

Article

AI-Assisted Multiview Cartoon Character Modeling: Physics-Consistent Inverse Rendering for Dynamic-Backlight Light-Field 3D Display

Wei Jiang

ART College, Xi'an University of Science and Technology, Xi'an 710054, China; jiangwei2026@xust.edu.cn

Abstract

Three-dimensional display is fundamentally a problem of reconstructing view-dependent optical rays rather than presenting planar images. Multilayer light field display provides an effective panel-based route for 3D scene reconstruction, but it remains constrained by two coupled bottlenecks: additional attenuation layers can improve angular reconstruction while reducing transmitted light, and higher spatio-angular density increases the cost of inverse synthesis. This study proposes a dynamic-backlight-enabled multilayer framework in which a pixel-addressable emission plane participates in light-field synthesis, thereby replacing one purely passive illumination stage with a controllable optical degree of freedom. The emitted rays of conventional- and dynamic-backlight configurations are represented by a unified non-negative tensor model, and layer synthesis is formulated as a weighted non-negative tensor factorization (NTF) problem. A physics-consistent convolutional neural network is further used to predict display-layer patterns through the same optical forward model and is extended from 5×5 to 10×10 viewpoint inputs. The quantitative evidence reported in this paper comprises NTF runtime and test-set PSNR, whereas the prototype photographs provide qualitative verification of viewpoint reconstruction. Because the conventional and dynamic prototypes use different LCD and OLED panels that were not luminance- or color-matched, their apparent brightness and contrast differences are not interpreted as calibrated measurements of architectural gain. Within this evidence boundary, the results support proof-of-concept feasibility of jointly designing the emission plane, attenuation layers, and inverse solver, rather than complete validation of a practical electronically addressable multilayer display.

Keywords: dynamic backlight; multilayer light field display; 3D display; non-negative tensor factorization; convolutional neural network; inverse rendering; computational display; large field of view; optical modulation; panel-based light field synthesis; AI character modeling



Received: 1 April 2026

Revised: 18 July 2026

Accepted: 29 July 2026

Published: 31 July 2026

Copyright: © 2026 by the author.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

AI-based cartoon-character modeling is increasingly expected to support view-dependent presentation rather than only the generation of a single frontal image. For a character to be displayed convincingly in an autostereoscopic or light-field system, its silhouette, shading, occlusion, and local texture must remain coherent as the observation direction changes. In this paper, the term character modeling therefore refers to multiview appearance-field modeling: the construction of a set of directional images that can be converted into physically realizable light rays. This scope differs from polygonal mesh creation,

skeletal animation, or semantic character design, although those processes may provide the source content used by the display pipeline. Research on multimodal three-dimensional content generation and spatially organized representation provides a broader methodological basis for connecting semantic descriptions with structured spatial outputs [1,2]. Related studies of intelligent design, neural-network-supported interaction, and AI-assisted evaluation further emphasize the importance of aligning learned representations with explicit tasks and measurable outputs [3–5].

Computational optics offers a particularly relevant technical foundation for such multi-view modeling. Optical field shaping, super-resolution beam formation, and depth-of-focus control demonstrate that spatial information can be redistributed through coordinated optical design and numerical reconstruction [6–8]. Multifocus generation, extended-depth-of-focus beam design, and compact structured illumination further show that the optical transfer process must be matched to the intended reconstruction objective [9–11]. Adjustable diffractive illumination, scan-free multicolor reconstruction, and high-fidelity optical datasets likewise highlight the importance of preserving spatial, spectral, and system-level consistency throughout the imaging pipeline [12–14]. Studies of dynamic optical fields also indicate that temporally or spatially varying light distributions require acquisition models that preserve the underlying field structure [15,16], while transient measurement and data-driven spectral decoding illustrate the value of combining physical signal formation with computational feature extraction [17,18].

Inverse imaging and computational sensing provide a closer methodological basis for the present display problem. Self-supervised restoration and geometry–intensity fusion show that incomplete or heterogeneous observations can be converted into coherent representations through model-guided computation [19,20]. Tomographic approaches governed by sensor geometry and field sensitivity further demonstrate the importance of incorporating the measurement process directly into reconstruction [21,22]. Motion-aware restoration, computational X-ray imaging, and signal-based state monitoring similarly indicate that reconstruction fidelity depends jointly on acquisition mechanisms, observed signals, and algorithmic processing [23–25]. These principles are consistent with the use of an explicit optical forward model in multilayer display synthesis.

Recent intelligent modeling methods also provide useful support for learning physically valid display patterns. Physics-informed visual learning and multimodal contrastive representation demonstrate that domain constraints can be embedded into feature learning rather than applied only after prediction [26,27]. Graph-guided and multi-receptive-field architectures show the value of structural context and multi-scale features when the input contains complex spatial relationships [28,29]. Multisource sequential modeling and multimodal estimation under limited observations further emphasize the importance of information fusion when the latent state cannot be inferred from a single channel [30,31]. Work on robust prediction, autonomous learning, and interpretable machine learning provides additional guidance for constructing models whose outputs remain traceable and task-consistent [32–34]. Broader discussions of AI governance also support the explicit definition of assumptions, validation boundaries, and evidential scope [35].

From the perspective of complex-system modeling, physically meaningful variables and constrained optimization are equally important. Distributed and fault-tolerant control studies demonstrate that coupled systems must retain explicit interaction relationships and stability conditions [36,37]. Trajectory planning and reduced-order modeling show how computational complexity can be controlled without discarding dominant geometric or dynamic relationships [38,39]. Transient-response and state-space formulations similarly illustrate the value of numerically stable representations for time-dependent physical processes [40,41]. Coordinated optimization and motion-system modeling further indi-

cate that compact computational formulations should remain connected to observable system behavior [42,43]. System-level performance analysis and mechanically constrained modeling provide broader support for optimization under multiple interacting physical variables [44,45]. These studies do not address light-field display directly, but they support the bounded, interpretable, and geometry-aware formulation used in this work.

Studies of hierarchical and layered physical systems also provide useful insights into multilayer optical modeling. Graded structures, defect reconfiguration, and structurally controlled synthesis demonstrate that system response can depend on interactions across several spatial levels [46–48]. Layered, amorphous, and supramolecular configurations further emphasize the importance of coupled component behavior and interface-dependent response [49–51]. Interfacial strengthening and multivariable formulation design illustrate how local structural parameters can be optimized with respect to a system-level objective [52,53]. Mechanism-based wear analysis, broadband response engineering, and graded structural design similarly show that geometry, material response, and operating conditions should be modeled jointly [54–56]. Research on hierarchical porous structures provides an additional example of how multiscale organization affects overall functional behavior [57]. For multilayer display, the corresponding requirement is to coordinate emission, attenuation, layer spacing, and inverse synthesis rather than optimize these elements independently.

Reliable implementation further requires robust mappings between encoded inputs and physical outputs. Information-preserving computational workflows and learning-based mitigation methods highlight the importance of maintaining data integrity under constrained or uncertain conditions [58,59]. Hardware-derived feature generation and sensor-array representations show that device-level variability must be characterized before it can be treated as a dependable computational signal [60,61]. Explicit trust and validation boundaries are likewise important when a computational model interacts with a physical system [62]. Interaction-aware control and coupled-system modeling demonstrate that predictions should remain consistent with measurable interactions [63,64], while mechanism-oriented process analysis illustrates the need to separate observed behavior from unsupported causal generalization [65]. Predictive modeling and multi-level validation studies further reinforce the importance of interpreting results within the limits of the available variables, data, and evaluation protocol [66–68]. These methodological considerations support the cautious treatment of reconstruction metrics and prototype photographs adopted in the present study.

Autostereoscopic and light-field displays provide a direct way to reproduce directional information without requiring wearable optics. Scalable dynamic multiview systems and interactive 360-degree light-field displays have demonstrated that sampled rays can generate parallax and viewpoint-dependent structure within flat or quasi-flat hardware configurations. However, these systems also reveal a basic constraint: finite display bandwidth must be shared among spatial resolution, angular resolution, viewing range, brightness, and hardware complexity. When more viewpoints are required, either more optical channels or more efficient computational encoding is needed. This trade-off motivates multilayer light-field displays, which synthesize directional emission by stacking several spatial modulation planes and solving for layer patterns whose combined optical action approximates a target light field.

Multilayer flat-panel displays are attractive because they keep the hardware compact while using computation to distribute light-field information across several layers. Layered optical modulation can be formulated as an optimization problem rather than as a fixed optical arrangement. Tensor-based display models further develop this idea by factorizing a target four-dimensional light field into structured display-layer patterns, making the low-

rank structure of the display model directly usable for synthesis. Reconstruction quality therefore depends strongly on the physical algebra of the stack, the non-negativity of the modulation patterns, the geometric relationship among the layers, and the selected optimization strategy. The inverse computation must remain tied to the actual optical geometry if the reconstructed views are expected to remain spatially and angularly consistent.

Several neighboring display systems have adopted similar optical–computational strategies, although their physical constraints differ from those of transmissive multilayer panels. Model-based rendering and computational holography demonstrate that virtual or physical light propagation can be embedded into an inverse optimization or learning process. These developments indicate that high-quality three-dimensional reconstruction increasingly depends on the joint design of photonic hardware and numerical synthesis. Nevertheless, they do not remove the central limitation of attenuation-based multilayer light-field displays: useful light must still pass through multiple lossy modulation stages.

This limitation cannot be treated as a purely optical or purely computational issue. Increasing angular density or viewing range enlarges the inverse problem, whereas increasing the number of attenuation layers reduces the available luminous flux. Illumination structure, layer geometry, modulation constraints, and computational throughput must therefore be considered together. For multilayer light-field display, the central design problem is to balance angular reconstruction accuracy against optical attenuation and inverse-solver cost.

For cartoon characters and other stylized objects, this trade-off becomes particularly relevant. Increasing the number of modulating layers provides more degrees of freedom for separating viewpoint-dependent silhouettes and texture cues, but each additional attenuation stage reduces transmitted light. Increasing the viewpoint count or field of view improves angular continuity around a character, but it also increases the dimensionality and computational cost of the inverse problem. An AI-assisted character-display framework must therefore coordinate optical architecture, multiview representation, and inverse synthesis rather than optimizing any one of them in isolation.

This paper proposes a physics-consistent AI framework for multiview cartoon-character and stylized-object modeling on a dynamic-backlight multilayer light-field display. The passive uniform backlight is replaced by a pixel-addressable emission plane that participates directly in ray synthesis. Conventional- and dynamic-backlight systems are represented through a unified non-negative tensor formulation, and the required display patterns are obtained either by weighted non-negative tensor factorization or by a convolutional neural network whose outputs are evaluated through the optical forward model. The method is extended from 5×5 to 10×10 viewpoint inputs to examine angular scalability. The available experiments use established multiview light-field scenes for numerical evaluation and a representative stylized-object visualization for qualitative explanation; consequently, the study demonstrates the transferable display-side modeling principle rather than a statistically complete cartoon-character benchmark.

2. Dynamic-Backlight Light-Field Representation for AI Character and Stylized-Object Display

2.1. Multiview Appearance-Field Representation and Display Objective

A cartoon character or stylized object is represented here by its directional appearance rather than by an explicit polygon mesh. The displayed content is perceived as three-dimensional when emitted light varies jointly with spatial position and viewing direction. The plenoptic formulation provides the general optical description, while the two-plane light-field parameterization supplies the lower-dimensional representation used by the

inverse model. This representation allows silhouettes, shading transitions, and viewpoint-dependent details to be encoded as a four-dimensional appearance field.

$$P(x, y, z; \theta, \varphi; \lambda; t) - I = 0 \quad (1)$$

Equation (1) expresses the plenoptic relation in residual form, where I is the received intensity, (x, y, z) is the observation position, (θ, φ) is the propagation direction, λ is the wavelength, and t is time. Equation (2) then defines the reduced four-dimensional light field by two plane intersections.

$$I(u, v, p, q) \equiv P(u, v; p, q) \quad (2)$$

Equation (2) parameterizes each ray by its intersections with two reference planes, where (u, v) and (p, q) are the plane coordinates. For AI character modeling, each angular sample corresponds to one directional appearance of the character or stylized object. The display objective is therefore to reproduce the target multiview appearance field through a finite set of emitting and attenuating planes.

2.2. Optical Modulation Logic of Multilayer Attenuation Display

A multilayer light-field display generates directional light by cascading attenuation layers in front of a backlight plane. The optical action of successive liquid-crystal modulators is sequential rather than additive: the polarization state is transformed by each modulator-analyzer stage, and the transmitted intensity from one stage becomes the incident intensity of the next. Under the scalar attenuation approximation adopted for the ray-synthesis model, each layer is represented by a non-negative transmittance $\tau_n \in [0, 1]$, where n indexes the layer, and the intensity is multiplied across a stack of N attenuation layers.

$$\frac{I_{\text{out}}}{I_{\text{in}}} = \prod_{n=1}^N \tau_n \quad (3)$$

Equation (3) expresses the normalized output intensity as the product of the sampled layer transmittances. For two layers, Equation (4) defines the total transmittance T_2 as ab . Equation (5) gives the equivalent logarithmic form; this numerical transformation does not imply physical additivity of polarization rotations or optical intensities.

$$T_2 \equiv \frac{I_{\text{out}}}{I_{\text{in}}} = ab \quad (4)$$

The logarithmic ratio in Equation (5) is used only to rewrite a multiplicative attenuation chain as a sum of log-transmittances. The physical forward process remains sequential and multiplicative.

$$\ln\left(\frac{I_{\text{out}}}{I_{\text{in}}}\right) = \ln a + \ln b \quad (5)$$

The layered stack therefore acts as a spatially distributed multiplicative modulation chain whose degrees of freedom are the pixel transmittances on each layer. This mechanism allows the display to synthesize viewpoint-dependent rays without mechanical scanning. Its main limitation is optical efficiency: each additional attenuation layer introduces light loss, which reduces the available luminance and makes the layer-count-versus-reconstruction-quality trade-off fundamental.

2.3. Proposed Dynamic-Backlight Architecture

The optical loss of a conventional multilayer display originates from two sources: repeated attenuation by the liquid-crystal stack and repeated loss on the polarizers. As the number of attenuation layers increases, crosstalk is reduced and the reconstructed light field becomes more accurate, but the useful luminous flux decreases at the same time. The proposed architecture resolves part of this contradiction by replacing the uniform backlight with a pixel-addressable dynamic backlight. The dynamic backlight functions simultaneously as a luminous source and as an additional controllable modulation plane, so the ordinary backlight plane and the final attenuation role are merged into one component.

The conventional stack uses a uniform backlight followed by multiple attenuation layers, whereas the proposed configuration drives a self-emissive backlight plane together with the front attenuation layers. This modification reduces the number of purely passive stages and increases the controllable degrees of freedom available for light-field synthesis without increasing geometric complexity. Under radiometrically matched conditions, this structure is expected to permit more efficient spatial allocation of emitted light; the present unmatched-panel prototype does not quantify that expectation.

An OLED panel was selected as the dynamic backlight in the prototype because each pixel is self-emissive and can be controlled independently without an additional backlight module. From the forward-model perspective, this architecture offers three potential engineering advantages. First, it replaces a fixed illumination term and one purely attenuating role with a controllable emission plane. Second, the emitted energy can be spatially allocated jointly with the front attenuation patterns, increasing the available synthesis degrees of freedom. Third, the structure remains compact and compatible with the same layered ray model used for ordinary multilayer synthesis. These are architecture-level properties rather than calibrated prototype measurements. The present hardware comparison uses different LCD and OLED panels, so any observed brightness or contrast difference also contains panel-specific luminance, spectral, and color-response effects.

2.4. Character-Oriented Evaluation Variables and Evidence Boundary

The evaluation variables reflect the requirements of multiview character and stylized-object presentation. Layer number controls the capacity available for separating silhouettes and view-dependent texture cues. Scene diversity tests whether the inverse model remains stable for different appearance structures. Field of view measures angular scalability, and interlayer spacing tests the sensitivity of the ray geometry. These variables evaluate the display-side modeling process; they do not evaluate semantic character identity, rigging quality, or animation realism.

The available evidence is separated into quantitative computational measures and qualitative optical observations. NTF runtime characterizes iterative-solver cost, and test-set PSNR characterizes the correspondence between the target light field and the light field synthesized from CNN-predicted layer patterns. The captured left, center, and right views are used to verify viewpoint-dependent content and prototype feasibility. They are not used to calculate luminance, contrast ratio, luminous efficiency, or angular crosstalk because the two backlight panels and the photographic acquisition were not radiometrically normalized. Accordingly, the manuscript distinguishes structural optical reasoning from measured performance and avoids treating apparent photographic differences as objective optical gains.

3. Non-Negative Tensor Modeling of Multiview Character and Stylized-Object Light Fields

3.1. Unified Tensor Representation of Multilayer Emitted Light Field

Before multiview character or stylized-object content can be synthesized, the emitted rays must be expressed in a factorized form. The ordinary multilayer display is modeled as a sequence of attenuation planes in front of a conventional backlight. For a ray parameterized by the two-plane coordinates (x, v) , its intensity is determined by the transmittance values sampled along the physical path. The following formulation directly encodes this multilayer ray geometry.

$$I(x, v; N) = \exp \left\{ \sum_{n=1}^N \ln \left[f^{(n)} \left(x + \frac{d_n}{d_r} v \right) \right] \right\} \quad (6)$$

Equation (6) rewrites the N -layer product as an exponential of summed log-transmittances. The function $f^{(n)}$ denotes the transmittance profile of layer n , and d_n/d_r maps the angular coordinate to the corresponding layer intersection.

$$\tilde{I}(x, v) = \exp[\ln f(\xi_1) + \ln g(\xi_2) + \ln h(\xi_3)] \quad (7)$$

$$d_r \xi_n = d_r x + d_n v \quad (8)$$

$$\tilde{I}(\xi_1, \xi_2, \xi_3) = \prod_{r=1}^3 \psi_r(\xi_r), \quad \psi_1 = f, \psi_2 = g, \psi_3 = h \quad (9)$$

$$\begin{bmatrix} \alpha & \beta & \gamma \end{bmatrix}^T = \begin{bmatrix} d_3 - d_2 & d_1 - d_3 & d_2 - d_1 \end{bmatrix}^T \quad (10)$$

Equations (7)–(10) describe the three-layer geometry in reparameterized form. The functions ψ_1 , ψ_2 , and ψ_3 correspond to the three layer profiles, while ξ_1 , ξ_2 , and ξ_3 are the ray intersections. The vector relation in Equation (10) collects the layer-distance differences that constrain physically realizable ray combinations.

$$\tilde{\mathcal{T}} = \mathbf{f} \otimes \mathbf{g} \otimes \mathbf{h} \quad (11)$$

$$\tilde{t}_{ijk} = f_i g_j h_k \quad (12)$$

$$\tilde{\mathcal{L}} = \mathcal{M}(\tilde{\mathcal{T}}) = \mathbf{W} \circledast \tilde{\mathcal{T}} \quad (13)$$

Equation (11) uses the outer product to form the rank-one layer tensor, Equation (12) states its elementwise factorization, and Equation (13) applies the sparse validity operator through the binary mask $\mathbf{W}$.

3.2. Weighted NTF Objective and Multiplicative Optimization

Equation (14) is written as a masked Frobenius-norm objective with the factor $\frac{1}{2}$ included for derivative convenience. The interval constraints preserve physically meaningful non-negative transmittances.

$$\begin{aligned} \min_{\mathbf{F}, \mathbf{G}, \mathbf{H}} \quad & \frac{1}{2} \| \mathbf{W} \circledast (\mathbf{L} - [\mathbf{F}, \mathbf{G}, \mathbf{H}]) \|_F^2 \\ \text{subject to} \quad & 0 \leq \mathbf{F}, \mathbf{G}, \mathbf{H} \leq 1 \end{aligned} \quad (14)$$

Equation (14) is the weighted objective used for the three-layer model. The weighting tensor ensures that the error is accumulated only on valid rays, and the bound constraints keep every transmittance within the physical interval of the attenuation layers. This

bounded non-negative formulation is essential: it prevents unphysical negative coefficients and preserves the optical interpretation of each factor.

$$\begin{aligned} \mathbf{F} &\leftarrow \Pi_{[0,1]} \left(\mathbf{F} \circledast \frac{[(\mathbf{W} \circledast \mathbf{L})_{(1)}](\mathbf{H} \odot \mathbf{G})}{[(\mathbf{W} \circledast [\mathbf{F}, \mathbf{G}, \mathbf{H}])_{(1)}](\mathbf{H} \odot \mathbf{G}) + \varepsilon} \right), \\ \mathbf{G} &\leftarrow \Pi_{[0,1]} \left(\mathbf{G} \circledast \frac{[(\mathbf{W} \circledast \mathbf{L})_{(2)}](\mathbf{H} \odot \mathbf{F})}{[(\mathbf{W} \circledast [\mathbf{F}, \mathbf{G}, \mathbf{H}])_{(2)}](\mathbf{H} \odot \mathbf{F}) + \varepsilon} \right), \\ \mathbf{H} &\leftarrow \Pi_{[0,1]} \left(\mathbf{H} \circledast \frac{[(\mathbf{W} \circledast \mathbf{L})_{(3)}](\mathbf{G} \odot \mathbf{F})}{[(\mathbf{W} \circledast [\mathbf{F}, \mathbf{G}, \mathbf{H}])_{(3)}](\mathbf{G} \odot \mathbf{F}) + \varepsilon} \right) \end{aligned} \quad (15)$$

Equation (15) expresses the three cyclic multiplicative updates with projection onto the physical interval $[0, 1]$ and a small stabilizer ε . The symbol $\circledast$ denotes elementwise multiplication, $\odot$ denotes the Khatri–Rao product, and the parenthesized mode subscript denotes tensor unfolding. Equations (16) and (17) give the corresponding arbitrary-layer update through the combined factor matrix $\mathbf{K}_n$.

$$\mathbf{F}^{(n)} \leftarrow \Pi_{[0,1]} \left(\mathbf{F}^{(n)} \circledast \frac{[(\mathbf{W} \circledast \mathbf{L})_{(n)}] \mathbf{K}_n}{[(\mathbf{W} \circledast \tilde{\mathcal{T}})_{(n)}] \mathbf{K}_n + \varepsilon} \right) \quad (16)$$

$$\mathbf{K}_n = \mathbf{F}^{(N)} \odot \dots \odot \mathbf{F}^{(n+1)} \odot \mathbf{F}^{(n-1)} \odot \dots \odot \mathbf{F}^{(1)} \quad (17)$$

Equations (16) and (17) extend the same update principle to an arbitrary number of layers. Only the display-coupled factorization is retained in the implementation, because generic discussions of unconstrained Tucker or polyadic decomposition do not change the reconstruction result of the multilayer display. The price of this physically transparent optimization is runtime: each sample requires repeated updates until convergence, and the cost increases with tensor order and layer count.

3.3. Dynamic-Backlight Extension of the NTF Model

A uniform backlight fixes the luminous source term and forces the front attenuation stack to encode the entire light field by transmittance alone. The proposed dynamic backlight introduces an additional controllable term that modulates the emitted light before it reaches the front layers. The forward model therefore changes from a purely attenuative stack to a mixed emission-and-attenuation stack, as expressed in Equations (18)–(20).

$$\frac{\tilde{l}(x, v)}{b(x, v)} = \prod_{n=1}^N f^{(n)} \left(x + \frac{d_n}{d_r} v \right) \quad (18)$$

$$\mathbf{b} = \text{vec}([\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_S]), \quad \mathbf{b}_s = [b_{1s}, b_{2s}, \dots, b_{As}]^T \quad (19)$$

$$\tilde{\mathcal{T}} = \bigotimes_{n=0}^N \mathbf{a}^{(n)}, \quad \mathbf{a}^{(0)} = \mathbf{b}, \quad \mathbf{a}^{(n)} = \mathbf{f}^{(n)} (n \geq 1) \quad (20)$$

Equation (18) isolates the dynamic emission term by normalizing the synthesized ray with $b(x, v)$. Equation (19) defines the vectorized backlight samples, and Equation (20) expresses the complete system as an outer product whose zeroth factor is the dynamic backlight.

3.4. Implementation Scope of the NTF Model

The factorization section is intentionally restricted to the components that directly determine the synthesized display patterns. Only the emitted-light-field parameterization, the weighted objective, and the multiplicative updates are retained because these elements are sufficient to reproduce the reconstruction pipeline and to explain the experimental results. Generic discussions of matrix-factorization history, uniqueness theory, second-order solvers, or broad tensor-decomposition taxonomies are not included in the working formulation.

This compact boundary is also methodologically appropriate. The optical conclusions of the paper do not depend on whether alternative unconstrained factorization variants are available; they depend on how accurately a physically valid sparse light-field tensor can be decomposed into layer transmittances under non-negativity constraints. The retained formulation therefore keeps the section aligned with computational display synthesis rather than general-purpose tensor analysis.

4. Physics-Consistent AI Inverse Modeling for Multiview Characters and Stylized Objects

4.1. AI Inverse-Rendering Formulation

The iterative NTF route solves each target independently, whereas a CNN can learn a reusable inverse mapping from multiview character or stylized-object appearances to display-plane patterns. The network is not trained as an unconstrained image generator. Its predicted planes are propagated through the same optical forward model, and the reconstructed multiview appearance is compared with the target field. This coupling makes the learned mapping physics-consistent and preserves the relationship between AI-generated content and realizable display rays.

$$\xi_1 = u - s \quad (21)$$

$$\xi_2 = u \quad (22)$$

$$\xi_3 = u + s \quad (23)$$

Equations (21)–(23) state the three ray-intersection coordinates directly. The middle layer is the reference plane, u is the spatial coordinate, and s is the angular offset.

$$\ln \tilde{L}(s, u) = \sum_{z \in \mathcal{Z}} \ln f(u + zs) \quad (24)$$

$$\ln \tilde{L}(s, t, u, v) = \sum_{z \in \mathcal{Z}} \ln f(u + zs, v + zt) \quad (25)$$

Equations (24) and (25) use log-domain accumulation for the 2D and 4D forward models. They are equivalent to multiplying the layer transmittances along each ray but provide a numerically compact expression for the physics-consistent loss.

$$\mathcal{L}(\Theta) = \frac{1}{|\Omega|} \sum_{(s,t,u,v) \in \Omega} [L(s, t, u, v) - \tilde{L}_{\Theta}(s, t, u, v)]^2 \quad (26)$$

Equation (26) defines the batch-averaged reconstruction loss over the valid spatio-angular sample set Ω . The parameter set Θ contains the CNN weights, and the reconstructed field $\tilde{L}_{\Theta}$ is obtained only after the predicted planes pass through the optical forward model.

4.2. Baseline CNN Architectures for 5×5 Multiview Character Inputs

For a 5×5 multiview input, the network receives 25 registered directional images of the same character or stylized object and predicts three display-plane patterns. Architecture 1 maps the 25 input channels to 64 feature channels, applies 19 hidden 64-channel convolutional layers with ReLU activation, and converts the final representation to the three physical output planes.

The remaining 5×5 architectures are controlled variants designed to test the width–depth trade-off while leaving the physical forward model unchanged. Architecture 2 uses 19 hidden layers and a symmetric 64–32–16–...–16–32–64 channel profile; Architecture 3 uses 14 hidden layers with 64 channels; and Architecture 4 uses 24 hidden layers with 64 channels.

These variants isolate the width–depth trade-off while leaving the physical forward model unchanged. All four architectures are trained with ReLU activations and the Adam optimizer, so any performance difference can be attributed to representational capacity rather than to a change in supervision or loss definition. This design makes the comparison directly meaningful for computational display, where the practical question is whether comparable layer synthesis quality can be achieved with a smaller model and, consequently, a lower computation cost.

4.3. Dynamic-Backlight and 10×10 Large-Angle Character Modeling

The dynamic-backlight learning model preserves the same inverse-rendering logic but changes the physical meaning of the output channels. One output channel drives the dynamic backlight plane and the remaining channels drive the front attenuation layers. Figure 1 illustrates this architecture-level modification, in which the emitted ray is jointly determined by the luminous backlight pixel and the transmittance values of the attenuation layers.

$$\tilde{L}(s, u) = \exp \left\{ \ln b(u) + \sum_i \ln f(\xi_i) \right\} \quad (27)$$

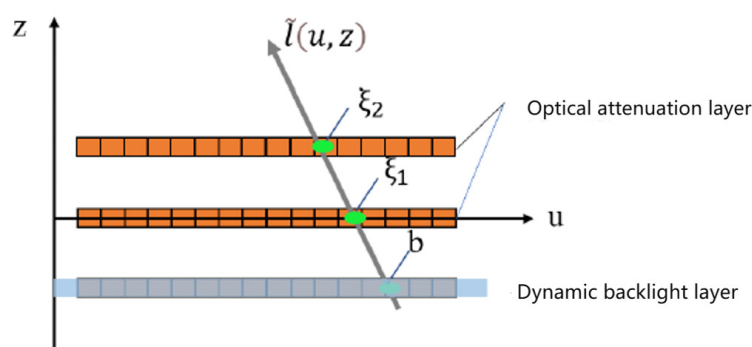


Figure 1. Dynamic-backlight multilayer light-field display model for CNN-based reconstruction.

Equation (27) rewrites the dynamic-backlight ray in log-exponential form. The term $b(u)$ is the self-emissive source sample, and $f(\xi_i)$ denotes the transmittance encountered at attenuation plane i .

The large-field-of-view extension increases the input viewpoint grid from 5×5 to 10×10 , so the number of input channels rises from 25 to 100. Architecture 5 retains the 19-layer, 64-channel hidden structure of Architecture 1; Architecture 6 uses the 14-layer, 64-channel structure of Architecture 3; and Architecture 7 uses the compressed 64–32–16–...–16–32–64 profile of Architecture 2. All three models produce the same 3-channel physical-plane output, while encoding a substantially richer angular input.

The 10×10 formulation is particularly relevant to AI character presentation because it increases the number of directional samples available for preserving silhouette continuity and view-dependent appearance. Once trained, the network replaces repeated sample-wise factorization with a feedforward prediction while remaining constrained by the same optical model. The experiments therefore examine whether the learned inverse remains numerically stable as the angular dimensionality of the modeled content increases.

4.4. Implementation Scope of the CNN Model

The learning section is restricted to the elements that directly determine reproducibility of the inverse-display pipeline: the physical forward model, the squared reconstruction loss, the convolutional architectures, ReLU activation, Adam optimization, and the retained training and test protocol. The network does not receive direct supervision on an assumed set of layer patterns; instead, its outputs are propagated through the display model and compared with the target light field. No auxiliary perceptual, adversarial, sparsity, or layer-pattern loss is included in the reported implementation.

This boundary keeps the method focused on computational display rather than general neural-network exposition. It also clarifies the evidential scope of the learning results. The archived experiments retain the source light fields, angular and spatial dimensions, optimizer settings, epoch counts, and test PSNR values, but they do not retain a reliable augmented-patch count, crop stride, or a scene-disjoint cross-validation record. The revised manuscript therefore reports the available protocol explicitly and does not reconstruct missing quantities by assumption.

4.5. Training Data, Loss, and Generalization Boundary for Character-Oriented Use

The CNN data were drawn from the five Stanford light fields used in the study: Butterfly, Dice, Dragon, Happy-Buddha, and Messerschmitt. Each source light field is a registered multi-view stack. For network training, dense spatial sampling was used to form 64×64 multi-view patches while preserving the angular channels: 25 channels for the 5×5 setting and 100 channels for the 10×10 setting. The retained experimental record documents dense patch extraction but does not preserve the exact crop stride or the final number of augmented patches. It also does not document additional rotations, flips, color jitter, or synthetic-view generation; the revised paper therefore does not claim that such transformations were used.

The 5×5 networks were trained for 200 epochs and the 10×10 networks for 500 epochs, with a batch size of 15, an initial learning rate of 0.001, ReLU activations, and the Adam optimizer. For each mini-batch, the implemented loss is the mean of the squared error in Equation (26), accumulated over the angular coordinates and spatial pixels after the predicted display planes have been passed through the optical forward model. Thus, the supervision is applied to the reconstructed light field rather than directly to the intermediate layer images, and no additional loss term is combined with Equation (26).

Testing was performed on 640×480 light-field images, and PSNR values are reported below. These results measure reconstruction beyond the 64×64 training-patch scale and assess the increase from 5×5 to 10×10 angular inputs. The retained dataset is not a dedicated cartoon-character benchmark and does not establish scene-disjoint generalization. The numerical evidence should therefore be interpreted as validation of the inverse modeling mechanism, while future work should include curated cartoon-character datasets with controlled poses, materials, silhouettes, and viewpoint splits.

5. Experimental Results and Character-Oriented Discussion

5.1. Experimental Setup, Datasets, and Reporting Protocol

The NTF and CNN experiments were implemented on separate hardware platforms because they emphasize different parts of the pipeline. The NTF experiments were executed on a personal computer with an AMD 3600X CPU (Advanced Micro Devices, Inc., Santa Clara, CA, USA) and an NVIDIA GTX660 GPU (NVIDIA Corporation, Santa Clara, CA, USA). The ordinary prototype used the LCD panel of a Redmi K30s as the backlight, whereas the dynamic prototype used the OLED panel of an iPhone 12. The attenuation planes were emulated by transparent films printed with the optimized layer patterns and separated by high-transmittance acrylic spacers. This static assembly was selected as a proof-of-concept for optical geometry and pattern transfer; it does not reproduce the refresh behavior, grayscale transfer, polarization response, or panel-to-panel registration of individually addressable liquid-crystal layers. A practical follow-up should replace the printed films with synchronized transmissive LCD panels. The Stanford light-field data used in the study comprised Butterfly, Dice, Dragon, Happy-Buddha, and Messerschmitt; the NTF inputs contained 7×7 viewpoints with a spatial resolution of 384×512 pixels per view. The NTF solver used multiplicative updates in MATLAB 2020a. Because the LCD and OLED panels were different devices, their peak luminance, emission spectrum, gamma response, white point, and native contrast were not controlled as a single experimental variable.

The static proof-of-concept hardware combines a phone display, printed attenuation films, and high-transmittance acrylic spacers. In the conventional configuration, the LCD phone panel acts as the illumination source; in the dynamic configuration, the OLED panel acts as a pixel-addressable emission plane behind the same type of printed attenuation stack. The assembly is consistent with the modeled ray geometry but is not an actively refreshed multilayer LCD system. The prototype description is therefore retained as a hardware-integration explanation rather than as a separate figure-based result.

The CNN experiments were performed on a server running Ubuntu 20.04 with an AMD Ryzen 9 7950X CPU (Advanced Micro Devices, Inc., Santa Clara, CA, USA) and an NVIDIA Titan X GPU (NVIDIA Corporation, Santa Clara, CA, USA), using Chainer version 7.8.1. The initial learning rate was 0.001 and the batch size was 15. The 5×5 models were trained for 200 epochs and the 10×10 models for 500 epochs. Dense spatial sampling generated 64×64 multi-view training patches, and testing was performed on 640×480 light-field images. Table 1 consolidates the retained training details, the exact loss interpretation, and the limits of the available generalization evidence.

The reporting protocol separates four evidence types. Layer number, scene, field of view, and interlayer spacing are used to examine reconstruction behavior. NTF runtime and CNN test PSNR are quantitative computational results. The target Stanford light fields serve as the numerical ground truth in the factorization objective and in the PSNR calculation. The prototype photographs are qualitative outputs and are not normalized to the target images or between the two backlight devices. Table 1 summarizes which aspects were controlled and which conclusions are supported; Table 2 summarizes the CNN training and evaluation protocol.

Table 1. Evidence scope and interpretive limits of the prototype comparison.

Evidence Item	Available Information	Supported Interpretation and Limitation
Backlight devices	Redmi K30s LCD for the conventional prototype; iPhone 12 OLED for the dynamic prototype.	The devices were not matched for luminance, spectrum, gamma, white point, or native contrast; panel-dependent photographic differences cannot be assigned solely to the architecture.
Optical stack and geometry	Printed attenuation films and high-transmittance acrylic spacers were used in both prototype types; the modeled interlayer geometry was applied to pattern synthesis.	Supports hardware integration and geometric feasibility, but not an isolated radiometric comparison.
Photographic acquisition	Captured left, center, and right views are available; calibrated exposure, RAW data, white-balance reference, and colorimetric metadata are not retained.	Supports qualitative viewpoint verification only; luminance, contrast ratio, efficiency, and color error cannot be recovered reliably.
Target light fields	Stanford light fields were used as the numerical targets; NTF inputs contain 7×7 views at 384×512 pixels, and CNN testing uses 640×480 images.	Supports numerical reconstruction objectives and PSNR evaluation; the target data do not normalize the two physical backlights.
Simulation-prototype correspondence	The same ray geometry and target-view structure are used in the forward model and prototype pattern generation.	Supports correspondence of viewpoint content and modeled geometry; photometric correspondence was not calibrated.
Quantitative results	NTF runtime and CNN test PSNR are reported.	Supports computational performance and reconstruction accuracy, not objective optical gain.

Table 2. Retained CNN training and evaluation protocol.

Protocol Item	5×5 Viewpoint Model	10×10 Viewpoint Model
Source corpus	Five Stanford light fields: Butterfly, Dice, Dragon, Happy-Buddha, and Messerschmitt.	Same retained source corpus.
Angular input	25 registered viewpoint channels.	100 registered viewpoint channels.
Training sample size	64×64 multi-view spatial patches.	64×64 multi-view spatial patches.
Augmentation	Dense spatial patch extraction is documented; crop stride and final patch count are not retained. No additional transformation is claimed.	Same documented augmentation boundary.
Test image size	640×480 light-field images.	640×480 light-field images.
Optimization	ReLU; Adam; initial learning rate 0.001; batch size 15; 200 epochs.	ReLU; Adam; initial learning rate 0.001; batch size 15; 500 epochs.
Loss	Batch-mean squared light-field reconstruction error after the predicted planes pass through the forward model in Equation (26); no auxiliary loss.	Same loss with 100-channel angular input.
Reported evaluation	Training PSNR curve and full-resolution test PSNR.	Training PSNR curve and full-resolution test PSNR.
Generalization boundary	No retained scene-disjoint split or repeated cross-validation; interpreted as within-corpus test performance.	Same boundary; additionally evaluates angular scalability to 100 input channels.

5.2. NTF Results I: Layer Count and Representative Stylized-Object Reconstruction

The first NTF study examines the layer count required to balance multiview fidelity and implementation cost. Numerical evaluation retains the original light-field scenes and field-of-view settings, while the representative visual example is interpreted as stylized-object content suitable for the same character-oriented pipeline. Adding attenuation layers increases the available degrees of freedom for separating directional appearance cues, but it also increases optical loss and iterative cost.

Figure 2 presents a representative multiview visualization of a stylized motorcycle object under two- and three-layer conventional- and dynamic-backlight configurations. The figure is used to illustrate viewpoint consistency, blur distribution, and relative reconstruction behavior across the four modeled conditions. It is not treated as a calibrated optical measurement or as evidence of semantic character-generation quality. The same pipeline can accept cartoon-character view sets because the inverse model operates on registered multiview appearance channels rather than on a particular object category.

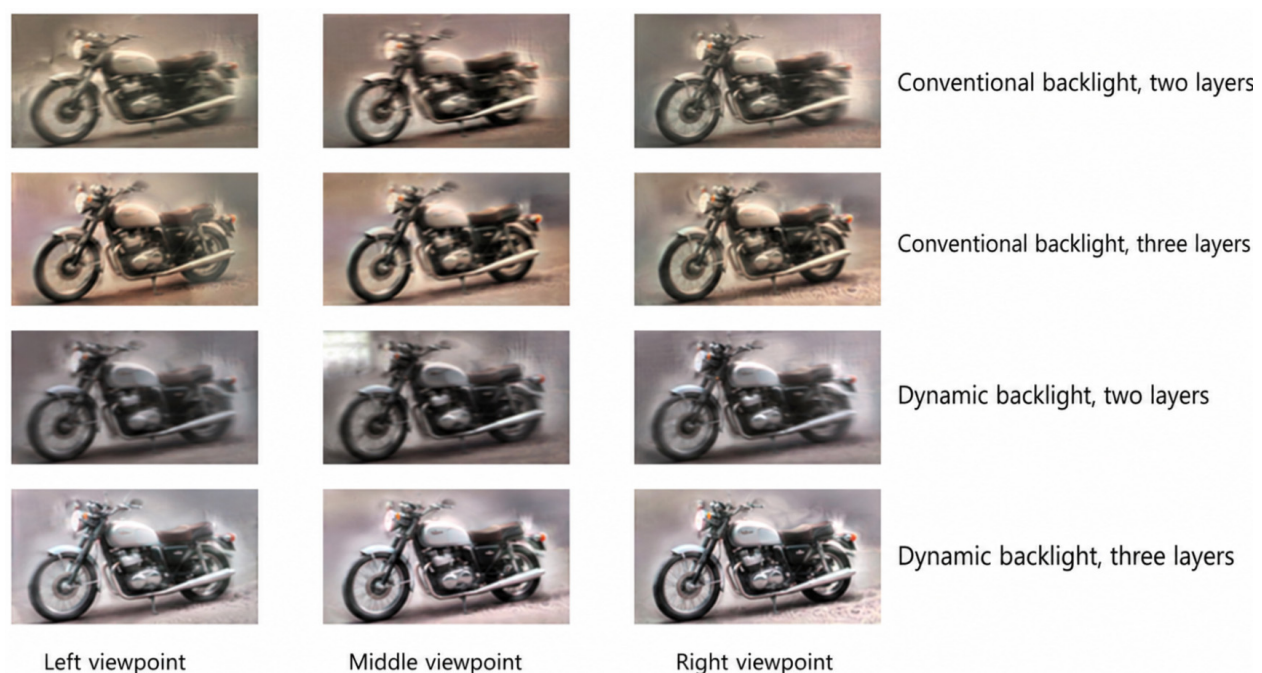


Figure 2. Representative multiview visualization of a stylized motorcycle object under conventional- and dynamic-backlight configurations with two and three layers. The figure illustrates viewpoint-dependent appearance and relative reconstruction behavior only; it is not a calibrated radiometric comparison.

The runtime trend confirms that additional layers increase computational cost as well as optical complexity. Table 3 reports the time required for 100 NTF iterations. The runtime increases from 41.54535 s for two layers to 42.5676 s for three layers and then to 71.89 s for five layers. Under the present hardware and tensor dimensions, the three-layer design therefore provides an intermediate compromise between reconstruction flexibility and iterative cost.

Table 3. Runtime of 100 NTF iterations for different numbers of attenuation layers.

Number of Attenuation Layers	Runtime for 100 Iterations (s)
2	41.54535
3	42.5676
5	71.89

5.3. NTF Results II: Content Diversity, Field of View, and Interlayer Spacing

Content diversity was evaluated using the retained Butterfly, Dice, Dragon, Happy-Buddha, and Messerschmitt light fields under the same three-layer NTF configuration. These scenes contain different contour, texture, depth, and occlusion structures and therefore provide transferable evidence for character-oriented multiview reconstruction. They do not, however, constitute a dedicated cartoon-character dataset; the results are used to validate the optical inverse model rather than semantic character fidelity.

Scene-dependent runtime variation was small. Table 4 lists the time required for 200 NTF iterations on the five scenes. The slowest case, Dice, required 44.345 s, whereas Butterfly, Dragon, Happy-Buddha, and Messerschmitt remained within the same overall range. This supports the interpretation that runtime is governed mainly by tensor dimensions and iteration count rather than by scene semantics.

Table 4. Runtime across different scenes for 200 NTF iterations.

Scene	Runtime for 200 Iterations (s)
Butterfly	40.57862
Dice	44.345
Dragon	41.8673
Happy-Buddha	43.7865
Messerschmitt	43.3784

Angular scalability was evaluated with the Messerschmitt scene at 10°, 20°, and 45° fields of view. The qualitative captures at 20° and 45° retained distinguishable viewpoint content, while reconstruction became less stable as the field of view increased because a fixed number of physical planes had to encode a wider angular range. These observations are used to describe the modeled angular trend and are not treated as quantitative evidence of practical large-field-of-view image quality or calibrated optical advantage.

The physical spacing between layers was varied to test geometric robustness. The three-layer system was analyzed at 2 mm, 8 mm, and 20 mm spacing with a fixed 20° field of view. The qualitative captures at 2 mm and 20 mm retained viewpoint-dependent content when the forward model was updated to the physical spacing. These conditions are reported as geometric feasibility checks; no quantitative photometric conclusion is drawn from the static-film prototype.

5.4. CNN Results I: 5×5 Multiview Character-Modeling Accuracy

The first CNN study uses the 5×5 viewpoint setting and compares Architectures 1–4 under the same physics-consistent loss. The retained training logs show that all four networks rise rapidly during the initial epochs and stabilize near 33 dB by the end of 200 epochs, indicating convergence under the recorded protocol. Because the training curves are intermediate optimization diagnostics, the architecture comparison is reported compactly through the independent full-resolution test PSNR values in Table 5.

Table 5. Test PSNR of different network architectures for 5×5 viewpoints.

Metric	Architecture 1	Architecture 2	Architecture 3	Architecture 4
Test PSNR (dB)	25.864534	24.534533	26.425335	24.864335

Table 5 reports PSNR on the retained 640×480 test images. Architecture 3 achieves 26.425335 dB, Architecture 1 reaches 25.864534 dB, and Architectures 2 and 4 obtain 24.534533 dB and 24.864335 dB, respectively. Within this test set, Architecture 3 provides the highest reconstruction accuracy, and the total spread is 1.890802 dB. These values

demonstrate performance beyond the 64×64 training-patch scale, but they are not presented as scene-disjoint cross-validation because the archived split composition and patch count are unavailable.

Architecture 3 is therefore the best-performing 5×5 variant within the retained test set. Separate prototype photographs are not used to rank the CNN architectures because the unmatched backlight panels and static attenuation films do not provide a more reliable comparison than the numerical PSNR results.

5.5. CNN Results II: 10×10 Character-View Scalability and Cross-Method Discussion

The 10×10 viewpoint setting increases the angular input dimensionality from 25 channels to 100 channels and provides a more demanding inverse problem. The retained training logs show that Architectures 5–7 converge and approach 30 dB by the end of 500 epochs. The lower terminal value relative to the 5×5 case is consistent with encoding a larger angular set using the same three-plane output structure. The quantitative comparison is therefore reported through the full-resolution test PSNR values in Table 6.

Table 6. Test PSNR of different network architectures for 10×10 viewpoints.

Metric	Architecture 5	Architecture 6	Architecture 7
Test PSNR (dB)	25.337373	24.687745	24.864355

Table 6 summarizes the 10×10 test results. Architecture 5 achieves 25.337373 dB, Architecture 7 reaches 24.864355 dB, and Architecture 6 reaches 24.687745 dB. Within the retained test set, Architecture 5 is favored when the angular channel count increases, while the other two variants remain within 0.649628 dB of the best result. These data support scalability of the physics-consistent loss to the larger viewpoint grid, but they do not substitute for scene-disjoint or external-dataset validation.

Table 6 indicates that the learned inverse remains numerically stable when the angular input is increased to 100 channels, with Architecture 5 providing the highest retained test PSNR. No separate photographic panel is used to claim practical large-field-of-view optical quality because the proof-of-concept hardware does not provide a calibrated or actively refreshed validation platform.

The combined NTF and CNN results reveal methodological complementarity. NTF is the more interpretable route because each factor is directly tied to a physical transmittance or emission profile and the effects of layer number, spacing, and backlight structure can be analyzed explicitly in the forward model. Its main limitation is iterative cost because each target light field requires a new optimization. The CNN route shifts this computation to offline training and produces the display patterns by feedforward prediction. Because the present study reports test PSNR and a single representative proof-of-concept capture rather than matched end-to-end latency and active-panel optical measurements, the CNN is described as an amortized inverse model with scalability potential rather than as a quantitatively proven real-time practical implementation. For both routes, calibrated gains in light utilization and contrast remain a subject for the matched-panel and active-LCD measurements specified in Section 6.2.

6. Conclusions

6.1. Conclusions for AI-Assisted Character and Stylized-Object Modeling

This study establishes a physics-consistent AI framework for converting multiview cartoon-character or stylized-object appearances into dynamic-backlight multilayer display patterns. The method treats character modeling as view-dependent appearance-field

reconstruction rather than explicit mesh generation. Weighted NTF provides an interpretable factorization in which each factor corresponds to an emission or transmittance plane, while the CNN learns an amortized inverse mapping constrained by the same optical propagation model.

The numerical results show that the inverse framework remains usable across different layer counts, content structures, fields of view, interlayer spacings, and angular input sizes. The 5×5 and 10×10 CNN results demonstrate the feasibility of increasing the directional sampling density needed for character-view continuity. Figure 2 provides a stylized-object visualization of the same multiview principle, but it is not used as a calibrated optical measurement. The conclusions are therefore limited to the display-side modeling mechanism and do not claim complete validation of semantic character generation, animation, or production-ready optical quality.

6.2. Limitations

The first limitation is the absence of a luminance- and color-matched optical comparison. The conventional prototype used a Redmi K30s LCD and the dynamic prototype used an iPhone 12 OLED; these panels differ in peak luminance, emission spectrum, gamma response, black level, color primaries, and viewing characteristics. The retained records do not include luminance in cd/m^2 , contrast ratio, luminous efficiency, angular crosstalk, camera RAW data, fixed-exposure metadata, or a calibrated white-balance reference. Figure 2 is therefore presented only as a stylized multiview visualization and is not used as quantitative proof of brightness, contrast, or character-display quality. A definitive validation should use radiometrically matched sources, fixed camera settings, angular crosstalk measurements, and energy-normalized comparison with the target light fields.

The second limitation concerns the gap between the retained benchmark content and the intended AI character-modeling application. The available corpus contains five general light-field scenes rather than a controlled cartoon-character dataset, and the archived protocol does not provide scene-disjoint folds. Future work should construct multiview cartoon-character data with controlled identity, pose, expression, material, silhouette complexity, and illumination, and should evaluate identity preservation, cross-view consistency, temporal stability, and perceptual quality in addition to PSNR.

The third limitation is the static optical stack. Printed transparency films and acrylic spacers do not reproduce electronically addressable liquid-crystal layers, synchronized refresh, calibrated grayscale transmittance, or the temporal demands of animated character content. A practical character-display prototype should use synchronized transmissive LCD panels, a matched dynamic emission plane, calibrated registration, and frame-sequential testing with animated multiview character sequences.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: AI-assisted language editing tools were used only for English language polishing. The author reviewed and edited the manuscript and takes full responsibility for the content of the publication.

Conflicts of Interest: The author declares no conflicts of interest.

References

1. Zhao, Y.; Chen, L.; Yu, H.; Zheng, Y. Intelligent generation of 3D disassembly processes driven by multimodal large language models. *J. Manuf. Syst.* **2026**, *86*, 1139–1160. [\[CrossRef\]](#)
2. Choi, H.S.; Cheng, T. Spatial cognition and emotion simulation in urban interior environments using agent-based modelling. *Int. J. Urban Sci.* **2026**, 1–24. [\[CrossRef\]](#)
3. Chen, J.; Shao, Z.; Zhu, H.; Chen, Y.; Li, Y.; Zeng, Z.; Yang, Y.; Wu, J.; Hu, B. Sustainable interior design: A new approach to intelligent design and automated manufacturing based on Grasshopper. *Comput. Ind. Eng.* **2023**, *183*, 109509. [\[CrossRef\]](#)
4. Chen, J.; Yin, H.; Zhang, K.; Ren, Y.; Zeng, H. Integration of neural networks in brain–computer interface applications: Research frontiers and trend analysis based on Python. *Eng. Appl. Artif. Intell.* **2025**, *151*, 110654. [\[CrossRef\]](#)
5. Xu, N.; Wu, F.; Liu, Z.; Zhan, Y. Evaluating the academic outcome of AI-powered joint support for at-risk students. *Comput. Educ. Artif. Intell.* **2025**, *9*, 100517. [\[CrossRef\]](#)
6. Xu, N.; Liu, G.; Tan, Q. High-fidelity far-field microscopy at $\lambda/8$ resolution. *Laser Photonics Rev.* **2022**, *16*, 2200307. [\[CrossRef\]](#)
7. Xu, N.; Liu, G.; Zhao, Y.; Tan, Q. Ultrahigh-aspect-ratio beam generation with super-resolution spot. *Appl. Phys. Lett.* **2021**, *119*, 094101. [\[CrossRef\]](#)
8. Xu, N.; Kong, Z.; Tan, Q.; Fu, Y. Multiring pure-phase binary optical elements to extend depth of focus. *Appl. Opt.* **2018**, *57*, 9643–9648. [\[CrossRef\]](#) [\[PubMed\]](#)
9. Xu, N.; Xiao, H.; Kong, Z.; Tan, Q. Axial multifocus beams formed by binary optical elements. *IEEE Photonics J.* **2019**, *11*, 6500410. [\[CrossRef\]](#)
10. Xu, N.; Liu, G.; Kong, Z.; Tan, Q. Creation of super-resolution hollow beams with long depth of focus using binary optics. *Appl. Phys. Express* **2020**, *13*, 012003. [\[CrossRef\]](#)
11. Liu, G.; Xu, N.; Yang, H.; Tan, Q.; Jin, G. Miniaturized structured illumination microscopy with diffractive optics. *Photon. Res.* **2022**, *10*, 1317–1324. [\[CrossRef\]](#)
12. Xu, N.; Liu, G.; Tan, Q. Adjustable super-resolution microscopy with diffractive spot-array illumination. *Appl. Phys. Lett.* **2020**, *116*, 254103. [\[CrossRef\]](#)
13. Xu, N.; Bohndiek, S.E.; Li, Z.; Zhang, C.; Tan, Q. Mechanical-scan-free multicolor super-resolution imaging with diffractive spot-array illumination. *Nat. Commun.* **2024**, *15*, 4135. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Xu, N.; Zhang, C.; Fu, Y.; Tan, Q. High-fidelity super-resolution microscopy datasets spanning multispectral to hyperspectral domains via diffractive optics. *Sci. Data* **2026**, *in press*. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Liu, J.; Guo, Z.; Wu, X.; Ren, Y.; Hu, M.; Zeng, H. Real-time observation of breathing vortex solitons in a Yb:CALGO laser. *Laser Photonics Rev.* **2026**, *20*, e01758. [\[CrossRef\]](#)
16. Liu, J.; Guo, Z.; Hu, M.; Zeng, H. Sub-100 fs vortex beams generated from a Yb:CALGO laser. *Laser Photonics Rev.* **2025**, *19*, e00819. [\[CrossRef\]](#)
17. Hu, M.; Wan, E.; Qiao, Y.; Xu, S.; Li, S.; Zhu, R.; Guo, J.; Chen, X.; Zeng, H. Transient interference-based spectroscopy for a molecular-bond-sensitive probe of lithium-ion batteries. *Energy Environ. Sci.* **2026**, *19*, 2889–2899. [\[CrossRef\]](#)
18. Jin, H.; Hu, M.; Wan, E.; Qiao, Y.; Duan, J.; Cheng, L.; Zeng, H. Density-clustering decoding of plasma micro-region dynamic molecular fingerprint spectra. *Laser Photonics Rev.* **2026**, e71451. [\[CrossRef\]](#)
19. Chu, Y.; Zhou, L.; Luo, G.; Kang, K.; Dong, S.; Han, Z.; Wu, L.; Meng, X.; Yang, C.; Guo, X.; et al. HorusEye: A self-supervised foundation model for generalizable X-ray tomography restoration. *Nat. Comput. Sci.* **2026**, *6*, 372–387. [\[CrossRef\]](#) [\[PubMed\]](#)
20. Yao, S.; Li, Z.; Guan, R.; Xia, M.; Hu, F.; Cao, K.; Xu, S.; Zhu, X.; Liu, R.W.; Ding, W. GIF-Calib: Geometry–intensity fusion for 4D radar–camera calibration in autonomous driving vehicles. *Inf. Fusion* **2026**, *136*, 104510. [\[CrossRef\]](#)
21. Li, X.; Lian, Y.; Bai, X.; Suo, P.; Xu, L.; Sun, J. Liquid-electrodes electrical impedance tomography sensor for single-cell imaging. *IEEE Sens. J.* **2026**, *26*, 4127–4138. [\[CrossRef\]](#)
22. Shen, M.; Sun, J.; Suo, P.; Zhang, X.; Yuan, Z.; Wang, Y.; Li, C.; Xu, L. Interface-sensitive electrical capacitance-ratio tomography for supercavity profile reconstruction in open-domain. *IEEE Trans. Instrum. Meas.* **2026**, *75*, 4502010. [\[CrossRef\]](#)
23. He, Y.; Liu, X.; Yang, R.; Chen, Q.; Deng, B.; Jin, M.; Wang, F. Deblur of infrared thermal images of in-service wind turbine blade based on rotational motion flow. *Measurement* **2026**, *287*, 122462. [\[CrossRef\]](#)
24. Zhong, Y.; Li, Y.; He, L.; Tang, Y.; Yang, Z.; Yu, X.; Zhang, J.; Li, W. Heavy-metal-free ZnSeTe quantum-dots-based liquid scintillators for scalable X-ray imaging. *Small* **2026**, *22*, e13925. [\[CrossRef\]](#) [\[PubMed\]](#)
25. Liu, F.; He, Y.; Deng, B.; Tang, L.; Zeng, Q.; Zhang, X.; Zhang, K. Detection and analysis of acoustic emission signal in lithium-ion batteries during charging and discharging. *IEEE Sens. J.* **2026**, *26*, 3751–3761. [\[CrossRef\]](#)
26. Yan, J.; Zhou, N.; Cheng, Y.; Zhang, F.; Wang, H.; Wang, M.; Jin, B.; Li, M.; Lu, Q.; Zhang, W. Application of machine-vision-driven physics-informed neural networks in pantograph–catenary system state detection. *Mech. Syst. Signal Process.* **2026**, *257*, 114577. [\[CrossRef\]](#)

27. Yang, H.; Hu, K.; Wang, H.; Hong, W.; Wang, X.; Wang, H.; Song, Y.; Liu, Z. BCLIP-ADer: A Bayesian prompt contrastive language–image pretraining method for catenary component anomaly detection in electrified railways. *IEEE Trans. Transp. Electrification*. **2026**, 1–13. [\[CrossRef\]](#)
28. Yang, H.; Liu, Z.; Liu, W.; Wang, H.; Zhang, Y.; Wang, H. Graph-MDETR: A graph-guided Mamba-DETR network for UAV catenary support component detection in electrified railways. *IEEE Trans. Intell. Transp. Syst.* **2026**, 27, 6319–6332. [\[CrossRef\]](#)
29. Duan, F.; Wang, H.; Yang, H.; Wei, C.; Zhang, C.; Song, Y.; Liu, Z. MRFM-IFCOS: An anchor-free interactive detector based on multi-receptive-field Mamba for detecting catenary support components. *IEEE Trans. Instrum. Meas.* **2025**, 74, 2553415. [\[CrossRef\]](#)
30. Sun, T.; Zhang, J.; Yang, H.; Yang, J.; Cheng, L. Bidirectional Mamba-based continuous prediction of human motion intention using multisource information fusion. *IEEE Trans. Autom. Sci. Eng.* **2026**, 23, 212–221. [\[CrossRef\]](#)
31. Gao, M.; Zhou, S.; Gu, W.; Fan, J.; Guan, A.; Zhu, H.; Wei, L.; Hu, Z. Enhancing distribution system state estimation under limited measurements: Leveraging large language model and multimodal information. *CSEE J. Power Energy Syst.* **2026**, 12, 622–631. [\[CrossRef\]](#)
32. Wu, X.; Dong, J.; Bao, W.; Zou, B.; Wang, L.; Wang, H. Augmented intelligence of things for emergency vehicle secure trajectory prediction and task offloading. *IEEE Internet Things J.* **2024**, 11, 36030–36043. [\[CrossRef\]](#)
33. Wu, X.; Wang, H.; Zhang, Y.; Zou, B.; Hong, H. A tutorial-generating method for autonomous online learning. *IEEE Trans. Learn. Technol.* **2024**, 17, 1558–1567. [\[CrossRef\]](#)
34. Chen, D.; Xu, C.; Yu, J.; Wang, Q.; Fang, H.; Yin, S.; Lookman, T.; Chen, R. Interpretable machine-learning framework for Nb–Si-based alloy design with enhanced fracture toughness. *Adv. Sci.* **2026**, e75815. [\[CrossRef\]](#) [\[PubMed\]](#)
35. Chen, L. Beyond external constraints: The missing dimension of AI governance. *SSRN Electron. J.* **2026**, 1–45. Available online: <https://ssrn.com/abstract=6449738> (accessed on 18 July 2026). [\[CrossRef\]](#)
36. Lv, J.; Wang, C.; Xie, L. Distributed output-feedback leader-following consensus of nonlinear multiagent systems. *Automatica* **2025**, 177, 112281. [\[CrossRef\]](#)
37. Xu, K.; Liu, J.; Huang, H.; Wang, C.; Zhao, W.; Ge, S.S. Multi-agent dynamic surface consensus control of a dual-motor steer-by-wire system under communication faults. *IEEE Trans. Ind. Electron.* **2026**, 1–12. [\[CrossRef\]](#)
38. Wang, R.; Guo, T.; Meng, D.; Wang, X.; Liang, B. Trajectory planning for continuum space manipulators based on generalized Jacobian operator. *IEEE Trans. Aerosp. Electron. Syst.* **2026**, 1–22. [\[CrossRef\]](#)
39. Yin, Y.; Liang, Y.; Zhang, X.; Qin, Z.; Han, X. A reduced-order Galerkin-projected flame transfer model for coupled dual-jet diffusion flames based on mixture fraction. *Combust. Flame* **2026**, 291, 115157. [\[CrossRef\]](#)
40. Bao, Y.; Wang, X.; Hu, H.; Zuo, Y.; Tu, S.; Zhang, P.; Qi, Q.; Deng, H. Transient vibration response analysis of underwater cylindrical shells based on precise integration method. *Ocean Eng.* **2026**, 358, 125881. [\[CrossRef\]](#)
41. Hu, H.; Wang, X.; Bao, Y.; Zuo, Y.; Zhang, P.; Qi, Q.; Deng, H. Transient analysis of cylindrical shells via a numerically stable discrete-time state-space technique. *Thin-Walled Struct.* **2026**, 231, 115357. [\[CrossRef\]](#)
42. Xu, B.; Lu, Q.; Gao, X.; Li, B.; Li, W.; He, J.; Gong, D.; Fan, Z. An optimized collaborative routing model for trucks and heterogeneous drones in delivery and pickup services. *IEEE Trans. Intell. Transp. Syst.* **2026**, 1–18. [\[CrossRef\]](#)
43. Yu, G.; Liu, J.; Hu, L.; Jiang, A.; Wang, L. Motion system modeling and acceleration analysis of five-axis hybrid machine tool. *Adv. Eng. Inform.* **2026**, 71, 104368. [\[CrossRef\]](#)
44. Wei, J.; Chen, Z.; Zhang, Y.; Zhang, W.; Li, J.; Lora, E.E.S.; Venturini, O.J.; Kozlov, A.N.; Isa, Y.M. An adjustable heat-to-power ratio combined heat and power system integrating biomass gasification, solid oxide fuel cell, and gas turbine: Energy, exergy, economic, and exergoeconomic (4E) analyses. *Int. J. Hydrogen Energy* **2026**, 244, 155706. [\[CrossRef\]](#)
45. Wang, L.; Zhang, B.; Li, Y.; Liu, B.; Tong, C.; Xiong, H.; Chen, G.; Hong, Z.; Zhang, C.; Tian, Y. Uplift resistance mechanism of pipes in lightweight backfill material of ceramsite. *Comput. Geotech.* **2026**, 192, 107927. [\[CrossRef\]](#)
46. Huang, X.; Kang, N.; Wang, Q.; El Mansori, M.; Guittonneau, F. Wear of directed energy deposited H13 steel as a function of its graded microstructure. *Tribol. Int.* **2025**, 207, 110617. [\[CrossRef\]](#)
47. Li, Y.; Qu, K.; Jiang, R.; Wang, H.; Zhao, X.; Yang, Z.; Tian, B.; Tao, J.; Chu, J.; Duan, C. Atomic-scale defect reconfiguration via thermally induced structural ordering for high-efficiency Sb₂Se₃ solar cells. *ACS Nano* **2025**, 19, 33460–33472. [\[CrossRef\]](#) [\[PubMed\]](#)
48. Dong, Y.; Bi, S.; Wang, M.; Yang, Z.; Tian, X.; Chen, Y.; Sun, Z.; Sun, R.; Ouyang, Y.; Bai, L.; et al. Plasma-driven material synthesis and structural chemical engineering of silicon-based anodes for lithium-ion batteries. *Chin. J. Struct. Chem.* **2026**, 45, 100907. [\[CrossRef\]](#)
49. Chen, X.; Huang, J.; Zhu, J.; Jing, J.; Hu, Z.; Bu, X.; Fan, J.; Li, S.; Song, W.L.; Han, X.; et al. An amorphous polyimide-based positive electrode for high-capacity and durable aluminum dual-ion batteries. *Angew. Chem. Int. Ed.* **2026**, 65, e2153127. [\[CrossRef\]](#) [\[PubMed\]](#)
50. Wen, L.; Hu, D.; Zhang, R.; Hu, X.; Luo, Y.; Chen, H.; Liang, Y.; Li, Y.; Wang, S. Two-dimensional SnSe₂/Bi₂Se₃ vertically stacked heterostructures for enhanced photoresponse through tuning layer coverage. *iScience* **2026**, 29, 115392. [\[CrossRef\]](#) [\[PubMed\]](#)

51. Su, Q.; Sun, Y.; Li, J.; Ma, B.; Zhang, X.; Tian, H.; Ju, Y.; Gao, S.; Liu, Z.; Zhang, T.; et al. Supramolecular engineering of a homo[2]catenane filler enables polymer composites with exceptional high-temperature capacitive energy storage. *Molecules* **2026**, *31*, 1691. [[CrossRef](#)] [[PubMed](#)]
52. Gu, G.; Xu, X.; Hong, Z.; Sheng, H.; Hou, D.; Fu, C. Nano-SiO₂-grafted PVA fibers for interfacial strengthening between PVA fiber and geopolymer matrix: Role of silane dosage. *Constr. Build. Mater.* **2026**, *528*, 146555. [[CrossRef](#)]
53. Hu, P.; Zhang, R.; Chen, L.; Li, L.; Tang, Q.; Hewitt, A.J. Atomization index to bridge the physicochemical properties of polymer-surfactant adjuvants and multivariate spray metrics: Guiding efficient pesticide formulation design. *Engineering* **2026**, in press. [[CrossRef](#)]
54. Yu, G.; Hu, L.; Qu, Y. Fretting slip wear mechanism of AlCrN-coated cemented carbide tool on titanium alloy TC4 under ultimate continuous cutting. *Int. J. Refract. Met. Hard Mater.* **2026**, *140*, 107859. [[CrossRef](#)]
55. Zhu, L.; Cui, C.; Jing, T.; Tan, S.; Liu, X.; Yu, M. Strong and broadband microwave absorption under thin thickness induced by multiple dielectric relaxation and multiple magnetic resonances through dual nanocrystalline phases in amorphous FeSiBCr flakes. *Acta Phys.-Chim. Sin.* **2026**, *42*, 100331. [[CrossRef](#)]
56. Zuo, Y.; Wang, X.; Gao, S. Ultra-low-frequency noise control in cruise modular cabin with thickness-graded star-shaped metamaterials. *Ocean Eng.* **2026**, *363*, 126490. [[CrossRef](#)]
57. Liu, X.; Wang, C.; Chen, C.; Pan, Z.; Gao, C.; Lai, W.; Zhao, J.; Tian, T.; Xiao, W. Recent advances in hierarchical porous materials for CO₂ capture and utilization. *Coord. Chem. Rev.* **2025**, *544*, 216927. [[CrossRef](#)]
58. Lai, C.-H.; Wu, T.-E.; Wang, C.-C. Enhancing information security in smart manufacturing through least significant bit steganography in engineering drawings. *J. Comput. Inf. Sci. Eng.* **2025**, *25*, 091006. [[CrossRef](#)]
59. Tang, D.; Tan, P.; Yan, Y.; Li, K.; Liang, W.; Qin, Z.; Zhang, J. Low-rate flow table overflow attacks in SDN: Model, analysis and a machine-learning-based mitigation approach. *IEEE Trans. Comput.* **2026**, *75*, 2253–2266. [[CrossRef](#)]
60. Han, X.; Wang, S.; Yu, J.; Wang, P.; Li, X. Design of PUF based on a Cr₂O₃-SnO₂ gas sensor for Internet of Things security. *Measurement* **2026**, *276*, 121395. [[CrossRef](#)]
61. Wang, S.; Zhang, X.; Wang, P.; Li, X. Design of magnetic sensor array-based PUFs for IoT security. *IEEE Trans. Instrum. Meas.* **2026**, *75*, 9517609. [[CrossRef](#)]
62. Wu, X.; Zou, B.; Lu, C.; Wang, L.; Zhang, Y.; Wang, H. Dynamic security computing framework with zero trust based on privacy domain prevention and control theory. *IEEE J. Sel. Areas Commun.* **2025**, *43*, 2266–2278. [[CrossRef](#)]
63. Li, H.; Gao, P.; Tan, H.; Li, H.; Zhu, H.; Jiang, M.; Ni, Y.; He, Y.; Huang, J.; Zhu, Y.; et al. Interaction-aware dexterous robot for minimally invasive transcanal inner ear interventions. *Nat. Commun.* **2026**, *17*, 5658. [[CrossRef](#)] [[PubMed](#)]
64. Lu, Y.-Z.; Sha, H.-Y.; Di, C.; Zhu, G.-C. A cross-kingdom survival strategy: A methanotroph recruits a microalga for synergistic detoxification of metronidazole. *J. Hazard. Mater.* **2026**, *503*, 141127. [[CrossRef](#)] [[PubMed](#)]
65. Wang, Z.; Wang, X.; Xu, S.; Zhou, R.; Zhang, M.; Li, W.; Zhang, Z.; Wang, L.; Chen, J.; Zhang, J.; et al. Off-site production of plasma-activated water for efficient disinfection: The crucial role of high-valence NO_x and new chemical pathways. *Water Res.* **2024**, *267*, 122541. [[CrossRef](#)] [[PubMed](#)]
66. Wang, H.; Wang, Y.; Li, Y.; Wang, X.; Chen, X.; Yu, S.; Wang, X. Spatial risk prediction and serological validation of emerging wild boar-borne diseases in Eastern Heilongjiang, China. *Pak. Vet. J.* **2025**, *45*, 1721–1732. [[CrossRef](#)]
67. Zeng, H.; Liu, X.; Liu, P.; Jia, S.; Wei, G.; Chen, G.; Zhao, L. Exercise's protective role in chronic obstructive pulmonary disease via modulation of the M1 macrophage phenotype through the miR-124-3p/ERN1 axis. *Sci. Prog.* **2025**, *108*, 00368504251360892. [[CrossRef](#)] [[PubMed](#)]
68. Liu, P.; Zhang, M.; Gao, H.; Han, S.; Liu, J.; Sun, X.; Zhao, L. Regulation of whole-transcriptome sequencing expression in COPD after personalized precise exercise training: A pilot study. *Respir. Res.* **2023**, *24*, 156. [[CrossRef](#)] [[PubMed](#)]

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.