

Article

Interactive Holographic Display System Based on Emotional Adaptability and CCNN-PCG

Yu Zhao ¹, Zhong Xu ¹, Ting-Yu Zhang ¹, Meng Xie ¹, Bing Han ¹ and Ye Liu ^{2,*}

¹ College of Information Engineering, Yangzhou University, Yangzhou 225127, China; zhaoyu@yzu.edu.cn (Y.Z.)

² Department of Creative Writing, Korea University, Sejong 30019, Republic of Korea

* Correspondence: listenyuran@korea.ac.kr

Abstract

Against the backdrop of the rapid advancement of intelligent speech interaction and holographic display technologies, this paper introduces an interactive holographic display system. This paper applies 2D-to-3D technology to acquisition work and uses a Complex-valued Convolutional Neural Network Point Cloud Gridding (CCNN-PCG) algorithm to generate a computer-generated hologram (CGH) with depth information for application in point cloud data. During digital human hologram building, 2D-to-3D conversion yields high-precision point cloud data. The system uses ChatGLM for natural language processing and emotion-adaptive responses, enabling multi-turn voice dialogs and text-driven model generation. The CCNN-PCG algorithm reduces computational complexity and improves display quality. Simulations and experiments show that CCNN-PCG enhances reconstruction quality and speeds up computation by over 2.2 times. This research provides a theoretical framework and practical technology for holographic interactive systems, applicable in virtual assistants, educational displays, and other fields.

Keywords: holographic display; color holography; image processing; computer-generated hologram; emotional adaptability



Academic Editor: Manohar Das

Received: 13 June 2025

Revised: 20 July 2025

Accepted: 24 July 2025

Published: 26 July 2025

Citation: Zhao, Y.; Xu, Z.; Zhang, T.-Y.; Xie, M.; Han, B.; Liu, Y.

Interactive Holographic Display System Based on Emotional Adaptability and CCNN-PCG.

Electronics **2025**, *14*, 2981.

<https://doi.org/10.3390/electronics14152981>

Copyright: © 2025 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license

(<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

1.1. Background

Currently, as technology advances at an unprecedented pace, consumer demand for personalized, intelligent, and immersive experiences continues to escalate. The integration of optical display and artificial intelligence (AI) interaction technologies has emerged as a focal point for both academia and industry. Compared to traditional two-dimensional (2D) display technologies, three-dimensional (3D) display technology has rapidly become a core pillar in domains such as 5G, big data, the metaverse, and the Internet of Things (IoT) [1–4]. Globally, holographic technology is widely recognized as the most promising ultimate solution for true three-dimensional displays. It not only accurately reproduces intricate details such as color and depth but also dynamically adjusts images based on the viewer's perspective, thereby creating an immersive three-dimensional visual experience. Interactive three-dimensional display technology has begun to gain prominence. Building on all the advantages of traditional holographic displays, this innovation allows for real-time interaction between users and three-dimensional images, delivering an unprecedented level of immersion.

Intelligent digital humans, a key application in this field, can replicate the appearance, personality, and behavioral traits of specific individuals based on their personal information and preferences. Moreover, through interactive three-dimensional holographic display technology, they enable real-time engagement with users, providing more authentic and deeply immersive experiences [5].

1.2. Related Studies

1.2.1. Research on Interactivity of Intelligent Digital Humans

In the landscape of digital transformation, intelligent digital humans have emerged as a novel medium for human–computer interaction, with breakthroughs in their interactive capabilities becoming a shared focus in both academia and industry [6,7]. Early digital human systems, constrained by traditional rule engines, primarily relied on keyword matching and predefined Q & A databases, leading to mechanical dialogs and poor contextual adaptability. With advancements in speech recognition and generative artificial intelligence, intelligent digital humans are gradually overcoming technical barriers, evolving toward multimodal and cognitively robust interaction paradigms.

In voice interaction, speech recognition systems based on HMM-DNN hybrid models and end-to-end deep learning architectures [8,9] equip digital humans with high-precision voiceprint recognition and speech-to-text capabilities. Leveraging breakthroughs in large language models (LLMs) [10], digital humans not only accurately parse voice commands but also acquire human-like interactive abilities—including semantic comprehension, contextual correlation, and creative responses—through training on massive corpora. This technological integration liberates digital humans from scripted dialogs, enabling them to autonomously generate contextually dynamic feedback, which significantly enhances the naturalness and intelligence of interactions.

In visual representation, innovations in 3D reconstruction technology have imbued digital humans with lifelike vitality. Tsinghua University's Unique3D framework [11] achieves precise single-view-to-3D model reconstruction via neural networks, while Nankai University's CAMixer architecture [12] addresses real-time rendering challenges for dynamic expressions and micro-motions.

Although the prior holographic voice interaction system applied [13] used salient object detection algorithms and a distributed computing architecture to achieve preliminary multichannel perception and real-time digital human interaction, current systems remain limited in deeper interactive dimensions such as contextual cognitive transfer and multimodal intent understanding.

1.2.2. Accelerating Computer-Generated Hologram Computation and Enhancing Quality

Significant global research efforts have been directed towards accelerating CGH computation. Initially, to address the slow generation of holograms, hardware acceleration emerged as a primary approach. International research teams explored the potential of graphic processing units (GPUs), leveraging their massive parallel computing capabilities to partition holograms. By assigning the light field computation of each micro-region to individual GPU cores—with one thread precisely managing per-pixel or per-point calculations—the computation time was significantly reduced [14,15].

To optimize memory usage and align with GPU access characteristics, researchers developed enhanced lookup table techniques [16,17], which substantially accelerated hologram computation. Shimobaba et al. introduced the wavefront recording plane (WRP) method [18], which significantly improved the efficiency of 3D scene hologram generation by incorporating virtual diffraction planes. Subsequently, Wang et al. integrated ray tracing with the WRP method for rapid hologram generation [19]. Shi et al. employed convolu-

tional neural networks (CNNs) trained on extensive datasets, reducing computation time to milliseconds—a revolutionary advancement published in *Nature* [20]. Peng et al. utilized camera-in-the-loop optimization, training high-quality holographic neural networks with optical images to simultaneously accelerate generation and enhance reconstruction quality [21]. To enhance the reconstruction quality of convolutional neural network-based methods, Zhong et al. introduced a complex-valued convolutional neural network that directly processes complex amplitudes to generate high-quality holograms while enhancing the network's representational capacity and computational efficiency [22]. Dong et al. developed a cross-domain super-resolution network that inputs low-resolution holograms and extracts inter-pixel information through multi-layer neural operations. This approach triples the speed compared to traditional methods while improving resolution, injecting new vitality into high-quality holographic displays [23]. Despite these algorithmic advancements, the computational speed remains inadequate for practical display applications.

Optimizing hologram generation for real-world objects represents another active research frontier. Li et al. deployed liquid crystal cameras equipped with fast-focusing elastic membrane lenses to capture real 3D scenes and generate holograms in real time [24]. However, as 3D capture technology advances, the demand for the personalized editing and processing of acquired data grows. Salient object detection—which filters out redundant background information and reduces processing complexity—is crucial for handling 3D data in holographic systems [25]. The authors previously proposed deep learning-based point cloud saliency segmentation and point cloud gridding (PCG) algorithms to accelerate CGH generation [26–29]. Nevertheless, challenges persist in real-time data acquisition and computation, along with insufficient reconstruction quality, which hinder practical deployment in intelligent 3D avatar interaction systems.

Substantial progress has also been made in enhancing CGH quality globally [30]. Some researchers have devised a diffraction model based on a split Lohmann lens, which synthesizes 3D holograms through single-step backpropagation. By integrating virtual digital phase modulation, this method significantly enhances the accuracy of 3D scene reconstruction, reduces computational costs, and boosts efficiency [31]. Sun et al. developed a pupil-aware gradient descent algorithm that combines multiple coherent sources and content-adaptive amplitude modulation in Fourier plane hologram spectra. Under the supervision of large-baseline target light fields, this innovation addresses the issues of poor image quality and short lifespan [32]. Despite breakthroughs, chromatic aberration remains a persistent challenge for color holography.

In summary, researchers worldwide have developed sophisticated algorithms addressing core holographic challenges: generation speed, reconstruction quality, and intelligent interaction. However, current computational efficiency and reconstruction quality for real-object holography still fall short of practical requirements for intelligent avatar interaction. To overcome these limitations, this paper presents an emotionally adaptive interactive holographic display system incorporating CCNN-PCG. The system employs 2D-to-3D technology for point cloud data acquisition. Interaction and emotionally adaptive responses are facilitated by ChatGLM. Furthermore, the CCNN-PCG algorithm reduces computational complexity while enhancing display quality. This framework provides a robust technical pathway for holographic intelligent interaction and offers novel perspectives for addressing related challenges.

2. Full-Color Holographic System

2.1. System Architecture

Building on the foundational research, objectives, and core scientific challenges outlined earlier, this paper delves deeply into optical signal acquisition and processing tech-

nologies. It integrates cutting-edge methodologies from multiple domains, including large language models (LLMs), CGH, holographic display technology, image processing algorithms, and holographic encryption strategies, to construct and optimize an interactive 3D holographic display system tailored for intelligent digital humans. The implementation blueprint adheres to the technical pathway illustrated in Figure 1, with the system comprising four core modules:

1. **Acquisition and Preprocessing Module:** This module utilizes the Unique3D framework and two-dimensional-to-three-dimensional technology to obtain depth multi-view images from single-view input, thereby generating 3D models. It performs point cloud sampling through Poisson sampling to improve the efficiency and accuracy of 3D model construction and point cloud data extraction, laying the foundation for subsequent holographic processing. Its main function is to efficiently build a digital human motion model library.
2. **Intelligent Voice Interaction Module:** To enable real-time voice interaction in the holographic system, this module integrates the ChatGLM large language model and an emotional adaptability analysis algorithm to improve fluency and accuracy. Utilizing Microsoft's Offline Speech Recognition API, it achieves speech-to-text conversion. The module constructs a point cloud model for digital humans that incorporates interactive textual information, enabling contextually appropriate responses through motion based on dialog content.
3. **Hologram Generation Module:** This module optimizes computational architecture to achieve high-quality hologram generation with significantly improved computational efficiency by using our proposed CCNN-PCG method. The point cloud data undergoes operations such as point removal, layering, and compression into an image. Subsequently, it is divided into three channels and input into the CCNN to obtain a three-channel output of the CGH. This advancement enables dynamic holographic display capabilities.
4. **Reconstruction Module:** This module innovatively adopts a double-layer verification mechanism. First, through high-precision numerical simulation, optical wave diffraction theory is used to simulate the hologram encoding data, and algorithms such as Fourier transform are used to verify the imaging effect in advance. Second, the module enters the optical reconstruction stage, relying on core devices such as spatial light modulators and lasers to convert digital signals into actual optical wave interference patterns. Through real-time monitoring and dynamic calibration, it ensures the spatial resolution and depth perception of color holographic images.

2.2. High-Quality Digital Human Model Generation and Processing

Automatically generating diverse and high-quality 3D models from single-view images is a fundamental task in 3D computer vision technology, with extensive applications across numerous fields. When collecting human figures with a single-depth camera, there are drawbacks such as insufficient depth accuracy, incomplete 3D information, poor real-time performance, weak anti-interference ability, limited application scenarios, and hardware cost bottlenecks [13]. However, 2D-to-3D conversion combined with Poisson sampling for point cloud collection can complement 3D information through multi-view image fusion, retain details with adaptive Poisson dense sampling, and simultaneously integrate RGB texture and depth data to improve accuracy, dynamically optimize data volume, and feature low hardware cost and strong scene adaptability. Therefore, this paper constructs the character depth information collection component in the intelligent digital avatar interaction and display system based on this technology.

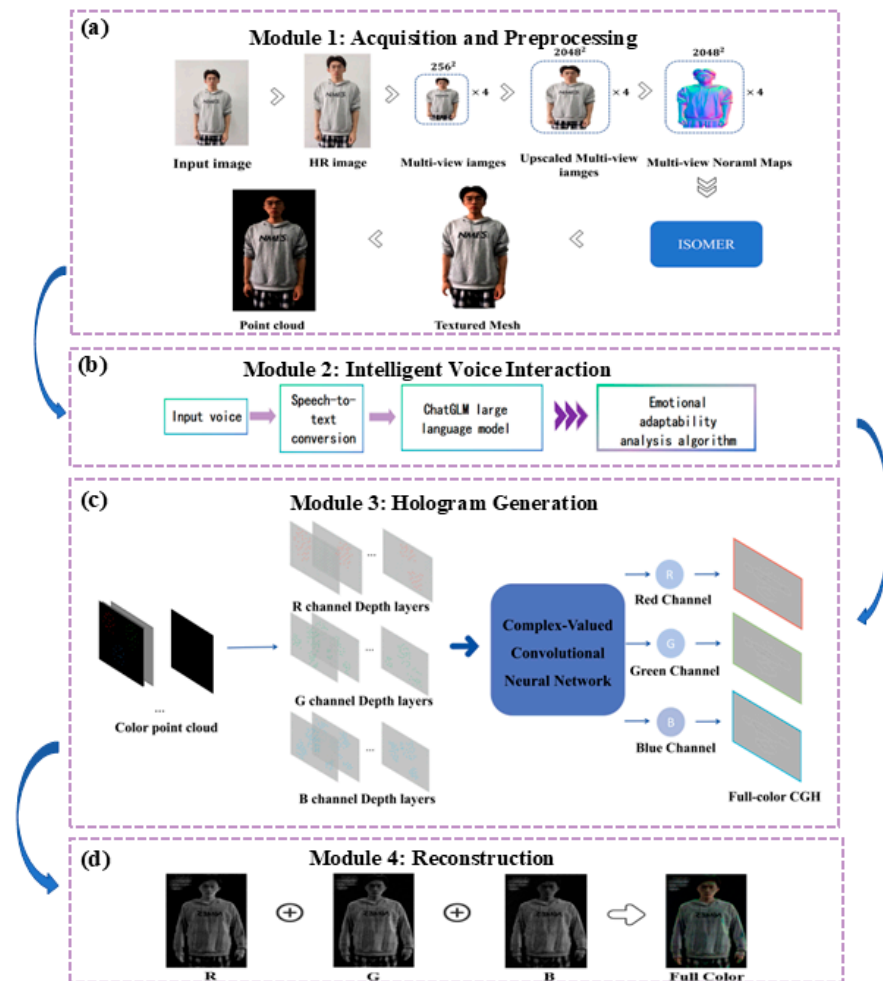


Figure 1. Modules of proposed holographic voice interaction system. (a) Acquisition and pre-processing module. (b) Intelligent voice interaction module. (c) Hologram generation module. (d) Reconstruction module.

In the acquisition phase, instead of using the depth camera from previous systems to capture character depth information [11], we adopt a method of generating high-quality 3D models from 2D images to further obtain point clouds for acquiring character depth information. In this part, this paper optimizes the process by using the Unique3D framework [9] to achieve the rapid acquisition of character depth information and generate high-quality digital human models.

We then employ a normal diffusion model to predict normal maps corresponding to multi-view color images. This section also adopts a multi-tier generation strategy, using a super-resolution model to increase resolution by 6–8 times. Finally, the normal diffusion model is used to predict normal maps for multi-view color images.

In the multi-view perceptual control mesh algorithm section, the process is divided into three parts: initial mesh estimation; mesh optimization; and explicit target optimization for multi-view inconsistency and geometric refinement. As shown in Figure 2, @ represents matrix multiplication, * represents element-wise multiplication, the initialization model undergoes model refinement to obtain a textured model, then proceeds to model coloring, and finally is converted into a point cloud model.

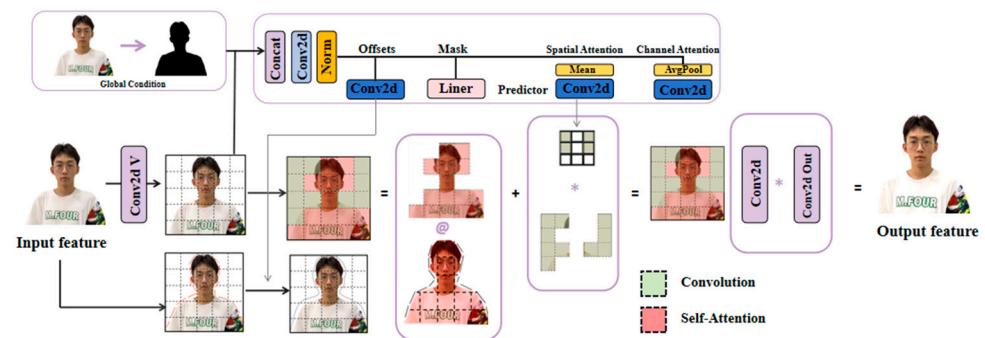


Figure 2. Point cloud generation process and single depth camera acquisition.

During initial mesh estimation, front rear views directly estimate the initial mesh. Orthographic maps from the front view are integrated to obtain depth maps (Equation (1)):

$$d(i, j) = \int_0^i \vec{n}(x) \cdot d\vec{x} \approx \sum_{t=0}^i n_x(t), \quad (1)$$

where $d(i, j)$ is the depth at coordinate (i, j) , $\vec{n}(x)$ is a normal vector, and $n_x(t)$ is the x-component normal at time t .

Pseudo-normal maps generated by diffusion cannot form rotation-free true normal fields. To address this, random rotation is introduced into normal maps before integration. This is repeated multiple times, and integration averages are used to calculate depth for reliable estimation. Then, the estimated depth is used to map pixels to spatial positions. Mesh models are created from the front/back views of objects. Next, smooth connections are ensured through Poisson reconstruction, and the models are simplified to less than 2000 faces for initialization. A 3D point cloud model of digital humans is then generated.

To further demonstrate the superiority of this method, we conducted a detailed comparison between this method and the point clouds collected by a single depth camera, as shown in Table 1, and the generation effects are presented in Figure 3.

Table 1. Comparison of 3D point cloud model acquisition.

	Number of Points	Viewing Angle	Point Cloud Distribution
Ours	200,000–900,000	360°	Uniform distribution
Depth Camera	50,000–300,000	180°	Concentrated distribution



Figure 3. Comparison of point cloud generation effects.

The point cloud generation method proposed in this study is capable of acquiring a full-view 3D point cloud model with a point count reaching 900,000, which significantly surpasses the maximum acquisition capacity of depth cameras. Furthermore, through

optimized sampling strategies, this method ensures the spatial uniformity of point cloud acquisition, effectively avoiding the problem of uneven data distribution caused by local point cloud aggregation during depth camera acquisition. This eliminates key factors affecting 3D reconstruction accuracy at the data quality level, providing a reliable data foundation for subsequent high-precision model construction.

2.3. Emotional Adaptability Analysis

Research indicates significant cognitive and behavioral differences between men and women in intimate relationships. Men and women exhibit distinct preferences in expansive postures: men tend to favor dominant gestures, while women prefer constrictive ones. In emotional expression, female digital characters more frequently adopt closed postures, such as lowering their heads with hunched shoulders, smiling while lowering their heads, or gently twisting their fingers. Male digital characters tend to display expansive and dominant postures, such as standing with their legs apart, clasping their hands behind their back, making outward or upward gestures, or pointing at others to reinforce authority.

Thus, this paper introduces an affective adaptation analysis algorithm module to enhance digital humans' non-verbal behavioral feedback following user emotion recognition. Through this mechanism, digital humans achieve more accurate emotional synchronization during voice interactions.

By pre-categorizing digital human emotions, this algorithm identifies label words in responses generated by large language models (LLMs), determines emotion categories based on recognized label words, triggers corresponding emotional actions, and ultimately generates a point cloud model of interactive movements to form action models incorporating interactive text. As shown in Figure 4, we apply this emotion analysis algorithm to classify the avatar's happiness, sadness, standby state, and other emotions with a classifier. After the LLM generates responses based on the user's vocal queries, this affective adaptation analysis algorithm automatically identifies the generated text content, extracts emotional keywords, and invokes the corresponding emotional interactive action point cloud model based on the emotion category of the keyword. This interactive action point cloud model is output alongside the LLM-generated text content, ultimately producing a digital human action model matching the emotional category of the text.

In the affective adaptation analysis module, we extract key information such as nouns and adjectives from the target domain. Descriptors constitute a set of label words representing high-probability words predicted by the PLM within a given context. Unlike methods relying on external knowledge, we extract domain-specific nouns and adjectives from the target domain using vocabulary annotation tools, avoiding noise and complexity associated with external sources.

Following the final construction of the descriptors, each word is associated with its corresponding token class. Implemented as cross-domain classification, this involves populating the "[MASK]" with each word from the label word space and computing its probability P . Subsequently, the predicted probability of each token word is mapped to its corresponding class label, and the average of these predicted probabilities is used for final classification. The predicted probability for the final class label can be expressed as $P([MASK]) = v \in V_t$. This is a function that converts token word probabilities into class label probabilities.

$$P(y|x) = f(P_M|v \in V_t) \quad (2)$$

where f is a function that transforms token probabilities into class label probabilities. In the experiments, the cross-entropy loss function is introduced, updating the parameters of the entire model during training as follows:

$$L = -\frac{1}{N_t} \sum \log p(y^*|x) + \alpha ||\theta||^2 \quad (3)$$

where $\sum$ represents the sum of losses across all samples, y^* denotes the true label, and α is a parameter regulating the impact of regularization on the total loss.

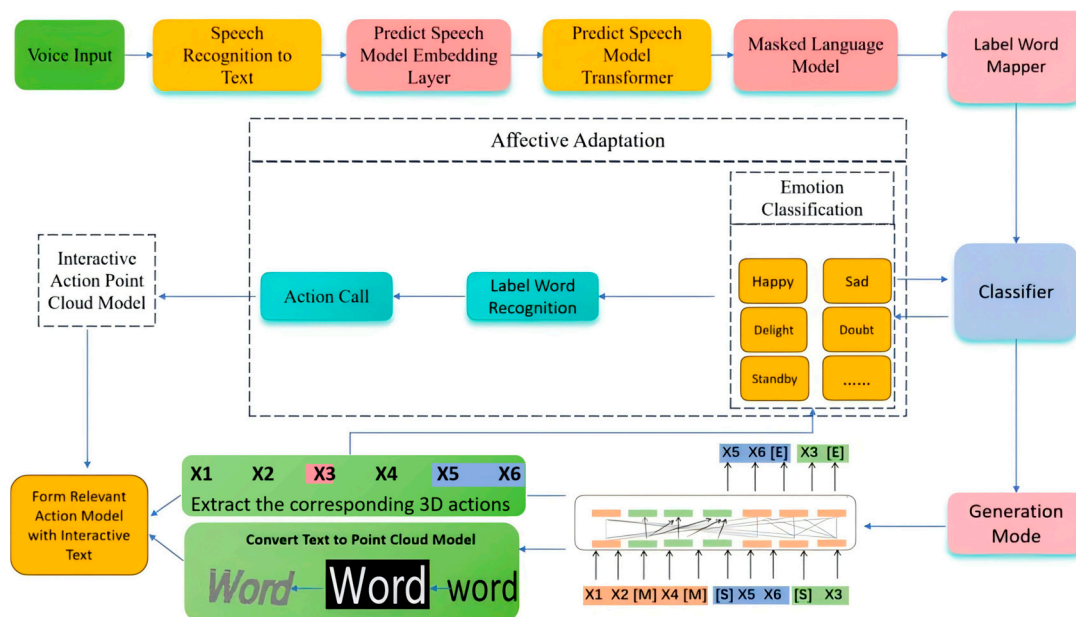


Figure 4. The affective adaptation analysis-based voice interaction module. The voice input is first converted to text and processed by the speech model, masked language model, etc. Then it enters affective adaptation, where label words are recognized, and emotions are classified. Combined with the interactive action point cloud model, features are extracted and converted into point clouds and output by the generation mode via the classifier.

As shown in Figure 5, the Speech Recognition and Text Conversion Module can utilize Microsoft's offline speech recognition interface. The system captures user voice input in real time and converts it into text. This process forms the foundation of voice interaction, ensuring the accurate transmission of user instructions. When large language models generate lengthy or complex textual content, the intelligent digital human that utilizes this emotion analysis algorithm can more accurately and effectively identify the emotional intent of the textual response and execute corresponding auxiliary motions.

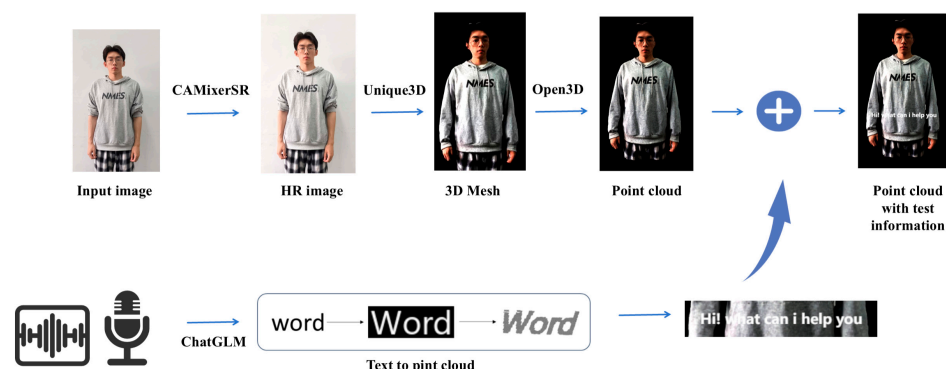


Figure 5. Extract corresponding 3D actions from input text.

As shown in Table 2, 14 sets of action modules were collected using our method for generating 3D point clouds from a single view, and an action library was established. Based on the corresponding point cloud modules output by the intelligent voice interaction module, we inserted the answer text point clouds. In small-sample scenarios, traditional algorithms achieved 90% recognition accuracy [13], whereas the emotion-adaptive algorithm proposed in this paper attains 97% accuracy, representing a substantial improvement over conventional methods. This enhancement underscores the algorithm's adaptability to emotional nuances in limited data environments.

Table 2. Common action modules for holographic voice interactive system.

Number	1	2	3	4	5	6	7
Models	Standby1	Standby2	Affirmation	Negation	Instructions	Fear	Surprise
Number	8	9	10	11	12	13	14
Models	Ponder	Uneasiness	Confidence	Anxiety	Fatigue	Laugh	Cry

2.4. Complex-Valued Convolutional Neural Network Point Cloud Gridding Algorithm

Currently, in the field of computer-generated holography, numerous methods can achieve hologram generation and enhance quality, but the process of generating high-quality computational holograms is generally time-consuming. In light of this, this study builds upon the point cloud meshing algorithm from the authors' prior research by incorporating a convolutional neural network to accelerate hologram generation and improve reconstruction quality.

The CCNN exhibits unique advantages in holographic phase preservation, which stem from its deep alignment with the physical properties of optical fields. The core of holography lies in the encoding of optical field complex amplitudes, and processes such as light diffraction are inherently complex-domain operations. The CCNN directly uses complex numbers for input, output, and weights, avoiding the information fragmentation caused by real-valued networks splitting complex amplitudes into real and imaginary parts.

This paper proposes a hierarchical computational hologram generation method based on a CCNN. By leveraging learnable initial phases and taking the complex amplitude of the computational hologram plane as input, the method predicts computational holograms, with input depth layers randomly selected during training. The model generates 2D reconstructed computational holograms through different depth layers and synthesizes complex longitudinal magnification to produce layered holograms. The CCNN encodes computational holograms with minimal time consumption and imposes no additional computational burden for multiple layers. Numerical reconstruction and optical experiments demonstrate that this method enables real-time hologram generation for layered scenes. The system generates holograms for the digital human's interactive point cloud model, achieving holographic 3D reconstruction with depth information. The network structure employed is illustrated in Figure 6. The CCNN also incorporates residual skip connections (SCs). The addition of SCs helps mitigate the issues of gradient vanishing and explosion, promotes feature reuse, and simplifies the learning process of the network to accelerate convergence.

Meanwhile, it enhances the model's ability to capture information at different scales and levels, thereby improving the model's representational power, performance, and generalization ability. The input to the complex-valued omni-dimensional dynamic convolution layer is the complex amplitude A of the SLM plane across multiple depth layers. The output is a complex amplitude A that contains more feature information and has a different number of channels, and the phase of this complex amplitude is used as the single-phase CGH.

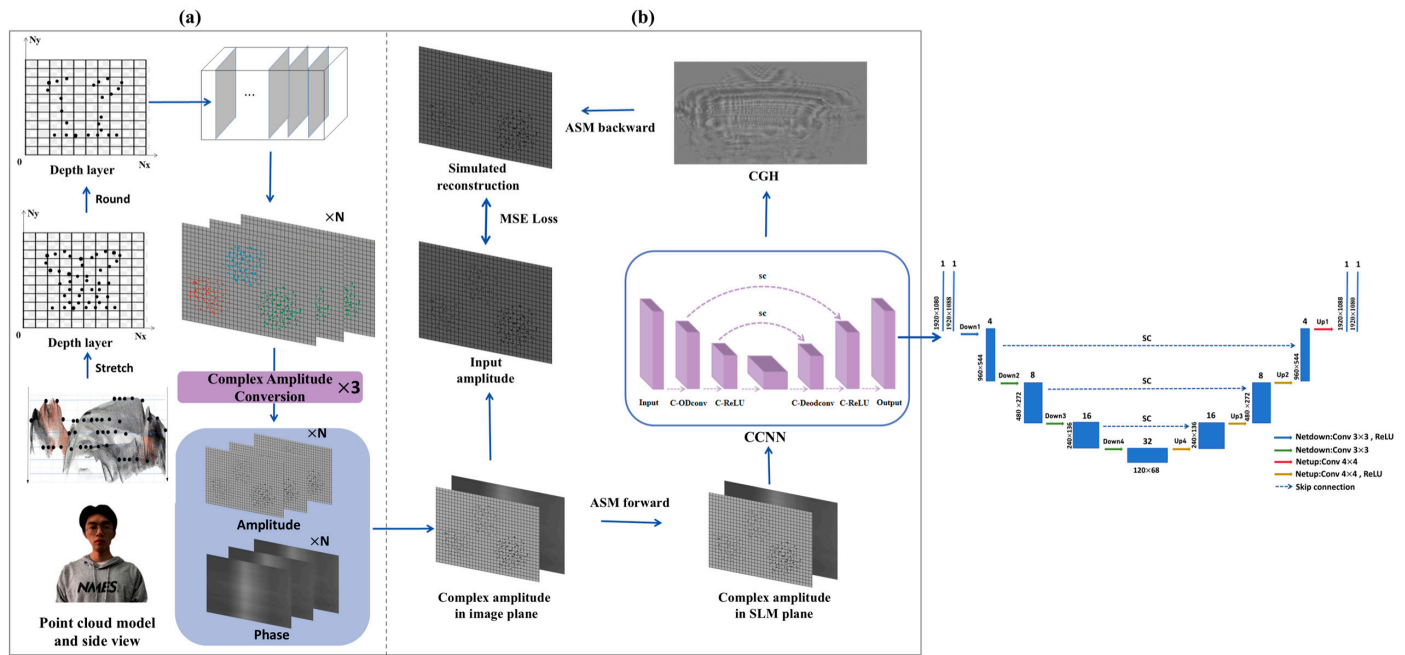


Figure 6. The overall process of the CCNN-PCG algorithm. (a) Point cloud processing module: The point cloud undergoes sampling, de-pointing, slice compression, and complex amplitude conversion to obtain the complex amplitude of the point cloud slice image. (b) The structure of the CCNN algorithm. Finally, the output CGH is obtained.

In this network, the number of point cloud grid layers and the distance between each grid layer must first be configured. For each grid layer, an initial phase set to zero is employed as a learnable parameter during training. For each input amplitude, its depth layer is randomly selected from different target grid layers. The corresponding initial phase is then assigned to the input amplitude. After the system performs segmentation from the image plane to the computational hologram plane using the Angular Spectrum Method (ASM), the CCNN is employed for encoding control. The numerically reconstructed amplitude is subsequently compared with the input amplitude. Within this network, input amplitude A is randomly assigned depth d , corresponding phase φ_d , and complex amplitude C_d as follows:

$$C_d = A \exp(i\varphi_d) \quad (4)$$

Next, the ASM calculates the propagation distance from the image plane to the SLM plane. The complex amplitude at the SLM plane is given by the following:

$$D_d = F^{-1}[F(C_d) \cdot H_d] \quad (5)$$

$$H_d = \begin{cases} \exp(ikz_d \sqrt{1 - \lambda^2 f_x^2 - \lambda^2 f_y^2}), & \text{if } \sqrt{f_x^2 + f_y^2} < \frac{1}{\lambda} \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

where F and F^{-1} represent the 2D fast Fourier transform and its inverse, k is the wavenumber, z_d denotes propagation distance, λ is the wavelength, and f_x and f_y are spatial frequencies along the x - and y -axes. Zero-padding interpolation and band-limited ASM propagation are applied for enhanced diffraction calculation accuracy. The *loss* function is defined as follows:

$$Loss = \frac{1}{Z} \sum_{i=1}^Z \|A_{R,i} - A_{T,i}\|^2 \quad (7)$$

where $A_{R,i}$ and $A_{T,i}$ represent the amplitude values of the reconstructed image and the target image at the i -th pixel, respectively, and Z is the total number of pixels. With

the addition of the attention mechanism, the network has stronger feature extraction capabilities, which can be reflected in the train loss values during the training process.

The CCNN emphasizes complex-valued operations to preserve phase information and strengthen nonlinear mapping. By integrating PCG with the CCNN, their complementary strengths lead to enhanced feature extraction and processing capabilities, improving point cloud grid hologram generation across multiple dimensions. CCNN-PCG further processes and fuses features in complex-valued space, enabling the network to learn more comprehensive and representative point cloud characteristics. This provides stronger support for generating high-quality mesh holograms.

During training, distinct parameters are set for different channels to optimize output. The processed point cloud data is then separated into three channels (RGB) and fed into the algorithm to generate single-channel holograms. Finally, holograms from all three channels are combined to produce a full-color hologram. In terms of training data, we used 100 layers of point cloud grids and 700 images from the public dataset DIV2K as the training set. In addition, we used 100 images from the DIV2K dataset, excluding the 700 images used for training, as the validation set. Comparisons were made with algorithms such as PCG, the CCNN, and Holo-Encoder under the same testing environment, thereby demonstrating the robustness and universal value of the proposed method.

To enhance the network's performance in generating a CGH from point cloud sliced images, we adopted a multimodal approach for the dataset. Specifically, we incorporated point cloud sliced images into the DIV2K training set, with the ratio of regular images to point cloud sliced images adjusted to 7:1. We set the number of training iterations to 800, the number of validation iterations to 100, the number of training epochs to 50, and the learning rate to 0.08.

3. Experiment and Results

3.1. Interactive Voice Experiment Verification

This section presents the digital human point cloud display results based on the emotion analysis algorithm for voice interaction and applies our proposed CCNN-PCG algorithm to generate digital human point cloud holograms. Simulation tests were conducted on a Windows 11 (64-bit) PC equipped with an NVIDIA GeForce RTX 4090 (24GB video memory, Santa Clara, CA, USA) and an Intel Core i9-13900 (32 GB RAM, Santa Clara, CA, USA) using MATLAB R2024a and Python 3.12.

In the human point cloud generation process, Unique3D constructs 3D models from input 2D images; Open3D combined with Poisson disk sampling generates uniformly distributed colored point clouds, with output quality surpassing that of our previous depth camera-based acquisition system. The effect is shown in Figure 7.

During system operation, users can select either a male or female digital human and activate standby motions. As depicted in Figure 8a,b, when a user issues requests for voice input, the system generates inquiry gestures and questions through its emotion analysis-based voice interaction module. When a user asks, "Could you recommend some books?", the system identifies this as a general instruction, performs emotion analysis on the response content, generates recommendation gestures, and provides suggested book titles. As illustrated in Figure 8c,d, our interactive system supports both Chinese and English. The experimental process and results of voice interaction in the color holographic display system demonstrate that the voice interaction system exhibits interactivity.

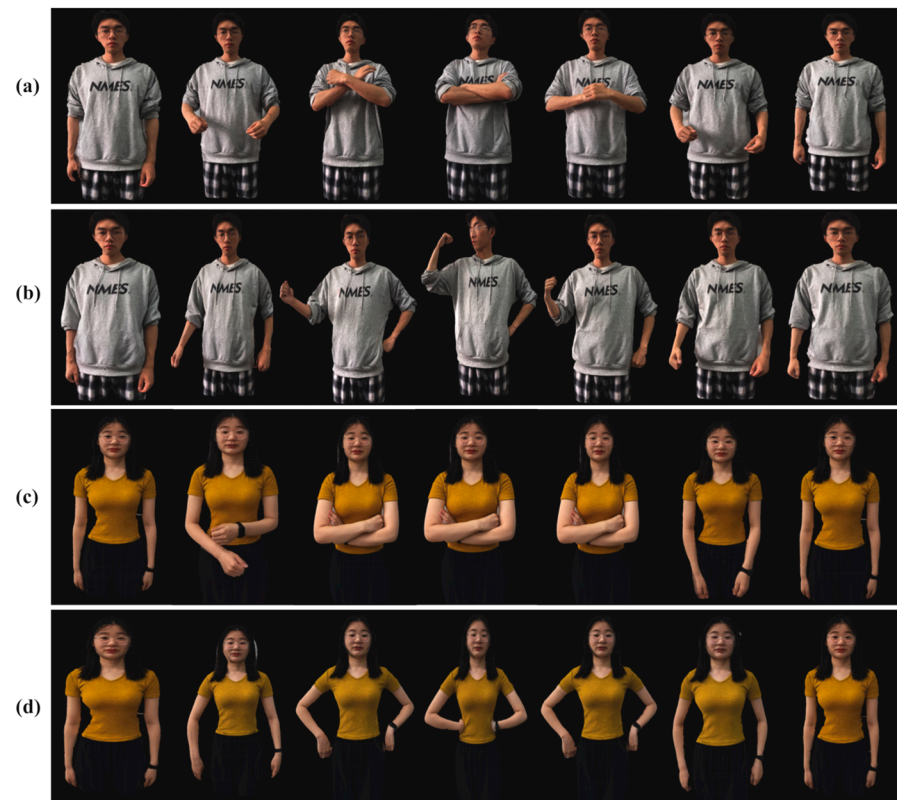


Figure 7. Digital human actions by gender: (a) male standby, (b) male delight, (c) female standby, and (d) female delight.



Figure 8. Interactive voice experiment verifies results: Voice input requests book recommendations from (a) male digital human and (b) female digital human. (c) Voice input requests book recommendations in Chinese and (d) introduction of some travel destinations in English.

3.2. Generation Speed and Reconstructed Image Quality Enhancement

To highlight the performance of the CCNN-PCG algorithm in generating 3D point cloud-based holograms, we compared it with the Holo-encoder [20] and PCG [11] algorithms. By comparing layered RGB point cloud images with corresponding simulated reconstruction images, average peak signal-to-noise ratio (PSNR) values were obtained. After dataset modification and network optimization, our method achieved improved PSNR metrics while maintaining generation speed when producing a 1920×1080 -resolution CGH. Simulated reconstructions of CCNN-PCG, Holo-encoder, and PCG are depicted in Figure 9a–c, respectively. For point cloud-based CGH generation, CCNN-PCG demonstrates significantly superior performance over the CCNN while simultaneously generating R, G, and B channel holograms. Figure 9d,e present comparative reconstructions of CCNN-PCG versus the CCNN without PCG. When GPU memory permits, CCNN-PCG generates full-color 30-layer CGHs in 0.066 s—5% faster than the CCNN (without PCG), 2.32 times faster than PCG, and 3.18 times faster than Holo-encoder. Compared to the CCNN (without PCG), CCNN-PCG achieves a 9% higher PSNR and 50% higher SSIM. The PSNR and generation time comparisons across all methods are shown in Figure 9f,g.

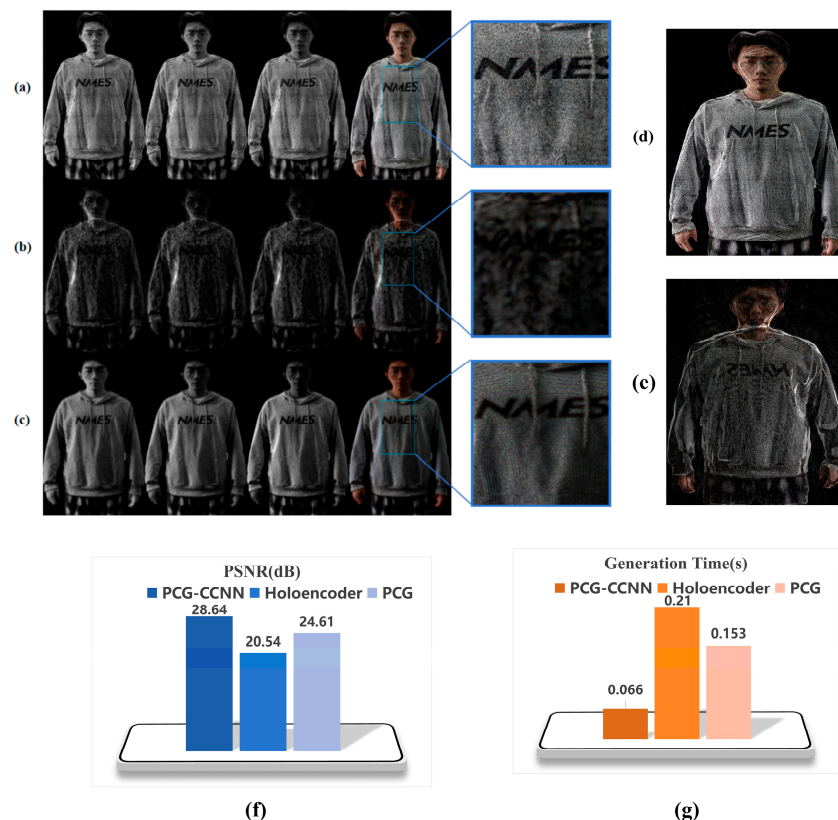


Figure 9. Numerical reconstruction in R, G, and B channels, and full-color reconstructed images of (a) our proposed CCNN-PCG, (b) Holo-encoder, and (c) PCG methods. Comparison of simulation reconstruction effects: (d) CCNN with PCG and (e) CCNN without PCG. Comparison of (f) PSNR and (g) generation times of holograms.

To more clearly demonstrate the improvements in CCNN-PCG over traditional CNNs, we conducted a structural comparison between CCNN-PCG and the traditional CNN of Holo-encoder, as shown in Table 3.

Table 3. Comparison between CCNN-PCG and Holo-Encoder architecture.

	Network Layers	Arity	Convolution Type	Normalization Module	Activation Function
CCNN-PCG	3	938	C-conv	Used	C-Relu
Holo-encoder	8	Million-level	Conv	Batch Normalization	Relu

As shown in Table 4, our proposed CCNN-PCG method achieves quality improvements while accelerating CGH generation, and RGB channels can be processed in parallel. When generating holograms over 400,000 point clouds, CCNN-PCG produces a color CGH in 0.063 s, approximately 3.1–3.6 times faster than the Holo-encoder method and 2.2–2.4 times faster than the PCG methods.

Table 4. Calculation of generation time of holographic system when resolution is 1920×1080 .

Object			Generation Time		
Name	Points	Layers	PCG	Holo-Encoder	CCNN-PCG
Figure 7a—1	450,429	30	0.154	0.227	0.063
Figure 7a—6	423,568	30	0.157	0.214	0.068
Figure 7b—1	414,298	30	0.152	0.221	0.064
Figure 7b—6	470,651	30	0.152	0.208	0.067
Figure 7c—1	438,386	30	0.153	0.231	0.064
Figure 7d—1	435,582	30	0.153	0.223	0.066

As shown in Table 5, in terms of reconstruction quality, the Holo-encoder algorithm exhibits insufficient performance in processing point cloud data. Our proposed CCNN-PCG method achieves an approximately 40–44% higher PSNR compared to the Holo-encoder and about a 15–20% higher PSNR than PCG. Additionally, our CCNN-PCG algorithm can encode complex-amplitude holograms with different wavelengths and reconstruction distances, demonstrating stronger adaptability and higher coding efficiency.

Table 5. Calculated PSNR of holographic system when resolution is 1920×1080 .

Object			PSNR		
Name	Points	Layers	PCG	Holo-Encoder	CCNN-PCG
Figure 7a—1	450,429	30	25.41	20.47	29.34
Figure 7a—6	423,568	30	24.18	20.63	29.32
Figure 7b—1	414,298	30	24.22	19.98	28.95
Figure 7b—6	470,651	30	23.46	20.33	29.01
Figure 7c—1	438,386	30	23.33	21.07	28.87
Figure 7d—1	435,582	30	23.74	19.54	28.28

As illustrated in Figure 10, we selected male and female digital humans and conducted diverse interactions to generate corresponding interactive motions and textual responses. This process validated the system's feasibility through both simulated and optical experiments. The optical setup for holographic display is illustrated in Figure 10a. To achieve color holographic display, a time division multiplexing experimental architecture employing RGB laser sources (250 mW, 12 V, 1 A) was implemented: 638 nm red/520 nm green/450 nm blue laser beams were attenuated, expanded, and projected onto a spatial light modulator (SLM). The full-color phase-only SLM (CAS Microstar, Xi'an, China) featured a resolution of 1920×1080 with a pixel pitch of $4.5 \mu\text{m}$. At a refresh rate of $3 \times 60 \text{ Hz}$, three-channel holograms were sequentially loaded onto the SLM. When the refresh rate exceeded the human eye's response time, the reconstructed images were perceived as color images. The reconstructed images of the R, G, and B channels and the color-simulated reconstructed images at different reconstruction distances are shown in Figure 10b–e. The

optical reconstruction effects of the R, G, and B channels and the color one are shown in Figure 10f–i.

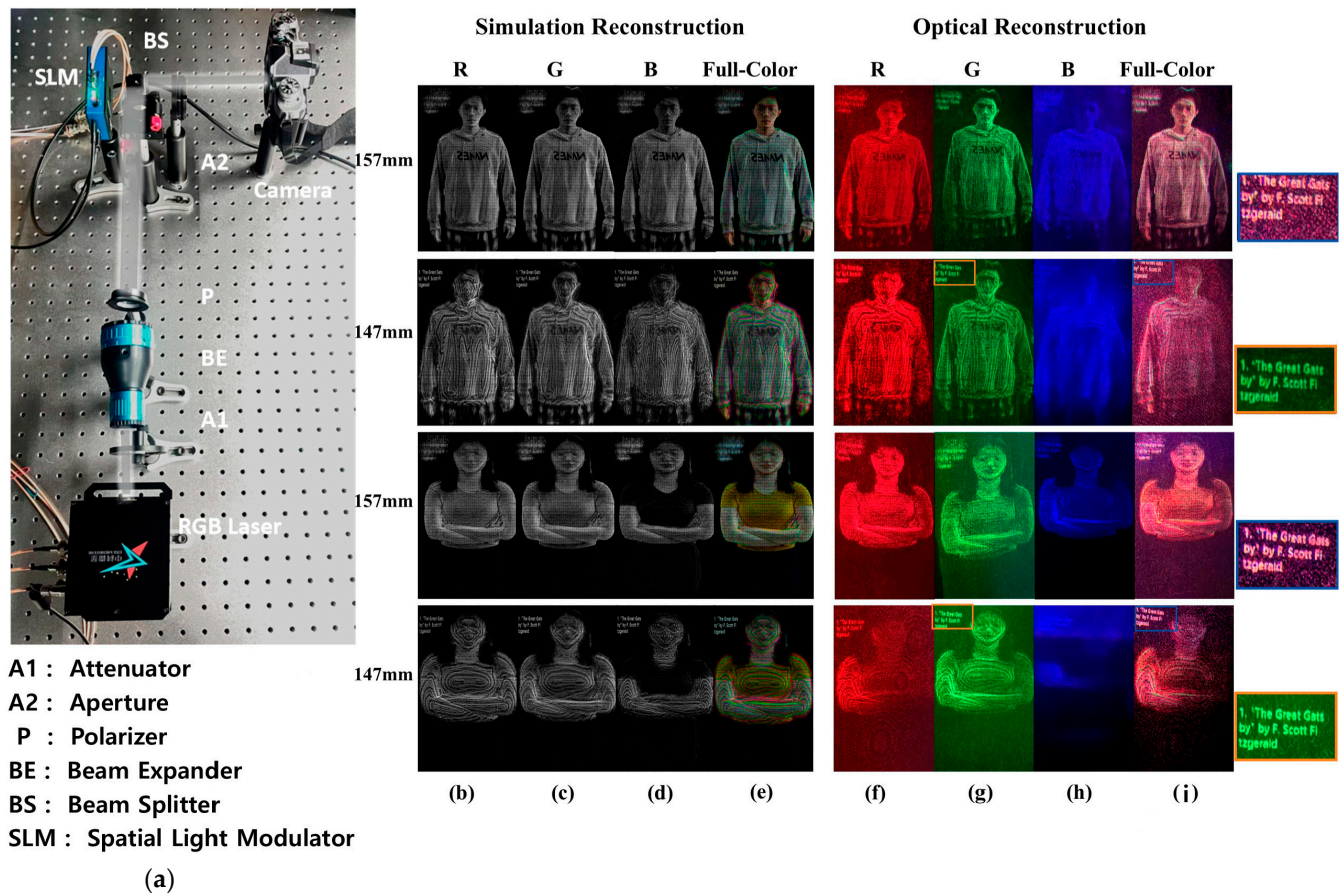


Figure 10. (a) Optical setup for holographic display, (b) red channel, (c) green channel, (d) blue channel, and (e) full-color simulated reconstructed images; (f) red channel, (g) green channel, (h) blue channel, and (i) full-color optical reconstructed images.

Optical reconstructed images are presented in Figure 10 f–i. After transmitting phase holograms containing digital human point cloud information and textual interactive information to the SLM, the camera-captured patterns clearly demonstrate the in-focus/out-of-focus effects in optical experiments. When the focusing distance was 147 mm, the textual interactive information in the optical reconstruction was also clearly visible. Both simulated and optical experiments validate the effectiveness of our proposed CCNN-PCG method.

4. Conclusions

This paper proposes a holographic voice interactive display system incorporating speech interaction and holographic modules. As the core voice interaction component, the ChatGLM model enables text interaction, emotional adaptation, and digital human motion control within the holographic system. For the holographic module, the CCNN-PCG algorithm is employed to generate a CGH from point clouds with enhanced efficiency and superior quality. The experimental results demonstrate measurable improvements in both generation speed and reconstruction quality compared to the PCG and Holo-encoder methods. This research pioneers approaches for integrating full-color holographic reconstruction with interactive technologies, offering innovative technical pathways and practical references for the advancement of human–computer interaction from 2D screens to 3D holographic environments.

However, since CCNN-PCG requires a large amount of video memory during the training process, this greatly limits the number of depth layers contained in the generated CGH. The maximum number of layers that can be trained with 24 GB of video memory is only about 200, which significantly restricts the demand for generating a CGH with depth information for large scenes. In future research, we will focus on improving the network's feature extraction capability to further increase the number of depth layers generated by the CGH, thereby making the network applicable to more scenarios.

Author Contributions: Conceptualization, Y.Z. and Y.L.; Methodology, Y.Z., Z.X. and Y.L.; Software, Y.Z., Z.X., T.-Y.Z. and M.X.; Validation, Z.X. and B.H.; Formal analysis, Z.X., T.-Y.Z. and M.X.; Investigation, Z.X. and B.H.; Resources, Z.X., T.-Y.Z., M.X., B.H. and Y.L.; Data curation, T.-Y.Z., B.H. and Y.L.; Writing—original draft, Y.Z.; Writing—review & editing, Y.L.; Visualization, M.X.; Supervision, Y.L.; Funding acquisition, Y.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Natural Science Foundation of China (No. 62205283) and Natural Science Foundation of Jiangsu Province (No. BE2023340).

Data Availability Statement: All data in this manuscript are available from the corresponding author upon reasonable request.

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

References

1. Cao, L.C.; He, Z.H.; Liu, K.X.; Sui, X.M. Progress and challenges in dynamic holographic 3D display for the metaverse. *Infrared Laser Eng.* **2022**, *51*, 267–281.
2. Chen, C.P.; Ma, X.; Zou, S.P.; Liu, T.; Chu, Q.; Hu, H.; Cui, Y. Quad-channel waveguide-based near-eye display for metaverse. *Displays* **2023**, *81*, 102582. [\[CrossRef\]](#)
3. Tang, Y.; Yi, J.; Tan, F. Facial micro-expression recognition method based on CNN and transformer mixed model. *Int. J. Biom.* **2024**, *16*, 463–477. [\[CrossRef\]](#)
4. Tan, F.; Zhai, M.; Zhai, C. Foreign object detection in urban rail transit based on deep differentiation segmentation neural network. *Heliyon* **2024**, *10*, e37072. [\[CrossRef\]](#)
5. Nagahama, Y. Interactive zoom display in a smartphone-based digital holographic microscope for 3D imaging. *Appl. Opt.* **2024**, *63*, 6623–6627. [\[CrossRef\]](#)
6. Li, J.; Lai, Y.; Li, W.; Ren, J.Y.; Zhang, M.; Kang, X.H.; Wang, S.Y.; Li, P.; Zhang, Y.Q.; Ma, W.Z.; et al. Agent Hospital: A simulacrum of hospital with evolvable medical agents. *arXiv* **2024**, arXiv:2405.02957. [\[CrossRef\]](#)
7. Durante, Z.; Huang, Q.; Wake, N.; Gong, R.; Park, J.S.; Sarkar, B.; Taori, R.; Noda, Y.; Terzopoulos, D.; Choi, Y.; et al. Agent AI: Surveying the horizons of multimodal interaction. *arXiv* **2024**, arXiv:2401.03568. [\[CrossRef\]](#)
8. Prabhavalkar, R.; Hori, T.; Sainath, T.N.; Schlüter, R.; Watanabe, S. End-to-end speech recognition: A survey. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2024**, *32*, 325–351. [\[CrossRef\]](#)
9. Yang, F.; Yang, M.; Li, X.; Wu, Y.; Zhao, Z.; Raj, B. A closer look at reinforcement learning-based automatic speech recognition. *Comput. Speech Lang.* **2024**, *87*, 101641. [\[CrossRef\]](#)
10. Boisseau, É. Imitation and large language models. *Minds Mach.* **2024**, *34*, 42. [\[CrossRef\]](#)
11. Wu, K.; Liu, F.; Cai, Z.; Yan, R.J.; Wang, H.Y.; Hu, Y.T.; Duan, Y.Q.; Ma, K.S. Unique3D: High-quality and efficient 3D mesh generation from a single image. *arXiv* **2024**, arXiv:2405.20343.
12. Wang, Y.; Zhao, S.J.; Liu, Y.; Li, J.L.; Zhang, L. CAMixerSR: Only Details Need More Attention. *arXiv* **2024**, arXiv:2402.19289. [\[CrossRef\]](#)
13. Zhao, Y.; Huang, Z.; Ji, J.; Xie, M.; Liu, W.; Chen, C.P. Holographic voice-interactive system with Taylor Rayleigh-Sommerfeld based point cloud gridding. *Opt. Lasers Eng.* **2024**, *179*, 108270. [\[CrossRef\]](#)
14. Pan, Y.; Xu, X.; Liang, X. Fast distributed large-pixel count hologram computation using a GPU cluster. *Appl. Opt.* **2023**, *52*, 6562–6571. [\[CrossRef\]](#)
15. Kwon, M.W.; Kim, S.C.; Kim, E.S. Three-directional motion-compensation mask-based novel look-up table on graphics processing units for video-rate generation of digital holographic videos of three-dimensional scenes. *Appl. Opt.* **2016**, *55*, A22–A31. [\[CrossRef\]](#)
16. Pi, D.; Liu, J.; Han, Y.; Khalid, A.; Yu, S. Simple and effective calculation method for computer-generated hologram based on non-uniform sampling using look-up-table. *Opt. Express* **2019**, *27*, 37337–37348. [\[CrossRef\]](#)

17. Cao, H.K.; Jin, X.; Ai, L.Y.; Kim, E.S. Faster generation of holographic video of 3-D scenes with a Fourier spectrum-based NLUT method. *Opt. Express* **2021**, *29*, 39738–39754. [\[CrossRef\]](#)
18. Shimobaba, T.; Masuda, N.; Ito, T. Simple and fast calculation algorithm for computer-generated hologram with wavefront recording plane. *Opt. Lett.* **2009**, *34*, 3133–3135. [\[CrossRef\]](#)
19. Wang, Y.; Sang, X.; Chen, Z.; Li, H.; Zhao, L. Real-time photorealistic computer-generated holograms based on backward ray tracing and wavefront recording planes. *Opt. Commun.* **2018**, *429*, 12–17. [\[CrossRef\]](#)
20. Shi, L.; Li, B.; Kim, C.; Kellnhofer, P.; Matusik, W. Towards real-time photorealistic 3D holography with deep neural networks. *Nature* **2021**, *591*, 234–239. [\[CrossRef\]](#)
21. Peng, Y.; Choi, S.; Padmanaban, N.; Wetzstein, G. Neural holography with camera-in-the-loop training. *ACM Trans. Graph.* **2020**, *39*, 185. [\[CrossRef\]](#)
22. Zhong, C.L.; Sang, X.Z.; Yan, B.B.; Li, H.; Chen, D.; Qin, X.J. Real-Time High-Quality Computer-Generated Hologram Using Complex-Valued Convolutional Neural Network. *IEEE Trans. Vis. Comput. Graph.* **2024**, *30*, 3709–3718. [\[CrossRef\]](#)
23. Dong, Z.X.; Jia, J.D.; Li, Y.; Ling, Y.Y. Divide-Conquer-and-Merge: Memory-and Time-Efficient Holographic Displays. In Proceedings of the 2024 IEEE Conference Virtual Reality and 3D User Interfaces (VR), Orlando, FL, USA, 16–21 March 2024.
24. Li, Z.S.; Liu, C.; Li, X.W.; Zheng, Y.; Huang, Q.; Zheng, Y.W.; Hou, Y.H.; Chang, C.L.; Zhang, D.W.; Zhuang, S.L.; et al. Real-time holographic camera for obtaining real 3D scene hologram. *Light Sci. Appl.* **2025**, *14*, 74. [\[CrossRef\]](#)
25. Chen, S.; Yu, J.; Xu, X.; Chen, Z.; Lu, L.; Hu, X.; Yang, Y. Split-guidance network for salient object detection. *Vis. Comput.* **2023**, *39*, 1437–1451. [\[CrossRef\]](#)
26. Chen, S.; Tang, H.; Huang, Y.; Zhang, L.; Hu, X. S2dinet: Towards lightweight and fast high-resolution dichotomous image segmentation. *Pattern Recognit.* **2025**, *164*, 111506. [\[CrossRef\]](#)
27. Zhao, Y.; Bu, J.W.; Liu, W.; Ji, J.H.; Yang, Q.H.; Lin, S.F. Implementation of a full-color holographic system using RGB-D salient object detection and divided point cloud gridding. *Opt. Express* **2023**, *31*, 1641–1655. [\[CrossRef\]](#)
28. Bu, J.W.; Zhao, Y.; Ji, J.H. Full-color holographic system featuring three-dimensional salient object detection based on a U2-RAS network. *J. Opt. Soc. Am. A* **2023**, *40*, B1–B7. [\[CrossRef\]](#)
29. Yang, Q.H.; Zhao, Y.; Liu, W.; Bu, J.W.; Ji, J.H. A full-color holographic system based on Taylor Rayleigh-Sommerfeld diffraction point cloud grid algorithm. *Appl. Sci.* **2023**, *13*, 4466. [\[CrossRef\]](#)
30. Sun, X.H.; Mu, X.Y.; Xu, C.; Pang, H.; Deng, Q.L.; Zhang, K.; Jiang, H.B.; Du, J.L.; Yin, S.Y.; Du, C.L. Dual-task convolutional neural network based on the combination of the U-Net and a diffraction propagation model for phase hologram design with suppressed speckle noise. *Opt. Express* **2022**, *30*, 2646–2658. [\[CrossRef\]](#)
31. Chang, C.L.; Ding, X.; Wang, D.; Ren, Z.Z.; Dai, B.; Wang, Q.; Zhuang, S.L.; Zhang, D.W. Split Lohmann computer holography: Fast generation of 3D hologram in single-step diffraction calculation. *Adv. Photonics Nexus* **2024**, *3*, 036001. [\[CrossRef\]](#)
32. Chao, B.; Gopakumar, M.; Choi, S.; Kim, J.; Shi, L.; Wetzstein, G. Large Étendue 3D holographic display with content-adaptive dynamic fourier modulation. In Proceedings of the SIGGRAPH Asia 2024 Conference Paper, Tokyo, Japan, 3–6 December 2024; pp. 1–12.

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.