International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Traitement du Signal
Vol. 40, No. 4, August, 2023, pp. 1621-1628
Journal homepage: http://iieta.org/journals/ts

Optimization of Deep Learning Algorithms for Image Segmentation in High-Dimensional Data Environments

Hua Sun

School of Software, Xinjiang University, Urumqi 830008, China

Corresponding Author Email: sunhua@xju.edu.cn

https://doi.org/10.18280/ts.400430

Received: 10 April 2023
Revised: 2 July 2023
Accepted: 18 July 2023
Available online: 31 August 2023

Keywords:
high-dimensional data flow, image segmentation, U-Net-inspired architecture, depth skip connections, lightweight modeling, BCE loss, Dice loss, small object precision segmentation

ABSTRACT

As image segmentation tasks become increasingly intricate within high-dimensional data flow environments, conventional segmentation techniques are challenged in delivering both efficiency and precision. In this context, the problem of image segmentation under high-dimensional data flux was examined. Depth skip connections, inspired by the *U-Net* architecture, were introduced, harnessing the superior feature extraction capabilities of deep encoders and enabling the formulation of a lightweight model structure. Furthermore, an equilibrium between *Binary Cross-Entropy* (*BCE*) loss and *Dice* loss was established, targeting enhanced accuracy in small object segmentation tasks within such data-intensive settings. These innovations not only augment algorithmic accuracy and resilience but also provide pivotal contributions to ongoing research in the image segmentation realm. The methodologies delineated herein present a refined approach to image segmentation, revealing significant potential for application in pivotal sectors, including medical image analysis and autonomous vehicular navigation.

1. INTRODUCTION

With society's burgeoning reliance on image information, image segmentation has emerged as a pivotal task in the realm of computer vision and is now extensively utilized across myriad sectors [1]. In realms such as medical imaging, autonomous driving, and industrial inspection, the significance of image segmentation technology in information extraction and analysis has been well-established [2-7]. However, conventional image segmentation approaches, in the face of escalating data volumes and dimensionality, have shown shortcomings in efficiency and accuracy, especially in high-dimensional data flow environments [8, 9]. It is observed that as these traditional methods grapple with expansive high-dimensional data, there is an exponential increase in computational resource and time consumption.

The examination of the design and optimization of deep learning image segmentation algorithms under high-dimensional data flow is perceived as essential for advancing the field of image segmentation [10-14]. Undertaking this task in such intricate data environments introduces multifaceted challenges, encompassing concerns like computational efficiency, model intricacy, and segmentation precision [15, 16]. By the application of specialized optimizations and innovations, it has been indicated that not only can segmentation accuracy be elevated, but also processing velocities can be accelerated, making them suitable for real-time or nearly real-time operational contexts [17-19]. Such advancements, in turn, have potential implications in sectors such as medical image analysis and autonomous driving, thus supporting more accurate and dependable decision-making processes.

While contemporary image segmentation techniques have garnered commendable success in multiple facets, deficiencies are observed when these methods are subjected to high-dimensional data flows. For instance, the conventional *U-Net* architecture, despite its prowess in numerous applications, shows potential areas of improvement, particularly in efficiency and precision for small object segmentation tasks [20]. Additionally, traditional network frameworks are often susceptible to overfitting in the presence of intricate high-dimensional data, potentially leading to an elongated optimization process and subsequently compromising the model's adaptability and generalization potential [21]. Such observations underscore the avenues available for refinement within the existing image segmentation methodologies.

This research is categorized into two primary facets. Initially, based on the *U-Net* architecture, modifications aiming for a lightweight design were introduced, with depth skip connections supplanting the traditional skip connections, underscoring the feature extraction capabilities of deep encoders. Such an evolution is deemed crucial for the model's performance in high-dimensional data scenarios, mitigating computational requirements while simultaneously enhancing accuracy. Subsequently, an equilibrium between *BCE* loss and *Dice* loss was instituted, aiming to adeptly address small object segmentation tasks in data-intensive settings. Such loss function architectures are designed to prioritize essential segments during the training phase, thus optimizing the identification and localization capacities for diminutive objects. Collectively, the methodologies detailed in this examination not only advance the technological frontier in image segmentation but also promise enhanced efficiency and dependability across sectors, highlighting substantial application potential and real-world value.

2. ENCODER DESIGN AND PARAMETER ANALYSIS FOR LIGHTWEIGHT IMAGE SEGMENTATION

In light of the computational resource demands posed by high-dimensional data streams, an imperative for lightweight image segmentation models based on the foundational *U-Net* network has been identified. The challenge lies in achieving a design optimized for minimal resource consumption, yet capable of extracting intricate coarse-grained semantic features, thereby ensuring precision in semantic segmentation.

Within this context, a lightweight image segmentation model tailored for high-dimensional data streams was developed. Grounded in the *U-Net* framework, this design sought efficiency in lightweight optimization. Incorporating both an encoder and a decoder, the interlink between these components was altered from the originally conceived flat skip connection to a more intricate deep skip connection. Through the introduction of these deep skip connections, it was observed that a synergy of the corresponding encoder layer's flat skip connections and skip connections from all subsequent deeper encoder layers, excluding the final layer, could be established. Within this framework, feature maps stemming from all underlying encoder layers, with the exception of the last, underwent convolution processes prior to their relay to the decoder layer. While features were transmitted downwards by the encoder via a downsampling mechanism, the decoder correspondingly relayed them upwards through upsampling, ensuring robust feature extraction and integral information reconstruction. The inclusion of these deep skip connections proved instrumental in encompassing a wider spectrum of coarse-grained semantic features, thus enabling refined semantic segmentation. By integrating convolution operations within the deep skip connections, gradient back-propagation was effectively facilitated, bolstering the stability during the model's training phase. The design's lightweight character was deemed instrumental in optimizing efficiency within high-dimensional data stream settings, fulfilling the criteria for instantaneous or near-immediate analyses, all while preserving exemplary segmentation precision. A depiction of this innovatively architected model is presented in Figure 1.

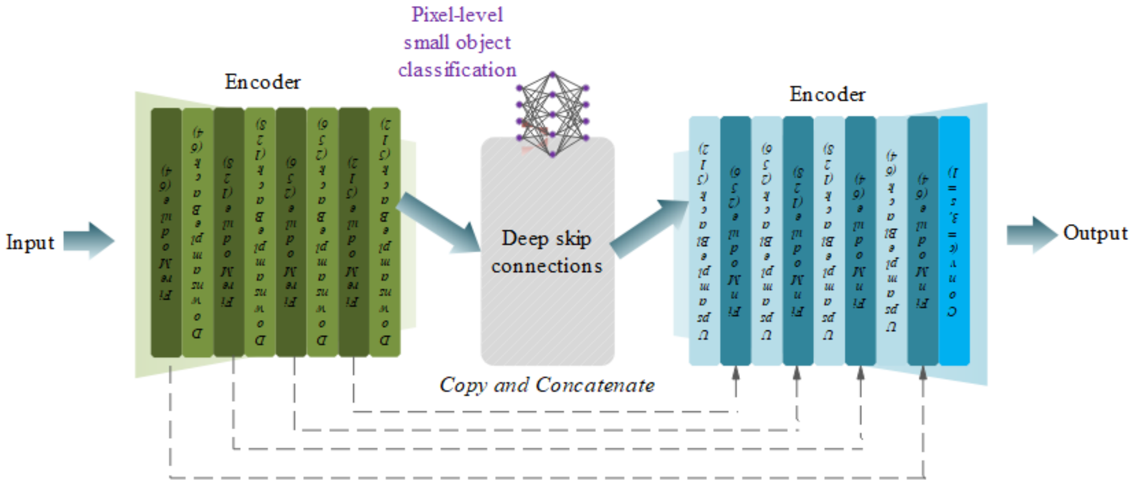


Figure 1. Proposed model architecture

In the quest for designing a lightweight image segmentation model tailored for high-dimensional data streams, appropriate parameter configurations have been identified as pivotal for achieving optimal performance. Within the encoder, multiple layers serve to extract distinguishing features from the initial image. This total layer count is denoted by B. Characteristically, each layer within the encoder undergoes a downsampling operation, aimed at both image dimension reduction and coarser feature extraction. Such downsampling levels are represented as *u.* A specific encoder layer, *i*, is symbolized as $Z^{u}{}_{Rb}$, and encompasses operations such as convolution, activation, and downsampling. These operations are crucial for distilling higher semantic features either directly from the input or from the preceding layer, (*i*-1). On the contrary, the decoder's i^{th} layer, expressed as $Z^{u}{}_{Fr,}$ typically integrates upsampling, convolution, and activation. A schematic representation delineating the model's construction principles is provided in Figure 2.

Assuming a feature aggregation mechanism consisting of two convolutions, batch normalization, and *ReLU* activation function is represented by $\Phi(\cdot)$, convolution operations are represented by $\Pi(\cdot)$. Upsampling operations are indicated by $I(\cdot)$, while fusion operations are represented by $[\cdot]$. The expression for $Z^{u}{}_{Fr}$ is given as:

$$Z_{Dr}^{u}=\begin{cases}\Phi\left(\left[\Pi\left(Z_{Rb}^{u}\right),\Pi\left(I\left(Z_{Rb}^{j}\right)\right)_{u+1}^{B-1},\Pi\left(I\left(Z_{Fr}^{u+1}\right)\right)\right]\right),\\ u=1,\dots,B-2\\ \Phi\left(\left[\Pi\left(Z_{Rb}^{u}\right),\Pi\left(I\left(Z_{Rb}^{u+1}\right)\right)\right]\right),\\ u=B-1\end{cases} \tag{1}$$

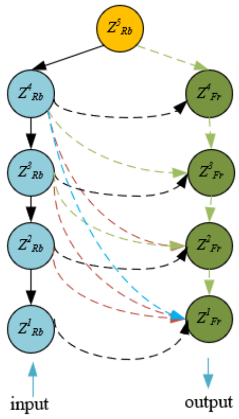


Figure 2. Schematic representation of model construction

Within this design, if an encoder is composed of five layers, the feature map corresponding to the decoder's second layer is designated as Z^2_{Fr}, while encoder feature maps for layers two, three, and four are represented by $[Z^2_{Rb},Z^3_{Rb},Z^4_{Rb}]$. The third decoder layer's feature map is described as Z^3_{Fr}. Notably, as one traverses through each layer, there is a discernible decrease in the image's spatial dimension, while the feature dimension experiences a proportional increase. As a result, Z^2_{Fr} is an amalgamation of $[Z^2_{Rb},Z^3_{Rb},Z^4_{Rb}]$ and Z^3_{Fr}.

It has been observed that the tailored lightweight image segmentation model for high-dimensional data flows offers distinct advantages in terms of model efficiency. Through the introduction of depth skip connections, the model can more effectively harness feature extraction capabilities intrinsic to deep encoder layers. This approach circumvents the need for handling intricate operations arising from full-scale feature maps of both shallow encoders and deep decoders, subsequently curtailing computational demands. By methodically applying downsampling and upsampling operations across encoder and decoder layers, it is possible to concurrently diminish spatial feature dimensions while augmenting their depth, achieving a balance between feature compression and essential information extraction. Such refinements not only curtail computational resource requirements but also amplify the model's generalization potential. Typically, lightweight designs correspond to a reduction in parameters and computational intricacies. By astutely configuring parameters like convolution kernel size and stride, substantial reductions in both model scale and computational overhead can be realized.

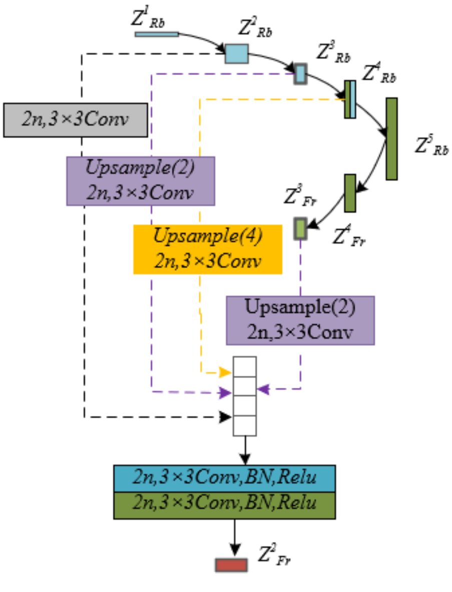


Figure 3. Composition principle of the aggregated feature map in the developed model

Figure 3 illustrates the composition principle of the aggregated feature map of the constructed model. Within the tailored lightweight image segmentation framework optimized for high-dimensional data flow, both Z^u_{Rb} and Z^u_{Fr} were identified to contain $2^{u-1}b$ channels. In the model's decoder, specifically at the u^{th} layer, B-u+3 convolutional operations were integrated. At any given i^{th} layer of the decoder, features were discernibly sourced not only from the immediate decoder layer but also from deeper encoder and decoder strata. Furthermore, in this decoder layer, convolution operations were deployed for an array of functions: extraction of features from the superior decoder layer, amalgamation with the analogously positioned encoder layer, adjustments in the channel count, and resolution alignment. A meticulous procedure was followed: each decoder layer was brought to align with the resolution and channel count of its predecessor using convolution operations. This was followed by an integration with features from the matched encoder layer, and subsequently, the channel number was adjusted again via convolutional methods. Such a procedural design ensured the intricate transfer and fusion of information within the decoder layer-by-layer.

Beyond the realm of convolutional operations, the significance of upsampling in the context of feature fusion was emphasized. It was through this mechanism that features derived from variegated layers could be amalgamated at a unified resolution. By leveraging channel accumulation, features from diverse origins were effectively integrated within the decoder. The subsequent convolutional process seamlessly blended these features, altering the channel count to fit the needs of the ensuing layers. It was observed that these convolutional procedures within the decoder bore an intrinsic connection to feature extraction processes within the encoder, culminating in a holistic bidirectional flow of data.

Given a scenario where the convolution kernel's size is denoted by j_r and the node's channel count is symbolized by $f(\cdot)$, the parameter count, O^u, for the encoder's u^{th} layer in the model can be deduced using a prescribed formula.

$$\begin{aligned} O^u &= j_r^2\left[\begin{array}{l} f\left(\sum_{j=i}^{B-1} Z^j_{Rb} + X^{u+1}_{FrX}\right) f\left(Z^u_{Fr}\right) \\ + f\left(\left(B-u+1\right) Zu^i_{Fr}\right) f\left(Z^u_{Fr}\right) + f\left(Z^u_{Fr}\right)^2 \end{array}\right] \\ &= \left(\sum_{j=u}^{B-1} 2^{j-u} + B - u + 4\right) 2^{2u-2} b^2 j_r^2 \end{aligned} \tag{2}$$

For a more nuanced exploitation of segmentation capabilities within the lightweight image segmentation model tailored for high-dimensional data streams, the encoder's structure was derived from the foundations of both $Vgg16$ and $ResNet34$. Assuming the input channel number is denoted by b_0, and the output class number, which corresponds to segmentation outcomes, is symbolized by b_v, the parameter count, O_c, relevant to the $Vgg16$-based model, can be articulated as:

$$\begin{aligned} O_c = &\left[b_0 f\left(Z^1_{Rb}\right) + f\left(Z^1_{Rb}\right)^2\right] j_r^2 + \sum_{u=1}^{4} O^u + b_c f\left(Z^1_{Rb}\right) \\ &+ \sum_{u=2}^{5}\left[f\left(Z^{u-1}_{Rb}\right) f\left(Z^1_{Rb}\right) + f\left(Z^1_{Rb}\right)^2\right] j_r^2 \end{aligned} \tag{3}$$

Within the ResNet framework, the count of Basicblock units in the u^{th} layer was hypothesized to be q_u. Thus, the parameter count, o_E, pertinent to the model resting upon ResNet34, can be deciphered through the ensuing equation:

$$\begin{aligned} O_e = &\, b_0 f\left(Z^1_{Rb}\right) j_r^2 \\ &+ \sum_{u=2}^{B}\left[f\left(Z^{u-1}_{Rb}\right) f\left(Z^1_{Rb}\right)\right] + \sum_{u=1}^{4} O^U + b_c f\left(Z^1_{Fr}\right) \\ &+ \sum_{u=2}^{B}\left[f\left(Z^{u-1}_{Rb}\right) f\left(Z^1_{Rb}\right) + \left(2q_u - 1\right) f\left(Z^1_{Rb}\right)^2\right] j_r^2 \end{aligned} \tag{4}$$

In the formulation of the lightweight image segmentation model for high-dimensional data streams, the determination of the initial encoder layer's output channel number emerged as a cornerstone. This choice delicately balanced the objectives of truncating computational intricacy and the volume of parameters against preserving a potent feature extraction capability. An optimal channel count is inherently reliant upon the intricacy of the input image coupled with the nature of features sought for extraction. For instances where targets are imbued with a plethora of textures and minutiae, an amplified channel count may be requisite to ensnare such data. It was discerned that a gradual amplification of the channel count, coinciding with the encoder's depth augmentation, affords the model the versatility to assimilate various feature hierarchies across its layers. By opting for fewer channels in the inaugural layer, the acquisition of more overarching features was facilitated, thereby curtailing computational requisites. Reverence was given to lithe neural network architectures like MobileNet, which have vindicated their mettle across a myriad of applications. Drawing insights from their preliminary encoder layer's channel count, it was aligned at 8, and concurrently, the convolution kernel size was ascertained to be 3.

3. DESIGN OF LOSS FUNCTIONS FOR HIGH-DIMENSIONAL TARGET SEGMENTATION

In the domains of computer vision and image processing, the significance of segmentation for minute targets within high-dimensional data streams has been recognized as a cornerstone technology. This form of segmentation is indispensable for discerning and identifying diminutive and nuanced entities within images. A gamut of applications, spanning medical image diagnostics, military reconnaissance, and precision agriculture, underscore its profound implications. With the amplification of data dimensions, an upsurge in the quantum of information within these high-dimensional data streams is concurrently observed. For gleaning invaluable insights and data facets, it becomes imperative that small targets within such data-rich images are segmented with precision, laying the groundwork for more sophisticated and nuanced data interpretations.

Yet, these high-dimensional data streams are often imbued with complex configurations and a profusion of information. The onus of extracting and processing such multitudinous features rests on exacting models and adept algorithms, thereby heralding challenges in computational and storage aspects. Furthermore, the heterogeneity in these high-dimensional streams-manifested in variegated scales and modalities-warrants state-of-the-art fusion methodologies and malleable model constructs. A confluence of exigencies arises when, for specific utilities such as autonomous navigation and drone reconnaissance, there's an imperative to achieve small target segmentation either in real-time or with minimal latency. Navigating this labyrinth within the confines of high-dimensional data streams emerges as a Herculean task.

To navigate the complexities inherent in segmentation exercises within these voluminous data streams, an amalgamation of *BCE* and *Dice* losses was adopted to penalize the network. Conventional loss functions were discerned to be suboptimal for diminutive target segmentation tasks. Whereas *BCE* loss is tailored to bolster pixel-level classification fidelity, *Dice* loss gravitates towards the holistic segmentation consequence. The synergy of both is perceived to fortify the loss function's robustness, capacitating the model to adapt and excel across a spectrum of scenarios. Concurrently, the confluence of *BCE* and *Dice* losses affords a judicious equilibrium between positive and negative samples, adeptly circumventing pitfalls associated with class disproportionality.

BCE, predominantly tailored for binary classification paradigms, has been harnessed for pixel-level image segmentation endeavours, categorizing each discrete pixel. By quantifying the variance between the likelihood of each pixel affiliating with the target class and its actual label—and then endeavouring to curtail this disparity—the nuanced recognition of segmentation targets is realized. Especially in the milieu of minuscule target segmentation within such data-rich streams, *BCE* facilitates a nuanced calibration between the weightages of positive and negative classes, steering the model's focus towards the less prevalent, yet pivotal, diminutive target categories. Assuming the sample count is articulated by *B*, the *softmax* output and *groundtruth* labels for class v are symbolized by $log(o_{u,v})\in[0,1]$ and $t_{u,}\ v$ respectively. With image taxonomies represented by *V* and the weighting metric denoted by α_y, the intricacies of the *BCE* loss function are elaborated in the subsequent formulation:

$$LOSS_{BCE}=\frac{1}{B}\sum_{u}-\alpha_y\cdot\sum_{v=1}^{V}t_{u,v}\log\left(o_{u,v}\right) \tag{5}$$

For the calculation of the weightage factor, α_y, the following formula is proposed:

$$\alpha_y=\begin{cases}\alpha & IF(t=1)\\ 1-\alpha & others\end{cases} \tag{6}$$

Originating from the *Dice* coefficient, *Dice* loss functions as an instrumental metric, probing the resemblance between two sample sets. Within the scope of image segmentation tasks, the *Dice* loss is employed to measure the congruence of spatial overlap between the predicted segmented regions and their true counterparts. By optimising the *Dice* coefficient-essentially minimizing the *Dice* loss-it has been observed that the segmentation boundaries inferred by the model closely mirror the authentic boundaries. Particularly for circumstances grappling with class imbalances, *Dice* loss has demonstrated pronounced efficacy, proving instrumental for lightweight small target segmentations within high-dimensional data flows. Its focus on spatial overlap of segmented regions attenuates the ramifications of disparities between positive and negative sample distributions, fortifying the model's adeptness at pinpointing diminutive targets. If the weight attributed to class *v* is delineated as Q_v, the intricacies of the *Dice* loss function can be unravelled in the following equation:

$$LOSS_{Dice}=1-\sum_{v=1}^{V}\frac{2Q_v\cdot\sum_{u=1}^{B}hy\left(v,u\right)\log o\left(v,u\right)}{\sum_{u=1}^{B}\left(hy\left(v,u\right)^2+\log o\left(v,u\right)^2\right)} \tag{7}$$

An amalgamation of *BCE* and *Dice* losses has been chosen for this endeavour. While the former predominantly targets classification accuracy, the latter underscores the spatial uniformity within segmented territories. Together, they bestow the model with a harmonized blend of target identification prowess and boundary delineation precision. Challenges such as class imbalances and intricate feature constructs, which persistently plague small target segmentation within high-dimensional data conduits, are aptly addressed by the tandem of *BCE* and *Dice* losses. Their symbiotic relationship ensures both a heightened sensitivity to small target identification and an unparalleled spatial acuity within segmented realms, leading to impeccable segmentation outcomes. By calibrating the relative emphases of these losses in the overarching objective function, and by making requisite adjustments contingent upon the unique nature of tasks and dataset attributes, an enhanced optimization of segmentation outcomes is anticipated. The loss function tailored for the devised model, where β symbolizes the proportionality coefficient, is elaborated upon in the succeeding equation:

$$LOSS_{SE} = \beta * LOSS_{BCE} + LOSS_{Dice} \quad (8)$$

4. EXPERIMENTAL FINDINGS AND INTERPRETATION

Ablation studies are conventionally employed as pivotal experimental paradigms, facilitating nuanced insights into the influence of individual elements on overarching model efficacy. By executing systematic removal or substitution of these constituents, their specific contributions can be discerned. Pertaining to the lightweight image segmentation model tailored for high-dimensional data streams discussed in this investigation, several comparative network architectures were devised:

(1) A foundational model harnessing the core *U-Net* framework, devoid of any lightweight adaptations.

(2) A modified *U-Net* structure incorporating solely the deep skip connections, with the *BCE* and *Dice* losses being conspicuously absent.

(3) A *U-Net* derivative, where only the *BCE* and *Dice* losses were integrated, eschewing the employment of deep skip connections.

(4) A holistic model amalgamating all the aforestated modifications, notably the synchronous integration of deep skip connections alongside the *BCE* and *Dice* losses.

For the segmentation endeavours within this high-dimensional lightweight image schema, the following objects of interest were delineated:

(1) Microbiological entities, encompassing bacteria and viral agents.

(2) Inconspicuous metallic denominations or diminutive artifacts juxtaposed against multifaceted backdrops.

(3) Subtle oncological or pathological demarcations discerned within medical imaging spectra.

(4) Miniscule aquatic entities or artefacts, discernible within deep marine imagery.

(5) Elusive pedestrian figures or compact vehicular entities embedded within convoluted urban terrains.

Table 1. Detailed empirical outcomes derived from the ablation assessment across the spectrum of formulated models

Network Model	Microorganisms	Coins	Minute Pathological Areas	Pedestrians	Cars	Underwater Objects	*MIoU* (%)	*PA* (%)	*F1* (%)	*Time* (%)
Baseline Model (*U-Net*)	65.24	52.36	64.21	24.36	52.11	41.22	51.23	66.39	66.33	0.18
Lightweight *U-Net* (Depthwise Skip Connections Only)	78.26	65.44	72.16	38.99	51.23	47.26	57.99	74.16	72.13	0.24
Lightweight *U-Net* (Only *BCE* and *Dice* Loss)	75.36	65.39	75.36	55.21	65.88	42.99	63.21	83.58	78.99	0.25
Complete Model	82.66	72.36	81.22	63.25	70.13	58.13	70.13	82.03	82.15	0.27

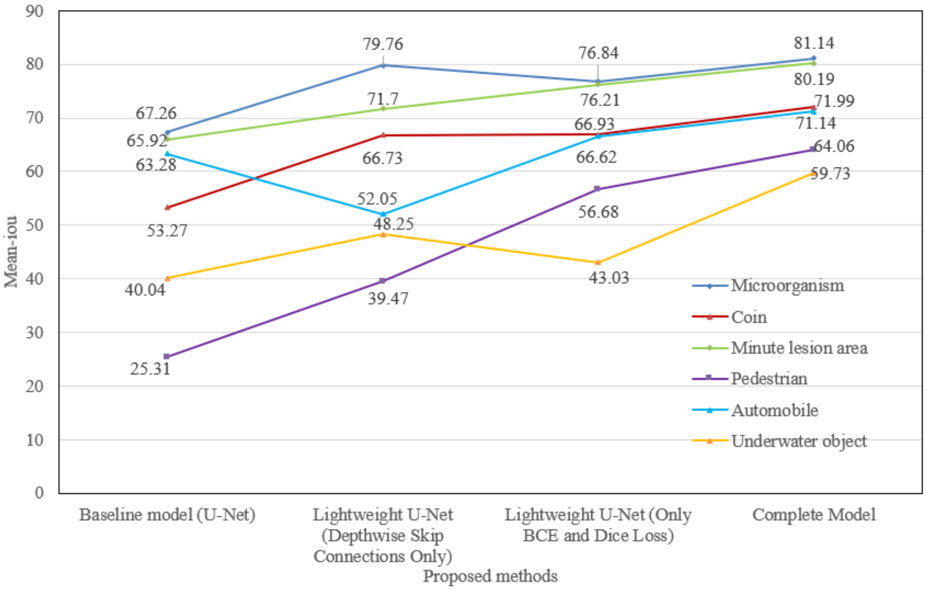


Figure 4. Mean *IoU* values of propounded models against validation set across diverse categories

Table 2. An analytical comparison of experimental outcomes across diverse model architectures

Network Model	Microorganisms	Coins	Minute Pathological Areas	Pedestrians	Cars	Underwater Objects	*MIoU* (%)	Parameter Volume (*M*)	Time (*s*)
FCN	82.31	64.88	76.31	53.26	34.58	42.03	58.62	6.32*M*	0.99
Mask R-CNN	75.23	63.99	74.26	61.44	57.69	36.25	62.02	3.15*M*	0.45
H-DenseUNet	74.28	67.81	78.36	55.36	71.02	47.26	65.32	8.26*M*	1.23
PSPNet	82.66	81.23	78.12	43.26	56.32	66.23	68.92	1.89*M*	1.36
The model of this study	81.36	72.15	81.55	63.25	72.15	58.91	72.13	1.32*M*	0.29

Table 3. A quantitative evaluation: Precision, recall, and F-score across network models

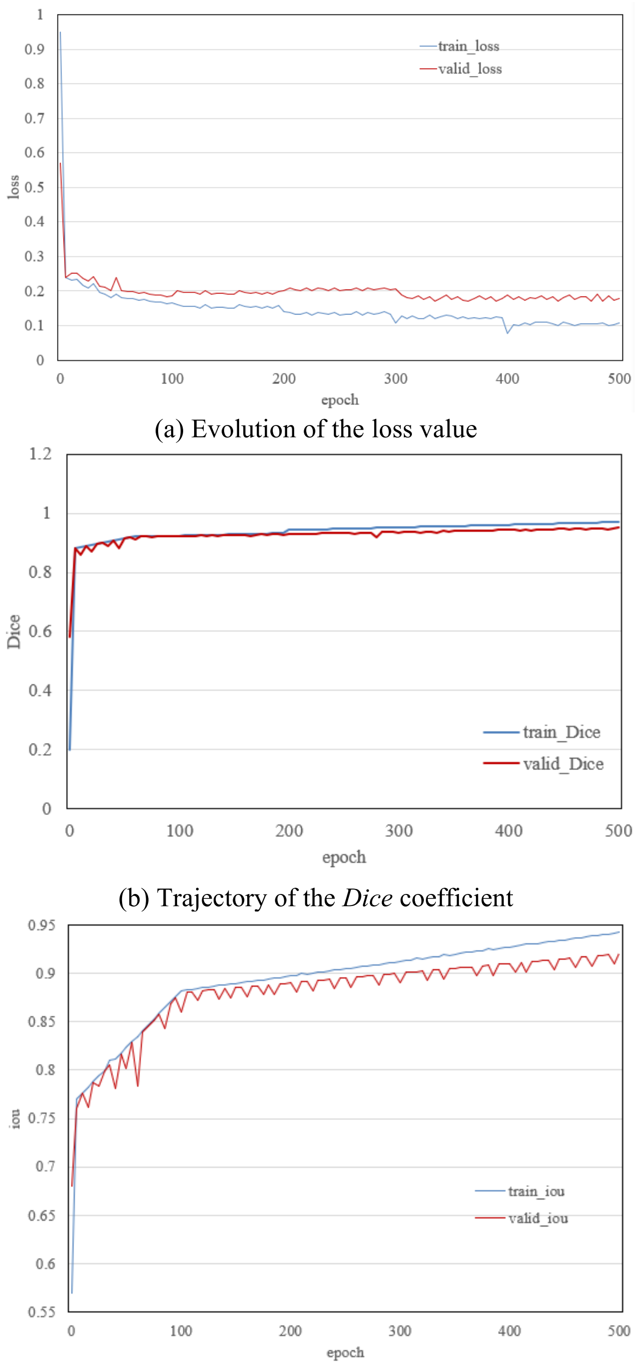
Network Model	*P* (%)	*R* (%)	*F*1 (%)
FCN	81.26	61.03	71.26
Mask R-CNN	73.25	73.21	73.20
H-DenseUNet	76.13	74.56	74.27
PSPNet	83.25	85.26	81.02
The model of this study	88.16	88.95	90.55

In Table 1, a thorough quantitative elucidation of the ablation study's outcomes for the array of developed models is delineated. An analysis of these outcomes suggests that the baseline model, rooted in the *U-Net* architecture, manifested moderate performance metrics across all experimental tasks, culminating in a diminished overall efficiency. Through the integration of deep skip connections, a tangible improvement in the performance metrics of the lightweight *U-Net* was observed, with pronounced elevation noted particularly within microorganism, coin, and small pathological area segmentations. A distinct model variant of the lightweight *U-Net*, solely capitalizing on the *BCE* and *Dice* loss functions, showcased augmented efficiency. This enhancement was conspicuously evident in minute pathological detection and pedestrian categorizations. Consistent superior performance metrics across all investigative tasks were exhibited by the comprehensive model, integrating both deep skip connections and the combined *BCE* and *Dice* loss methodologies. It was inferred that the introduction of deep skip connections boosted the lightweight *U-Net*'s efficacy, a trend particularly discernible within the realm of microorganism and petite pathological segmentation. Additionally, the concomitant implementation of *BCE* and *Dice* losses augmented the capability matrix for small target segmentation tasks. Notable prowess in this context was exhibited within minute pathological and pedestrian detection parameters. Despite its augmented performance metrics, it must be noted that the refined model introduced in this research, which amalgamates both deep skip connections and loss functions, did register a marginal uptick in computational overhead.

Figure 4 delineates the average Intersection over Union (*IoU*) metrics of the formulated models across varied diminutive object categories subjected to high-dimensional data flows. An analytical interpretation of this data posits the baseline *U-Net* model as a rudimentary benchmark with general performance metrics. Enhanced outcomes, particularly for microorganisms, coins, and aquatic entities, were registered for the Lightweight *U-Net* augmented with depthwise skip connections. However, a perceptible dip was noted within the vehicular domain. Meanwhile, the Lightweight *U-Net* variant, relying exclusively on *BCE* and *Dice* loss functions, manifested remarkable prowess within the realms of petite lesion areas and pedestrian categorizations. Such results corroborate the hypothesis that tailored loss functions enhance model performance for specific tasks. The comprehensive model, with its all-encompassing architecture, achieved near-optimal or optimal metrics across all delineated categories, underscoring its robust and versatile nature. These empirical findings accentuate the merit of integrating depthwise skip connections and tailored loss functions within the comprehensive model. This model exhibited exemplary performance attributes across all testing categories, underscoring the proposed methodology's reliability and efficacy in lightweight image segmentation tasks for diminutive entities within high-dimensional data contexts.

In Table 2, a comparative evaluation is conducted across five distinct network models. Metrics such as average Intersection over Union (*IoU*), parameter volume, and computational time were meticulously assessed across six lightweight image small-object categories. It was observed that FCN and Mask R-CNN, despite showcasing commendable average *IoU* values, were encumbered by substantial parameter volumes. Moreover, FCN was particularly characterized by an extended computational duration. In the automobile category, H-DenseUNet's performance was notably superior, yet when assessing overall average *IoU* and parameter volume, its distinctions remained unremarkable. While PSPNet demonstrated appreciable results for coins and underwater objects, a lapse in efficacy was detected in the pedestrian category. Conversely, the model introduced in the present study emerged as an archetype of comprehensive performance. It was found to be endowed with superior average *IoU* scores and boasted notable advantages in terms of lightweight architecture and computational expediency. The ability to simultaneously maintain heightened accuracy and ensure computational efficiency is essential, especially when considering the demands of lightweight image segmentation for petite entities under high-dimensional data environments.

When delving into Table 3, the precision (P), recall (R), and F-score metrics were meticulously examined for five disparate network models. From this analysis, it was deduced that the model articulated in the current study consistently outstripped its counterparts concerning precision, recall, and the F1 score. Such elevated metrics emphasize the heightened accuracy and reliability inherent in this model for lightweight image segmentation challenges presented by high-dimensional data streams. Although PSPNet's metrics were laudable, they were overshadowed by the superior scores of the model described herein. FCN's performance, especially concerning recall, was found lacking. The unparalleled efficacy of the model delineated in this study can potentially be traced back to its architectural decisions, including but not limited to its choice of loss functions and feature fusion strategy. Across all evaluated metrics, the model consistently showcased exemplary results, thereby reiterating its profound efficacy and dominance in the realm of lightweight image segmentation tasks, particularly those involving minuscule entities under high-dimensional data landscapes.



(a) Evolution of the loss value

(b) Trajectory of the *Dice* coefficient

(c) Progression of Intersection over Union (*IoU*)

Figure 5. Training and validation metrics for the lightweight image segmentation model

Within Figure 5a, the encountered training loss (train_loss) and validation loss (valid_loss) throughout the training epoch of the lightweight image segmentation model are depicted. From this representation, it is observed that a consistent decrease in loss values is manifested over the training phase, indicating a stable training process. Notably, no significant overfitting was detected. An alignment in the decline of validation loss with that of the training loss was recorded, which can be interpreted as an attestation of the model's robust generalization capabilities. Such stable convergence behavior, observed during both training and validation stages, suggests the potential adoption of an optimal training strategy and hyperparameter configurations.

The *Dice* Coefficient, a pivotal metric in the realm of image segmentation tasks, typically oscillates between 0 (indicating absolute mismatch) and 1 (representing flawless alignment). As depicted in Figure 5b, the training *Dice* coefficient (train_*Dice*) and its validation counterpart (valid_*Dice*) were evaluated throughout the model's training trajectory. Superior learning capabilities were exhibited by the lightweight image segmentation model, as evidenced by a consistent ascent in both training and validation *Dice* coefficients with the progression of epochs. Despite occasional oscillations in the validation *Dice* coefficient, the predominant trajectory was upwards, bolstering confidence in the model's adeptness at predicting on unobserved data.

In Figure 5c, an insightful portrayal of the training *IoU* (train_iou) juxtaposed against the validation *IoU* (valid_iou) is presented. Both metrics were found to surge progressively as the epochs unfolded, underscoring the model's commendable learning dynamics during the training regimen. Parallels in performance were detected between training and validation datasets, hinting at a well-balanced generalization capability devoid of conspicuous overfitting phenomena. While sporadic fluctuations in the validation *IoU* were noted, the overarching pattern remained ascendant, reinforcing the model's competence in handling novel datasets.

5. CONCLUSION

In the realm of this research, a focus was placed on the design and meticulous evaluation of a lightweight image segmentation model, with aspirations to yield an efficient and precise segmentation methodology suitable for a gamut of tasks. Prevalent image segmentation networks such as FCN, Mask R-CNN, H-DenseUNet, PSPNet, among others, were subjected to comprehensive analysis. From this scrutiny, a novel lightweight image segmentation model was devised. It was observed that the proposed model outperformed contemporaneous models in terms of Precision, Recall, and F1 scores, all while maintaining impressive efficiency. Fluctuations in metrics such as loss values, *Dice* coefficients, and *IoU* offered insights into the model's learning trajectory. Consistency was discerned in training and validation dynamics, with notable convergence and an absence of overfitting phenomena. Further experimental evidence suggested that the model, as introduced, necessitates fewer parameters and operates with reduced runtime, emphasizing its inherent lightweight attributes.

Taking into account the collective experimental findings, the proposed lightweight image segmentation model revealed exceptional proficiency in an array of image segmentation challenges. In comparison with the prevailing image segmentation paradigms, the model did not solely stand out in accuracy metrics but also in terms of reduced computational footprint and enhanced operational efficiency. A marked stability in the training process, coupled with formidable generalization capabilities, underscores its potential merit in real-world applications. Such results bolster the contention that embracing lightweight designs in image segmentation does not necessitate a compromise in performance, suggesting that even with scaled-down model complexity and computational demands, results akin to, or even surpassing, more intricate models are attainable.

To encapsulate, the contributions of this research extend a novel, efficient, and streamlined solution to the discipline of image segmentation. Its implications appear to be most pronounced in environments with limited computational resources, emphasizing its potential as an instrumental tool in advancing the broader field.

REFERENCES

[1] Tan, H.Y., Gu, D.H. (2022). Application of medical image segmentation algorithm based on genetic algorithm in intelligent medical nursing system. In International Conference on Applications and Techniques in Cyber Intelligence, pp. 568-575.

[2] Ru, K.L., Li, X.L. (2023). A dermoscopic image segmentation algorithm based on U-shaped architecture. In 2023 2nd International Conference on Big Data, Information and Computer Network (BDICN), Xishuangbanna, China, pp. 1-5. https://doi.org/10.1109/BDICN58493.2023.00044

[3] Bennai, M.T., Guessoum, Z., Mazouzi, S., Cormier, S., Mezghiche, M. (2023). Multi-agent medical image segmentation: A survey. Computer Methods and Programs in Biomedicine, 232: 107444. https://doi.org/10.1016/j.cmpb.2023.107444

[4] Deng, J.L., Gong, H.G., Liu, M. (2023). Medical image segmentation based on object detection. Journal of the University of Electronic Science and Technology of China, 52(2): 254-262. https://doi.org/10.12178/1001-0548.2022081

[5] Liu, X., Liu, P., Wang, J., Wang, Q., Guo, Q., Tang, R. (2023). UT-MT: A semi-supervised model of fusion transformer for 3D medical image segmentation. In 2023 8th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA), Chengdu, China, pp. 190-196. https://doi.org/10.1109/ICCCBDA56900.2023.10154641

[6] Han, C., Huang, X., Zhang, Y., Wang, M. (2023). ECU-Net: Multi-scale salient boundary detection and contrast feature enhancement U-Net for breast ultrasound image segmentation. Signal, Image and Video Processing, 17(5): 2287-2295. https://doi.org/10.1007/s11760-022-02445-3

[7] Jiang, Z., He, Y., Ye, S., Shao, P., Zhu, X., Xu, Y., Yang, G. (2023). O2M-UDA: Unsupervised dynamic domain adaptation for one-to-multiple medical image segmentation. Knowledge-Based Systems, 265: 110378. https://doi.org/10.1016/j.knosys.2023.110378

[8] Xin, Y., Zhou, Z., Xia, Y. (2023). Scene separation & data selection: Temporal segmentation algorithm for real-time video stream analysis. arXiv preprint arXiv: 2308.00210. https://doi.org/10.48550/arXiv.2308.00210

[9] Van-Ha, T.D., Thanh-An, N. (2022). 3D-FaultSeg-UNet: 3D fault segmentation in seismic data using Bi-stream U-Net. In International Conference on Future Data and Security Engineering, Ho Chi Minh City, Vietnam, pp. 477-488. https://doi.org/10.1007/978-981-19-8069-5_32

[10] Shu, Y., Zhang, J., Xiao, B., Li, W. (2021). Medical image segmentation based on active fusion-transduction of multi-stream features. Knowledge-Based Systems, 220: 106950. https://doi.org/10.1016/j.knosys.2021.106950

[11] Kalapos, A., Gyires-Tóth, B. (2022). Self-supervised pretraining for 2D medical image segmentation. In European Conference on Computer Vision, Tel Aviv, Israel, pp. 472-484. https://doi.org/10.1007/978-3-031-25082-8_31

[12] Pirabaharan, R., Khan, N. (2022). Improving interactive segmentation using a novel weighted loss function with an adaptive click size and two-stream fusion. In 2022 IEEE Eighth International Conference on Multimedia Big Data (BigMM), Naples, Italy, pp. 7-12. https://doi.org/10.1109/BigMM55396.2022.00009

[13] Chen, S., Wei, D., Gu, S., Yang, Z. (2023). Blood flow characterization in nailfold capillary using optical flow-assisted two-stream network and spatial-temporal image. Biomedical Physics & Engineering Express, 9(4): 045023. https://doi.org/10.1088/2057-1976/acdb7c

[14] Ye, Z., Zhang, W. (2023). A dynamic few-shot learning framework for medical image stream mining based on self-training. EURASIP Journal on Advances in Signal Processing, 2023(1): 49. https://doi.org/10.1186/s13634-023-00999-z

[15] Shu, Y., Zhang, J., Xiao, B., Luan, X., Liu, L., Hu, C. (2020). Aft-net: Active fusion-transduction for multi-stream medical image segmentation. In 2020 IEEE 32nd International Conference on Tools with Artificial Intelligence (ICTAI), Baltimore, MD, USA, pp. 753-760. https://doi.org/10.1109/ICTAI50040.2020.00120

[16] Kang, M.S., Park, R.H., Park, H.M. (2022). Efficient two-stream network for online video action segmentation. IEEE Access, 10: 90635-90646. https://doi.org/10.1109/ACCESS.2022.3201208

[17] Liu, G., Ding, W., Shu, J., Strauss, A., Duan, Y. (2023). Two-stream boundary-aware neural network for concrete crack segmentation and quantification. Structural Control and Health Monitoring, 2023: 3301106. https://doi.org/10.1155/2023/3301106

[18] Elghazy, H.L., Fakhr, M.W. (2021). Multi-modal multi-stream unet model for liver segmentation. In 2021 IEEE World AI IoT Congress (AIIoT), Seattle, WA, USA, pp. 28-33. https://doi.org/10.1109/AIIoT52608.2021.9454216

[19] Sun, L., Zhang, Y., Ge, H., Liu, T., Zhao, Y., Sun, J., Qi, X. (2022). TSINet: Efficient breast lesion segmentation with three-stream interactive neural network on magnetic resonance images. In 2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Las Vegas, NV, USA, pp. 813-816. https://doi.org/10.1109/BIBM55620.2022.9995065

[20] Chen, G., Li, L., Zhang, J., Dai, Y. (2023). Rethinking the unpretentious U-net for medical ultrasound image segmentation. Pattern Recognition, 142: 109728. https://doi.org/10.1016/j.patcog.2023.109728

[21] Wang, J., Lv, W., Wang, Z., Zhang, X., Jiang, M., Gao, J., Chen, S. (2023). Keyframe image processing of semantic 3D point clouds based on deep learning. Frontiers in Neurorobotics, 16: 988024. https://doi.org/10.3389/fnbot.2022.988024