Edge Detection Using Clip ReLU-Based Enhanced Hybrid Network

Wazir Lakra and Rajeshwar Dass*

Department of Electronics and Communication Engineering, Deenbandhu Chhotu Ram University of Science and Technology, Murthal, Haryana, India
Email: lakrawazir@gmail.com (W.L.); rajeshwardass.ece@dcrustm.org (R.D.)
*Corresponding author

***Abstract*—Edge detection is a crucial step in computer vision, serving as a foundation for different applications like object detection, segmentation and scene understanding. Traditional edge detection methods often fail to capture complex boundaries in natural images. This study proposes a novel deep learning-based architecture, Clip-MobileNetV2-UNet, that integrates the lightweight, efficient MobileNet encoder with the segmentation capabilities of U-Net and the stabilizing properties of the Clip Rectified Linear Unit (ReLU) activation function. The MobileNet backbone significantly reduces computational cost and model size, making it suitable for edge detection on resource-constrained devices such as mobile and embedded devices. The Clip ReLU activation, a clipped version of the standard ReLU, is employed throughout the network to ensure bounded activations, helping prevent gradient explosions and stabilizing training. This modification preserves fine-grained features to improve model generalisation, overfitting, and edge sharpness. The U-Net decoder with skip connections recovers spatial details lost during downsampling. The training and validation datasets have been artificially increased using the rotation data augmentation technique. The Berkeley Segmentation Data Set and Benchmarks 500 (BSDS500) dataset has been used to train and evaluate the performance of Clip-MobileNetV2-U-Net using 3200 training, 1000 validation, and 500 testing images. The proposed model achieved competitive performance, with a mean Dice Coefficient (mDC) of 0.9256 and a mean Intersection over Union (mIoU) of 0.8419, and outperformed other deep architectures in edge detection. The proposed edge detection model, cross-validated on the Barcelona Images for Perceptual Edge Detection (BIPED) dataset, offers a reliable, precise, and scalable solution with low computational complexity and high accuracy, making it ideal for real-time computer vision applications.**

***Keywords*—U-Net, Berkeley Segmentation Data Set and Benchmarks 500 (BSDS500), Barcelona Images for Perceptual Edge Detection (BIPED) datasets, Clip-MobileNetV2-U-Net, encoder-decoder, edge detection**

I. Introduction

Edge detection is one of the most fundamental steps in image processing and the computer vision field [1–4]. It plays a key role in identifying object boundaries, shapes, and essential structures in an image [5]. Extracting edges can simplify image information while preserving the necessary features needed for higher-level tasks such as segmentation, recognition, and tracking. As a result, edge detection is widely utilised in fields such as medical image analysis, remote sensing, traffic monitoring, and autonomous driving [1–3].

Traditional edge detection methods, such as Sobel, Prewitt, Laplacian, and Canny operators, are widely known for their simplicity and speed. These methods are sensitive to noise, lighting variations, and complex textures due to their handcrafted filters and fixed thresholds. Therefore, they often produce inaccurate edges in real-world images involving different conditions.

U-Net and other encoder-decoder models improve performance by combining local and global context. U-Net variants address efficiency, accuracy, and fine edge preservation. Achieving accurate, noise-free, and computationally efficient edge detection remains a challenge, even with these advancements. Current research focuses on lightweight, practical models that balance accuracy and speed. Edge detection depends on the encoder, which uses pre-trained models and a hybrid approach with the clip Rectified Linear Unit (ReLU) function [6]. These models are pre-trained architectures that learn significant features from the ImageNet dataset, even with limited training data [7]. This process, known as transfer learning, allows the model to generalise more effectively and learn more quickly [8].

Training becomes more efficient with the introduction of non-linearity by the ReLU activation function [9]. The ReLU function helps the network learn complex patterns and avoid gradient vanishing. The skip connection between encoder and decoder layers preserves spatial features and improves edge map accuracy [10]. The model's performance in this work has been evaluated through a visual and quantitative assessment. Visual evaluation showed that the proposed model produced more accurate and cleaner edge maps than traditional methods. For quantitative evaluation, the Berkeley Segmentation Data Set and Benchmarks 500 (BSDS500) and Barcelona Images for Perceptual Edge Detection (BIPED) datasets are crucial for measuring performance using metrics such as Intersection over Union (IoU) and Dice Coefficient (DC) [11]. This research introduces a

Manuscript received July 15, 2025; revised July 31, 2025; accepted September 23, 2025; published February 27, 2026.

structured method for improving edge detection by integrating a modified activation function into a variant of the U-Net architecture. The steps taken in the proposed work are as follows:

(1) The study starts with an introduction and explains the motivation behind the edge detection.
(2) The problem statement discusses conventional activation functions' limitations and proposes a clipped ReLU. Next, the proposed method's key innovations are discussed.
(3) Edge detection research is reviewed in the related work section to identify gaps that this study fills.
(4) The core of this work is described in the proposed framework, which is divided into four phases. Phase 1 covers the selection of a benchmark dataset, its bifurcation, and the balancing strategy employed. Phase 2 describes data augmentation techniques, including image rotation, as well as the separation of data into training, validation, and testing sets. Phase 3 introduces the architectural design of the modified U-Net, focusing on integrating the clipped ReLU function. Phase 4 outlines the performance evaluation metrics used to assess the model's effectiveness.
(5) The BSDS500 dataset is used for training, validation & testing, and the BIPED dataset is used for cross-validating the performance of the proposed model.
(6) The results and discussion section analyses the statistical outcomes, comparing them against existing methods. The visual interpretations section provides graphical comparisons to support the quantitative findings.
(7) Finally, the conclusion and future prospects summarized the research contributions and suggested future directions for continued advancement in edge detection techniques.

II. Literature Review

A summary of the literature review, based on the latest work on edge detection, is presented in Table I. This section supports the background section by providing evidence for the proposed hypothesis.

TABLE I. A Brief Review of the Literature Based on the Latest Work on Edge Detection

Investigator (s)	Year	Method / Model	Dataset (s) Used	Outcomes
Wang and Zhang [1]	2024	BDCN	BSDS500	ODS
Elharrouss *et al*. [2]	2023	CHRNet	BSDS500	ODS, OIS, and AP
Xie *et al*. [3]	2025	EdgeFormer	ABC dataset	Better contextual
Wang *et al*. [4]	2023	multiscale TransUNet ++	Prostate MR and liver CT image	Combining the U-Net
An *et. al* [5]	2024	exclusive U-net	BSDS500	ODS F-Score
Ronneberger *et al*. [6]	2015	U-Net (Original)	ISBI Challenge, BSDS500	Encoder-decoder
Zhou *et al*. [7]	2018	U-Net++	DRIVE, CHASE_DB1	Nested and dense skip pathways
Oktay *et al*. [8]	2018	Attention U-Net	DRIVE, ISIC	Relevant edge features
Zhang *et al*. [9]	2018	ResUNet	BSDS500, Medical Datasets	Combines ResNet
Yu *et al*. [10]	2016	DRIVE, SK-LARGE	DRIVE, SK-LARGE	Dense blocks integrated into the U-Net
Syed *et al*. [11]	2025	Multi-Scale U-Net	CoFly-WeedDB	IoU
Qin *et al*. [12]	2020	U^2-Net	DUTS, ECSSD	U-Net structure
Chen *et al*. [13]	2021	TransUNet	Synapse, NYUDv2	U-Net
Pacot *et al*. [14]	2025	Hybrid multi-stage	BSDS500, NYUDv2	Structure recognition
Ahmadian *et al*. [15]	2023	U-Net and RGB-DSM Image	WHU and IEEE Data Fusion Dataset	Footprint extraction
Muntarina *et al*. [16]	2023	MultiResEdge	BSDS500, NYUDv2	Modified U-Net
Soria *et al*. [17]	2023	TEED	BIPED	Tiny and Efficient
Zhou *et al*. [18]	2023	UAED	BSDS500, NYUDv2	Uncertainty-Aware
Su *et al*. [19]	2021	PiDiNet	BSDS500, ImageNet	Pixel difference networks
Guo *et al*. [20]	2025	SFIA-MSNet	NYUDv2	Spatial-frequency attention
C. Liu *et al*. [21]	2024.	CPDC	BSDS500, NYUDv2, BIPED & CID	Cycle pixel
Y. Ye *et al*. [22]	2024.	DiffusionEdge	NYUDv2	Utilises diffusion
Liu *et al*. [23]	2024	Msmsfnet	BSDS500, NYUDv2, SAR images	Fusion network
J. Smith *et al*. [24]	2024	Multiscale Edge Detection	2DeteCT & Extended MNIST	Hybrid
Liu *et al*. [25]	2023	Edge Detection in SAR Images	SAR datasets	Ratio operator
Jain and Singh [26]	2025	Training-free Quantum-Inspired	BIPED, Multicue & NYUD	Enhanced accuracy
Han *et al*. [27]	2024	Edge-UNet	Satellite imagery	U-Net with edge extraction
Jonnala *et al*. [28]	2025	DSIA U-Net	Water body	Deep-shallow interaction mechanism

From Table I, it can be concluded that edge detection has garnered significant attention using deep learning. It is noted that multi-scale features, context learning, and attention mechanisms are used to detect thin, complex, and weak boundaries. Models are increasingly designed to reduce texture noise and more accurately highlight fine edges. To address the imbalance between edge and non-edge pixels, researchers have proposed advanced loss functions to mitigate this discrepancy. The research community is also moving toward validation on multiple datasets, such as BIPED and NYUDv2, to ensure that models are robust across different domains. According to the literature review, numerous deep learning models have been employed for edge detection. Classical U-Net and its variants remain popular because they offer encoder–decoder architectures that strike a balance between detail and context. Jonnala *et al*. [29] introduce AER U-Net, an attention-enhanced multi-scale residual U-Net integrated with clip ReLU activation for efficient water body segmentation from Sentinel-2 images. The approach improves stability, feature learning, and segmentation accuracy, outperforming conventional models in remote sensing and environmental monitoring tasks.

Accordingly, in this work, six different models based on Unet and its variants have been used for image edge detection. The BSDS500 benchmark dataset is used to identify edges in images. The model has been evaluated using two benchmark metrics: IoU and Dice Coefficient. The U-Net variant based on residual and clip ReLU functions typically contains 20–30 million parameters, which can be reduced in lightweight variants for real-time or mobile applications. Overall, parameter tuning in architecture and training is critical for enhancing edge sharpness, reducing noise, and improving model robustness across domains such as remote sensing and urban planning. Data augmentation and regularization techniques (e.g., dropout and batch normalisation) are used to prevent overfitting.

To strengthen the robustness and generalizability of the proposed model, the BIPED dataset is used as an external validation benchmark. A subset of 50 images is carefully selected and tested to verify the performance of the proposed model Clip-MobileNetV2-UNet beyond the BSDS500 dataset. This additional evaluation confirms that the model is not overfit to a single dataset and maintains consistent accuracy and edge detection quality across different modalities and Domains.

A. Research Trends and Gaps

Recent developments in encoder-decoder architectures, such as U-Net and its variants, have had a significant impact on edge detection. Some research gaps persist despite these trends. Previously, there was a lack of comprehensive research on the applicability of hybrid activation functions related to encoder-decoder contexts. The absence of a direct benchmark dataset makes it difficult to provide reliable comparisons. Lastly, research that focuses primarily on accuracy often overlooks computational cost, scalability, and resource requirements, and fails to offer any practical implications for deployment.

The results demonstrate that the ReLU-based hybrid U-Net model outperforms existing methods, producing high-quality edge maps, handling noise effectively, and yielding consistent results [12]. These studies have shown that pre-trained backbones and ReLU activation significantly improve the performance [13].

B. Problem Statement

In computer vision, edge detection is crucial because it enables the extraction of structural information from images by identifying critical intensity transitions and object boundaries. Although Sobel, Prewitt, and Canny are practical and straightforward to use, traditional gradient-based edge detection algorithms frequently perform poorly in complex environments with noise, texture clutter, fluctuating lighting, or low contrast. The flexibility and resilience of these approaches are further constrained by their heavy reliance on manually adjusted parameters and their inability to learn from data.

Deep learning-based techniques that learn features directly from data have significantly improved edge detection accuracy in recent years. To recognised fine and coarse edges, deep convolutional networks are used in architectures. However, these models are frequently computationally intensive and huge and they are not feasible for implementation on embedded or mobile devices.

In this study, consolidating feature extraction and reconstruction into a single framework is the prime goal in designing and evaluating an encoder-decoder-based architecture. Following the traditional model, the hybrid model uses a lightweight encoder and a modified decoder to reduce computational cost while preserving fine structural details. Hence, this work is unique in advancing encoder-decoder design with hybrid activation (Clipped ReLU) and lightweight backbones, while also providing quantitative benchmarking against these methods to ensure fairness and objectivity. The proposed framework addresses a distinct need in efficient segmentation and edge detection by serving as a supplement to existing frameworks rather than competing with them.

The benchmark datasets for edge detection tasks, namely BSDS500 and BIPED, play a vital role in training and evaluating the proposed model's behaviors. For training and validation of the proposed work, BSDS500 and BIPED provide a varied and demanding environment with intricate natural images annotated with high-quality human-drawn edge maps.

C. Major Contributions of Research

In this work, several new features have been introduced within the simple architecture of the popular U-Net framework. The Clipped ReLU is integrated into the pre-trained encoder-decoder network, enhancing edge preservation and improving gradient stability. This work is robust and prioritises cross-validation with BIPED. The significant contributions of this research are as follows:

- The lightweight model based on MobileNetV2 has been adopted as an incredibly effective convolutional neural network. The computational complexity and parameter count are significantly reduced, while the high feature-extraction capability, appropriate for edge detection, is maintained.
- Clipped ReLU, a bounded activation function that restricts outputs to a predetermined maximum, is utilised in place of the conventional ReLU activations found in MobileNetV2. This change enhances feature sparsity, reduces the likelihood of overactivation, and improves the model's ability to identify faint or subtle edges accurately.
- Based on the U-Net framework, a bespoke decoder is created, combining high-level semantic features with skip connections that combine low-level spatial information from previous layers. This ensures precise localisation and restores fine edge details that are often lost during downsampling.
- The cross-dataset validation has been performed on BIPED to show robustness across different domains.

- The BSDS500 dataset, a well-known benchmark for natural image edge recognition, is used to train and assess the proposed model.
- IoU and Dice Coefficient have been evaluated in pixel-wise prediction tasks and are used to calculate their performance.
- The outcomes show that the model can produce precise, high-resolution edge maps that closely match the ground truth contours.

D. Key Innovations

To strengthen the evaluation, the external BIPED dataset was used for validation. By comparing the suggested model to other datasets, we can evaluate its performance beyond the training set. The current work demonstrates that the proposed model maintains high performance across diverse data conditions, as evidenced by results on the external BIPED dataset. The standard evaluation metrics used include IoU and DC. These results demonstrate the competitiveness and efficiency of the proposed approach. The suggested Clip-MobileNetV2-U-Net model is more credible, practical, and relevant to real-world situations. The major contributions of this work are as follows:

- Clipped ReLU activation replaces standard ReLU architecture, which uses clipped ReLU, which limits activation values to a predetermined upper limit, in place of the widely used ReLU activation. Particularly in low-contrast regions, this enhances the network's ability to identify delicate and precise edges while suppressing noisy responses.
- Hybrid architecture (MobileNetV2 Encoder + U-Net Decoder): The model enables high-resolution edge map reconstruction through skip connections and fast multi-scale feature learning by combining a U-Net style decoder with the lightweight and practical MobileNetV2 encoder.
- The model's compact decoder design and depth-wise separable convolutions enable real-time applications on mobile and embedded devices, providing high edge detection accuracy with minimal computational load.
- The model is well-suited for real-time and resource-constrained applications due to its high accuracy and low computational cost.
- The proposed method outperforms previous U-Net versions in efficiency, robustness, and boundary detection.

III. Materials and Methods

The proposed edge detection framework uses a hybrid U-Net with a clip ReLU Function [16]. The skip connections and edge refinement module capture fine-grained edge details for precise boundary extraction [17]. The benchmark dataset, BSDS500, is used to train the model for robust performance across domains. The proposed work on Clip-MobileNetV2-UNet combines the efficiency of the MobileNetV2 encoder with the stability advantages of Clipped ReLU. Clipped ReLU introduces bounded nonlinearity, distinct from standard activation variants. The cross-dataset validation using the BSDS500 and BIPED datasets ensures the proposed model's stability. As a result, this work is distinguished from other hybrid U-Net variants by its practical applicability, strong validation, efficiency-oriented design choices, and innovative architectural design. The phase-by-phase development of the proposed framework is as follows.

A. Phase 1—Benchmark Dataset and Bifurcation

The BSDS500 is widely used for edge detection and image segmentation. It contains 500 natural images, split into 400 for training and 100 for testing. Each image in the dataset has been labelled by multiple experts, meaning each image has multiple ground truth masks. The sample images, along with numerous ground-truth images, are shown in Fig. 1.

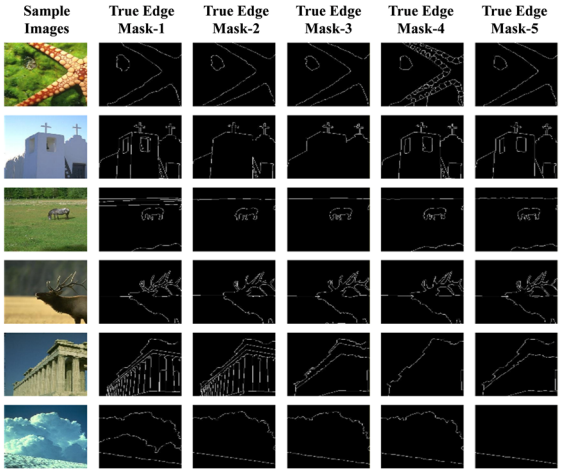

Fig. 1. The sample image with multiple ground truth images.

A single image has been seen to have five to six ground truths, with multiple object detections marked by an expert. These masks reflect subjective variations in edge perception and provide a rich benchmark for evaluating algorithm performance. BSDS500 supports both binary edge maps and probabilistic ground truth images. The proposed edge-detection framework is shown in Fig. 2.

The BSDS500 training set comprises 400 images, accompanied by approximately 2100 ground-truth edge maps annotated by multiple human subjects, and 100 test images with 500 ground-truth annotations [29–32]. Each image contains several ground-truth masks that capture subjective interpretations of object boundaries. This diversity helps to train models that generalise well across varying edge definitions and natural scenes. Balancing the BSDS500 dataset with the same number of natural images, i.e., for the training set, 2100 natural images with 2100 corresponding ground truth images, and for the testing set, 500 natural images with 500 corresponding ground truth images, ensures uniform learning and avoids model bias. This involves selecting or generating a single consistent ground truth per image (e.g., averaging or selecting the most consistent annotation). It simplifies training, enhances stability, and improves generalisation in edge detection.

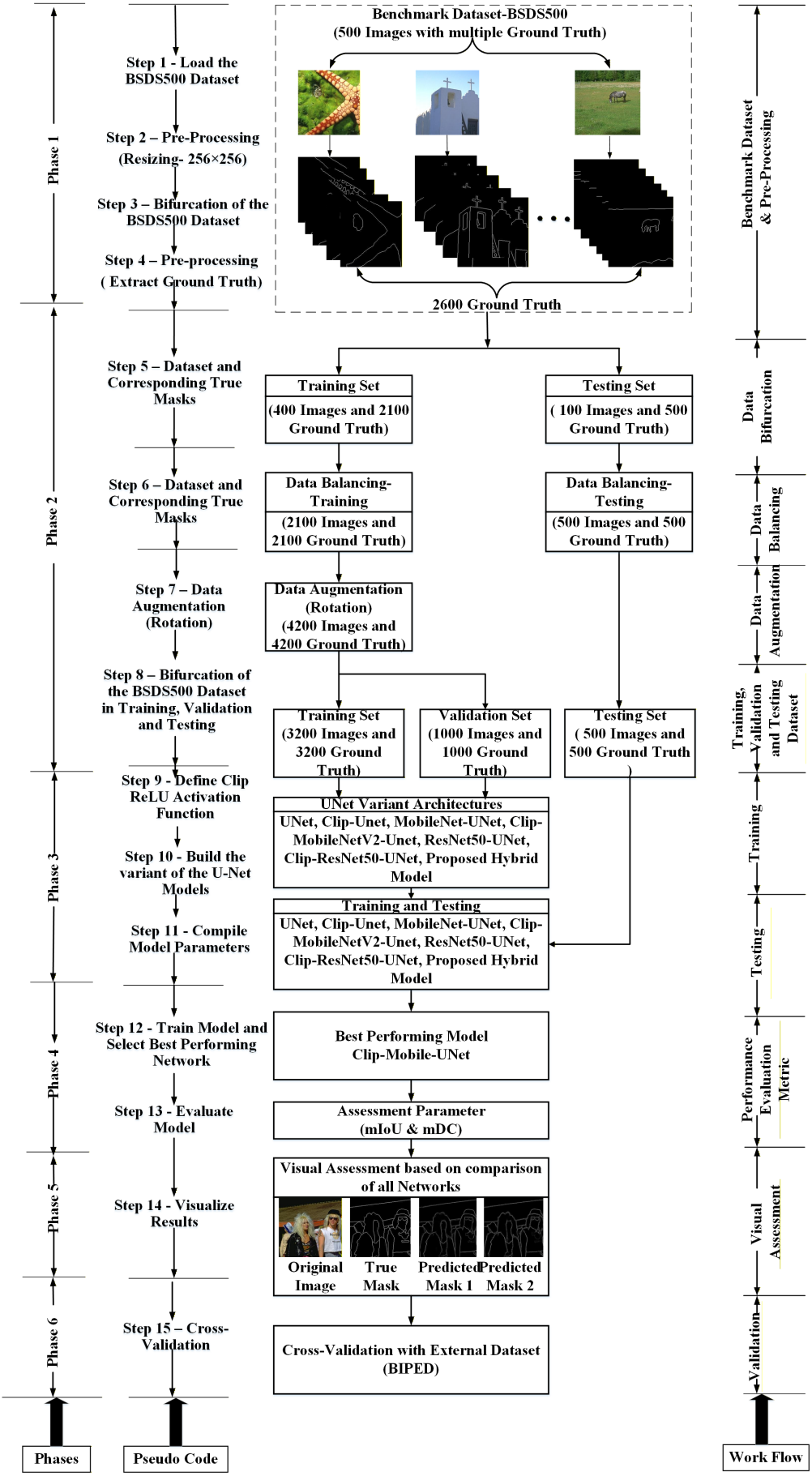

Fig. 2. The proposed framework adopted for edge detection.

B. Phase 2—Data Augmentation with Rotation and Training, Validation and Testing Dataset Bifurcation

Data augmentation via a 45° rotation enhances the diversity of training samples, enabling the model to learn edge patterns across different orientations. This technique enhances robustness and generalisation by exposing the model to rotated versions of the original images and their corresponding masks, thereby improving edge detection accuracy. This work utilises data augmentation to artificially expand the dataset by applying a 45° rotation. The sample images, including rotated natural images and ground truth images, are shown in Fig. 3.

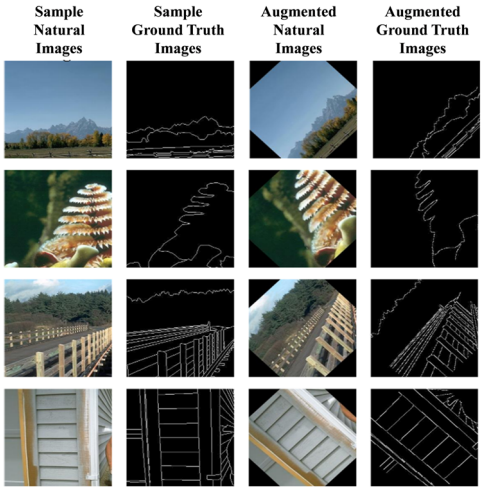

Fig. 3. The sample images with rotation of natural images and ground truth images.

The dataset comprises 4200 natural images with corresponding 4200 ground truth edge maps, ensuring a one-to-one ratio for balanced training. Further, the dataset is bifurcated into 3200 images for training and 1000 images for validation, allowing the model to learn effectively from diverse edge patterns while reserving sufficient data for performance evaluation.

Each ground-truth mask is carefully aligned with its corresponding image, ensuring consistency and reliability. This structure supports supervised learning approaches by providing clear input-output pairs. The balanced, well-divided dataset enhances the model's ability to generalise and precisely detect edges in unseen natural image scenarios.

C. Phase 3—Architectures of the Variants of U-Net with Model Specifications

The basic U-Net architecture, with an input size of 256×256×3 (for RGB images), maintains its encoder-decoder framework with symmetric skip connections. The input image is gradually downsampled via convolution and max-pooling layers in the encoder to capture deep features, then upsampled in the decoder to reconstruct the output edge map. MobileNet-U-Net is a lightweight and efficient variant of the traditional U-Net architecture, in which the encoder is replaced with MobileNet v2. It is designed for real-time edge detection tasks, particularly on mobile or low-power devices, without compromising accuracy too much. ResNet50-U-Net is a powerful edge detection model that combines the deep feature extraction capabilities of ResNet-50 (as an encoder) with the spatial reconstruction capabilities of U-Net (as a decoder) [33–44]. It's widely used for tasks that require accurate boundary detection and deep contextual understanding, such as medical imaging, remote sensing, and urban scene understanding. The model's specifications are shown in Table II.

TABLE II. THE MODEL'S SPECIFICATION

Model (s)	Model specification (s)	Encoder–Decoder Blocks
U-Net	Conv + ReLU + Maxpooling	4 Upsampling and downsampling blocks
Clip-U-Net	Conv + Clip ReLU + Maxpooling	4 Upsampling and downsampling blocks
ResNet50-U-Net	Conv + ReLU + Maxpooling	4 Upsampling and downsampling blocks with residual
Clip-ResNet50-U-Net	Conv + Clip ReLU + Maxpooling	4 Upsampling and downsampling blocks with residual
MobileNet-U-Net	Conv + ReLU + Maxpooling	4 Upsampling and downsampling blocks with depthwise separable
Clip-MobileV2-U-Net (Proposed Hybrid Model)	Conv + Clip ReLU + Maxpooling	4 Upsampling and downsampling blocks

Note: Output Framing−8×8 → 16×16 → 32×32 → 64×64 → 128×128 → 256×256.

The U-Net architectural design is widely used for edge detection due to its encoder-decoder framework with skip connections that preserve spatial details. MobileNet-U-Net replaces the standard encoder with a lightweight MobileNet, making it efficient for real-time applications on low-power products. ResNet50-UNet integrates a deep ResNet50 encoder, enhancing feature extraction through residual learning and improving performance in complex scenes. All variants share a standard decoder that upsamples features to generate precise edge maps. These networks support inputs such as 256×256 images and utilise sigmoid activations for binary edge maps. They are trained on annotated datasets and fine-tuned using loss functions such as edge-aware loss.

The ReLU clip, or clipped rectified linear unit, offers several benefits in DL models, such as U-Net, MobileNet-U-Net, and ResNet50-U-Net. Capping the activation output within a defined range (e.g., 0–6) prevents activation explosion and enhances training stability. This helps to avoid extreme values disrupting gradient flow and slowing convergence. It also acts as a regularity, reducing overfitting by controlling feature intensity. The ReLU clip supports efficient computation, especially in resource-constrained environments, and is well-suited for quantised models due to its bounded output. Overall, it contributes to the robust and efficient performance of deep neural networks. The basic parameters used to train the model are presented in Table III.

TABLE III. THE BASIC TRAINING PARAMETERS/SPECIFICATIONS USED TO TRAIN THE MODEL

Parameter (s)	Variated	Selected Specification
Optimizer	Adam and RMSprop	Adam
Learning Rate	0.01–0.0001	0.0001
Batch Size	8–32	32
Weight Initialization	ImageNet	ImageNet
Epochs	20–30	30
Loss Function	Binary cross-entropy and dice loss	Binary cross-entropy
Data Augmentation	Rotation, Flipping and Translation	Rotation with 90°
Early Stopping	Constant 3 to 5 accuracy on the validation dataset	5 Constant validation accuracy that saves the model
GPU	NVIDIA GEFORCE RTX 4060 with 8 GB RAM	

The selection of training parameters is crucial for effective model learning and accurate edge detection. A learning rate between 0.01 and 0.0001 ensures stable convergence without overshooting. The Adam optimiser is commonly used due to its adaptive learning capabilities. The batch size, typically 8 to 32, is chosen based on available GPU memory and training speed. The epochs range from 20 to 30, depending on the dataset size and convergence level. Binary cross-entropy or edge-aware loss functions are used to optimise boundary detection. Early stopping prevents overfitting by monitoring validation loss during training.

The basic Clip-MobileNet-UNet architecture is shown in Fig. 4.

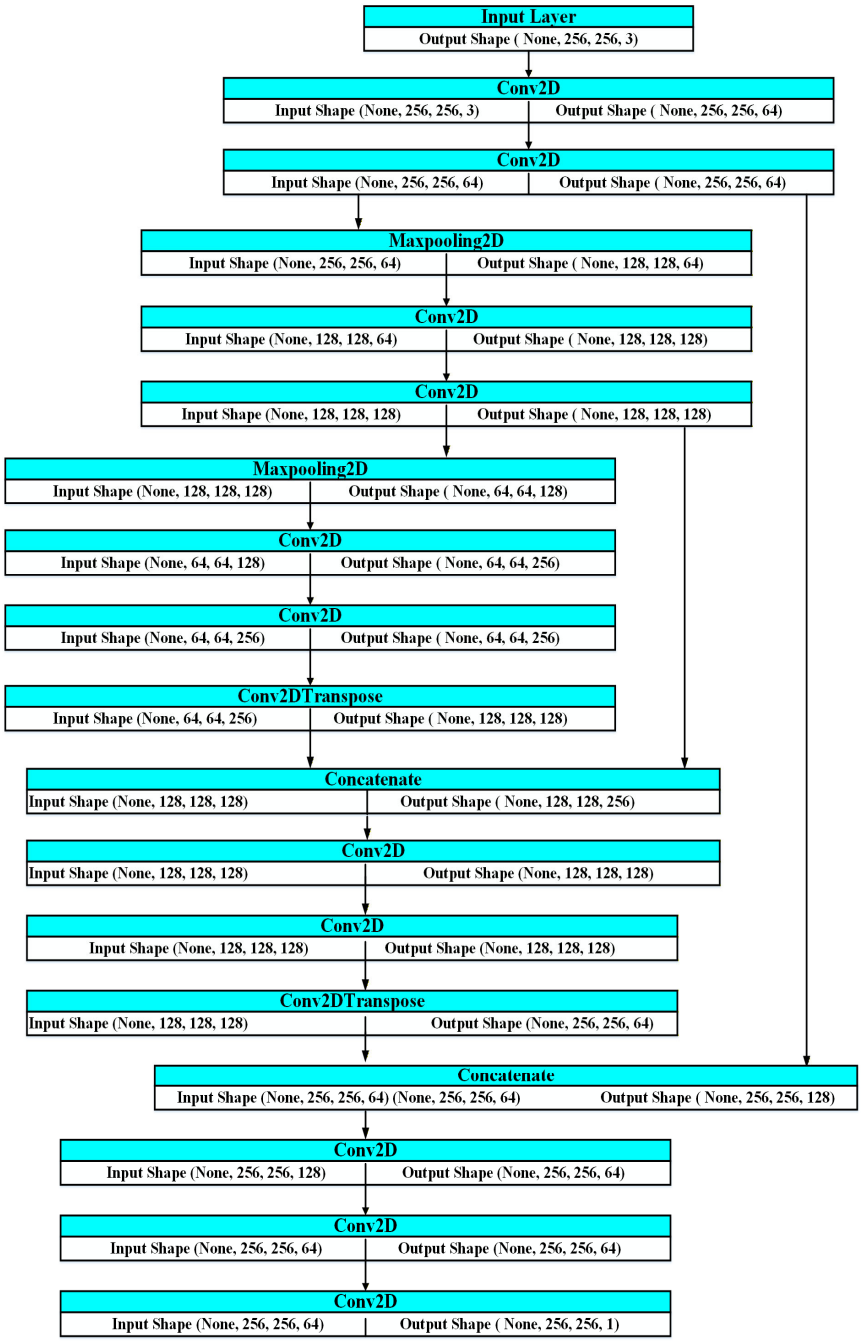

Fig. 4. The basic Clip-MobileNet-U-Net architecture.

D. Clip ReLU Function and U-Net Architecture

The Clip Rectified Linear Unit (ReLU) is a modified activation function designed to address limitations of the standard ReLU. While ReLU outputs 0 for negative inputs and the input value for positive inputs, Clip ReLU limits the activation to a predefined range, typically 0 to 1. The formula for Clip ReLU is Eq. (1) [30]:

$$Clip\ ReLU(x) = Min(Max(0, x), \theta) \tag{1}$$

where θ is the clipping threshold.

This bounded nature prevents very high activations, stabilises training, and reduces overfitting by maintaining consistent gradient flow during backpropagation. It combines the sparsity advantage of ReLU with added regularizations, making it particularly useful in deep architectures for edge detection. The U-Net architecture is a symmetric encoder-decoder model widely used in segmentation tasks [29, 31–42]. It is composed of a contraction path (encoder) that utilises convolution and downsampling to collect context, and an expansion path (decoder) that employs skip connections and upsampling to enable precise localisation. Each encoder block comprises two convolutional layers, followed by an activation function (Clip ReLU in this model) and a max pooling layer. The bottleneck lies between the encoder and the decoder, where data is compressed. The decoder mirrors the encoder, using upsampling layers and concatenating features from the encoder to recover spatial information by integrating Clip ReLU into the U-Net. The model benefits from controlled activations, sharper edge predictions, and improved generalisation. This combination maximizes the model's ability to detect fine edges in noisy or complex images, making it particularly suitable for edge detection tasks.

E. Phase 4—Performance Evolution Metrics Parameters

A brief description of the exhaustive literature survey on selecting performance evolution metrics is shown in Table IV.

TABLE IV. BRIEFLY DESCRIBES THE EXHAUSTIVE LITERATURE SURVEY ON SELECTING PERFORMANCE EVOLUTION METRICS

Investigator	Year	Model	Dataset	Assessment Metrics
Xie and Tu [45]	2017	CNN-based HED	BSDS500	ODS, OIS, IoU
Liu *et al*. [46]	2018	RCF-Net	BSDS500, NYUD	F-measure, IoU
Qin *et al*. [47]	2019	U-Net with Residual	DUTS, ECSSD	IoU
Zhao *et al*. [48]	2019	Enhanced Edge Guidance	BSDS500, PASCAL	IoU
Bao *et al*. [49]	2023	PDUF	BSDS500 & NYUDv2	F-Score
Li *et al*. [50]	2024	UHNet	BSDS500, NYUD, and BIPED	OIS and F Measure
Jie *et al*. [51]	2024	Trans-based Edge Detector	BSDS500 and NYUDv2	OIS and ODS

Note: PDUF-Pixel difference unmixing feature networks.

Table IV reveals that deep learning-based edge detection models have increasingly used IoU and Dice Coefficient (DC) as key performance metrics. These parameters evaluate the overlap and accuracy between the predicted edges and ground truth masks. Early models, recent U-Net variants, MobileNet-UNet, and Transformer-based networks incorporate IoU and DC for comprehensive assessment [29–34, 52]. Dice is instrumental in fine-structure tasks due to its sensitivity to small boundaries. The basic formulas used here for calculating IoU and Dice Coefficient are as follows [31]:

$$IoU = \frac{|G_m \cap P_m|}{|G_m \cup P_m|} \tag{2}$$

$$DC = \frac{2 \times |G_m \cap P_m|}{|G_m| + |P_m|} \tag{3}$$

where: G_m = Ground Truth Mask (binary mask: 1 = edge, 0 = background); P_m = Predicted Mask (binary mask).

IV. RESULTS AND DISCUSSION

The comparison result based on the statistical performance metrics of various U-Net variants is shown in Table V.

TABLE V. THE COMPARATIVE RESULTS BASED ON THE STATISTICAL PERFORMANCE METRICS OF VARIOUS U-NET VARIANTS

Model	mIoU	mDC
U-Net	77.45	84.78
ResNet50-U-Net	82.15	90.16
MobileNet-U-Net	80.92	88.36
Clip-U-Net	79.22	86.37
Clip-ResNet50-U-Net	82.73	89.36
Clip-MobileNetV2-U-Net (Proposed Hybrid Model)	84.19	92.56

In this work, the proposed Clip-MobileNetV2-Unet achieved the best performance among all tested models, clearly demonstrating its effectiveness in edge detection. It outperformed the others with a mIoU of 84.19 and a mean Dice Coefficient (mDC) of 92.56, highlighting its strong ability to capture spatial features while maintaining computational efficiency. The lightweight yet robust MobileNetV2 backbone, combined with the Clip mechanism, enabled accurate boundary detection and robust feature extraction. The Clip-ResNet50-U-Net achieved the second-best performance, with mIoU of 82.73 and mDC of 89.36. This indicates its capability, albeit with a slightly lower accuracy than Clip-MobileNetV2-Unet. These findings verify that Clip-MobileNetV2-Unet provides superior efficiency and accuracy in segmentation tasks.

This study's findings may have limited generalizability if applied only to the BSDS500 dataset. Using a second dataset, BIPED, this work evaluates the effectiveness of addressing this deficiency. In contrast to BIPED's more varied offerings, which include variations in lighting, texture, and object boundaries, BSDS500 primarily features images of natural scenes. Therefore, the model can be tested across various contexts, thereby enhancing

the reliability of the evaluation. The validation is strengthened and our approach is made more practical for real-world applications by including BIPED as a secondary benchmark dataset. This revision enhances the credibility of the suggested framework. Together, the two datasets prove that the proposed Clip-MobileNetV2-UNet is domain-agnostic and not limited to processing a single image type. Testing on more difficult images does not detract from the model's high accuracy, preservation of acceptable boundaries, or efficiency, as shown on BIPED. This proves that the proposed method is flexible and resilient.

To enhance the evaluation and cross-validation of the best-performing model, i.e., Clip-MobileNetV2-UNet, 50 images were selected for validation from the BIPED dataset. The inclusion of complex object boundaries, lighting variations, and diverse edge characteristics in this dataset made it an ideal complement to BSDS500. More extensive and varied testing of the proposed Clip-MobileNetV2-UNet demonstrated the model's generalizability across different image domains, thereby boosting confidence in the results. To further illustrate the model's generalizability across diverse image domains and increase the reliability of the results, the proposed Clip-MobileNetV2-UNet is subjected to additional validation under more challenging, varied conditions. The performance of the proposed network on the BIPED dataset is shown in Table VI.

TABLE VI. THE PERFORMANCE OF THE PROPOSED NETWORK ON THE BIPED DATASET

Dataset (BIPED)	mIoU	mDC
Clip-MobileNetV2-UNet (Proposed hybrid model)	81.3	89.4

The training and validation losses, based on accuracy and epoch, are shown in Fig. 5.

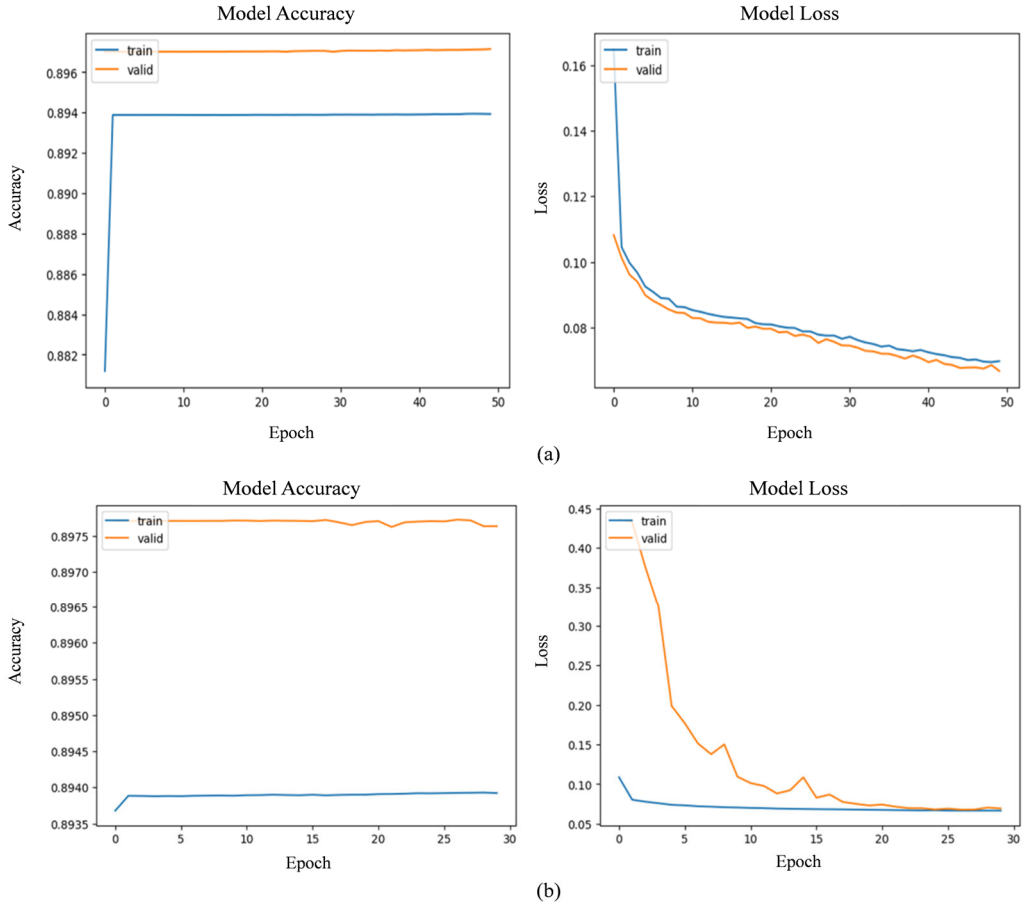

Fig. 5. The training and validation losses are based on accuracy and the number of epochs. (a) Training and validation loss for Clip-ResNet50-UNet. (b) Training and validation loss for Clip-MobileNet-UNet.

A. Phase 5—Visual-Based Analysis

In the visual interpretation of results, three images are commonly displayed side by side: the original image, the ground truth mask, and the predicted mask. The original image depicts natural scenes, including people, landscapes, and objects. The ground truth mask is often displayed as white contours on a black background, representing consensus boundaries labelled by annotators. The predicted mask, output by the U-Net model, is a grayscale or binary map highlighting the edges the model has detected.

The Clip-MobileNetV2-U-Net model produces a predicted mask with contours that closely align with the ground truth. The mask shows thin, continuous lines around object boundaries, capturing fine details such as outlines of faces, trees, or buildings. In areas where the

prediction is uncertain or the model underperforms, edges may appear blurred, broken, or thickened. These visual discrepancies enable researchers to evaluate the model's performance qualitatively [42–44, 53, 54]. Colour overlays can further enhance visual interpretation. Overall, the visual interpretation of a predicted mask on the BSDS500 dataset provides intuitive insight into the model's edge detection capability and complements quantitative metrics mIoU and mDC for model evaluation. The visual interpretations based on the original, ground truth, and predicted masks for Clip-MobileNetV2-UNet and Clip-ResNet50-UNet are shown in Fig. 6.

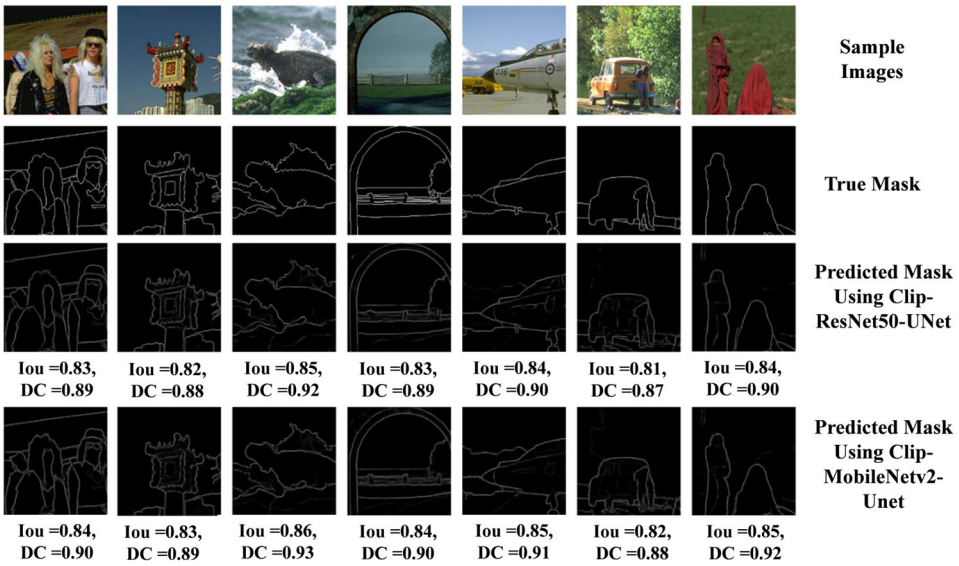

Fig. 6. The visual interpretation is founded on the original ground truth and predicted mask for Clip-ResNet50-UNet and Clip-MobileNetV2-U-Net.

The visual results from the BIPED dataset for the Clip-MobileNetV2-UNet model are shown in Fig. 7.

The visual results obtained on the BIPED dataset using the Clip-MobileNetV2-UNet model demonstrate remarkable accuracy in detecting fine edges and structural boundaries. The predicted outputs clearly highlight object contours with sharpness and continuity, reducing background noise and avoiding fragmented edges that are commonly observed in conventional approaches. The model effectively captures both global and local structural details, enabling the preservation of intricate object shapes without losing contextual information. Compared with other models, Clip-MobileNetV2-UNet produces more refined and consistent edge maps, emphasising its ability to generalise well on challenging natural images. These results confirm the robustness, reliability, and efficiency of the proposed method, making it highly suitable for edge detection and related segmentation tasks where precision and clarity are essential. The Comparisons of the proposed model with recent state-of-the-art models are shown in Table VII.

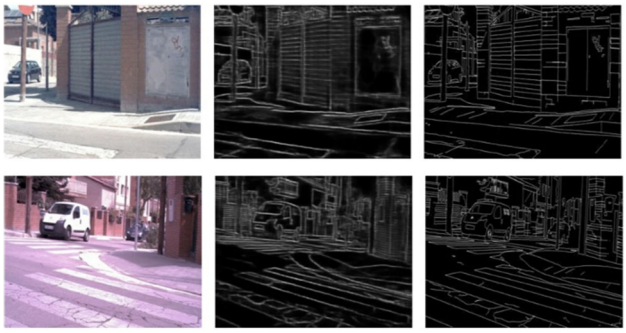

Fig. 7. The visual result of the BIPED dataset for the Clip-MobileNetV2-U-Net.

TABLE VII. COMPARISON WITH RECENT STATE-OF-THE-ART MODELS

Features	Model				
	HED [53]	PiDiNet [54]	Edge Former [44]	ResNet50–U-Net [43]	Proposed Clip-MobilenetV2-UNet
Dataset	BSDS500	BSDS500 BIPED	3D point data	BSDS500 BIPED	BSDS500 BIPED
Pre-training	ImageNet	no	No or EdgeLM	ImageNet	ImageNet
ODSF-measure/F1-Score/Dice Coefficient	OSD-0.78	OSD-0.807	Cannot compare the 3D point data to the 2D point data.	BSDS500 [F1-Score: 0.83] BIPED [F1-Score: 0.59]	BSDS500 [IoU: 0.8419; mDC: 0.9256] BIPED [IoU: 0.813, mDC: 0.894]

Compared to recent state-of-the-art 2D edge detectors such as PiDiNet and other methods, the proposed model achieves competitive metrics on the BSDS500 & BIPED datasets. At the same time, EdgeFormer operates in 3D space, so it is not directly comparable to 2D benchmarks.

B. Phase 6–ANOVA Statistical Test

To evaluate whether the models differ significantly in the DC, a one-way ANOVA is performed on the results of two experiments conducted on the BSDS500 and BIPED datasets. Dice score values are organised into two

groups to test the null hypothesis that all group means are equal, and the *p*-value is computed. The *p*-value for the DC is 0.8206. The *p*-value ≥ 0.05, indicating no significant difference between the two experiments. The ANOVA diagram for the DC, including between-group and within-group variance, is shown in Fig. 8.

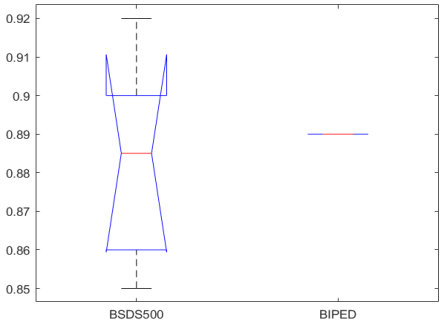

Fig. 8. The ANOVA diagram, between-group and within-group variance for the DC.

V. Conclusion and Future Prospects

In computer vision, edge detection is a vital task that lays the groundwork for more complex tasks, including object detection, image understanding, and segmentation. This study suggested a deep learning-based edge detection framework that incorporates the Clip ReLU activation function and a variant of the U-Net architecture. Deep learning models based on Clip ReLU have been tested on the benchmark BSDS500 dataset. This study integrates the Clip ReLU activation function into the U-Net model, a significant contribution. A ReLU clip limits activation values to a predefined range (e.g., 0 to 6). This prevents activation values from exploding and promotes stable gradient flow, leading to better convergence during training. This non-linearity in the U-Net encoder and decoder blocks improved edge representations, especially in gently transitioning regions.

The proposed model demonstrated strong performance on the BSDS500 test set, achieving a Dice Coefficient (DC) of 0.9256 and an Intersection over Union (IoU) of 0.8419. These findings demonstrate the architecture's robustness by showing high agreement between the predicted and ground-truth edge maps. The proposed U-Net with Clip ReLU improved edge detection accuracy and ensured stability and computational feasibility. The architecture is well-suited for both academic research and practical deployment in real-time systems, including mobile and embedded devices.

Future work may incorporate attention mechanisms or transformer blocks to refine edge localisation further and address challenges in more cluttered scenes or in real-time video edge-detection scenarios. The researchers may concentrate on minimizing information loss by implementing adaptive clipping in the activation function. The accuracy of edge detection can be increased by incorporating multi-scale feature fusion or attention techniques into the model. To manage complex datasets more efficiently while preserving accuracy, hybrid architectures and self-supervised learning can also be explored. The *p*-value indicates that the experiments using the BSDS500 and BIPED datasets are identical and show no difference.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Wazir Lakra conducted the research, conceptualization, methodology design, data curation, software implementation, and writing—original draft preparation. Rajeshwar Dass analysed the data and contributed equally to conceptualisation, methodology design, data curation, and writing—original draft preparation, supervision, validation, and formalisation—and wrote a review and edited it. All authors had approved the final version.

References

[1] F. Wang and M. Zhang, "Deep learning-based edge detection algorithm for noisy images," in *Proc. 6th Int. Conf. Artif. Intell. Pattern Recognit.*, 2023, pp. 465–472. doi: 10.1145/3641584.3641653

[2] O. Elharrouss, Y. Hmamouche, A. K. Idrissi *et al.*, "Refined edge detection with cascaded and high-resolution convolutional network," *Pattern Recognit.*, vol. 138, 109361, 2023. doi: 10.1016/j.patcog.2023.109361

[3] Y. Xie, Z. Tu, T. Yang *et al.*, "EdgeFormer: Local patch-based edge detection transformer on point clouds," *Pattern Anal. Appl.*, vol. 28, no. 1, p. 11, 2025. doi: 10.1007/s10044-024-01386-6

[4] B. Wang, F. Wang, P. Dong *et al.*, "Multiscale TransUNet++: Dense hybrid U-Net with transformer for medical image segmentation," *Signal Image Video Process.*, vol. 16, no. 6, pp. 1607–1614, 2022. doi: 10.1007/s11760-021-02115-w

[5] Y. An, J. Jing, X. Li, J. Zhang, and J. Bao, "An exclusive U-Net for fine and crisp edge detection," *Multimed. Tools Appl.*, vol. 83, no. 18, pp. 54657–54672, 2024. doi: 10.1007/s11042-023-17706-7

[6] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*, 2015, pp. 234–241. doi: 10.1007/978-3-319-24574-4_28

[7] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A nested U-Net architecture for medical image segmentation," arXiv Print, arXiv:1807.10165, 2018. https://arxiv.org/abs/1807.10165

[8] O. Oktay, J. Schlemper, L. L. Folgoc *et al.*, "Attention U-Net: Learning where to look for the pancreas," arXiv Print, arXiv:1804.03999, 2018. doi.org/10.48550/arXiv.1804.03999

[9] Z. Zhang, Q. Liu, and Y. Wang, "Road extraction by deep residual U-Net," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 5, pp. 749–753, 2018.

[10] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," arXiv Print, arXiv:1511.07122, 2016. doi.org/10.48550/arXiv.1511.07122

[11] A. Syed, B. Chen, A. A. Abbasi, S. A. Butt, and X. Fang, "MSEA-Net: Multi-scale and edge-aware network for weed segmentation," *AgriEngineering*, vol. 7, no. 4, p. 103, 2025. doi: 10.3390/agriengineering7040103

[12] X. Qin, Z. Zhang, C. Huang *et al.*, "U2-Net: Going deeper with nested U-structure for salient object detection," *Pattern Recognit.*, vol. 106, 107404, 2020. doi: 10.1016/j.patcog.2020.107404

[13] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, and Y. Zhou, "TransUNet: Transformers make strong encoders for medical image segmentation," arXiv Print, arXiv:2102.04306, 2021. doi.org/10.48550/arXiv.2102.04306

[14] M. P. Pacot, J. Juventud, and G. Dalaorao, "Hybrid multi-stage learning framework for edge detection: A survey," arXiv Print, arXiv:2503.21827, 2025. doi.org/10.48550/arXiv.2503.21827

[15] N. Ahmadian, A. Sedaghat, N. Mohammadi, and M. Aghdami-Nia, "Deep-learning-based edge detection for improving building footprint extraction from satellite images," *Environmental*

Sciences Proceedings, vol. 29, no. 1, p. 61, 2023. doi: 10.3390/ECRS2023-16615

[16] K. Muntarina, R. Mostafiz, F. Khanom, S. B. Shorif, and M. S. Uddin, "MultiResEdge: A deep learning-based edge detection approach," *Intell. Syst. Appl.*, vol. 20, 200274, 2023. doi: 10.1016/j.iswa.2023.200274

[17] X. Soria, Y. Li, M. Rouhani, and A. D. Sappa, "Tiny and efficient model for the edge detection generalization," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, 2023, pp. 1364–1373.

[18] C. Zhou, Y. Huang, M. Pu, Q. Guan, L. Huang, and H. Ling, "The treasure beneath multiple annotations: An uncertainty-aware edge detector," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2023, pp. 15507–15517.

[19] Z. Su, W. Liu, Z. Yu *et al*., "Pixel difference networks for efficient edge detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, 2021, pp. 5117–5127.

[20] Y. Guo, B. Li, W. Zhang, and W. Dong, "Multi-scale image edge detection based on spatial-frequency domain interactive attention," *Front. Neurorobot.*, vol. 19, 1550939, 2025. https://doi.org/10.3389/fnbot.2025.1550939

[21] C. Liu, W. Zhang, Y. Liu *et al*., "Cycle pixel difference network for crisp edge detection," *Neurocomputing*, vol. 619, 129153, 2025. doi: 10.1016/j.neucom.2024.129153

[22] Y. Ye, K. Xu, Y. Huang, R. Yi, and Z. Cai, "DiffusionEdge: Diffusion probabilistic model for crisp edge detection," in *Proc. the AAAI Conference on Artificial Intelligence*, vol. 38, no. 7, 2024, pp. 6675–6683. doi: 10.1609/aaai.v38i7.28490

[23] C. Liu, C. Wang, F. Dong *et al*., "MSMSFNet: A multi-stream and multi-scale fusion net for edge detection," arXiv Print, arXiv:2404.04856, 2024. doi.org/10.48550/arXiv.2404.04856

[24] J. Smith, D. Williams, E. Brown *et al.*, "Enhanced multiscale edge detection using deep learning techniques," *Edge Detection*, 2024. doi: 10.13140/RG.2.2.10860.78727

[25] C. Liu, C. Wang, X. Yang *et al*., "Edge detection in SAR images with deep learning: The combination of a ratio operator and fully convolutional neural networks," in *Proc. 2024 IEEE International Geoscience and Remote Sensing Symposium*, 2023. https://hal.science/hal-04362770

[26] A. Jain and R. Singh, "A training-free quantum-inspired image edge extraction method," arXiv Print, arXiv:2501.18929, 2025. doi.org/10.48550/arXiv.2501.18929

[27] R. Han, X. Fan, and J. Liu, "EUNet: Edge-UNet for accurate building extraction and edge emphasis in Gaofen-7 images," *Remote Sensing*, vol. 16, no. 13, p. 2397, 2024. doi: 10.3390/rs16132397

[28] N. S. Jonnala, R. C. Bheemana, K. Prakash *et al*., "DSIA U-Net: Deep shallow interaction with attention mechanism UNet for remote sensing satellite images," *Sci. Rep.*, vol. 15, no. 1, p. 549, 2025. https://doi.org/10.1038/s41598-024-84134-4

[29] N. S. Jonnala, S. Siraaj, Y. Prastuti *et al.*, "AER U-Net: Attention-enhanced multi-scale residual U-Net structure for water body segmentation using Sentinel-2 satellite images," *Scientific Reports*, vol. 15, no. 1, 16099, 2025. https://doi.org/10.1038/s41598-025-99322-z

[30] MathWorks. ClippedReLULayer. *Deep Learning Toolbox Documentation*. [Online] Available: https://www.mathworks.com/help/deeplearning/ref/nnet.cnn.layer.clippedrelulayer.html

[31] N. Yadav, R. Dass, and J. Virmani, "Assessment of encoder-decoder-based segmentation models for thyroid ultrasound images," *Med. Biol. Eng. Comput.*, vol. 61, no. 8, pp. 2159–2195, 2023. https://doi.org/10.1007/s11517-023-02849-4

[32] Berkeley Segmentation Dataset and Benchmark (BSDS). (2021). [Online]. Available: https://www2.eecs.berkeley.edu/Research/Projects/CS/vision/bsds/

[33] R. Dass, "A hybrid model for brain tumor classification via deep feature extraction and optimized classifier weighting," *Biomedical Physics & Engineering Express*, vol. 11, no. 5, 055029, 2025. doi: 10.1088/2057-1976/adf8ef

[34] N. Yadav, R. Dass, and J. Virmani, "Deep learning-based CAD system design for thyroid tumor characterization using ultrasound images," *Multimedia Tools and Applications*, vol. 83, no. 4, 43071–43113, 2023. https://doi.org/10.1007/s11042-023-17137-4

[35] N. Yadav, R. Dass, and J. Virmani, "Despeckling filters applied to thyroid ultrasound images: A comparative analysis," *Multimedia Tools and Applications,* vol. 81, no. 6, pp. 8905–8937, 2022. https://doi.org/10.1007/s11042-022-11965-6

[36] N. Yadav, R. Dass, and J. Virmani, "A systematic review of machine learning based thyroid tumor characterisation using ultrasonographic images," *Journal of Ultrasound,* vol. 27, no. 2, pp. 209–224, 2024. https://doi.org/10.1007/s40477-023-00850-z

[37] N. Yadav, R. Dass, and J. Virmani, "Objective assessment of segmentation models for thyroid ultrasound images," *Journal of Ultrasound,* vol. 26, no.3, pp. 673–685, 2023. https://doi.org/10.1007/s40477-022-00726-8

[38] N. Yadav, R. Dass, and J. Virmani, "Machine learning-based CAD system for thyroid tumour characterisation using ultrasound images," *Int. J. Med. Eng. Inform.,* vol. 16, no. 6, pp. 547–559, 2024. doi: 10.1504/IJMEI.2024.141790

[39] N. Yadav, R. Dass, and J. Virmani, "Texture analysis of liver ultrasound images," in *Proc. Emergent Converging Technologies and Biomedical Systems 2021*, 2022, pp. 575–585. doi: 10.1007/978-981-16-8774-7_48

[40] R. Dass and N. Yadav, "Image quality assessment parameters for despeckling filters," *Procedia Comput. Sci.*, vol. 167, no. 1, pp. 2382–2392, 2020. doi: 10.1016/j.procs.2020.03.291

[41] D. Saproo, A. N. Mahajan, S. Narwal, and N. Yadav, "Deep feature extraction and classification of diabetic retinopathy images using a hybrid approach," *Eng., Technol. Appl. Sci. Res.*, vol. 15, no. 2, pp. 21475–21481, 2025. doi: 10.48084/etasr.10188

[42] S. Sarita, R. Dass, and J. Saini, "Assessment of de-noising filters for brain MRI T1-weighted contrast-enhanced images," in *Proc. Emergent Converging Technologies and Biomedical Systems 2021,* 2022, pp. 607–613. doi: 10.1007/978-981-16-8774-7_50

[43] Wazir and R. Dass, "Assessment of deep learning models for image edge detection," *International Journal of Basic and Applied Sciences,* vol. 14, no. 5, pp. 45–57, 2025. doi: 10.14419/zgt6cp83

[44] Y. Xie, Z. Tu, T. Yang *et al*., "EdgeFormer: Local patch-based edge detection transformer on point clouds," *Pattern Analysis and Applications*, vol. 28, no. 11, 2025. doi: 10.1007/s10044-024-01386-6

[45] S. Xie and Z. Tu, "Holistically-nested edge detection," *Int. J. Comput. Vis.*, vol. 125, no. 1–3, pp. 3–18, 2017. doi: https://doi.org/10.1007/s11263-017-1004-z

[46] Y. Liu, M. Cheng, X. Hu, K. Wang, and X. Bai, "Richer convolutional features for edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 8, pp. 1939–1946, 2018. doi: 10.1109/TPAMI.2018.2878849

[47] X. Qin, Z. Zhang, C. Huang, M. Dehghan, O. Zaiane, and M. Jagersand, "BASNet: Boundary-aware salient object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2019, pp. 7479–7489.

[48] H. Zhao, Y. Zhang, S. Liu *et al*., "EGNet: Edge guidance network for salient object detection," *IEEE Transactions on Image Processing*, vol. 29, pp. 8226–8236, 2019. doi:10.1109/ICCV.2019.00887

[49] S. S. Bao, Y. R. Huang, J. C. Xu and G. Y. Xu, "Pixel difference unmixing feature networks for edge detection," *IEEE Access*, vol. 11, pp. 52370–52380, 2023. doi: 10.1109/ACCESS.2023.3279276

[50] F. Li and C. Lin. "UHNet: An ultra-lightweight and high-speed edge detection network," arXiv Print, arXiv:2408.04258, 2024. doi.org/10.48550/arXiv.2408.04258

[51] J. Jie, Q. Wang, J. Wu *et al*., "Efficient Transformer-based edge detector," in *Proc. 2024 International Joint Conference on Neural Networks (IJCNN)*, 2024, pp. 1–7.

[52] Z. Zhang, Q. Liu, and Y. Wang, "Road extraction by deep residual U-Net," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 5, pp. 749–753, 2021.

[53] S. Xie and Z. Tu, "Holistically-nested edge detection," *International Journal of Computer Vision,* vol. 125, no. 1–3, pp. 3–18, 2017.

[54] Z. Su, W. Liu, Z. Yu *et al*., "Pixel difference networks for efficient edge detection," *in Proc. the IEEE/CVF International Conference on Computer Vision (ICCV),* 2021, pp. 507–517.