backlight distribution across varying luminance conditions. Instead of imposing hard clipping, this mechanism suppresses the influence of extreme values through a luminance compression function.

Let B_{\min} and B_{\max} denote the global minimum and maximum values of the current backlight grid, respectively. The constrained luminance of each partition is obtained by

$$B_{i,j}^{con} = B_{\min} + (B_{\max} - B_{\min}) \cdot \frac{B_{i,j}^{(0)} - B_{\min}}{B_{\max} - B_{\min} + \varepsilon} \quad (7)$$

where ε is a stabilization term introduced to prevent numerical amplification under extreme conditions. This mapping compresses the dynamic range while preserving relative contrast relationships among partitions, maintaining luminance hierarchy between bright and dark regions.

Through the two-stage constraint process, the resulting backlight distribution exhibits improved spatial continuity and smoother luminance transitions, effectively suppressing flicker and halo artifacts in local dimming. The proposed constraint strategy is entirely based on software-level statistical and structural modeling, facilitating parameter tuning and scene adaptivity, and providing robust support for the stable performance of SwinLightNet in video display applications.

2.4. Optimal Backlight Decision

The optimal backlight decision module is designed to perform temporal coordination and optimization of backlight distributions in video sequences. Its objective is to suppress flicker and abrupt luminance transitions caused by inter-frame variations while preserving responsiveness to image content. Operating on block-level backlight representations, the module models the variation trends of backlight distributions across consecutive frames and applies adaptive temporal correction to the current backlight estimation, thereby generating a temporally continuous and stable final backlight output.

The module processes backlight data on an $r_b \times c_b$ partition grid, and the resulting backlight distribution is directly forwarded to the pixel compensation stage, making it particularly suitable for display optimization in dynamic scenes.

2.4.1. Change-Aware Inter-Frame Backlight Modeling

During temporal decision making, the backlight estimation of the current frame serves as the primary information source. Let B_t^{cur} denote the spatially refined backlight estimation of the current frame, and B_{t-1}^{final} denote the final backlight result of the previous frame. The difference between these two reflects the intensity of scene luminance variation. Accordingly, a block-level difference-based variation metric is introduced:

$$D_t = \frac{1}{r_b c_b} \sum_{i,j} | B_t^{\text{cur}}(i,j) - B_{t-1}^{\text{final}}(i,j) | \quad (8)$$

where D_t characterizes the overall magnitude of backlight change between adjacent frames. When no previous frame is available, D_t is set to zero by default, indicating that no temporal constraint is applied.

2.4.2. Adaptive Incremental Temporal Adjustment

To prevent excessive reliance on historical information under rapid scene changes, a dynamic constraint mechanism is introduced to regulate temporal fusion. Let g_t denote a dynamic upper-bound function, and the final temporal fusion weight is defined as

$$\omega_t = \min(s_t, g_t) \quad (9)$$

where ω_t is further constrained within the interval $[\omega_{\min}, \omega_{\max}]$ to balance temporal stability and responsiveness.

Based on this weight, temporal modulation is applied to the current backlight estimation, yielding the fused backlight distribution:

$$B_t^f = \omega_t \cdot B_t \quad (10)$$

This fusion strategy preserves content adaptivity of the current frame while introducing implicit temporal continuity, resulting in smoother and more natural backlight transitions, particularly in video sequence processing.

2.4.3. Output Normalization and Module Analysis

To ensure numerical stability and reliability, a unified normalization process is applied to the final backlight result:

$$B_t^{\text{out}} = \mathcal{N}(B_t^{\text{final}}) \quad (11)$$

where $\mathcal{N}(\cdot)$ denotes normalization and range-mapping operations that constrain backlight intensities to a valid and stable range.

After processing by this module, the output backlight grid exhibits strong temporal smoothness while maintaining the ability to rapidly adapt to scene luminance variations. By integrating change-aware modeling, incremental temporal updating, and numerical normalization into a unified decision pipeline, the proposed module effectively mitigates flicker and luminance discontinuities in video display, providing critical support for achieving high visual quality with SwinLightNet under complex dynamic content.

3. Simulation Design

In the simulation experiments, the block-level backlight results generated by the backlight optimization algorithm are first subjected to spatial smoothing to construct a continuous backlight distribution. Subsequently, a pixel-level compensation mechanism is applied to perform luminance fusion and correction on the smoothed backlight data, effectively alleviating luminance discontinuities and color imbalance caused by block boundaries, and ensuring visual consistency and naturalness in the final output images. During evaluation, multiple objective image quality metrics are employed to quantitatively analyze the simulation results of different methods, enabling a comprehensive assessment of their performance in luminance reconstruction accuracy and perceptual quality.

3.1. Pixel Compensation

The pixel compensation module serves as the final display mapping stage of the SwinLightNet framework. Its primary function is to convert block-level optimized backlight results into continuous pixel-level modulation signals and to perform luminance-consistent adjustment with the original image, thereby enhancing local contrast and overall dynamic range. Designed with perceptual consistency as the objective, this module generates the final output images suitable for Mini-LED display systems through multi-scale propagation, luminance reconstruction, and constraint-aware fusion.

Since the output of the backlight decision module is represented as a low-resolution backlight grid, direct upsampling may introduce perceptible discontinuities at block boundaries. To address this issue, instead of employing simple interpolation, a continuous propagation-based pixel-domain backlight reconstruction strategy is adopted.

Let $B_t^{\text{final}}(i, j)$ denote the final backlight grid. It is first mapped to a sparse guidance representation, and a continuous backlight distribution is then generated in the pixel domain via distance-weighted propagation:

$$B^{\text{cont}}(x, y) = \frac{\sum_{i,j} B_t^{\text{final}}(i, j) w_{i,j}(x, y)}{\sum_{i,j} w_{i,j}(x, y)} \quad (12)$$

where the weighting function $w_{i,j}(x, y)$ is determined by the spatial distance between the pixel location and the center of the corresponding backlight partition, ensuring smooth transitions between neighboring regions. This formulation avoids explicit block boundaries and yields a spatially continuous backlight distribution in the pixel domain.

After obtaining the continuous backlight distribution, it is interpreted as a pixel-level transmittance modulation factor of the display system rather than a direct luminance addition term. To enhance the display dynamic range, a nonlinear remapping is applied:

$$T(x, y) = \left(\frac{B^{\text{cont}}(x, y)}{B_{\text{max}}} \right)^p \quad (13)$$

where ρ denotes the transmittance adjustment exponent that controls the balance between highlight expansion and dark-region compression. This nonlinear modeling more closely reflects the coupling between backlight intensity and liquid crystal transmittance in real display systems, contributing to improved highlight clarity while preserving dark-region details.

The final output image is obtained by applying pixel-level transmittance modulation to the original input image. Let $I(x,y)$ denote the luminance component of the input image, then the compensated luminance is given by

$$I^{\text{out}}(x,y) = I(x,y) \cdot T(x,y) \quad (14)$$

This multiplicative modulation avoids contrast compression that may arise from linear weighted fusion, allowing backlight adjustment to directly influence pixel luminance representation and thereby enhance local contrast.

To ensure display safety and numerical stability, a global constraint is applied to the output:

$$I^{\text{final}}(x,y) = \mathcal{C}(I^{\text{out}}(x,y)) \quad (15)$$

where $\mathcal{C}(\cdot)$ denotes luminance clipping and dynamic range mapping operations, ensuring that the final image remains within the valid luminance range of the display system.

By interpreting the backlight result as a continuous transmittance modulation signal rather than a simple luminance compensation term, the proposed pixel compensation module more faithfully simulates the collaborative relationship between backlight control and pixel response in Mini-LED display systems. This approach effectively suppresses block artifacts, significantly enhances local contrast, and maintains smooth luminance transitions in dynamic scenes, providing high-quality, perceptually consistent visual output for the SwinLightNet framework.

3.2. Experimental Results

In this experiment, several backlight extraction algorithms are employed for performance comparison, including the maximum-value method, mean-value method, root mean square (RMS) method, standard deviation-based method, error correction method, and the cumulative distribution function (CDF) thresholding method [19–23]. Since subjective visual assessment inherently involves a certain degree of randomness, objective image quality metrics are required to provide a comprehensive and quantitative evaluation of the processed images.

Accordingly, peak signal-to-noise ratio (PSNR), information entropy (IE), and structural similarity index (SSIM) are adopted as objective evaluation metrics in this study [24]. These metrics are used to quantitatively assess image fidelity before and after processing, contrast enhancement effectiveness, power-saving potential, information richness, and structural similarity.

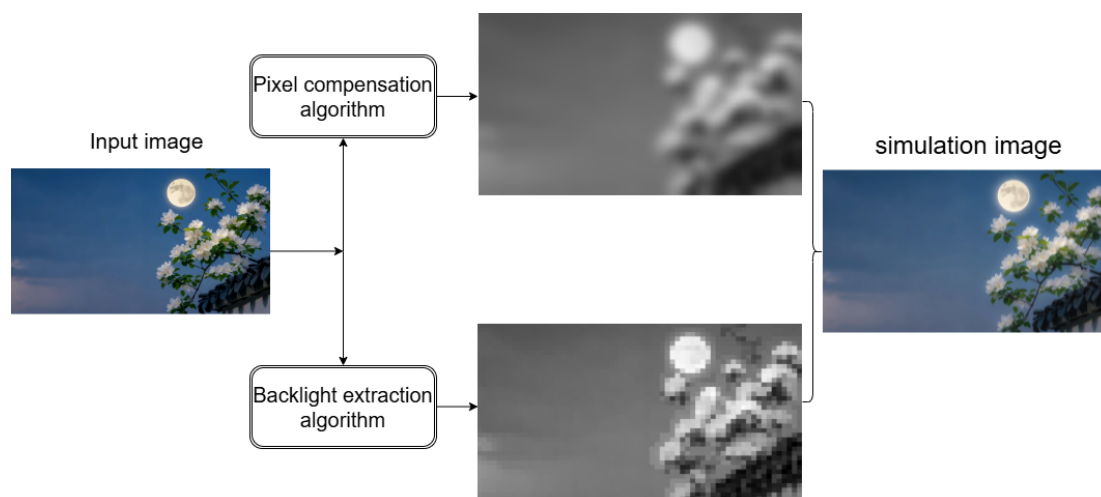


Figure 3. Schematic illustration of the algorithm simulation.

PSNR measures the noise level between the original image I and the processed image K , and is widely used to evaluate image fidelity. Information entropy (IE) reflects the richness of image information and is employed to assess the degree of detail preservation. SSIM evaluates the structural similarity between the original image x and the processed image y by jointly considering luminance, contrast, and structural components. In SSIM computation, μ_x and μ_y denote the mean intensities, σ_x^2 and σ_y^2 represent the variances, and σ_{xy} denotes the covariance between the two images. The stabilizing constants are defined as $c_1 = (k_1L)^2$ and $c_2 = (k_2L)^2$, where $L = 255$, $k_1 = 0.01$, and $k_2 = 0.03$. In local dimming applications, an SSIM value closer to 1 indicates higher structural fidelity.

PSNR, SSIM and IE are computed as follows:

$$\begin{aligned} \text{PR} &= \left(1 - \frac{\sum_{m,n} k_{m,n}}{r_b \times c_b}\right) \times 100\%, \\ \text{IE} &= - \sum_{i=0}^{255} p(i) \log_2 p(i), \\ \text{SSIM}(x, y) &= \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}. \end{aligned} \quad (16)$$

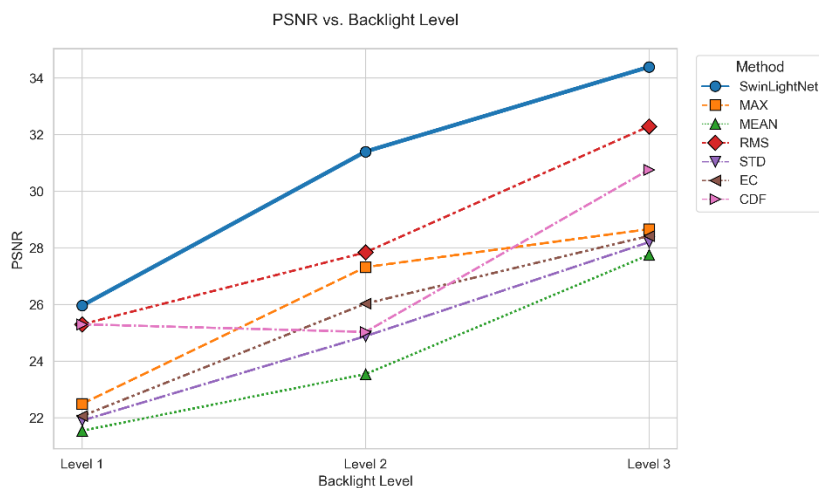


Figure 4. Line Chart of PSNR Improvement Rate for Different Algorithms.

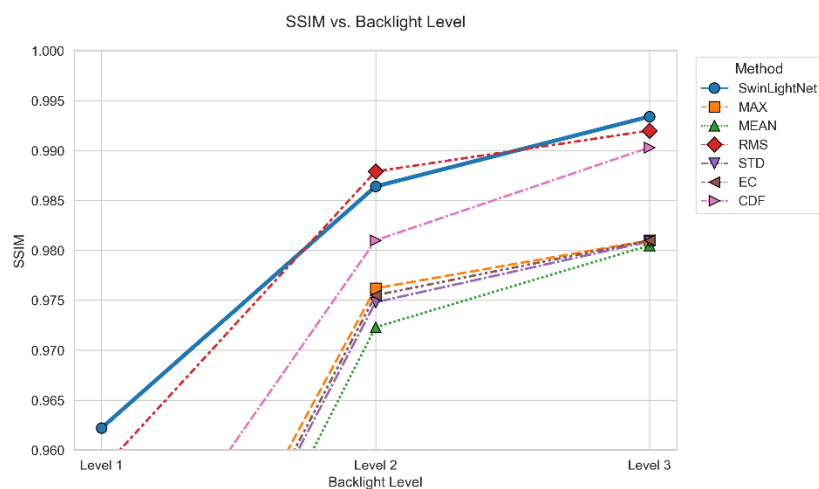


Figure 5. Line Chart of SSIM Improvement Rate for Different Algorithms.

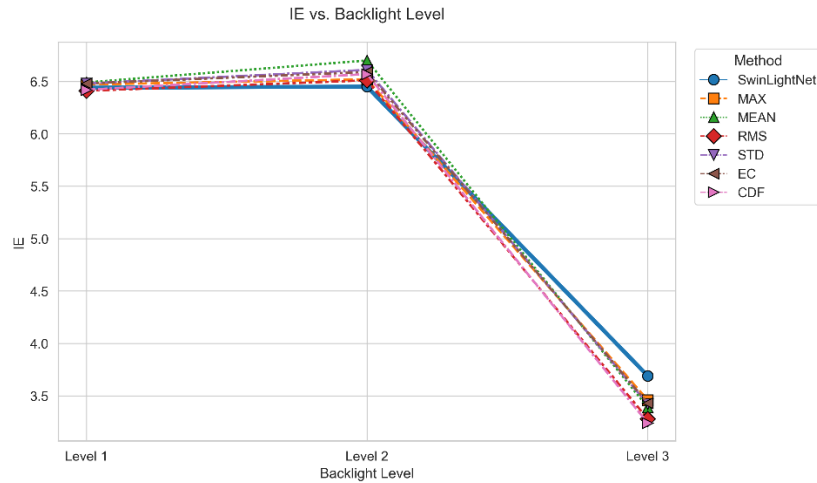


Figure 6. Line Chart of IE Improvement Rate for Different Algorithms.

Quantitative results on three representative test scenarios demonstrate that the proposed SwinLightNet consistently outperforms the compared methods in terms of overall performance and stability. In terms of PSNR, SwinLightNet achieves the highest or near-highest values across all three test sets, with particularly pronounced advantages in the second and third scenarios involving complex luminance variations, indicating improved robustness in structural fidelity and luminance reconstruction accuracy.

Regarding the perceptual ratio (PR), SwinLightNet shows a clear advantage over traditional statistical backlight methods in the first test scenario and maintains stable performance in the remaining cases, reflecting its ability to enhance visual contrast while effectively suppressing distortion. Meanwhile, the entropy distribution produced by SwinLightNet remains well balanced, avoiding the luminance noise and instability introduced by over-enhancement in methods such as MEAN and MAX.

Overall, SwinLightNet achieves a favorable trade-off among structural accuracy, perceptual enhancement, and luminance stability, validating its effectiveness in generating high-quality Mini-LED backlight distributions under diverse content conditions.

As shown in Table 1, the proposed Swin-LightNet achieves the best overall performance on the low-contrast image set. Specifically, Swin-LightNet obtains the highest PSNR value (21.22), indicating superior luminance reconstruction accuracy compared with traditional statistical backlight methods. In terms of structural similarity, Swin-LightNet reaches an SSIM of 0.96, which is the highest among all compared methods and equal only to RMS, demonstrating its strong ability to preserve image structures under low-contrast conditions.

Table 1. Performance Comparison of Different Algorithms.

Backlight Method	SwinLightNet	MAX	MEAN	RMS	STD	EC	CDF
PSNR	21.22	17.63	17.19	20.69	17.41	17.44	19.34
IE	6.87	6.83	6.80	6.78	6.80	6.82	6.75
SSIM	0.96	0.93	0.92	0.96	0.92	0.92	0.92

Moreover, Swin-LightNet maintains a high information entropy (IE = 6.87), reflecting effective detail preservation without introducing excessive noise. In contrast, methods such as MAX and MEAN exhibit lower PSNR and SSIM values, suggesting limited capability in handling low-contrast content. Overall, these results confirm that Swin-LightNet provides a more balanced and robust backlight optimization strategy for low-contrast images.

As shown in Table 2, the proposed SwinLightNet achieves the best objective performance in the backlight control task. Specifically, SwinLightNet attains a PSNR of 34.39 dB, which is higher than

those of References [13,15], and [14] by 4.89 dB, 9.08 dB, and 0.52 dB, respectively, and exceeds the baseline model without the Swin Transformer by 3.93 dB. In terms of structural similarity, SwinLightNet achieves an SSIM of 0.99, maintaining the highest level among all compared methods.

Table 2. Performance Comparison of Different Algorithms.

Backlight Method	SwinLightNet	Reference [13]	Reference [15]	Reference [14]
PSNR	34.39	29.50	25.31	33.87
SSIM	0.99	0.87	0.90	0.99

The observed performance gains can be primarily attributed to the joint local–global modeling capability introduced by the Swin Transformer, which enables more accurate recovery of structural information and fine details in high-contrast and complex-texture regions. Moreover, while preserving a lightweight network architecture, SwinLightNet achieves high-quality backlight reconstruction, providing an efficient and practical solution for HDR backlight optimization.

4. Conclusions

This paper proposes an adaptive backlight optimization network, SwinLightNet, based on the Swin Transformer, to address the challenges of insufficient dark-region detail, pronounced halo artifacts, and the trade-off between power consumption and stability in Mini-LED partitioned backlight displays. The proposed method employs joint local–global modeling to predict backlight luminance distributions and incorporates luminance-gradient constraints together with temporal fusion strategies, enabling coordinated optimization of spatial smoothness and dynamic-scene stability.

Experimental results demonstrate that SwinLightNet consistently outperforms conventional local dimming approaches and existing deep learning-based methods in terms of PSNR and SSIM across various representative scenarios. In particular, the proposed network exhibits superior structural preservation in regions with complex textures and high contrast, while achieving favorable energy-saving performance. Furthermore, FPGA-based implementation verifies that SwinLightNet maintains high display quality with good real-time performance and hardware feasibility. Overall, SwinLightNet provides an efficient and stable dimming solution for high-resolution Mini-LED backlight systems, offering substantial engineering application potential.

Author Contributions: Conceptualization, J.J. , P.R. , J.L. and M.Z.; methodology, J.J. , P.R. and J.L.; software, J.J. , P.R. and J.L.; validation, J.J. , P.R. and J.L.; formal analysis, J.J. , P.R. and J.L.; investigation, J.J. , P.R. and J.L.; resources, J.J. , P.R. and J.L.; data curation, J.J. , P.R. and J.L.; writing—original draft preparation, J.J. , P.R. and J.L.; writing—review and editing, M.Z.; supervision, M.Z.; funding acquisition, M.Z. All authors have read and agreed to the published version of the manuscript.

Funding: Guiding project of Scientific Research Plan of Hubei Provincial Department of Education, Design of Application-Specific Integrated Circuit for Voice Command Recognition Based on Compute-in-Memory Architecture (No. B2020046); Cooperation Agreement on the Project of Design and Process Manufacturing Technology of 3D Stacked Code Flash Memory Chip (No. 2021802); PhD Research Foundation Project of Hubei University of Technology, Research on Sun Tracking Control and System State Monitoring of Butterfly Concentrating Photovoltaic Based on Deep Learning Image Analysis (No. 00185).

Data Availability Statement: The DIV2K dataset is open to the public and can be obtained from New Trends in Image Restoration and Enhancement website: <https://data.vision.ee.ethz.ch/cv1/DIV2K/>.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Yang Z.; Hsiang E.L.; Qian Y.; et al. Performance comparison between mini-LED backlit LCD and OLED display for 15.6-inch notebook computers[J]. *Applied Sciences*, 2022, 12(3): 1239. [[CrossRef](#)]
2. Gao Z.; Ning H.; Yao R.; et al. Mini-LED backlight technology progress for liquid crystal display[J]. *Crystals*, 2022, 12(3): 313. [[CrossRef](#)]
3. Chen E.; Guo J.; Jiang Z.; et al. Edge/direct-lit hybrid mini-LED backlight with U-grooved light guiding plates for local dimming[J]. *Optics Express*, 2021, 29(8): 12179-12194. [[CrossRef](#)]
4. Schmidt, M.; Grüning, M.; Ritter, J.; et al. Impact of High-Resolution Matrix Backlight on Local Dimming Performance and Its Characterization. *J. Inf. Display* 2019, 20(2), 95–104. [[CrossRef](#)]
5. Lei J.; Zhu H.; Huang X.; et al. Mini-LED Backlight: Advances and Future Perspectives[J]. *Crystals*, 2024, 14(11): 922. [[CrossRef](#)]
6. Kwon J U.; Bang S.; Kang D.; et al. 65-2: The Required Attribute of Displays for High Dynamic Range[C]//SID Symposium Digest of Technical Papers. 2016, 47(1): 884-887. [[CrossRef](#)]
7. Kang, S.J.; Bae, S. Fast Segmentation-Based Backlight Dimming. *J. Display Technol.* 2015, 11(5), 399–402. [[CrossRef](#)]
8. Zhu R.; Sarkar A.; Emerton N.; et al. 81-3: Reproducing High Dynamic Range Contents Adaptively based on Display Specifications[C]//SID Symposium Digest of Technical Papers. 2017, 48(1): 1188-1191. [[CrossRef](#)]
9. Chen E.; Fan Z.; Zhang K.; et al. Broadband beam collimation metasurface for full-color micro-LED displays[J]. *Optics Express*, 2024, 32(6): 10252-10264. [[CrossRef](#)]
10. Zheng X.; Guo W.; Tong C.; et al. Origin of the inhomogeneous electroluminescence of GaN-based green mini-LEDs unveiled by microscopic hyperspectral imaging[J]. *ACS Photonics*, 2022, 9(11): 3685-3695. [[CrossRef](#)]
11. Tong C.; Yang H.; Zheng X.; et al. Luminous characteristics of RGBW mini-LED integrated matrix devices for healthy displays[J]. *Optics & Laser Technology*, 2024, 170: 110229. [[CrossRef](#)]
12. Song, S.J.; Kim, Y.I.; Bae, J.; et al. Deep Learning Based Pixel Compensation Algorithm For Local Dimming Liquid Crystal Displays of Quantum-dot Backlights. *Opt. Express* 2019, 27(11), 15907–15917. [[CrossRef](#)]
13. Chia, T.L.; Syu, Y.Y.; Huang, P.S. A Novel Local Dimming Approach by Controlling LCD Backlight Modules via Deep Learning. *Information* 2025, 16(9), 815. [[CrossRef](#)]
14. Zheng, Y.; Wang, W. Lightweight CNN-Based Local Backlight Dimming for HDR Displays. In *Proc. 37th Chinese Control and Decision Conference (CCDC); IEEE: 2025; pp. 365–370.* [[CrossRef](#)]
15. Zhang, T.; Wang, H.; Du, W.; et al. Deep CNN-Based Local Dimming Technology. *Appl. Intell.* 2022, 52(1), 903–915. [[CrossRef](#)]
16. Han B A. ; Yang J.J. Research on adaptive job shop scheduling problems based on dueling double DQN[J]. *Ieee Access*, 2020, 8: 186474-186495. [[CrossRef](#)]
17. Salh A.; Audah L.; Alhartomi M.A.; et al. Smart packet transmission scheduling in cognitive IoT systems: DDQN based approach[J]. *IEEe Access*, 2022, 10: 50023-50036. [[CrossRef](#)]
18. Salh A.; Audah L.; Kim K.S.; et al. Refiner GAN algorithmically enabled deep-RL for guaranteed traffic packets in real-time URLLC B5G communication systems[J]. *IEEE Access*, 2022, 10: 50662-50676. [[CrossRef](#)]
19. Lin F.C.; Huang Y.P.; Liao L.Y.; et al. Dynamic backlight gamma on high dynamic range LCD TVs[J]. *Journal of Display Technology*, 2008, 4(2): 139-146. [[CrossRef](#)]
20. Zhang, X.B.; Wang, R.; Dong, D.; et al. Dynamic Backlight Adaptation Based On The Details Of Image For Liquid Crystal Displays. *Journal of Display Technol.* 2012, 8(2), 108–111. [[CrossRef](#)]
21. Seetzen H.; Whitehead L.A.; Ward G.; 54.2: A high dynamic range display using low and high resolution modulators[C]//SID Symposium Digest of Technical Papers. Oxford, UK: Blackwell Publishing Ltd, 2003, 34(1): 1450-1453. [[CrossRef](#)]
22. Cho H.; Kwon O.K.; A backlight dimming algorithm for low power and high image quality LCD applications[J]. *IEEE Transactions on Consumer Electronics*, 2009, 55(2): 839-844. [[CrossRef](#)]
23. Yan-zhong LIU.; Xue-ren Z.; Jian-bin C.; Dynamic backlight signal extraction algorithm based on threshold of image CDF for LCD-TV and its hardware implementation[J]. *Chinese Journal of Liquid Crystals and Displays*, 2020, 25(3): 449-453. [[CrossRef](#)]

24. Song, S.; Du, B.; Jiang, Z.; et al. A Dynamic Synchronization Adjustment Method for High-Speed Camera Exposure Time Based on Improved Information Entropy. *J. Ordnance Equipment Eng.* 2025, 46(11), 345–352. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.