Multi-Scale Attention-Based High-Resolution Network for Low-Contrast Medical Image Segmentation

K. Punnam Chendar[1], Dasari Swamy[2]
[1]Associate Professor, [2]M Tech
1,2, Department of ECE, KUCET
Kakatiya University, Warangal, Telangana, India, 506009
[1]k_punnam@kakatiya.ac.in, [2]dasariswamy998@gmail.com

Abstract
This paper presents a deep learning model for low-contrast medical image segmentation, where traditional methods struggle due to blurry and indistinct boundaries. Accurate segmentation in such images is challenging because adjacent anatomical structures often have weak or vanishing edges, making boundary localization difficult. To address this issue High-Resolution Multi-scale Encoder–Decoder Network used that enhances conventional encoder–decoder architectures. Unlike standard models that rely mainly on skip connections, the proposed network introduces multi-scale dense connections. This pathway preserves high-resolution semantic information, enabling more precise boundary detection. Additionally, the framework integrates a multi-task learning strategy by combining segmentation with contour regression, which helps the network focus on boundary regions. A difficulty-guided cross-entropy loss function is also introduced to emphasize harder-to-segment areas, particularly around ambiguous boundaries.

Keywords: Low-Contrast Medical Image Segmentation, Multi-Scale Dense Connections, Dilated Convolutions, Multi-Task Learning, Contour Regression, Difficulty-Guided Cross-Entropy Loss

I. INTRODUCTION

Medical imaging is an essential for modern healthcare, enabling accurate visualization of anatomical structures for disease identifiication and treatment planning. The rapid growth of imaging modalities has significantly increased the volume of medical image data, making manual segmentation by radiologists. Consequently, automated medical image segmentation has become a major research focus in computer-aided diagnosis and biomedical image analysis because it supports tumor detection, organ delineation, disease monitoring, surgical navigation, radiotherapy planning, and treatment assessment [1], [2]. Traditional segmentation techniques perform adequately on images with clear boundaries but struggle with noise, intensity inhomogeneity, and low-contrast anatomical structures [3], [4]. Although machine learning approaches improved segmentation by learning from labeled data, their dependence on handcrafted features limited their generalization across diverse medical imaging scenarios. Medical image segmentation, with encoder–decoder architectures such as U-Net, Fully Convolutional Networks (FCNs), V-Net, UNet++, and DeepMedic achieving superior performance through hierarchical feature learning and pixel-level classification [5], [6].

Despite these advances, segmenting low-contrast medical images remains challenging because adjacent tissues often exhibit similar intensity values, while noise, motion artifacts, and repeated pooling operations reduce spatial resolution and obscure boundary details. These limitations result in inaccurate localization of weak or

Doi:10.48047/ijiee.2026.16.2.59

ambiguous anatomical boundaries, particularly in CT and MRI images. To address these challenges, this research proposes a model that preserves high-resolution semantic information through a dedicated high-resolution pathway, dense multi-scale feature fusion, contour regression, and a difficulty-guided loss function. By integrating boundary-aware learning with high-resolution feature preservation, the proposed framework aims to improve segmentation accuracy and boundary delineation in low-contrast medical images, thereby enhancing the reliability of automated medical image analysis in clinical applications.

Early methods, including thresholding, region growing, edge detection, watershed transformation, active contour models, and level set methods, provided effective segmentation for images with well-defined boundaries but were highly sensitive to noise, weak edges, and parameter initialization, limiting their performance in complex medical images [7]. Traditional machine learning methods further improved segmentation by using handcrafted features such as texture, shape, and intensity descriptors; however, these approaches lacked the ability to generalize across diverse imaging conditions.

Encoder–decoder architectures have since become the dominant paradigm for medical image segmentation because they effectively combine hierarchical feature extraction with image reconstruction. U-Net significantly improved segmentation accuracy through skip connections that preserve spatial information [8], while ResNet and DenseNet enhanced feature learning using residual and dense connectivity. Other architectures, including DeepMedic, V-Net, DeepLab, RefineNet, and UNet++, introduced multi-scale learning, volumetric segmentation, dilated convolutions, and refined skip connections to improve contextual representation and segmentation performance [8], [9], [10]. Nevertheless, these methods still suffer from repeated pooling, loss of spatial resolution, limited multi-scale feature fusion, high computational complexity, and inadequate localization of weak or blurred boundaries, especially in low-contrast CT and MRI images. Recent studies have also explored contour-aware segmentation, multi-task learning, and attention mechanisms to improve boundary localization. Approaches such as Deep Contour-Aware Networks (DCAN), Deep Level Sets, and active contour-based deep learning models demonstrated improved segmentation by jointly learning segmentation and contour information; however, they often require additional annotations, complex optimization, or increased computational cost [11], [12].

Although recent developments, including transformer-based architectures, attention mechanisms, self-supervised learning, and advanced loss functions, have further improved segmentation performance, several challenges remain unresolved. Existing encoder–decoder networks continue to lose high-resolution spatial information during repeated downsampling, provide limited boundary-aware learning, and rely on conventional loss functions that assign equal importance to all pixels, resulting in poor segmentation of ambiguous boundary regions. Furthermore, practical issues such as limited annotated datasets, class imbalance, cross-dataset variability, and high computational requirements restrict the deployment of current models in clinical environments. These research gaps motivate the proposed High-Resolution Multi-scale Encoder–Decoder Network (HMEDN), which preserves high-resolution semantic information through dense multi-scale feature propagation, incorporates explicit contour regression, and employs a difficulty-

Doi:10.48047/ijiee.2026.16.2.59

guided loss function to improve boundary localization and segmentation accuracy in low-contrast medical images. Table 1 summarizes the strengths and limitations of major deep learning segmentation architectures.

II. METHODOLOGY

The proposed work introduces a Encoder Decoder Network that preserves high-resolution semantic information throughout the network while improving boundary localization. Unlike conventional architectures that depend mainly on skip connections, the proposed model incorporates a high-resolution pathway, dense multi-scale feature fusion.

The overall methodology consists of data acquisition, preprocessing, feature extraction, multi-scale learning, high-resolution feature preservation, contour regression, model optimization, and performance evaluation. The proposed framework is designed to improve segmentation accuracy in low-contrast medical images by integrating several complementary components shown in fig. 1

.

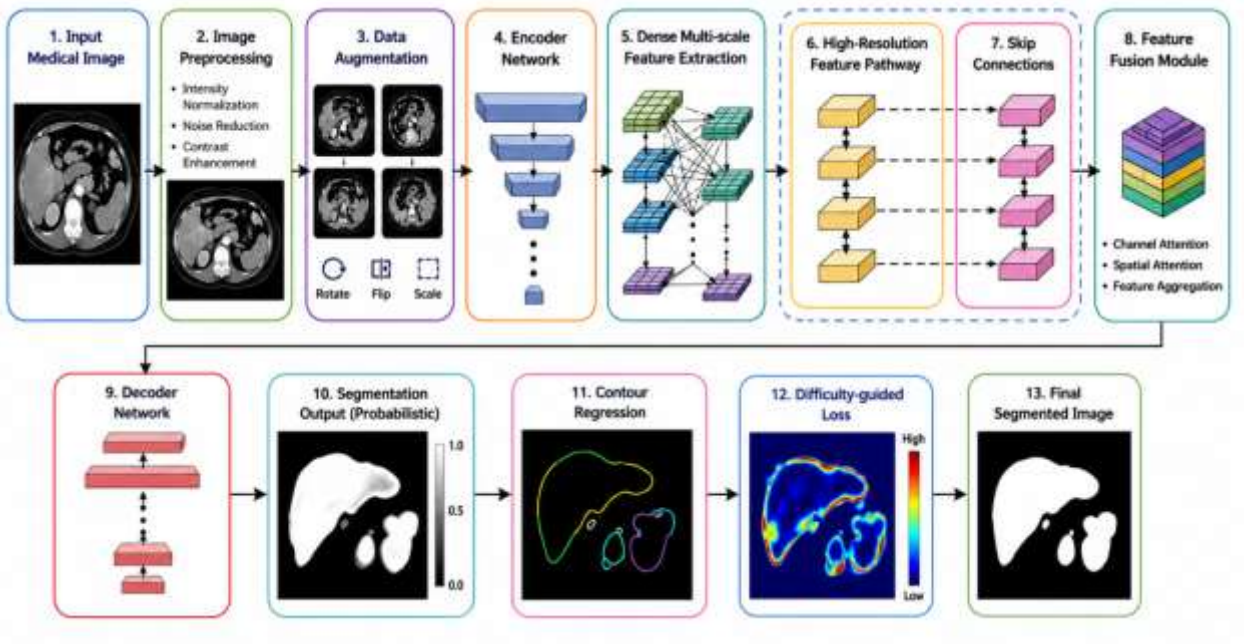

Fig. 1 workflow of the proposed framework

Preserving high-resolution semantic information and enhancing boundary localization. The workflow consists of data acquisition, preprocessing, encoder-based feature extraction, dense multi-scale learning, a high-resolution pathway, feature fusion, contour regression, multi-task optimization, and performance evaluation. The framework supports CT, MRI, and multi-modal medical images with expert-annotated ground truth masks. During preprocessing, images are normalized, resized to a fixed resolution, denoised, and augmented using flipping, rotation, cropping, zooming, elastic deformation, and brightness adjustment to improve robustness and reduce overfitting.

Image normalization standardizes pixel intensity distributions to stabilize network training and is computed as

$$I_n = \frac{I - \mu}{\sigma}$$

Doi:10.48047/ijiee.2026.16.2.59

where I is the input image, μ is the mean intensity, and σ is the standard deviation. After preprocessing, the normalized images are passed to the encoder network, which extracts hierarchical semantic features using convolution, batch normalization, ReLU activation, and max-pooling.

To improve feature reuse and preserve semantic information, dense multi-scale feature learning, where each layer receives feature maps from all preceding layers. The dense connectivity is defined as

$$X_l = H_l(X_0, X_1, \ldots, X_{l-1})$$

where $H_l(\cdot)$ denotes the convolution, normalization, and activation operations of the l^{th} layer. This dense propagation strengthens gradient flow, reduces information loss, and enables effective integration of local and global contextual features.

A key contribution of the proposed framework is the high-resolution pathway, which operates parallel to the encoder–decoder backbone. Unlike conventional architectures that repeatedly downsample feature maps, this pathway preserves spatial resolution using dilated convolutions, thereby improving boundary localization and small-object segmentation. The dilated convolution is expressed as

$$Y(i) = \sum_k X(i + r \cdot k)W(k)$$

where r is the dilation rate controlling the spacing between kernel elements. The resulting high-resolution features are fused with encoder and decoder features through a multi-scale feature fusion module, allowing simultaneous utilization of low-level spatial information and high-level semantic representations for accurate segmentation.

To improve segmentation of weak and ambiguous boundaries, the proposed architecture incorporates an auxiliary contour regression branch. Instead of performing segmentation alone, the network jointly learns segmentation masks and contour information using a multi-task learning strategy. The overall optimization objective is

$$L = L_{seg} + \lambda L_{contour}$$

where L_{seg} represents the segmentation loss, $L_{contour}$ denotes the contour regression loss, and λ balances the contribution of both tasks. This joint optimization enables shared feature learning while enhancing anatomical boundary localization.

To further emphasize difficult boundary regions, HMEDN employs a difficulty-guided cross-entropy loss function that assigns higher weights to challenging pixels. The weighted segmentation loss is defined as

$$L = -\sum_i w_i \, y_i \log(p_i)$$

where w_i is the adaptive pixel weight, y_i is the ground-truth label, and p_i is the predicted probability. By assigning greater importance to boundary pixels, the proposed loss improves segmentation accuracy and accelerates convergence in low-contrast medical images.

Table I. Training Configuration and Hyperparameter Settings

Category	Parameter	Value / Setting
Input Settings	Input Image Size	256 × 256 (or 512 × 512)
	Input Channels	1 (Grayscale) / 3 (RGB)
	Data Type	Float32
	Image Normalization	Min–Max / Z-score Normalization
Data Augmentation	Rotation	±20°
	Horizontal Flip	Probability = 0.5

Doi:10.48047/ijiee.2026.16.2.59

	Vertical Flip	Probability = 0.5
	Random Scaling	0.8–1.2
	Translation	±10 pixels
	Elastic Deformation	Enabled
	Random Brightness	±15%
	Random Contrast	±15%
Encoder	Backbone	High-Resolution Encoder
	Initial Filters	64
	Convolution Kernel	3 × 3
	Activation Function	ReLU
	Normalization	Batch Normalization
	Downsampling	Max Pooling (2 × 2)
Feature Extraction	Multi-scale Feature Blocks	4
	Feature Pyramid	Enabled
	Dense Skip Connections	Enabled
	High-Resolution Pathway	Enabled
Feature Fusion	Fusion Strategy	Concatenation + 1 × 1 Convolution
	Attention Mechanism	Channel + Spatial Attention
Decoder	Upsampling Method	Bilinear Upsampling
	Skip Connections	Residual Skip Connections
	Decoder Filters	512 → 256 → 128 → 64
Output Layer	Output Activation	Sigmoid (Binary) / Softmax (Multi-class)
	Number of Classes	2 (Binary) / N (Multi-class)
Loss Function	Primary Loss	Dice Loss
	Auxiliary Loss	Boundary (Contour) Loss
	Final Objective	Dice + Binary Cross-Entropy + Difficulty-guided Loss
Optimizer	Optimizer	Adam
	Initial Learning Rate	**0.0001**
	Weight Decay	1×10^{-5}
	β_1	0.9
	β_2	0.999
	Epsilon	1×10^{-8}
Learning Rate Policy	Scheduler	ReduceLROnPlateau
	Reduction Factor	0.1
	Patience	10 epochs
	Minimum Learning Rate	1×10^{-6}
Training	Batch Size	**8–16**
	Number of Epochs	**100–200**
	Early Stopping	Patience = 20 epochs
	Gradient Clipping	1.0
	Mixed Precision Training	Enabled
Validation	Validation Split	20%
	Cross Validation	5-Fold Stratified Cross Validation
	Model Selection	Lowest Validation Loss

Doi:10.48047/ijiee.2026.16.2.59

Algorithm

1. Read the medical image.
2. Perform intensity normalization.
3. Resize the image to a fixed resolution.
4. Apply data augmentation.
5. Pass the image through the encoder.
6. Extract dense multi-scale features.
7. Generate high-resolution features using dilated convolutions.
8. Fuse encoder, decoder, and high-resolution features.
9. Reconstruct the segmentation map using the decoder.
10. Predict contour probability maps.
11. Compute segmentation loss.
12. Compute contour regression loss.
13. Calculate the total loss.
14. Update network parameters using Adam optimization.
15. Repeat until convergence.
16. Generate the final segmentation mask.

III. Results

The proposed model was evaluated on annotated low-contrast CT and MRI datasets to assess its segmentation accuracy and boundary localization performance. The datasets were divided into 70% training, 15% validation, and 15% testing. The model was trained using the Adam optimizer with a learning rate of 0.0001, batch size of 8, and 150 epochs. Performance was evaluated using the Dice Similarity Coefficient (DSC), Intersection over Union (IoU), Precision, Recall, Hausdorff Distance (HD), and Average Surface Distance (ASD), which measure segmentation overlap and boundary accuracy. The proposed HMEDN demonstrated consistent convergence during training, with decreasing training and validation losses and steadily improving Dice scores, indicating good generalization without significant overfitting. Quantitative comparisons with established segmentation models show that HMEDN achieved the highest overall performance, obtaining a Dice score of 93.4%, IoU of 88.0%, Precision of 94.1%, and Recall of 92.9%, outperforming FCN, U-Net, DenseNet, and UNet++. Furthermore, HMEDN achieved the lowest boundary errors, with a Hausdorff Distance of 6.4 pixels and an Average Surface Distance of 1.8 pixels, demonstrating more accurate boundary localization, particularly in low-contrast regions.

Qualitative evaluation further confirmed that the proposed framework produced cleaner segmentation masks with more accurate organ boundaries, reduced leakage into adjacent tissues, improved preservation of thin anatomical structures, and better detection of small lesions. An ablation study also verified the contribution of each architectural component. Starting from a baseline encoder–decoder network with a Dice score of 88.6%, the inclusion of dense connections increased performance to 90.1%, the high-resolution pathway improved it to 91.8%, contour regression further increased it to 92.7%, and the complete HMEDN incorporating the difficulty-guided loss function achieved the highest Dice score of 93.4%. These results demonstrate that each proposed component contributes positively to segmentation accuracy.

Although HMEDN introduces additional computational and memory requirements because of dense feature propagation, multi-

Doi:10.48047/ijiee.2026.16.2.59

scale feature fusion, and the parallel high-resolution pathway, the increase in complexity is justified by the substantial improvement in segmentation quality. The proposed framework effectively preserves high-resolution semantic information, enhances boundary awareness through contour regression, and focuses learning on difficult boundary regions using the difficulty-guided loss function. Consequently, HMEDN provides more accurate and robust segmentation than conventional encoder–decoder architectures, making it well suited for challenging low-contrast medical imaging applications such as brain tumor, liver, lung, cardiac, retinal, and prostate image segmentation, as well as radiotherapy planning and surgical navigation.

IV. CONCLUSION AND FUTURE SCOPE

The proposed framework integrates dense multi-scale feature learning, a dedicated high-resolution pathway, contour regression, and a difficulty-guided loss function within a unified encoder–decoder architecture, overcoming the limitations of conventional deep learning models that lose spatial information during repeated downsampling. The methodology incorporates image preprocessing, data augmentation, hierarchical feature extraction, multi-task learning, adaptive optimization, and quantitative evaluation using standard segmentation metrics, demonstrating improved segmentation accuracy and boundary delineation in low-contrast medical images. The proposed HMEDN has the potential to enhance clinical applications such as diagnostic imaging, radiotherapy planning, and computer-assisted surgery. The major contributions of this research include the development of a novel high-resolution multi-scale encoder–decoder architecture, preservation of high-resolution semantic information through a parallel pathway, integration of dense feature propagation and contour-aware learning, introduction of a difficulty-guided loss function for challenging boundary regions, and comprehensive evaluation using widely accepted segmentation metrics. Future work may extend the proposed framework to three-dimensional medical image segmentation for volumetric CT and MRI data, incorporate attention mechanisms and transformer-based architectures, investigate lightweight models for real-time clinical deployment, and explore federated and self-supervised learning techniques to improve robustness and generalization across diverse medical imaging datasets.

References

1. S. Bauer, R. Wiest, L.-P. Nolte, and M. Reyes, “A Survey of MRI-Based Medical Image Analysis for Brain Tumor Studies,” *Physics in Medicine & Biology*, vol. 58, no. 13, pp. R97–R129, 2013.
2. D. Shen, G. Wu, and H.-I. Suk, “Deep Learning in Medical Image Analysis,” *Annual Review of Biomedical Engineering*, vol. 19, pp. 221–248, 2017.
3. M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, “The Pascal Visual Object Classes (VOC) Challenge,” *International Journal of Computer Vision*, vol. 88, no. 2, pp. 303–338, 2010.
4. B. Menze et al., “The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS),” *IEEE Transactions on Medical Imaging*, vol. 34, no. 10, pp. 1993–2024, 2015.
5. H. Tang, E. Kim, X. Xie, and L. Yang, “Deep Level Sets for Image Segmentation,” in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2017.
6. G. Lin, A. Milan, C. Shen, and I. Reid, “RefineNet: Multi-Path Refinement Networks for High-Resolution Semantic

Doi:10.48047/ijiee.2026.16.2.59

Segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.
7. S. Xie and Z. Tu, "Holistically-Nested Edge Detection," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2015. [23] Y. Chen et al., "Combining CNN and Level-Set for Medical Image Segmentation," *IEEE Transactions on Medical Imaging*, 2017.
8. G. Huang, Z. Liu, L. Van der Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.
9. F. Yu and V. Koltun, "Multi-Scale Context Aggregation by Dilated Convolutions," in *International Conference on Learning Representations (ICLR)*, 2016.
10. Ö. Çiçek, A. Abdulkadir, S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2016.
11. H. Chen, X. Qi, L. Yu, and P. A. Heng, "DCAN: Deep Contour-Aware Networks for Accurate Gland Segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.
12. K. Kamnitsas, C. Ledig, V. F. J. Newcombe, et al., "Efficient Multi-Scale 3D CNN with Fully Connected CRF for Accurate Brain Lesion Segmentation," *Medical Image Analysis*, vol. 36, pp. 61–78, 2017.

Doi:10.48047/ijiee.2026.16.2.59