photonics

MDPI

Article

MAENet: A Multi-Scale Attention Efficient Network for Coherent Noise Suppression in Digital Holographic Microscopy

Yifan Zhu, Jing Yu, Zihao Zhang, Ming Kong, Yushuo Feng, Feixue Hou, Zihan Tang and Wei Liu *

College of Metrology Measurement and Instrument, China Jiliang University, Hangzhou 310018, China; p23020854149@cjlu.edu.cn (Y.Z.); 2300201403@cjlu.edu.cn (Z.Z.); s24020804014@cjlu.edu.cn (Y.F.)
* Correspondence: liuw@cjlu.edu.cn

Abstract

Coherent noise in digital holographic microscopy (DHM) seriously degrades the accuracy of quantitative phase imaging, limiting its applications in fields such as nondestructive testing. However, traditional numerical denoising methods struggle to achieve an ideal balance between noise suppression, detail preservation, and computational efficiency. To address this challenge, we propose a multi-scale attention efficient network (MAENet). This network employs a dual-encoder architecture to achieve complementary extraction of multi-scale features. To efficiently integrate the features from these two branches, a dual-branch dense attention fusion (DDAF) module is designed. It performs a weighted fusion of features from the dual branches via an adaptive attention mechanism and enhances feature representation via dense residual connections, significantly boosting the model's denoising performance. Furthermore, a hierarchical fusion strategy is adopted to preserve high-frequency details in the shallow layers of the network while performing feature fusion in the deeper layers, thereby maximizing protection of image textures while effectively suppressing noise. To address the lack of paired training data in real-world scenarios, a DHM simulation system capable of simulating the key physical characteristics of coherent noise was constructed. Extensive experiments on the simulated dataset show that MAENet achieves a PSNR of 33.25 dB and an SSIM of 0.93042, outperforming various mainstream denoising algorithms and demonstrating its excellent performance in suppressing coherent noise, providing an effective solution for denoising in coherent imaging systems.

Keywords: digital holography; coherent noise; deep learning; attention mechanism

Check for updates

Received: 14 January 2026
Revised: 10 February 2026
Accepted: 5 March 2026
Published: 20 March 2026

Correction Statement: This article has been republished with a minor change. The change does not affect the scientific content of the article and further details are available within the backmatter of the website version of this article.

Copyright: © 2026 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.

1. Introduction

Digital holographic microscopy (DHM) is a powerful quantitative phase-imaging technique that can recover the amplitude and phase distribution of an object from a single recorded hologram. With the advantages of being non-contact, label-free, and wide-field, DHM has become one of the core methods in numerous research fields, such as biomedical applications, wavefront sensing, and so on. However, coherent noise arising from the light source's coherence is a critical challenge in DHM. Coherent noise in digital holographic microscopy (DHM) severely degrades the accuracy of quantitative phase imaging, often introducing artifacts that cause nanometer-scale measurement fluctuations. In applications such as surface morphology and non-destructive testing, these artifacts can be indistinguishable from genuine micro-features. This noise severely influences

measurement accuracy and reconstructed image quality, thereby constraining its further development in various applications.

Methods to suppress noise in coherent imaging systems are broadly categorized into optical and numerical methods. Optical methods primarily aim to suppress noise by reducing the coherence of the light source in the time or spatial domain, but this often increases the complexity and cost of the optical system. In contrast, numerical methods have been more extensively studied due to their flexibility and effectiveness. Numerical methods can be further divided into three classes: spatial domain methods (e.g., mean filter [1], median filter [2], and non-local means filter [3]), transform domain methods (e.g., block-matching 3-D filtering (BM3D) [4] and windowed Fourier filtering [5]), and algorithms based on deep learning [6–13]. While transform-domain methods strike a good balance between denoising performance and detail preservation, their image decomposition and reconstruction processes often incur high computational cost. Although spatial-domain filtering is computationally fast, its performance may be severely limited when dealing with coherent noise. In contrast, learning-based methods exhibit significant advantages in both noise suppression and inference efficiency.

Recently, deep learning algorithms, such as convolutional neural networks (CNNs), have become a mainstream research direction in image denoising. However, many denoising models are designed under the assumption of Gaussian noise, rendering them ineffective against coherent noise. As demonstrated in prior research, the point spread function (PSF) of DHM systems causes the statistical distribution of composite real-world noise to significantly deviate from the Gaussian distribution [14]. To this end, a DHM simulation system was constructed to accurately characterize the key properties of coherent noise, including its non-Gaussian statistics, non-stationarity, and spatial correlation. The high consistency between the simulated noise and experimental noise validates the accuracy and effectiveness of our simulation system. While increasing the depth of neural networks can yield better performance, it also leads to substantial increases in parameter count and inference time, thereby hindering the model's application in real-world scenarios. To address this issue, researchers have explored alternative architectures, such as dual-encoder designs [15–18], to improve representation ability while avoiding the high computational costs of extreme depth. If one simply turns to computationally intensive widening structures, it can also degrade inference speed, which poses a constraint in dynamic scenarios. Therefore, our aim was to seek a computationally efficient solution. Inspired by the dual-branch structure and guided by a theoretical analysis of coherent noise characteristics, we propose a multi-scale MAENet to efficiently remove coherent noise in holograms.

Overall, the main contributions of our method are summarized as follows:

(1) We propose a dual-branch encoding architecture composed of an enlarged scale efficient (ESE) encoder and a basic scale detail (BSD) encoder. This design aims to efficiently extract complementary multi-scale features without excessively increasing network depth. The ESE encoder uses depthwise separable convolutions with a 7×7 kernel to capture broad noise information and spatially correlated noise patterns at low computational cost. The BSD encoder employs standard 3×3 convolutions, focusing on preserving high-frequency details and fine image textures. This dual-path parallel design enables the model to achieve a robust balance between effective global noise modeling and high-fidelity local detail reconstruction.

(2) We present a dual-branch dense attention fusion (DDAF) module for efficiently integrating complementary features from the dual-branch encoders. The core functionality of DDAF is collaboratively achieved by two branches: an adaptive fusion (AF) module that utilizes an attention mechanism to dynamically assign weights to features from each branch, thus generating an optimal combination. The fused features are then

processed by a dense residual enhancement (DRE) module, which significantly improves the model's denoising performance by promoting feature reuse and enhancing representational capability.

The remainder of this paper is arranged as follows. Section 2 reviews related denoising methods. Section 3 details the proposed MAENet architecture. Section 4 analyzes the experimental results. This paper is summarized in Section 5.

2. Related Work

2.1. Deep Neural Networks for Image Denoising

Before the advent of convolutional neural networks, noise suppression in DHM primarily relied on traditional techniques [1–5], which are usually categorized into spatial-domain and transform-domain methods. Spatial-domain-based methods are computationally faster but can cause image blurring in high-density regions [13,14]. In contrast, transform domain methods, such as windowed Fourier filtering (WFF) [5], while better at preserving details, are limited by high computational cost and artifacts. In summary, traditional methods often struggle to achieve an ideal balance between denoising performance and computational efficiency, resulting in insufficient robustness in DHM denoising applications.

Deep learning-based methods have provided a new solution for the limitations of traditional methods, rapidly becoming a major research focus. Among them, convolutional neural network models, thanks to their end-to-end feature learning, have achieved breakthrough performance across various quantitative evaluation metrics. In particular, the DnCNN proposed by K. Zhang et al. [19] has surpassed the denoising performance of classic traditional methods such as BM3D [4] and WNNM [20]. By incorporating residual learning [21], DnCNN effectively handles blind denoising problems while simultaneously promoting network performance and accelerating training convergence. However, a study reports that it performs poorly on repeated textures [22]. This limitation is particularly relevant in DHM because structured parasitic fringes frequently exhibit periodic texture-like patterns, making them harder to suppress. Furthermore, the U-Net architecture [23] has been shown to effectively eliminate coherent noise due to its efficient feature learning. However, when the U-Net is directly employed as a denoising model, its standard skip-connection fusion introduces an inherent trade-off [24]. While skip connections preserve fine textures and structural details, shallow encoder features often contain substantial noise. Without effective suppression during feature fusion, these noise-corrupted features are propagated to the decoder, thereby compromising denoising performance. The success of these works indicates that deep learning-based methods have become a mainstream research direction in image denoising. Deep learning-based noise suppression methods extract image features from training data, achieving excellent denoising performance and a lower computational cost than traditional methods.

It is also crucial to explicitly define the specific challenges posed by coherent noise suppression in digital holographic microscopy (DHM), consisting of both speckle noise and structured parasitic fringes. While numerous deep learning denoising models exist for suppressing speckle noise or parasitic fringes, these single-purpose models are often insufficient for coherent noise suppression in DHM. A model designed only for speckle may fail to remove the circular parasitic fringes, while a de-striping algorithm may not address the global speckle noise. Therefore, a successful denoising approach for DHM must address both noise types simultaneously. As models addressing this specific hybrid noise are rare, we designed MAENet to fill this gap and selected the U-Net architecture as a baseline, thereby demonstrating the need for specialized designs over single-purpose or generic solutions.

2.2. Attention Mechanism and Skip Connections

One of the core challenges for convolutional neural networks in image denoising tasks lies in how to effectively extract and preserve key feature information [25], especially when dealing with coherent noise [26]. To this end, researchers have introduced attention mechanisms [27,28], which enable the network to dynamically focus on the most informative parts of the feature maps, thereby extracting more discriminative features [29–31]. For example, an attention block is utilized to specifically capture noise information in the ADNet [32]. Meanwhile, in RIDNet, the feature attention module improves denoising performance by emphasizing the weights of the most informative features [33]. Although attention mechanisms can strengthen feature reweighting and improve denoising quality, attention modules often introduce additional computation and model complexity, motivating lightweight attention designs for broader vision tasks [34].

In parallel, to address the issue of detail loss during layer-by-layer propagation in deep networks, skip connections provide critical technical support for information flow and feature reuse. They provide a pathway for information to flow across deep layers of the network, effectively preventing vanishing or exploding gradients as the network becomes deeper. For instance, the symmetric design of DSNet [35] and the dense connections in the RDN [36] both validate the significant advantages of skip connections in preserving fine image details and stabilizing model training. However, in denoising networks, concatenative skip connections inevitably expand channel dimensionality after feature fusion, thereby increasing computational burden and memory consumption. Furthermore, dense connections repeatedly aggregate intermediate features, which further amplifies memory usage and computational complexity. Without careful design constraints, densely connected denoising models may become inefficient and resource-intensive, limiting their practical applicability [37]. In summary, the investigation of attention mechanisms and skip connections has greatly pushed the performance of image denoising networks. From the DHM perspective, these prior works collectively suggest that an effective coherent noise suppressor should combine multi-scale representation with selective fusion, so that coherent noise is suppressed without sacrificing fine details.

3. The Proposed Denoising Model

3.1. Architecture of MAENet

The overall architecture of the proposed MAENet is shown in Figure 1. This architecture extends the classic encoder–decoder structure, adopting an asymmetric design with a dual-branch encoder and a single decoder. It aims to achieve end-to-end mapping, learning from the noisy image Y to the corresponding clean image $\hat{X}$ using effective extraction and fusion of multi-scale features. Specifically, MAENet comprises three core components: a dual-branch encoder for parallel extraction of multi-scale features, a dual-branch dense attention fusion (DDAF) module for adaptively enhancing and fusing features, and a decoder for refining image reconstruction.

To capture image features under different receptive fields, the encoder is designed with two parallel branches. The first branch (BSD) uses a 3×3 kernel to extract fine local texture information. The second branch (ESE) uses depthwise separable convolutions with a 7×7 kernel, effectively expanding the receptive field and capturing a broader range of image information while significantly reducing computational complexity.

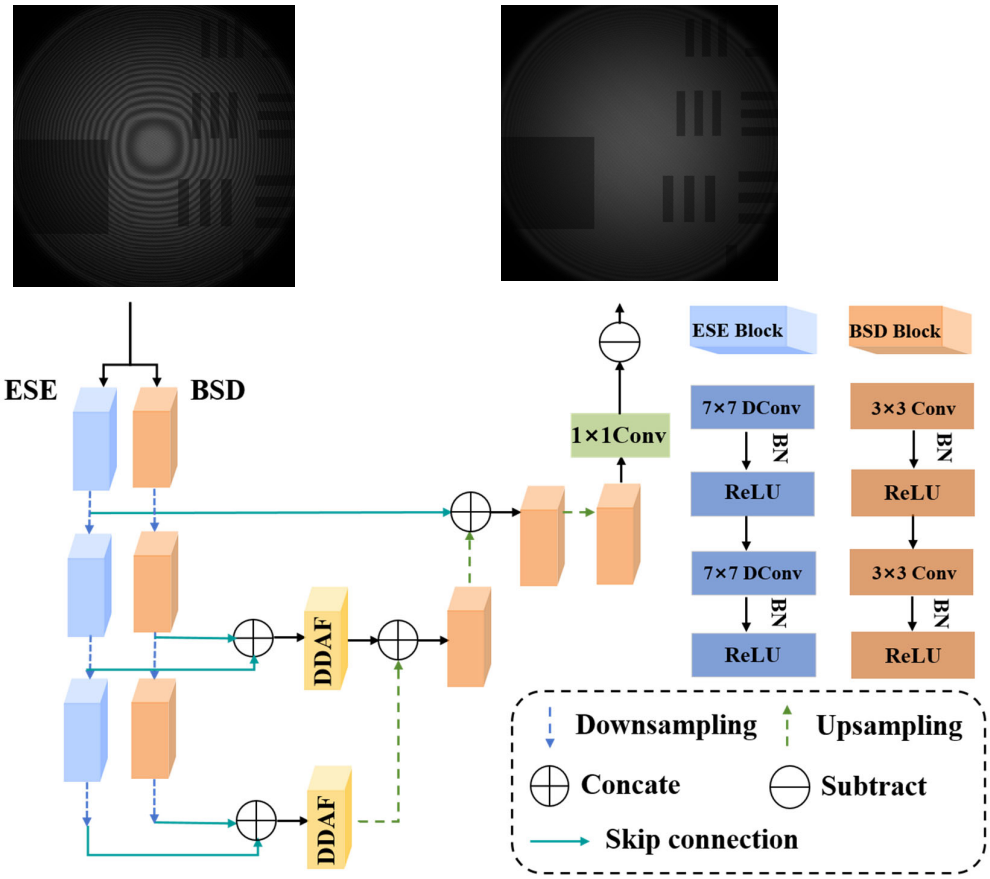


Figure 1. Network architecture of the proposed MAENet, composed of four parts: an enlarged-scale efficient (ESE) encoder, a basic-scale detail (BSD) encoder, a decoder, and a dual-branch dense attention fusion (DDAF).

At each encoder level i $\in \{1,2,3\}$, the input feature map F_{in}^{i-1} is fed into the two branches, processed by feature extraction blocks H_{BSD}^{i} and H_{ESE}^{i}, respectively, generating two sets of feature maps F_{BSD}^{i} and F_{ESE}^{i} with different emphases:

$$F_{BSD}^{i} = H_{BSD}^{i}(F_{BSD}^{i-1}) \tag{1}$$

$$F_{ESE}^{i} = H_{ESE}^{i}(F_{ESE}^{i-1}) \tag{2}$$

where $F_{ESE}^{i}, F_{BSD}^{i} \in \mathbb{R}^{C\times H\times W}$, C, H, and W represent the number of channels, height, and width of the feature map, respectively. Each feature-extraction block consists of two convolutional operations and one max-pooling operation, which not only progressively reduce spatial dimensions but also effectively increase the receptive field of subsequent feature layers.

The decoder progressively restores the image's spatial resolution through a series of upsampling operations. At each upsampling level i $\in \{1,2,3\}$, the decoder concatenates the upsampled output from the previous level F_{up}^{i+1} with features F_{DDAF}^{i} from the DDAF module (or F_{bsd}^{i} from the shallowest BSD branch) and then processes them through convolutional layers. This effectively integrates high-level semantic information from the encoding stage with low-level fine features. This process can be expressed as follows:

$$F_{dec}^{i} = H_{dec}^{i}([U(F_{dec}^{i+1}), F_{DDAF}^{i}]) \tag{3}$$

where $U()$ represents an upsampling operation using transposed convolution or bilinear interpolation, and $[\cdot,\cdot]$ denotes feature concatenation. Finally, after the last upsampling

layer and processing by the output convolution layer $H_{out}()$, the network directly outputs the predicted clean image $\hat{X}$:

$$\hat{X} = H_{out}(F_{dec}^{i}) \tag{4}$$

Through this end-to-end mapping, the network directly learns the complex nonlinear transformation from the noisy image Y to the clean image $\hat{X}$. The entire model is optimized to minimize the discrepancy between the predicted result $\hat{X}$ and the ground-truth clean image X, thereby achieving competitive quality for image denoising.

3.2. Dual-Branch Encoder

The conventional single-path encoder structure exhibits inherent limitations in capturing the multi-scale features of coherent noise [38]. While standard convolutional networks with small kernels excel at extracting local features, they often struggle to capture long-range dependencies efficiently due to the limited receptive fields. Conversely, architectures employing large-kernel convolutions can capture global information but may risk losing fine-textural details. Relying on a single strategy thus falls short of achieving an optimal balance between coherent noise suppression and detail preservation. To overcome this problem, we propose a dual-branch encoder architecture for encoding image information. By incorporating two parallel pathways with distinct receptive fields, this design enables synchronous perception and encoding of image features across multiple scales, thereby ensuring comprehensive feature extraction. The encoder stage of MEANet consists of two parallel sub-networks: an enlarged-scale efficient (ESE) encoder and a basic-scale detail (BSD) encoder.

3.3. Dual-Branch Dense Attention Fusion (DDAF)

To efficiently integrate the complementary information extracted by the dual-branch encoder at different scales, we design a dual-branch dense attention fusion (DDAF) module, as shown in Figure 2. This module performs weighted fusion of feature maps from the two branches through an adaptive attention mechanism. Specifically, it first concatenates the feature maps F_{ESE}^{i} and F_{BSD}^{i} from the two branches and compresses the global spatial information into a channel using global average pooling. A lightweight network composed of two 1×1 convolutional layers then generates the respective channel attention weights w_{ESE}^{i} and w_{BSD}^{i} for the two branches. After these weights are activated by the Sigmoid function, they are used for normalized weighted fusion of the original feature maps. The process can be expressed as follows:

$$F_{fused} = \frac{F_{BSD}^{i} \cdot w_{BSD}^{i} + F_{ESE}^{i} \cdot w_{ESE}^{i}}{w_{BSD}^{i} + w_{ESE}^{i} + \epsilon} \tag{5}$$

where ϵ is a small value to ensure numerical stability. This normalization operation adaptively balances the contributions of the two branches while maintaining feature scale consistency.

Furthermore, to further enhance the representational capability of the fused features, the DDAF module also integrates a context enhancement module internally. This module consists of a densely connected residual attention block. It contains multiple densely connected convolutional layers and an Efficient Channel Attention (ECA) layer (as shown in Figure 2), capable of effectively refining and strengthening critical noise information without introducing excessive parameters [39].

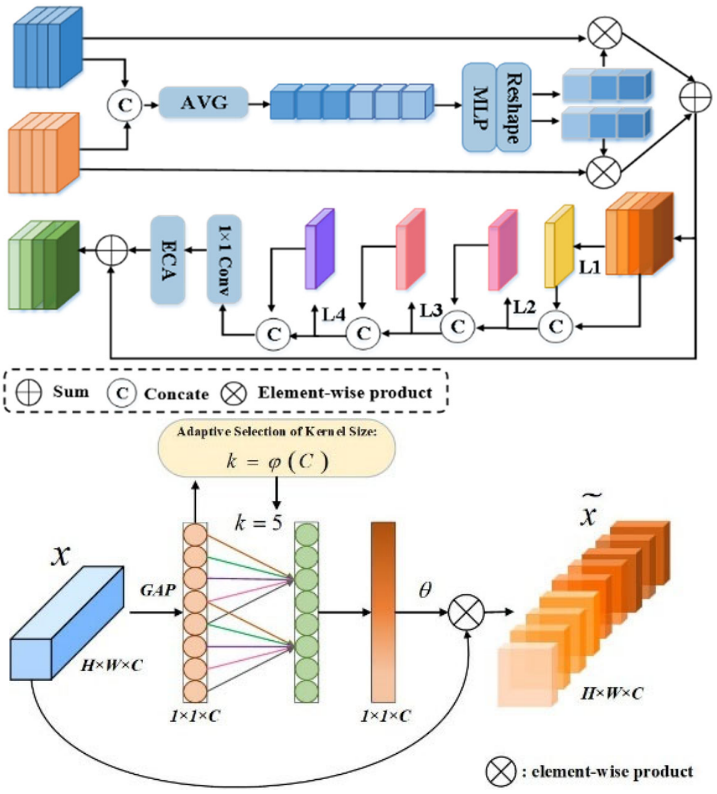


Figure 2. Architecture of DDAF and ECA. The dual-branch dense attention fusion (DDAF) is composed of an adaptive fusion (AF) module and a dense residual enhancement (DRE) module. ECA generates channel weights through a fast 1D convolution of size k, where k is adaptively derived from the channel dimension C.

In the network structure, a hierarchical fusion strategy is adopted to maximize information utilization efficiency. For the shallowest layer of the encoder $(i = 1)$, to maximally preserve the original high-frequency texture details of the image, we choose to directly pass the local features F^i_{BSD} extracted by the BSD branch to the corresponding level of the decoder via a skip connection, without fusion. This is represented as follows:

$$F^i_{\text{skip}} = F^i_{BSD} \tag{6}$$

At deeper levels of the network $(i > 1)$, the outputs F^i_{ESE} and F^i_{BSD} from the two branches are fed into the DDAF module for fusion, and the resulting information-enhanced features F^i_{DDAF} are then fed to the corresponding stage of the decoder via skip connections. The features of the deepest layer of the network are represented by F^L_{DDAF}, where L is the total number of encoder layers.

3.4. Loss Function

In deep learning, the choice of a loss function is crucial for optimizing the parameters of the image denoising networks. To achieve this, we train our network using the L_1 loss function, which is defined as follows:

$$L_1 = \sum_{i=1}^{n} |y_i - f(x_i)| \tag{7}$$

where y_i denotes the target pixel value, and $f(x_i)$ stands for the pixel value predicted by the network. L_1 parametric tends to minimize the absolute difference between the target and estimated images. Outliers do not cause particularly large losses, and their fluctuations are small and stable, thus making it more robust.

4. Experimental Results and Discussion

4.1. Dataset

Since clean holograms are difficult to obtain in real-world scenarios, most networks use simulated datasets for training, but the denoising results are not ideal when applied to real images. Accordingly, acquiring large-scale, effective datasets for deep learning-based denoising algorithms remains a significant challenge that warrants further investigation. To address this critical data challenge and facilitate the development of robust denoising models for real noise images, we developed a high-fidelity DHM simulation system. This simulation system achieved a full-link simulation from the light source to the recording plane, thereby constructing a high-fidelity modeling platform suitable for the entire holographic microscopy recording process. The simulated optical path is shown in Figure 3. The simulation system is built using MATLAB R2024b software. The system construction approach is as follows: The resampled band-limited angular spectrum method is adopted as the fundamental model for light wave propagation. During propagation along the optical path, the corresponding transfer function or modulation effect is introduced at each optical element in the sequence of the optical path. Specifically, the light wave emanates from the source, propagates freely to each optical element, and continues to propagate after superimposing its modulation effect. This process is iterated until it reaches the camera receiving surface. Finally, a hologram is generated by extracting the intensity information of the complex amplitude at the receiving surface. We used the constructed platform to simulate the holographic recording process of the USAF 1951 resolution test target, setting the standard target line pair height to 150 nm, the light source wavelength to 632.8 nm, the initial source radius to 10 mm, the objective's numerical aperture (NA) to 0.25, the magnification to 10×, the working distance to 10.6 mm, and the receiver resolution to 3000 × 4096 with a pixel size of 3.45 μm × 3.45 μm. Since the propagation medium is air (refractive index n = 1.0) and the NA is 0.25, the system's theoretical resolution can be calculated as 1.544 μm. From the simulated hologram, the line width of the fourth element in Group 7 corresponds to two pixels, and the individual pixel width is 0.801 μm, yielding an actual resolution of 1.603 μm. The calculated absolute difference between the simulated actual resolution and the theoretical resolution is only 0.059 μm, with a relative error of approximately 3.82%. Through an experiment using the USAF 1951 resolution test target, the system's accuracy in object wavefield simulation and imaging was validated. Building upon this foundation, the platform was further employed to simulate the generation of coherent noise in holographic microscopy measurements. This system generates our training dataset by accurately simulating key physical characteristics of coherent noise, allowing for the production of customized output datasets with controllable speckle noise and parasitic fringes.

In DHM, coherent noise is a significant issue encountered in coherent imaging systems that use light sources with a large coherence length, such as single-longitudinal-mode lasers. This noise degrades the quality of reconstructed images and is primarily a mixture of parasitic fringes and speckle noise. Speckle noise originates from coherent light scattering from the sample's surface topography and internal refractive index fluctuations. Specifically, when coherent light illuminates an optically rough surface, its microscopic irregularities cause each scattering point on the surface to emit a wavelet with a random phase. These wavelets undergo coherent superposition upon reaching the detector plane, forming a granular random interference pattern. Parasitic fringes originate from imperfections on the surfaces or within optical components (e.g., microscope objectives). The laser undergoes multiple reflections at these interfaces, and the resulting reflected waves interfere, ultimately forming chaotic interference fringes superimposed on the main signal.

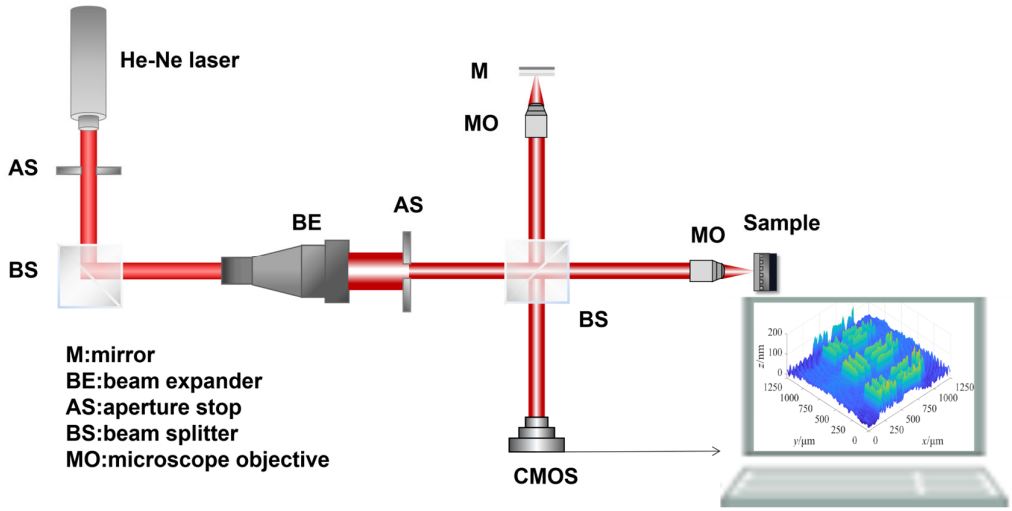


Figure 3. Schematic diagram of the DHM experimental setup, serving as both the physical acquisition system and the optical model for simulation.

In coherent holography, the bright–dark variation on a hologram is jointly governed by two factors: intensity and phase. The overall brightness is primarily determined by reflectance-related intensity, whereas phase variations modulate the interference fringes, thereby altering local bright–dark patterns. When coherent light illuminates an optically rough surface with roughness comparable to the wavelength, each surface point emits a scattered wavelet with random amplitude and random phase. These wavelets interfere in space, producing randomly distributed bright and dark grains, referred to as speckle. Under the standard assumption for sufficiently rough surfaces, the amplitudes and phases of scattered wavelets are statistically independent, and the phase is approximately uniformly distributed in the interval from minus π to π.

Recent research [40] has statistically demonstrated a high degree of similarity between Digital Holographic Hybrid Phase Noise (DHHPN) and Perlin noise. These correlations were quantitatively established using robust metrics, including the Maximum Information Coefficient (MIC) and the Pearson correlation coefficient, across noise distributions, spectral characteristics, and autocorrelation functions.

In this study, we validated these findings by integrating the methodology into our simulation framework. Our empirical results corroborate that Perlin noise exhibits a significant correlation with speckle noise and that, by modulating parameters such as frequency and amplitude, it can effectively simulate its primary characteristics. Consequently, we adopt Perlin noise as the model for speckle simulation. To reflect the signal-dependent multiplicative nature of speckle, we inject speckle fluctuations into the complex object wavefield in a multiplicative form, i.e., noise complex amplitude multiplied by the wave complex amplitude. In implementation, the Perlin-based speckle field is used to represent amplitude fluctuations induced by sample-plane reflectance variations and the corresponding reflected-phase fluctuations, and it is multiplied with the complex object field during object-wave formation. To better approximate the fine-grained fluctuation characteristics of real optical speckle, we further superimpose multiplicative high-frequency perturbations on the Perlin base, compensating for the Perlin model's lack of micro-texture details. Therefore, speckle is introduced at the sample plane and naturally propagates through the entire optical pipeline to the CCD recording plane; after coherent superposition, the hologram is obtained by intensity measurement rather than adding speckle to the reconstructed image at the reconstruction stage. It should be emphasized that the statistical characteristics of the simulated speckle are determined by the adopted multiplicative noise function.

The generation process of Perlin noise primarily relies on local random gradient vectors and weighted interpolation mechanisms. Specifically, for an arbitrary input point, the algorithm locates the corresponding grid cell, computes the relative vectors from the point to the cell vertices, and performs dot products with the pre-defined gradient vectors to obtain responses. An ease curve is then used to compute distance-based interpolation weights, and the final noise value is produced via weighted interpolation, yielding a continuous noise field. The ease curve employed is defined as follows:

$$s(t) = 6t^5 - 15t^4 + 10t^3 \tag{8}$$

During the practical simulation, given that speckle noise in optical systems exhibits complex multiplicative characteristics and interference effects, we observed that superimposing multiplicative high-frequency perturbations onto the Perlin noise base results in a more precise representation. This improvement effectively compensates for the lack of fine-grained details in the foundational Perlin model, thereby successfully bridging the gap between simulated and real speckle noise and satisfying the stringent requirements for reproducing its intricate optical characteristics.

To accurately simulate the formation of parasitic fringes, we analyzed the lens's optical characteristics. Assuming the object surface to be an ideal focal plane, the optical path difference (OPD) between the reflected light from the front and rear surfaces can be computed through the objective lens's focal length and the 2D spatial coordinate grid (x, y) on the lens plane. The OPD is then converted into optical path lengths at each point in the optical field matrix, ultimately resulting in parasitic fringe patterns. The mathematical expression for the OPD matrix is given by the following:

$$\varphi_f(x, y) = n_1 \delta_0 \cdot \left(\sqrt{(x^2 + y^2 + f^2)} / f \right) \tag{9}$$

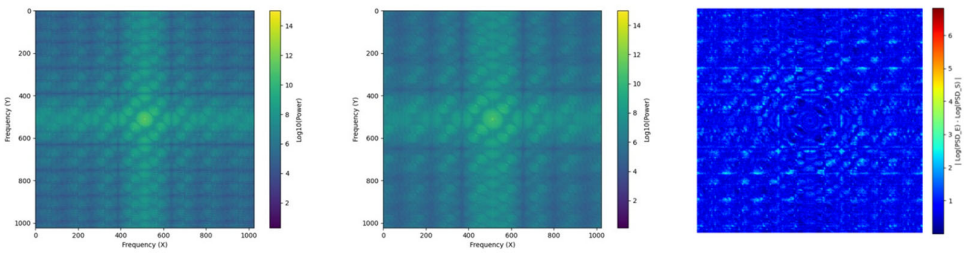
In the above equation, f represents the focal length of the lens, n_1 is the refractive index of the transparent medium, δ_0 is the thickness of the transparent medium, and x and y are the coordinates on the observation plane of the lens.

To rigorously validate the effectiveness of this simulation system, we constructed a DHM experimental setup based on a Michelson interferometer (see Figure 3 for the schematic diagram of the optical system) and performed measurements on the USAF 1951 resolution test target in a transparent medium. The experimental system uses a Daheng DH-HN250 He-Ne laser (China Daheng Group, Inc., Beijing, China) as the light source. Its operating mode is TEM_{00}, the wavelength is 632.8 nm, the beam diameter is <0.7 mm, and the laser power is 12.0 mW. The objective lens is a critical component for system imaging, and its parameters and quality directly affect the imaging resolution and clarity. The parameters primarily considered here that influence imaging quality are focal length, magnification, and numerical aperture. The focal length is the distance from the objective's optical center to the focal point when the light is parallel to the incident; magnification is the ratio of the image size to the corresponding object size during imaging; and the numerical aperture (NA) describes the angular range of light that the objective can collect and determines the objective's spatial resolution. Based on experimental requirements for magnification, resolution, etc., the parameters of the selected microscope objective are as follows: NA = 0.25, magnification = 10×, working distance (WD) = 10.6 mm, focal length = 18 mm, and objective field number (OFN) = 22 mm. In digital holography, CCD or CMOS cameras are commonly used to record holograms. The digital camera used in this system is the OMRON STC-MCS1242POE, with the following main parameters:

photosensitive area size of 10.35 mm × 14.13 mm, resolution of 3000 × 4096, pixel size of 3.45 µm × 3.45 µm, and applicable spectral range of 400 nm–900 nm.

To rigorously and quantitatively validate the effectiveness of our simulation system, we performed a detailed spectral analysis of simulated noise relative to experimental noise. For this comparison, we selected regions from both the experimental and simulated images that contained only the background and coherent noise.

First, we computed the 2D Power Spectral Density (PSD) for both the S and E noise regions. As shown in Figure 4a,b, the 2D spectral maps demonstrate a high degree of structural agreement. Both clearly reproduce the ring-shaped parasitic fringes and the diffuse high-frequency components attributed to speckle noise. Note that the spectra are displayed on a logarithmic scale, which compresses the dynamic range and may visually mask subtle local discrepancies. To provide a quantitative assessment, we computed the root-mean-square error (RMSE) between the logarithmically transformed 2D PSDs of the experimental and simulated noise, yielding an RMSE of 1.0368. This result indicates that the simulated and experimental spectra are statistically consistent but not identical. In addition, Figure 4c presents a heat map of the absolute difference between the two log-PSDs, showing that the residuals are randomly distributed in the frequency domain, without forming structured patterns, suggesting that the simulation reproduces the deterministic noise structure while maintaining the stochastic independence of speckle-related components. Moreover, we further validated the simulated speckle under constant illumination to improve reproducibility. Specifically, we selected a uniformly illuminated region and plotted the histogram of the normalized speckle intensity. As shown in Figure 5, the histogram follows the expected negative exponential distribution, which is a canonical statistical signature of speckle noise. This verification complements the PSD-based spectral comparison by providing a statistical perspective, and it further supports the physical plausibility and reproducibility of our speckle simulation.



(a) (b) (c)

Figure 4. Comparison of logarithmic 2D Power Spectral Density (PSD) and spectral difference analysis for experimental and simulated coherent noise. (**a**) Log-PSD of experimental (E) noise; (**b**) log-PSD of simulated (S) noise; (**c**) heat map of the absolute difference between the logarithmic PSDs of experimental and simulated noise.

For a precise quantitative comparison, we further computed the Radially Averaged PSD. This metric integrates the 2D spectral power radially, revealing the average power distribution across spatial frequencies. As depicted in Figure 6, the experimental (E) curve and the simulated (S) curve show a nearly perfect match. This confirms two key points: Our simulation accurately reproduces the dominant spatial frequencies of the parasitic fringes, visible as the primary peaks in the curves. The decay trends of both curves are highly consistent across the entire high-frequency range, validating the accurate characterization of the speckle noise statistics.

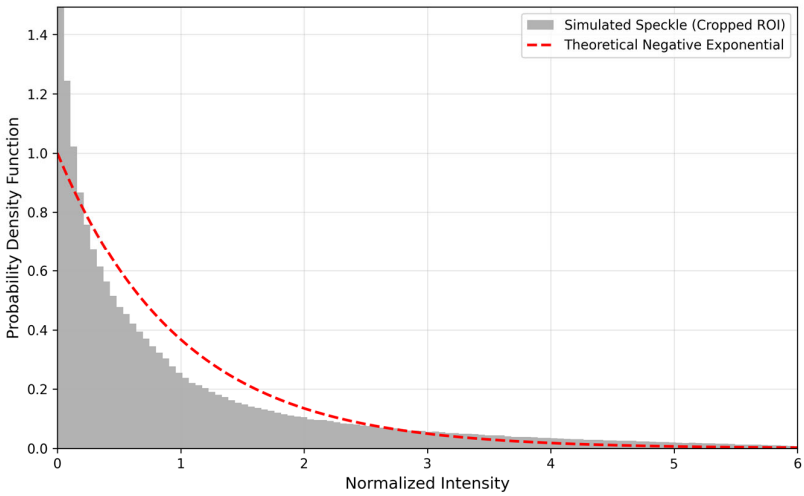


Figure 5. Speckle intensity histogram under constant illumination.

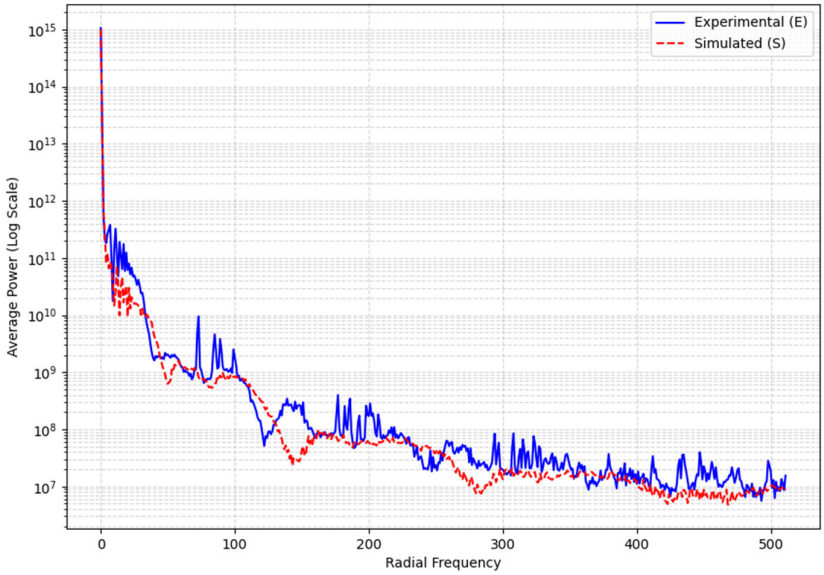


Figure 6. Quantitative comparison of Radially Averaged PSD for experimental (E) and simulated (S) coherent noise, plotted as a solid blue line and a dashed red line, respectively.

In summary, this spectral analysis provides strong quantitative evidence that the coherent noise generated by our DHM simulation system is highly consistent with the real experimental data in both its statistical properties and spectral characteristics.

Our simulation system enhances model generalization by generating diverse, coherent noise. This is achieved by adjusting the parasitic fringe density and contrast, and by controlling the spatial distribution of speckle noise. To ensure the diversity of the generated simulation data, three core physical parameters in the simulation system are set to vary randomly within specific ranges, with the specific parameter ranges shown in Table 1: the medium thickness Z_Z is used to control the spatial frequency of parasitic fringes; the propagation distance Z_1 is perturbed within a certain range to introduce defocus effects; and the variation in the angle between the object and reference beams θ is used to adjust the signal fringe density. The partial simulation data is shown in Figure 7. To clarify the effect of different parameter values on coherent noise patterns, Figure 7 is arranged in a controlled comparison layout: images (a)–(f) correspond to θ = 0.06 rad and (g)–(l) correspond to θ = 0.08 rad, where θ denotes the off-axis angle θ. Within each row, (a, b), (c, d), and (e, f) form three paired comparisons with Z_Z = 0.008, 0.012, and 0.016 m, respectively, and only Z_1 varies between 0.090 and 0.095 m within each pair.

Table 1. Key parameters of simulation system.

Parameter	Symbol	Range	Unit
Medium thickness	Z_Z	[0.008, 0.0016]	m
Propagation distance	Z_1	[0.09, 0.095]	m
Angle	θ	[0.06, 0.08]	rad

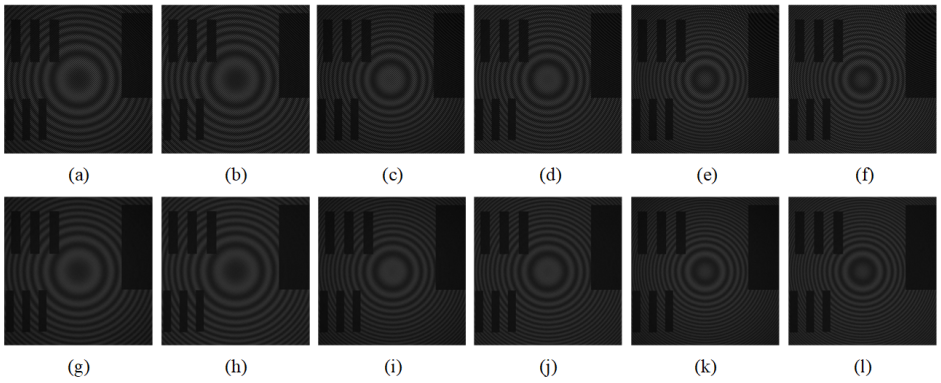


Figure 7. Representative samples of the generated simulation dataset. Images (**a**–**l**) correspond to 12 distinct parameter combinations.

To explicitly address parameter-dependent differences, we now clarify the observable effects of each parameter in the generated holograms: increasing Z_1 strengthens defocus blur and alters the spatial distribution of fringes; increasing Z_Z increases the effective optical path difference and typically shifts parasitic fringes toward higher spatial frequencies (denser fringe patterns); and varying θ changes the fringe density in the recorded holograms. Crucially, our controllability refers to these physical source parameters (e.g., Z_Z, Z_1, θ), rather than targeting a specific, pre-defined statistical value (e.g., VAR, SNR). The resulting SNR and noise distribution are emergent properties of this physics-based simulation, ensuring high fidelity to experimental noise, as validated in Figure 6. Finally, the spectral consistency with experimental coherent noise is quantitatively verified in Figure 6, while the parameter-driven diversity of generated samples is visualized in Figure 7; together, Figures 5–7 provide complementary evidence for reproducibility, fidelity, and diversity of the proposed simulation dataset.

4.2. Metrics

In our research, the quality of predicted clean images was assessed using the PSNR [41] and the SSIM [42].

PSNR directly reflects the relative level of the image signal to background noise. A higher PSNR value suggests less image distortion and better denoising performance. We employed SSIM to focus on preserving structural information.

The SSIM value ranges from -1 to 1, used to measure the structural similarity of images. A higher value closer to 1 indicates a higher similarity between the two images and better denoising performance.

4.3. Experimental Setup

All experiments are performed on a PC running Windows 11, an Intel Core i5-14600KF processor, 32 GB of RAM, and an NVIDIA GeForce RTX 4070 GPU (12 GB). The MAENet model is built and trained using the PyTorch 2.4.1 framework with Python 3.10.4. We trained our model for 20 epochs.

We chose the synthetic data as the training data for network training. For the synthetic noise-removal task, a simulated dataset of 500 high-resolution holograms is used. To

construct the noise-clean training image pairs, 400 images are randomly selected for the training dataset, and the remaining 100 images serve as the test set. To ensure adequate training and optimize computational efficiency, we crop high-resolution holograms into 64×64 patches and set the batch size to 32. The MAENet is trained with the ADAM [26] optimizer using $\beta 1 = 0.9$, $\beta 2 = 0.999$, and $\varepsilon = 10^{-8}$. The learning rate is initially set to 1×10^{-4} and is halved every five epochs. We monitored the training and validation losses and observed that the validation loss stabilized around the 17th epoch, suggesting that 20 epochs are sufficient for the model to fully converge. The convergence curve during model training is shown in Figure 8.

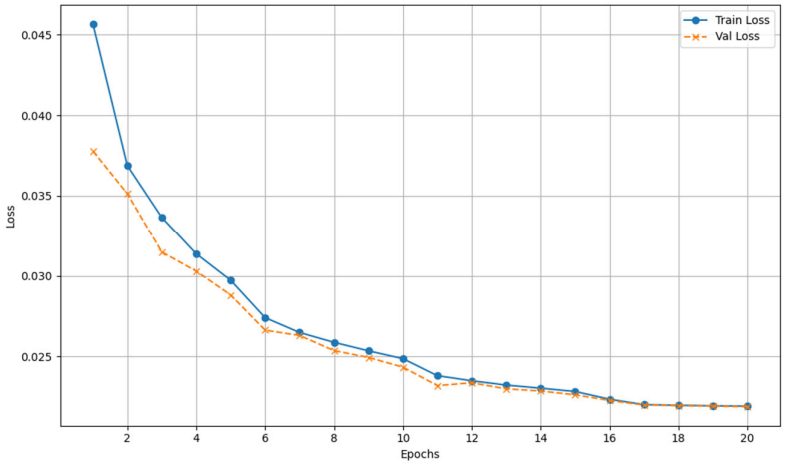


Figure 8. Convergence curves of the training and validation loss.

4.4. Ablation Research

In this section, we measure the effectiveness of the main modules of the proposed network architecture. Table 2 shows the denoising performance of models with different structures, illustrating the impact of the enlarged-scale efficient (ESE) encoder and dual-branch dense attention fusion (DDAF) on denoising results. All networks use the same settings. "−" means that the module is not included, and "+" means that the module is included. The metrics used to measure the denoising performance of different networks in Table 2 are PSNR and SSIM.

Table 2. Comparison of denoising performance after ablation of different modules.

Models	Dual-Encoder		DDAF		Performance	
	ESE Encoder	BSD Encoder	AF Module	DRE Module	PSNR	SSIM
$MAENet_{v0}$	√				31.58	0.90880
$MAENet_{v1}$		√			32.83	0.92628
$MAENet_{v2}$	√	√			32.91	0.92683
$MAENet_{v3}$	√	√	√		33.11	0.92866
MAENet	√	√	√	√	33.25	0.93042

The "√" in the table denotes that the specific module is integrated into the model.

As shown in Table 2, the entire MAENet model achieves the best denoising performance among the five models. In terms of average PSNR, the complete MAENet model, which incorporates a dual-encoder architecture, surpasses the single-encoder MAENetv0 (with only the ESE encoder) by 1.67 dB and MAENetv1 (with only the BSD encoder) by 0.42 dB. This verifies that the basic scale detail (BSD) encoder is indispensable and enhances the network's learning ability. The dual branches effectively address the problem of insufficient local feature extraction, which often arises when the ESE encoder focuses exclusively on global information. The BSD branch focuses on capturing image texture details, and complementary features play a crucial role in enhancing the model's denoising performance.

Table 2 shows that removing the entire DDAF module results in a performance degradation more significant than that caused by removing only its DRE module. Comparing $MAENet_{v3}$ and $MAENet_{v2}$, the inclusion of the adaptive fusion (AF) module results in a PSNR improvement of 0.2 dB. Furthermore, the complete DDAF module achieves a higher PSNR than the DDAF lacking the DRE module, with an improvement of 0.14 dB, confirming that each part contributes to the network's denoising ability. In the DDAF, the effect is best when both the DRE and AF modules are added, indicating that the DRE module promotes feature reuse and enhances fused feature information, while the AF module facilitates the network's effective integration of feature information from different encoders. This process provides higher-quality features to the decoder, thereby enabling more accurate image reconstruction.

4.5. Simulated Image Denoising

To verify the reliability and effectiveness of the proposed algorithm, we conducted a rigorous quantitative comparison between our MAENet and several state-of-the-art denoising techniques, including U-Net, VNet, DnCNN, and FFDNet. Specifically, we use PSNR and SSIM for comprehensive quantitative evaluation [33]. In addition, to quantify model complexity and training scale, we computed the number of trainable parameters for all compared methods and reported them in Table 3.

Table 3. Quantitative denoising performance and model complexity on simulated holograms.

Method	PSNR	SSIM	Params/M
DnCNN [19]	29.25	0.86539	12.89
VNet [43]	33.01	0.92826	11.33
FFDNet [44]	28.87	0.86192	13.01
Unet [22]	32.83	0.92628	9.34
Ours	33.25	0.93042	11.85

As demonstrated in Table 3, the proposed MAENet achieves the best performance on both PSNR and SSIM metrics. Compared with other learning-based methods, our approach yields PSNR gains of 4.00 dB, 0.24 dB, 4.45 dB, and 0.90 dB over DnCNN, VNet, FFDNet, and UNet, respectively. Regarding parameter efficiency, MAENet has 11.85M trainable parameters. It uses fewer parameters than DnCNN (12.89 M) and FFDNet (13.01 M) while achieving higher PSNR and SSIM. Although MAENet has slightly more parameters than U-Net (9.34 M) and VNet (11.33 M), it still provides better quantitative performance; considering the achieved gains, this modest increase in parameters is reasonable. Overall, MAENet offers a favorable balance between the number of parameters and denoising accuracy.

The substantial disparity in performance, as shown in Table 3, can be attributed to the models' varying adaptability to the unique characteristics of coherent noise. We acknowledge that models such as DnCNN, VNet, and FFDNet are primarily designed for generic, randomly distributed noise, like additive white Gaussian noise (AWGN). Our purpose in this comparison was not to conduct a like-for-like performance race, but rather to demonstrate that these state-of-the-art, general-purpose denoisers struggle significantly when faced with the unique, hybrid nature of DHM coherent noise.

The quantitative results in Table 3 demonstrate that these AWGN-focused models (specifically DnCNN, VNet, and FFDNet) exhibit inherent limitations in this context, achieving significantly lower PSNR and SSIM scores. This performance gap validates the need for a specialized architecture. Furthermore, as the visual comparisons in the following figures show, U-Net and other general-purpose architectures can also introduce visual flaws, such

as residual artifacts and over-smoothing. In contrast, our MAENet is specifically designed to suppress this complex, hybrid coherent noise, leveraging its dual-branch architecture to effectively separate coherent noise from object information and achieve a better balance between noise suppression and detail preservation, as evidenced by both quantitative metrics and visual results.

The visual comparison of different methods is shown in Figure 9, which showcases the performance of MAENet and U-Net on simulated holograms of a USAF 1951 resolution test target and a star-like pattern. Specifically, the green boxes in Figure 9 highlight randomly selected regions, and the corresponding red boxes provide magnified views to facilitate a clearer observation of fine textures and structural details. Figures 9a–d and 9e–h display the denoising results for two randomly selected regions of the USAF 1951 resolution test target, respectively, while Figure 9i–l correspond to the results for the star-like pattern. By testing on these varied morphologies, we demonstrate that the trained neural network can effectively analyze and denoise objects with different shape information. Given that the visual clarity of the grayscale holograms is limited by contrast issues, we further present the 3-D reconstruction results in Figure 10. These reconstructions provide a clearer, more perceptible comparison, thereby enabling a more effective assessment of each method's ability to preserve boundary features and suppress artifacts. As illustrated in Figures 8 and 9, although parasitic fringes and dotted speckle noise are somewhat suppressed, the U-Net's denoising performance remains limited. It fails to completely eliminate distinct ring-like parasitic fringes, compromising structural integrity and leading to blurred edges and distortions. In contrast, the proposed MAENet obtains a visually appealing result. It not only effectively eliminates the noise but also maintains the sharp edges and intricate textures of the original input. This demonstrates that our method strikes a superior balance between noise suppression and detail preservation, achieving better results both subjectively and objectively.

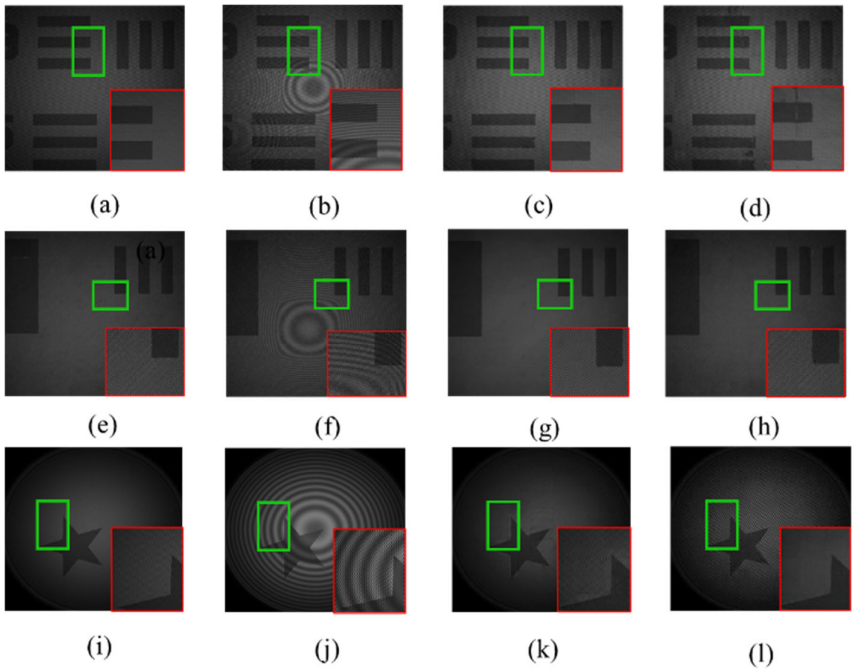


Figure 9. Denoised result of each method for simulated images. (**a**) Ground-truth image; (**b**) Noisy image; (**c**) MAENet; (**d**) U-net; (**e**) Ground-truth image; (**f**) Noisy image; (**g**) MAENet; (**h**) U-net; (**i**) Ground-truth image; (**j**) Noisy image; (**k**) MAENet; (**l**) U-net.

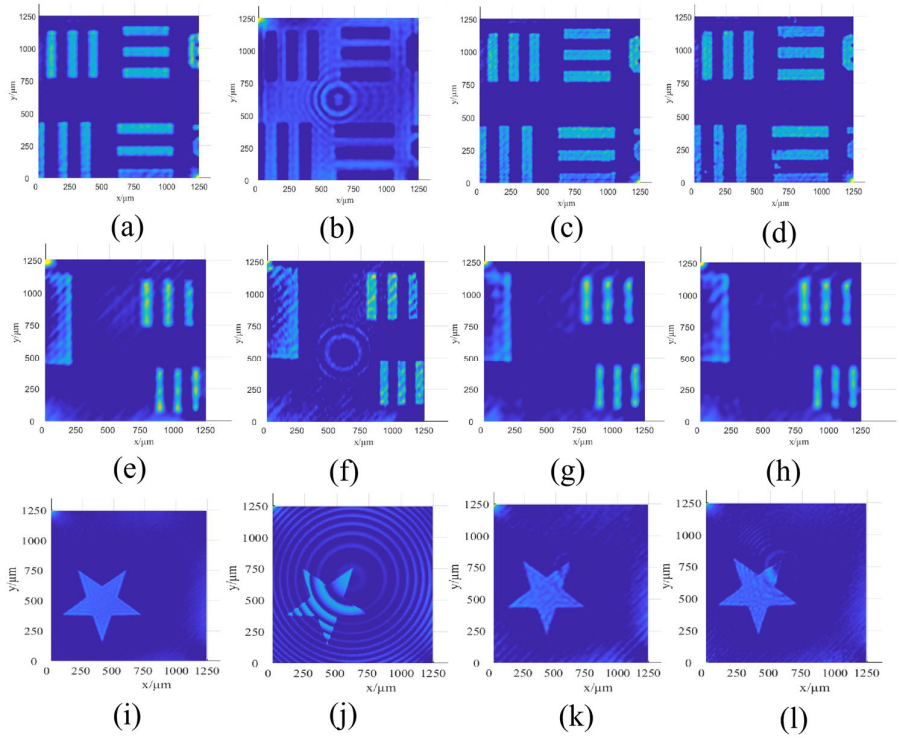


Figure 10. The 3-D reconstruction results corresponding to Figure 8. (**a**) Ground-truth image; (**b**) Noisy image; (**c**) MAENet; (**d**) U-net; (**e**) Ground-truth image; (**f**) Noisy image; (**g**) MAENet; (**h**) U-net; (**i**) Ground-truth image; (**j**) Noisy image; (**k**) MAENet; (**l**) U-net.

4.6. Real-World Image Denoising

To further validate the practical applicability and generalization capability of our proposed MAENet, we applied the model trained purely on the simulated dataset to denoise real-world holograms. Specifically, these holograms were captured using the DHM setup constructed and illustrated in Figure 3, as detailed in Section 4.1.

Figure 11 presents a qualitative comparison of denoising performance between the U-Net and our proposed MAENet. To strictly evaluate the model's robustness under varying optical conditions, we selected three distinct regions of the USAF 1951 resolution test target (Figure 11a–c). Notably, these figures are arranged in ascending order of interference fringe density. This increase in density is directly attributed to the progressively larger angle between the object and reference beams, effectively simulating a range of different optical operating conditions. The original noisy holograms are heavily corrupted by the coherent noise, characterized by distinct ring-like parasitic fringes and severe speckle noise, which significantly impacts image reconstruction.

The denoising results processed by U-Net are shown in Figure 11d–f. Although U-Net effectively suppresses background noise, it exhibits notable limitations. The object information is visibly damaged, resulting in the loss of valid information; specifically, the structural information originally masked by the parasitic fringes is missing in the denoised result. Furthermore, it yields poor detail preservation, struggling to balance noise reduction with feature retention. It fails to fully preserve the object structure and edges, resulting in dense interference fringes that blur and leave the object regions relatively indistinct.

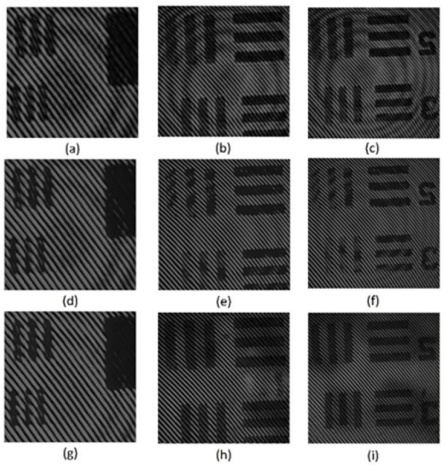


Figure 11. Denoising results on real-world holograms. (**a**–**c**) Three original noisy holograms captured by the DHM system. (**d**–**f**) The corresponding denoising results from the U-net. (**g**–**i**) The corresponding denoising results from our MAENet.

In contrast, the corresponding results from MAENet, shown in Figure 11g–i, demonstrate visually superior performance. Despite the variations in fringe density and object information across the three holograms, our model effectively suppresses coherent noise and significantly reduces parasitic fringes while robustly retaining high-frequency details and the integrity of the object structure. Compared to U-Net, the experimental results indicate that MAENet not only possesses superior denoising capabilities but also effectively preserves edge information, resulting in the best visual quality.

This experiment strongly demonstrates the excellent generalization capability and robustness of the proposed MAENet. It proves that the model can effectively handle different optical parameters and diverse target locations. Furthermore, it validates that our high-fidelity simulation system successfully captures the key physical characteristics of real-world coherent noise, making it an effective and practical training resource.

5. Conclusions

In this paper, we propose a novel dual-branch multi-scale residual attention network (MAENet) for reducing coherent noise in DHM. This network employs a dual-encoder architecture, which uses an enlarged scale-efficient (ESE) encoder with a large receptive field to capture global features and a basic scale detail (BSD) encoder with a small receptive field to preserve high-frequency textures, achieving complementary extraction of multi-scale features. In addition, the dual-branch dense attention fusion (DDAF) module effectively integrates feature information from both branches via an adaptive attention mechanism and dense residual connections, thereby significantly improving the denoising performance and representation ability of MAENet. To address the challenge of insufficient paired training data in real-world scenarios, we constructed and validated a DHM simulation system that accurately models the key physical characteristics of coherent noise. Extensive experiments have demonstrated MAENet's excellent performance. On our simulated dataset, MAENet achieved a PSNR of 33.25 dB and a SSIM of 0.93042, outperforming various denoising algorithms, including DnCNN, FFDNet, VNet, and U-Net. Visual evaluations and 3-D reconstructions further confirm the model's robustness in preserving sharp boundaries and suppressing artifacts across different object shapes. Crucially, validation on real-world holograms demonstrates exceptional generalization capability. MAENet effectively suppresses complex parasitic fringes and speckle under varying optical conditions while maintaining object structure. Notably, although the simulation pipeline does not explicitly incorporate a camera readout noise model, real experimental holograms inevitably contain

camera readout noise as well as other practical complex noise components. The denoising results on real holograms show that MAENet consistently improves reconstruction quality under these unmodeled noise conditions, indicating robustness to such noise sources. Ablation studies further validated the effectiveness and necessity of each component within the dual-encoder architecture and the DDAF module for improving model performance. In summary, the MAENet achieves an ideal balance among noise suppression, detail preservation, and computational efficiency, providing an effective and robust solution for denoising problems in coherent imaging systems. This method not only improves the quality and accuracy of DHM imaging but also demonstrates its great potential for applications in precision measurement fields, such as non-destructive testing and surface morphology. In future work, we will incorporate a camera readout noise model into the simulation system, conduct more systematic evaluations under mixed-noise conditions, and further explore its applicability to other imaging techniques.

Author Contributions: Conceptualization, Y.Z.; methodology, Y.Z., J.Y. and Z.Z.; software, Y.Z. and M.K.; validation, J.Y., Y.F., F.H. and Z.T.; formal analysis, Y.Z. and J.Y.; investigation, Y.Z., J.Y., Z.Z., M.K., Y.F., F.H. and Z.T.; resources, W.L.; data curation, Y.Z. and F.H.; writing—original draft preparation, Y.Z.; writing—review and editing, Y.Z. and W.L.; visualization, Y.Z. and Z.T.; supervision, W.L.; project administration, W.L.; funding acquisition, W.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the National Key Research and Development Program of China (Grant No. 2024YFF0617002); National Natural Science Foundation of China (52076200).

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Acknowledgments: The authors thank the reviewers for their helpful comments and suggestions, which have substantially improved the paper.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study, the collection, analysis, or interpretation of data, the writing of the manuscript, or the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

AF	Adaptive Fusion
DDAF	Dual-Branch Dense Attention Fusion
BSD	Basic Scale Detail
ESE	Enlarged Scale Efficient
DRE	Dense Residual Enhancement
DHM	Digital Holographic Microscopy

References

1. Gupta, G. Algorithm for image processing using improved median filter and comparison of mean, median and improved median filter. *Int. J. Soft Comput. Eng. (IJSCE)* **2011**, *1*, 304–311.
2. Pitas, I. *Digital Image Processing Algorithms and Applications*; John Wiley & Sons: Hoboken, NJ, USA, 2000.
3. Chen, R.; Yu, W.; Wang, R.; Liu, G.; Shao, Y. Interferometric phase denoising by pyramid nonlocal means filter. *IEEE Geosci. Remote Sens. Lett.* **2013**, *10*, 826–830. [CrossRef]
4. Dabov, K.; Foi, A.; Katkovnik, V.; Egiazarian, K. Image denoising with block-matching and 3D filtering. In Proceedings of the SPIE Conference, Image Processing: Algorithms and Systems, Neural Networks, and Machine Learning, San Jose, CA, USA, 15–19 January 2006; Volume 6064, pp. 354–365.
5. Kemao, Q. Two-dimensional windowed fourier transform for fringe pattern analysis: Principles, applications and implementations. *Opt. Lasers Eng.* **2007**, *45*, 304–317. [CrossRef]

6. Wu, J.; Tang, J.; Zhang, J.; Di, J. Coherent noise suppression in digital holographic microscopy based on label-free deep learning. *Front. Phys.* **2022**, *10*, 880403. [CrossRef]
7. Yin, D.; Gu, Z.; Zhang, Y.; Gu, F.; Nie, S.; Feng, S.; Ma, J.; Yuan, C. Speckle noise reduction in coherent imaging based on deep learning without clean data. *Opt. Lasers Eng.* **2022**, *133*, 106151. [CrossRef]
8. Yan, K.; Yu, Y.; Huang, C.; Sui, L.; Qian, K.; Asundi, A. Fringe pattern denoising based on deep learning. *Opt. Commun.* **2019**, *437*, 148–152. [CrossRef]
9. Lin, B.; Fu, S.; Zhang, C.; Wang, F.; Li, Y. Optical fringe patterns filtering based on multi-stage convolution neural network. *Opt. Lasers Eng.* **2020**, *126*, 105853. [CrossRef]
10. Montresor, S.; Tahon, M.; Laurent, A.; Picart, P. Computational denoising based on deep learning for phase data in digital holographic interferometry. *APL Photonics* **2020**, *5*, 030802. [CrossRef]
11. Fang, Q.; Xia, H.-T.; Song, Q.; Zhang, M.; Guo, R.; Montresor, S.; Picart, P. Speckle denoising based on deep learning via a conditional generative adversarial network in digital holographic interferometry. *Opt. Express* **2022**, *30*, 20666–20683. [CrossRef]
12. Tahon, M.; Montresor, S.; Picart, P. Towards reduced CNNs for denoising phase images corrupted with speckle noise. *Photonics* **2021**, *8*, 255. [CrossRef]
13. Zhang, J.; Huang, L.; Chen, B.; Yan, L. Accurate extraction of the +1 term spectrum with spurious spectrum elimination in off-axis digital holography. *Opt. Express* **2022**, *30*, 28142–28157. [CrossRef]
14. Lin, Z.; Jia, S.; Zhou, X.; Zhang, H.; Wang, L.; Li, G.; Wang, Z. Digital holographic microscopy phase noise reduction based on an over-complete chunked discrete cosine transform sparse dictionary. *Opt. Lasers Eng.* **2023**, *166*, 107571. [CrossRef]
15. Fu, Z.; Li, J.; Hua, Z. DEAU-Net: Attention Networks Based on Dual Encoder for Medical Image Segmentation. *Comput. Biol. Med.* **2022**, *150*, 106197. [CrossRef] [PubMed]
16. Liang, B.; Tang, C.; Zhang, W.; Xu, M.; Wu, T. N-Net: An UNet Architecture with Dual Encoder for Medical Image Segmentation. *Signal Image Video Process.* **2023**, *17*, 3073–3081. [CrossRef] [PubMed]
17. Li, H.; Zhai, D.-H.; Xia, Y. ERDUnet: An Efficient Residual Double-Coding Unet for Medical Image Segmentation. *IEEE Trans. Circuits Syst. Video Technol.* **2024**, *34*, 2083–2096. [CrossRef]
18. Yao, Z.; Bi, J.; Deng, W.; He, W.; Wang, Z.; Kuang, X.; Zhou, M.; Gao, Q.; Tong, T. DEUNet: Dual-Encoder UNet for Simultaneous Denoising and Reconstruction of Single HDR Image. *Comput. Graph.* **2024**, *119*, 103882. [CrossRef]
19. Zhang, K.; Zuo, W.; Chen, Y.; Meng, D.; Zhang, L. Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising. *IEEE Trans. Image Process* **2017**, *26*, 3142–3155. [CrossRef]
20. Gu, S.; Xie, Q.; Meng, D.; Zuo, W.; Feng, X.; Zhang, L. Weighted nuclear norm minimization and its applications to low level vision. *Int. J. Comput. Vis.* **2017**, *121*, 183–208. [CrossRef]
21. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016.
22. Yadav, M.; Tiwari, N. Performance comparison of image restoration techniques using CNN and their applications. In Proceedings of the 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 8–10 April 2021; pp. 1146–1151.
23. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*; Springer: Berlin/Heidelberg, Germany, 2015; pp. 234–241.
24. Yao, Z.; Hao, L.; Qin, L.; Chen, J.; Li, W. Efficient seismic data denoising via multi-scale attention network with depthwise separable and residual dilated convolutions. *Sci. Rep.* **2025**, *16*, 2818. [CrossRef]
25. Li, J.; Lu, G.; Zhang, B.; You, J.; Zhang, D. Shared linear encoder-based multikernel Gaussian process latent variable model for visual classification. *IEEE Trans. Cybern.* **2021**, *51*, 534–547. [CrossRef]
26. Li, Y.; Chen, X.; Zhu, Z.; Xie, L.; Huang, G.; Du, D.; Wang, X. Attention-guided unified network for panoptic segmentation. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 7026–7035.
27. Du, B.; Wei, Q.; Liu, R. An improved quantum-behaved particle swarm optimization for endmember extraction. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 6003–6017. [CrossRef]
28. Woo, S.; Park, J.; Lee, J.; Kweon, I.S. CBAM: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.
29. Zhang, Y.; Li, K.; Li, K.; Wang, L.; Zhong, B.; Fu, Y. Image super-resolution using very deep residual channel attention networks. In Proceedings of the Computer Vision—ECCV 2018: 15th European Conference, Munich, Germany, 8–14 September 2018; pp. 294–310.
30. Dai, T.; Cai, J.; Zhang, Y.; Xia, S.-T.; Zhang, L. Second-order attention network for single image super-resolution. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 11065–11074.

31. Lyn, J.; Yan, S. Non-local second-order attention network for single image super resolution. *Mach. Learn. Knowl. Extr.* **2020**, *12279*, 267–279.
32. Tian, C.; Xu, Y.; Li, Z.; Zuo, W.; Fei, L.; Liu, H. Attention-guided CNN for image denoising. *Neural Netw.* **2020**, *124*, 117–129. [CrossRef] [PubMed]
33. Anwar, S.; Barnes, N. Real image denoising with feature attention. In Proceedings of the IEEE International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 3155–3164.
34. Li, G.; Fang, Q.; Zha, L.; Gao, X.; Zheng, N. Ham: Hybrid attention module in deep convolutional neural networks for image classification. *Pattern Recognit.* **2022**, *129*, 108785. [CrossRef]
35. Peng, Y.; Zhang, L.; Liu, S.; Wu, X.; Zhang, Y.; Wang, X. Dilated residual networks with symmetric skip connection for image denoising. *Neurocomputing* **2019**, *345*, 67–76. [CrossRef]
36. Zhang, Y.; Tian, Y.; Kong, Y.; Zhong, B.; Fu, Y. Residual dense network for image super-resolution. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 2472–2481.
37. Song, Y.; Zhu, Y.; Du, X. Dynamic residual dense network for image denoising. *Sensors* **2019**, *19*, 3809. [CrossRef]
38. Choi, Y.; Park, S.; Kim, S. Development of Point Cloud Data-Denoising Technology for Earthwork Sites Using Encoder-Decoder Network. *KSCE J. Civ. Eng.* **2022**, *26*, 4380–4389. [CrossRef]
39. Wang, Q.; Wu, B.; Zhu, P.; Li, P.; Zuo, W.; Hu, Q. ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 14–19 June 2020.
40. Tang, J.; Chen, B.; Yan, L.; Huang, L. Continuous Phase Denoising via Deep Learning Based on Perlin Noise Similarity in Digital Holographic Microscopy. *IEEE Trans. Ind. Inf.* **2024**, *20*, 8707–8716. [CrossRef]
41. Montrésor, S.; Picart, P. On the assessment of de-noising algorithms in digital holographic interferometry and related approaches. *Appl. Phys. B* **2022**, *128*, 59. [CrossRef]
42. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [CrossRef]
43. Reyes-Figueroa, A.; Flores, V.H.; Rivera, M. Deep neural network for fringe pattern filtering and normalization. *Appl. Opt.* **2021**, *60*, 2022–2036. [CrossRef]
44. Zhang, K.; Zuo, W. FFDNet: Toward a fast and flexible solution for CNN-based image denoising. *IEEE Trans. Image Process.* **2018**, *27*, 4608–4622. [CrossRef]

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