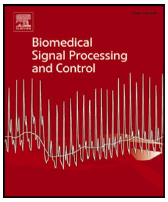
Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc



An EffcientNet-encoder U-Net Joint Residual Refinement Module with Tversky–Kahneman Baroni–Urbani–Buser loss for biomedical image Segmentation

Do-Hai-Ninh Nham [a], Minh-Nhat Trinh [b], Viet-Dung Nguyen [a], Van-Truong Pham [b,*], Thi-Thao Tran [b,*]

[a] *School of Applied Mathematics and Informatics, Hanoi University of Science and Technology, Viet Nam*

[b] *Department of Automation Engineering, School of Electrical and Electronic Engineering, Hanoi University of Science and Technology, Viet Nam*

ARTICLE INFO

Keywords:
Biomedical Image Segmentation
EfficientNet
Residual Refinement Module
Baroni–Urbani–Buser coefficient
Tversky–Kahneman probability-weighting function

ABSTRACT

Quantitative analysis on biomedical images has been on increasing demand nowadays and for modern computer vision approaches. While recently advanced procedures have been enforced, there is still necessity in optimizing network architecture and loss functions. Inspired by the pretrained EfficientNet-B4 and the refinement module in boundary-aware problems, we propose a new two-stage network which is called EffcientNet-encoder U-Net Joint Residual Refinement Module and we create a novel loss function called the Tversky–Kahneman Baroni–Urbani–Buser loss function. The loss function is built on the basement of the Baroni–Urbani–Buser coefficient and the Jaccard–Tanimoto coefficient and reformulated in the Tversky–Kahneman probability-weighting function. We have evaluated our algorithm on the four popular datasets: the 2018 Data Science Bowl Cell Nucleus Segmentation dataset, the Brain Tumor LGG Segmentation dataset, the Skin Lesion ISIC 2018 dataset and the MRI cardiac ACDC dataset. Several comparisons have proved that our proposed approach is noticeably promising and some of the segmentation results provide new state-of-the-art results. The code is available at https://github.com/tswizzle141/An-EffcientNet-encoder-U-Net-Joint-Residual-Refinement-Module-with-TK-BUB-Loss.

1. Introduction

Important advances in computer vision algorithms have been adopted to various fields, including biomedical image analysis. The application of convolutional neural networks (CNNs) [1] for several tasks of detection, classification, segmentation has consequential utilization in the medical field, where laborious assignments could be replaced by automatic systems. Deep CNNs have evidenced increasing usage in biomedical fields, for instance, organ, nuclei, brain tumor and skin segmentation [2,3]; however, the current CNN-based approaches still have some shortcomings in accuracy and training time optimization.

Automatic nuclei segmentation of microscopy images is an urgent task due to the subjectivity of manual segmentation. Experimentally, nuclei segmentation approaches need to be instance-aware to appropriately detach adjoining nuclei, and, those approaches required to tackle various problems like high cell density, low contrast, intensity inhomogeneity, weak boundaries, strong gradients inside the nuclei… Conventionally, nuclei segmentation has been addressed by traditional computer vision approaches such as thresholding [4], filtering [5], shape-matching [6], active contours [7], … Those approaches though work well on numerous benchmarks but seem to deteriorate in some problematic instances. Besides, U-Net [8] and Mask R-CNN [9] have been the preeminent deep-learning architectures applied in nuclei segmentation. These two models have been combined into one architecture by Vuola et al. [10] in order to integrate the properties of segmentation and bounding-box prediction for better learning important information about the nucleus shapes. Attention mechanism has also been employed, for example, in [11], so that multi-scale features could be received from original input and the receptive field can be strengthened with multi-scale convolutions. Zhang et al. [12] have used the FCN model to execute coarse nuclei images segmentation; before integrating with GANs model with splitting branches in the discriminator structure to improve performance accuracy. Nevertheless, these methods seem to be cumbrous and they have not consider the importance of boundary refinement in small cell segmentation.

With regard to brain tumor analysis, recently, deep-learning methods on automatic brain segmentation have evolved. Improvement of

[*] Corresponding authors.
E-mail addresses: truong.phamvan@hust.edu.vn (V.-T. Pham), thao.tranthi@hust.edu.vn (T.-T. Tran).

https://doi.org/10.1016/j.bspc.2023.104631
Received 2 September 2022; Received in revised form 8 December 2022; Accepted 25 January 2023
Available online 4 February 2023

networks yielding satisfactory brain LGG segmentation could potentially admit for the tumor genomic identification automatization process through MRI that is cost-effective and liberating inter-reader variability [13]. Dong et al. [14] has put forward non-invasive magnetic resonance techniques as an identifying tool for brain tumor detection without the risk associated with ionizing radiation with the use of U-Net. Brosch and his colleagues [15] have adopted a fully convolutional network with skip connections for sclerosis lesion segmentation. A modified U-Net has been introduced by Isensee [16] for brain tumor segmentation with the use of thorough data augmentation to successfully avoid overfitting problem. The Cui et al. [17] model employs of two distinct FCN models and one of them needs only forward computation; thus improving in core tumor category. However, accurate and effective segmentation of tumors remains a problematic task on account of different occurrences in brain regions and shapes and sizes variety.

In skin disease diagnosis, automated melanoma segmentation is challenging because of huge variations artifacts like color calibrations, hole and shrink. To tackle this challenge, Sarker et al. [18] have displayed a skin segmentation architecture using negative log-likelihood and end-point-error loss functions to retain pertinent contours. Li and his colleagues [19] have built up a dense deconvolutional model utilizing hierarchical supervision for occupying locally and globally contextual features for skin segmentation. In addition, if vanilla U-Net has been popularly employed in biomedical image segmentation, various advancement of U-Net have been designed for better segmentation veracity; for instance, an extended form of the U-Net MCGU-Net [20] is proposed by combining BConvLSTM [21] in the skip connections and employing Squeeze-and-Excited module in the decoding path before re-using information with dense convolutions for higher resolution features. Nazi et al. has combined U-Net with a model of Deep Convolutional Neural Network integrating with Support Vector Machine [22] to produce a joint classification and segmentation network for skin problems diagnosis. Similarly, these networks are high on the parameters number without putting boundary optimization in consideration.

Deep learning-based methods have also been applied for MRI cardiac segmentation. Wang et al. citecsrnet have proposed regression component for segmenting the left ventricle more accurately though the cardiac structures are variable. Zhang et al. have proposed BLU-Net [23] and Compressed Dense Blocks for fewer connections between the input and the inner layers. Pure Dilated Residual U-Net (PDR U-Net) [24] has been also proposed to segment the femur and tibia bones from X-ray images automatically and correctly. Although these deep learning-based architectures above have confirmed their performance efficiency, there are still certain drawbacks that their architectures are fairly complex. Furthermore, some of them could not leverage the usage of pretrained models which help the network converge faster. As well, some architectures have not seriously considered on preserving the edges of features. Motivated by these weaknesses, we have proposed a Modified EffcientNet-encoder U-Net Joint Residual Refinement Module network to progressively encourage biomedical image segmentation result reliability.

Applying appropriate loss function helps further improving segmentation model competency. The Mean Squared Error (MSE) and Cross Entropy (CE) loss function have been widespreadly adopted for extracting features from specific regions. Though practical experiments indicate that these two loss functions could perform classification and segmentation task well, there are still valid weaknesses in highly-unbalanced class training, because of their assumption on identical importance of distribution of labels. Recently, there has been an rising enthusiasm in exploiting the active contour models as loss functions for training the neural networks. For example, in active contour models, Mumford–Shah functional [25], Active-Contour loss [26], level-set methods [27], and proximal methods [28] have produced undoubtable segmentation performances. If Mumford–Shah and Active-Contour loss function pay attention on edge-preserving filtering method, the Dice Loss function [29] effectuates the mathematical representation of segmented object; however because of its non-convexity, there could be degradation in attaining desirable results. Region-based Tversky [30] and Focal Tversky loss functions [31] tend to control the information flow implicitly through pixel-level affinity and solve class-imbalanced problem. Notwithstanding, their convergence speed is witnessed to be not good enough.

In terms of binary similarity coefficients, the Baroni–Urbani–Buser coefficient is famous for limiting the impact of negative matches. It significantly down-weights the quantity similar to the true-negative (TN) quantity relative to a in its numerator [32]. The Baroni–Urbani–Buser coefficient is frequently utilized in the problems of detecting fingerprints similarity [33], clustering methods in biological systematics [34], . . . To the best of our knowledge, the Baroni–Urbani–Buser coefficient has not been investigated in segmentation study. This motivates us to propose a new region-based TK-BUB (Tversky–Kahneman Baroni–Urbani–Buser) loss function, with the help of the Tversky–Kahneman probability weighting function [35], so that not only address class-imbalanced tissues, but also intensively promote the model convergence rate.

In this paper, our fundamental contributions are:

- Proposing a novel model called the Modified EffcientNet-encoder U-Net Joint Residual Refinement Module to improve the overall biomedical image segmentation performance.
- Creating a new loss function called the Tversky–Kahneman Baroni–Urbani–Buser (TK-BUB) loss function and some versions of this loss function, for better network convergence speed.
- Experimenting on four datasets for evidencing the effectiveness of our proposed architecture and loss function over other loss functions applying in different methods. Due to this respect, experiments are executed on the four popular datasets: the 2018 Data Science Bowl Cell Nucleus Segmentation dataset, the Brain Tumor LGG Segmentation dataset and the Skin Lesion ISIC 2018 dataset, without external data usage. The TK-BUB loss function is proved to express better performances in almost cases.

2. Related work

2.1. EfficientNet

Tan et al. [36] have introduced a novel network scaling approach, namely EfficientNet, that utilizes a primary but exceedingly efficient compound coefficient to scale up CNNs in a more structured manner. If traditional methods that whimsically scale model dimensions (width, depth and resolution), the novel algorithm consistently scales each element with a certain set of scaling coefficients. Denote three dimensions of an input is d, w, r as d standing for the depth; w standing for the width and r standing for the resolution. In [36], the Compound Scaling method could be presented as follows:

$$d = \alpha^{\phi}; \quad w = \beta^{\phi}; \quad r = \gamma^{\phi}; \quad \alpha \times \beta^2 \times \gamma^2 \approx 2 \qquad \alpha, \beta, \gamma \geq 1 \tag{1}$$

with α, β, γ are constants that represent how much to scale the individual dimensions by, and ϕ is a variable that represents how much additional computational resources.

In terms of FLOPs, the cost of convolution scales linearly with d, but quadratically with w and r, that is if the depth is doubled, the computation cost is doubled too, however when the width or resolution is doubled, the computation cost is quadrupled. Putting the math together, this means that scaling a convolutional network with the parameter ϕ results in a new network that has a computational cost of approximately 2^{ϕ}. In the original paper [36], the specific values the authors get from running experiments on the EfficientNet architecture are $\alpha = 1.2, \beta = 1.1, \gamma = 1.15$. This compound scaling approach is reported to uniformly increment the network accuracy for scaling up some networks, for instance, MobileNet (+1.4% ImageNet accuracy), and ResNet (+0.7% ImageNet accuracy), in comparison with traditional scaling approaches.

2.2. Normalization methods

Normalization has always been a demanding concern in the deep-learning area, as normalization techniques could undeniably decrease training time. These techniques help normalize features in order to preserve the feature contribution, which leads our model to be unbiased. As well, normalization reduces Internal Covariate Shift, which adjusts in the contribution of model activations because of the alternation in model parameters during training process. Batch Normalization [37] allows smoother loss surface due to the fact that it tightly encircles the gradients magnitude. It further incites the model with better optimization on account of the fact that weights are not allowed to explode and restricted to a valid range. Weight Normalization [38] splits the weight vector from its direction, which provides an identical impact as in batch normalization with variance. As for the mean, mean-only batch normalization and weight normalization are combined for gaining pertinent outcomes even in small mini-batches; which means they subtract out the mean of the mini-batch but do not divide by the variance. Different from Batch Normalization, Layer Normalization [39] normalizes input across the features. Instance Normalization [40] and Layer Normalization are the same to some extent, but the variation is that Instance Normalization normalizes across each channel in each training instance, contrasting to Layer Normalization, which normalizes across input features in an training instance. Different from Batch Normalization, the Instance Normalization layer is employed during the testing process, by reason of non-dependency of the mini-batch. This algorithm is initially conceived for style transfer, as Instance Normalization tries to overcome the hurdle of agnostic model to the original image contrast. Group Normalization [41] normalizes over channels group for each training instances; thus Group Normalization is to some extent in between Instance Normalization and Layer Normalization. Batch-Instance Normalization is an interpolation between Batch Normalization and Instance Normalization, as Instance Normalization might completely erase style information despite of its own merits, it could be noticeably problematic in the conditions where contrast matters.

2.3. Refinement module

Lin et al. [42] has tackle the deficiency of feature maps downscaling and expensive computation cost of ResNet and Atrous Convolution by proposing a refinement network. It contains a residual block which is used but with batch normalization removed; multi-resolution fusion for merging the feature maps using element-wise summation; and a Chained Residual Pooling module fusing output feature map together with the input feature map through summation of residual connections; in order to capture background context of an image. To produce truly photorealistic results, the Cascaded Refinement Network [43] has synthesized images by progressive refinement, and going up an octave in resolution amounts to adding single refinement modules. There are three feature layers in each module M_i: the input layer, an intermediate layer, and the output layer. The downsampled semantic layout c-channel L and a bilinearly upsampled feature layer d_{i-1}-channel F_{i-1} concatenate into the input layer. The interactive refinement architecture in [44] contradicts to others that four extreme points are extricated for each training samples to train the segmentation model to counter the weakness of not providing the challenging segmented area and requiring more interactive to complete segmentation task like traditional interactive methods.

2.4. Jaccard/Tanimoto coefficient

Jaccard/Tanimoto coefficient, which is the ratio of the intersection of two objects to their union, is one of the most primary and ubiquitous similarity measurement to compare biological presence-absence data. Due to its generality, the Jaccard/Tanimoto Coefficient is exploited to a variety of applications in binary data, for example, from genomics, biochemistry, and other areas of science [33]. Given two vectors y_i and y_j displaying two distinct objects, we have $T(y_i, y_j) = \frac{y_i \cap y_j}{y_i \cup y_j}$ where $T(y_i, y_j)$ is the Jaccard/Tanimoto similarity coefficient. Nevertheless, this coefficient sometimes lacks probabilistic interpretations or statistical error controls; so as to deal with this dispute, Chung et al. [33] have divided the coefficient calculation process into two cases: If y_i and y_j are independent, $T(y_i, y_j) = \frac{p_i \times p_j}{p_i + p_j - p_i p_j}$ with p_i, p_j are the corresponding occurrence (e.x. success or gain) probabilities, $p_i, p_j \in [0, 1]$; else $T(y_i, y_j) = \frac{y_i \cap y_j}{y_i \cup y_j}$.

3. Methodology

3.1. The proposed model

As U-Net [8] has been vastly employed on segmenting biomedical images, we design our model based on U-Net backbone to come in for its advantages. We pretrain the encoder with EfficientNet [36]; because it has caused waves in the deep-learning field with a new scaling method called Compound Scaling. If we scale the dimensions by a certain amount consistently at the same time, we can produce model with better outcome with EfficientNet pre-trained model.

In this study, we propose a network encoder with EfficientNet-B4, based on its high-accuracy when being experimented on ImageNet. The EfficientNet-B4 initial architecture is indicated in Fig. 1. As can be seen from Fig. 1a, there are totally 7 MBConv block modules, each module comprises of different number of MBConv blocks and each MBConv block contains 4 phases: the Expansion phase, the Depthwise Convolution phase, the Squeeze-and-Excitation phase and the Output phase. Between the Depthwise Convolution phase and the Squeeze-and-Excitation phase is the BatchNorm layer, as illustrated in Fig. 1b. To further connect some of the pre-trained EfficientNet-B4 layers to the decoder through the skip connections, four layers of the pre-trained EfficientNet-B4 encoder are selected, so as to accurately fit the size in the decoder corresponding layer. We have chosen the 342nd, 154th, 94th and 30th layers to attach to the skip connection. To be more detailed, the 342nd layer is the BatchNorm layer of the first MBConv block in the 6th MBConv block module (block6a_bn); the 154th layer is the BatchNorm layer of the first MBConv block in the 4th MBConv block module (block4a_bn); the 94th layer is the BatchNorm layer of the first MBConv block in the 3rd MBConv block module (block3a_bn) and finally, the 30th layer is the BatchNorm layer of the first MBConv block in the 2nd MBConv block module (block2a_bn). These skip connections between pretrained encoder and decoder help preserve the spatial information in the encoder features. Although there are several ways to choose the origin of the skip-connections in the pretrained encoder, and our choice of layers is a new choice, we have considered choosing specific layers from the EfficientNetB4 for some reasons. Firstly, we focus on matching the height and width resolution of the feature maps in the encoder layer and the corresponding layer in the decoder to perform concatenating operations. Secondly, we notify the channel value that after deconvolutional operations, the difference between the input channel and the output channel is not so large, because in that case the convolutional operation might cause features loss. Lastly, when we selected specific layers, we focused on picking from the 7 MBConv blocks of the pretrained EfficientNetB4, to come in for the most salient features from the encoder.

In the second stage, feature maps are fed into a modified Residual Refinement Module as inspired from Qin et al. [45], as we are the first to combine boundary refinement enhancement module for segmentation method. Notably, our new point is modifying Batch Normalization layers to Mean-Variance Normalization (MVN) [46], as long as Batch Normalization could calculate windowed statistics and switch between accumulating or using fixed statistics, MVN simply centers and

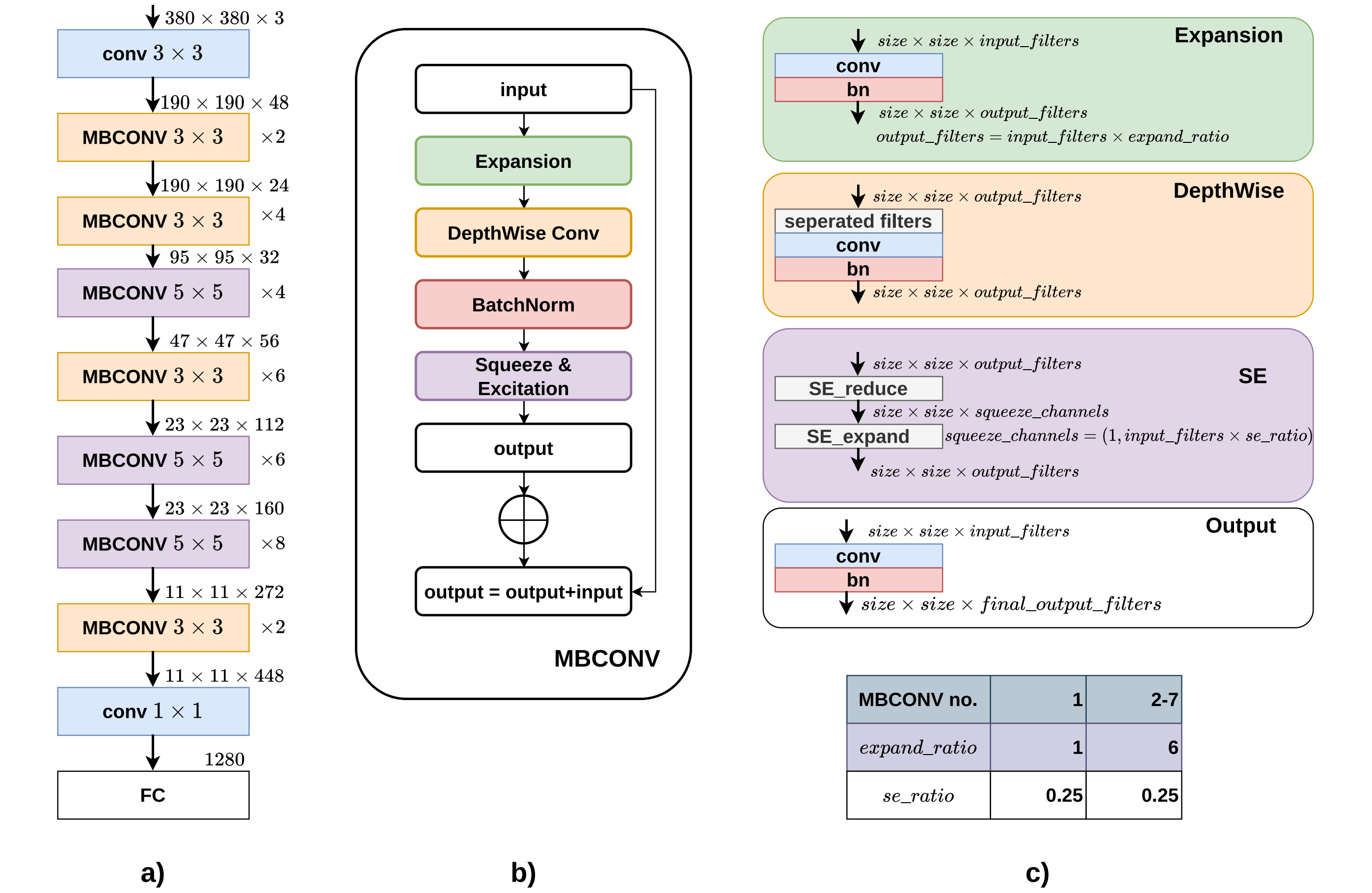


Fig. 1. Original Architecture of the EfficientNetB4. (a) the overall architecture of the EfficientNet B4; (b) the detail of each MBConv block; (c) indicates the designation of each phase in a MBConv block [36].

standardizes a single batch at a time. Additionally, Batch Normalization operations cost parameters, which results in network parameter increase; while MVN operations do not cost parameters. This normalization technique is also applied in the decoder for uniformity. We all maintain the core Residual Refinement Module (RRM) from [45], as the predicted coarse saliency maps S_{coarse} is adjusted by studying the residuals $S_{residual}$ between the saliency maps and the ground truth:

$$S_{refined} = S_{coarse} + S_{residual} \tag{2}$$

To refine regional and boundary deficiency in coarse maps, this RRM exploits the residual encoder–decoder architecture, consisting of an input layer, an encoder, a bridge, a decoder and an output layer. Each contains 64 3 × 3 filters, followed by a MVN layer and a ReLU nonlinearity. Non-overlapping MaxPooling2D manipulates downsampling in the encoder and bilinear interpolation manipulates upsampling in the decoder. The output of this RRM is the final resulting saliency map of our model; before going through a Sigmoid or Softmax activation layer (depend on the segmentation dataset) to obtain predicted segmented output. Our proposed network is demonstrated as Fig. 2 as follows.

3.2. The proposed loss function

3.2.1. The Baroni–Urbani–Buser (BUB) loss function

Firstly, we have an introduction on the Baroni–Urbani–Buser coefficient before going deeper into our proposed loss functions. Starting from the Jaccard/Tanimoto Coefficient described in Section 2.4, similarity measures are determined from the contingency table, containing four events: (a) 1–1 (interaction existing in both cases), (b) 1–0 (interaction existing in the first case and missing in the second case), (c) 0–1 (interaction missing in the first case but existing in the second), and (d) 0–0 (interaction missing both cases) [47]. With those parameters, a lot of similarity measures have been determined, for example:

$$\text{The Baroni–Urbani–Buser Coefficient:}\ \ \text{BUB} = \frac{\sqrt{a \times d} + a}{\sqrt{a \times d} + a + b + c} \tag{3}$$

Similarly the Jaccard/Tanimoto coefficient could be written as:

$$\text{JT} = \frac{a}{a + b + c} \tag{4}$$

From the expression of the BUB and the JT coefficients, it is obvious that the Jaccard/Tanimoto coefficient considers only double presences (a), whereas the Baroni–Urbani–Buser coefficient incorporates double absences (d). If the Jaccard/Tanimoto Coefficient is displayed in confusion matrix elements, it would be re-written as:

$$\text{JT} = \frac{TP}{TP + FP + FN} \tag{5}$$

with TP standing for true positive rate; FP standing for false positive rate and FN standing for false negative rate. Thus, it is understandable that $(a) = TP; (b) = FP; (c) = FN$ and further, (d) would be TN with TN is the true negative rate. Therefore, the Baroni–Urbani–Buser coefficient could be demonstrated in another way as follows:

$$\text{BUB} = \frac{TP + \sqrt{TP \times TN}}{TP + FP + FN + \sqrt{TP \times TN}} \tag{6}$$

To the best of our knowledge, we are the first ones applying the Baroni–Urbani–Buser coefficient for segmentation problems; and we will perform two different versions of Baroni–Urbani–Buser loss functions to target the class-imbalanced issue. The new loss function for biomedical

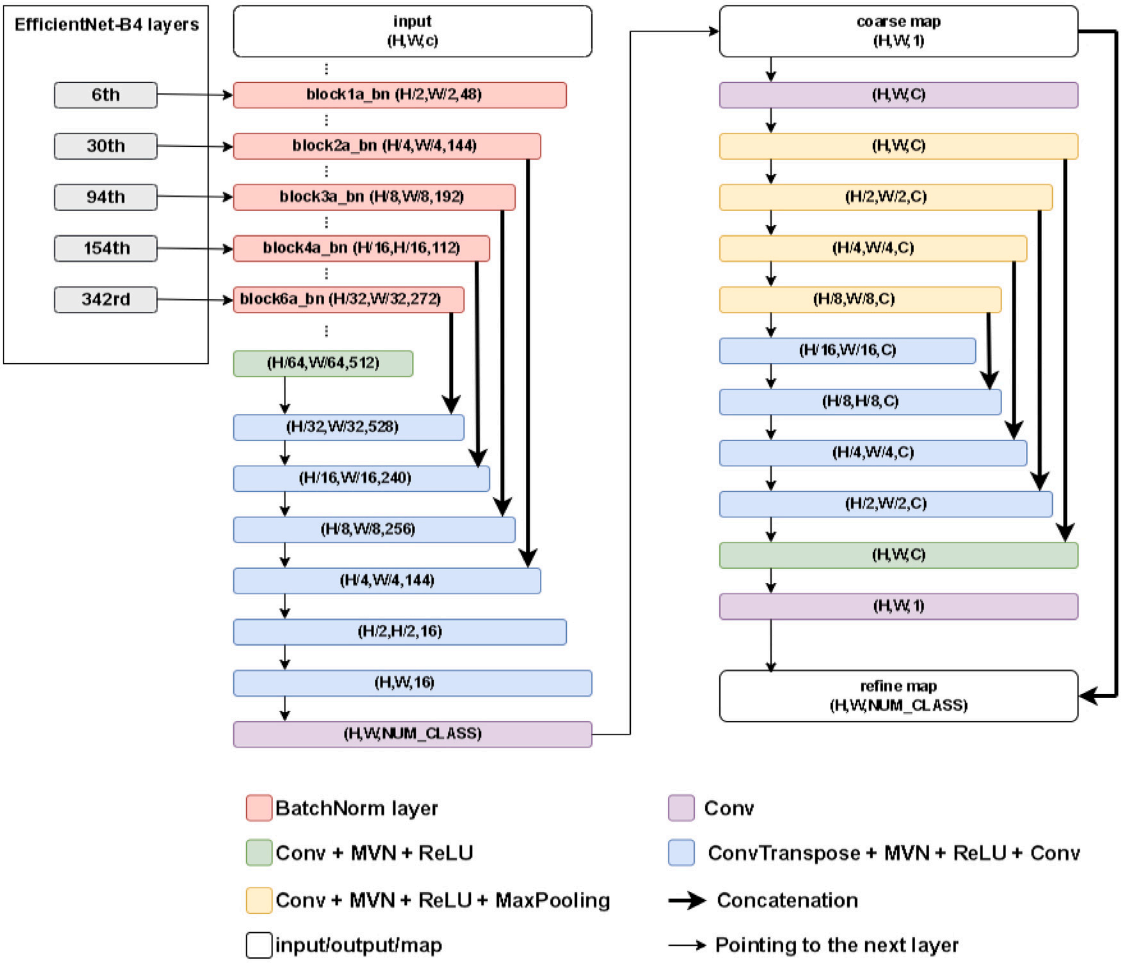


Fig. 2. Our Proposed two-stage Model. For generalization, we denote the resolution as $H \times W$. Here the '...' icon in the part of pretrained EfficientNet-B4 encoder denotes the layers between 2 layers appearing on the Figure. The variable 'C' in the Residual Refinement Module displays the filters number in a layer, and here C = 64; while c stands for the input channel, which is 1 or 3 depending on the dataset used.

image segmentation we create is the BUB (Baroni–Urbani–Buser) loss function (version 1), which could be presented as:

$$l_1 = 1 - BUB = \frac{FP + FN}{TP + FP + FN + \sqrt{TP \times TN}} \tag{7}$$

We could also replace TP by $\sqrt{TP \times TN}$ and keep the TN quantity, thus gaining the second version of the BUB loss function:

$$l_2 = \frac{FP + FN}{TN + FP + FN + \sqrt{TP \times TN}} \tag{8}$$

It could be easily seen that $0 < l_1, l_2 < 1$. Practices show us that the value of TN always exceeds the value of TP by certain times; then, if we replace TP by $\sqrt{TP \times TN}$ in l_1, it means we try to increase the quantity TP, which makes the denominator value decreases then l_1 increases. On the other hand, if we replace TN by $\sqrt{TP \times TN}$ in l_2, it means we try to decrease the quantity TN, which makes the denominator value increases then l_2 decreases.

3.2.2. Formulation with the Tversky–Kahneman function

We propose to have some modifications in order to increase the model convergence speed; therefore we propose to import our BUB loss function into the Tversky–Kahneman probability weighting function [35] for enhanced convergence speed. This probability weighting function is established as:

$$\omega(x) = \frac{x^\gamma}{[x^\gamma + (1 - x)^\gamma]^{\frac{1}{\gamma}}} \tag{9}$$

where $x \in [0, 1]$ is the cumulative probability distribution of gains or losses in economical fields. To generate the TK-BUB loss function, we have:

$$L_{TK-BUB-v1} = \frac{l_1^\gamma}{[l_1^\gamma + (1 - l_1)^\gamma]^{\frac{1}{\gamma}}} \tag{10}$$

$$L_{TK-BUB-v2} = \frac{l_2^\gamma}{[l_2^\gamma + (1 - l_2)^\gamma]^{\frac{1}{\gamma}}} \tag{11}$$

Above we have proposed two versions of the TK-BUB loss function. As regard to the parameter γ, experiments have been conducted with high values of γ till it indicates that the overall result displays the best when $\gamma \in (1, 2)$ and in addition, the best performance is confirmed with $\gamma = \frac{4}{3}$. Thus we train all experiments in case of $\gamma = \frac{4}{3}$. Carefully taking a look into both versions of the TK-BUB loss function, if the nominator value is powered by base $\gamma = \frac{4}{3} > 1$, the denominator value is powered by base $\approx \gamma \times \frac{1}{\gamma} = 1$, as a result, the nominator value tends to converge faster than the denominator value. As can be seen that $l_1 > l_2$, it follows that $L_{TK-BUB-v2}$ has a better convergence speed than $L_{TK-BUB-v1}$. As a consequence, we choose $L_{TK-BUB-v2}$ for our end-to-end training process. Tables in later sections will prove that our proposed $L_{TK-BUB-v2}$ gaining the best convergence speed as well as the most accurate segmentation results.

3.3. Evaluation metrics

In biomedical image analysis, while the Dice Similarity Coefficient (DSC) statistically measures the similarity between segmentation maps, the Intersection over Union index (IoU) statistically gauges the similarity and diversity of sample pixel sets. They are determined by:

$$\text{DSC} = \frac{2 \times TP}{2 \times TP + FP + FN} \tag{12}$$

$$\text{IoU} = \frac{TP}{TP + FP + FN} \tag{13}$$

with TP, TN, FP, FN have been determined as in the previous section.

In the simplest terms, the precision is the ratio between the true positives out of the total positives; while the recall is the measure of the model accurately identifying true positives. Mathematically:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{14}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{15}$$

F-score (also F_1) is a measure that is the harmonic mean of precision and recall:

$$\text{F-score} = \frac{2.\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{16}$$

Hausdorff Distance (HD) [48] is one of the most informative and useful validation criteria. For two point sets X and Y, the Hausdorff Distances from X to Y and from Y to X are calculated as (17) and (18):

$$HD(X, Y) = \max_{x \in X} \min_{y \in Y} \|x - y\|_2 \tag{17}$$

and

$$HD(Y, X) = \max_{y \in Y} \min_{x \in X} \|y - x\|_2 \tag{18}$$

4. Experiment

4.1. Datasets

4.1.1. The 2018 data science Bowl Cell Nuclei segmentation dataset

Cell nuclei identification is the original point for lots of analyses by reason of almost human body cells comprise of a nucleus full of DNA, the genetic code programming a body cell. Cell nuclei identification let researchers figure out each individual cell in an instance, and by assessing cell reaction to different medical operations, the expert could understand the elemental biological processes. Several cell nuclei algorithms allow apprehending structural and functional features of biological model systems. As cell nuclei makes development to seize such systems in greater detail and as the advancement of novel assays declares more compound characteristics of living organisms, the necessity for robust as well as easy to occupy microscopy image analysis approaches have become crucial to answer a much broader collection of biological questions. The Data Science Bowl 2018 [49] has held a competition on proposing an effective solution for automatic nuclei segmentation and detection. There are 670 nuclear images and respective pixel-level segmentation masks in the 2018 Data Science Bowl dataset, in our experiment, 80% samples for training and the remainder 20% samples for testing have been randomly chosen, resizing images to 256 × 256.

4.1.2. The Brain Tumor LGG Segmentation dataset

Lower-grade gliomas (LGG) consists of WHO grade II and grade III brain tumors. While grade I are always curable by surgical resection, grade II and III are infiltrative and tend to recur and evolve to higher-grade lesion. This LGG dataset [13] consists of brain MR images together with manual FLAIR abnormality segmentation masks. Images were taken from The Cancer Imaging Archive (TCIA); and they correspond to 110 patients with totally 3929 images contained in The Cancer Genome Atlas (TCGA) lower-grade glioma collection with at least fluid-attenuated inversion recovery (FLAIR) sequence and genomic cluster data available. In order to judge our proposed approach on this dataset, we split this into 80:20 that 80% of the total utilized for training and the remainder utilized for testing, resizing images to 256 × 256.

4.1.3. The Skin Lesion ISIC 2018 dataset

The next dataset we use for evaluational purpose is the Lesion Boundary Segmentation dataset from the ISIC 2018 competition [50,51], comprising 2594 dermoscopy pictures of skin lesions with expert annotations from diverse anatomic locations and institutions. Each image is resized to the shape of 256 × 256 to balance between complexity and training time, with 80% of the dataset adopted for training and 20% left adopted for testing respectively, due to the lack of the official testing dataset.

4.1.4. The MRI cardiac ACDC dataset

The ACDC dataset [52] is collected from 100 patient 4D cine-CMR scans, each comprising of segmentation labels for the left ventricle (LV), the myocardium (Myo) and the right ventricle (RV) at the end-systolic (ES) and end-diastolic (ED) phases of each patient. The training set, valid set and testing set are splited with ratio 70:10:20. All images have been resized to 128 × 128.

4.2. Implementation details

We have performed the proposed network, Modified EffcientNet-encoder U-Net Joint Residual Refinement Module, with our proposed customized Baroni–Urbani–Buser loss layer to segment multiple biomedical images. Our model is trained with cost minimization on several epochs (base on different cases), performed by using NADAM optimizer [53] with an original learning rate of 0.001. Learning rate is divided by half every 7 epochs, before reaching 0.00001 and being constantly kept through the remainder training period with 120 epochs for all the four datasets. The training time for our model is approximately 40 min to an hour at maximum on a workstation with NVIDIA Tesla P100 16 GB GPU.

4.3. Experimental results

Our proposed algorithm is compared with several algorithms to evaluate the competency of our new approach and write down the mean value of each metric index into several ablation test results tables. Method results with "*" are re-implemented by the released source codes; results elements with "–" denote the corresponding ones are not provided publicly. Models with code are trained with conditions in the Implementation Details and with the $L_{TK-BUB-v2}$ loss function; while methods without public code have results extracted from cited papers. In each metric index column, the best and the second best values are emphasized in **bold** and *italic* correspondingly. The column named "Params" denotes the number of parameters in each method.

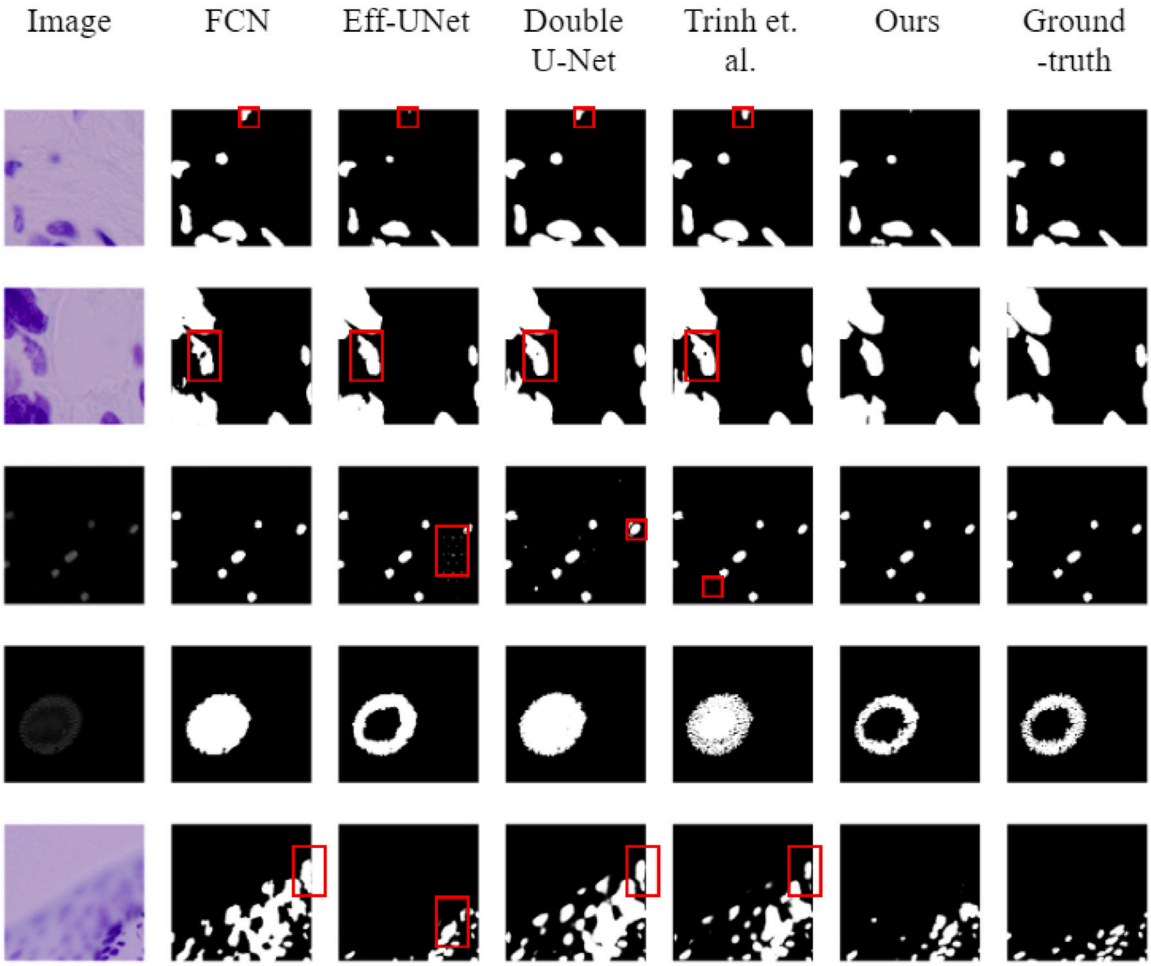


Fig. 3. Qualitative results on top-5 the best segmentation performances on the 2018 Data Science Bowl Cell Nucleus Segmentation dataset. Important regions are marked with appropriate boxes to easier visualize the difference between predictive results and ground truths.

4.3.1. Evaluation on the 2018 Data Science Bowl Cell Nucleus Segmentation dataset

To evaluate the persuasiveness of our proposed Modified EffcientNet-encoder U-Net Joint Residual Refinement Module, firstly experiments are conducted on the 2018 Data Science Bowl Cell Nucleus Segmentation dataset. The respective quantitative outcomes are shown in Fig. 3 and the comparison outcomes between various methods are indicated in Table 1. From Table 1, several remarks can be considered: firstly, those approaches having the integration of attention or gate mechanisms and residual module, such as FANet [54], MFRS-Net [55], Xie et al. [56], are obviously superior on segmenting the 2018 Data Science Bowl Cell Nucleus Segmentation dataset with the mean DSCs are all over 0.9. Secondly, two-stage models like Double U-Net [57] and ours have the dominant capability in solving nuclei segmentation task. Trinh et al. [58] also has promising performance on this dataset (0.9152 DSC, 0.8446 IoU); however, our proposed approach with utilization of pretrained EfficientNet B4 carries out the best scores on almost all evaluation metrics (0.9257 DSC, 0.8619 IoU and 0.9408 Recall) and shows a more precise outcome than the existing baselines, as shown in Fig. 3. Therefore, these comparative outcomes have demonstrated the efficacy of the proposed Modified EffcientNet-encoder U-Net Joint Residual Refinement Module for automated nuclei segmentation.

4.3.2. Evaluation on the Brain Tumor LGG Segmentation dataset

Additionally, we evaluate our proposed Modified EffcientNet-encoder U-Net Joint Residual Refinement Module to experiment on brain tumor segmentation. The respective quantitative results are illustrated in Fig. 4 and the comparison results of evaluation metrics are displayed in Table 2. New observations could be obtained as it could be confirmable that the scores of the evaluation metrics belonging to our proposed method far surpasses the scores outcoming from other methods (0.9251 DSC, 0.8458 IoU). Particularly, the performance of our proposed method distinctly outperforms the previous baselines, notably with the improvement of 1.51% in terms of the DSC from the second highest value of Eff-UNet [63]. This improvement demonstrates that the proposed method immensely comes in for the pretrained model properties as well as the encoder–decoder architecture and skip connections of U-Net, which enforces studying the globally context and peculiar features to discriminate the tumor area from the neighboring structures.

4.3.3. Evaluation on the ISIC 2018 dataset

To present an another evaluation of our proposed Modified EffcientNet-encoder U-Net Joint Residual Refinement Module and the proposed Tversky–Kahneman Baroni–Urbani–Buser loss function, additional experiments have been operated to compare our approach with some others on the ISIC 2018 dataset. The respective quantitative results are illustrated in Fig. 5 and ablation testing results are recorded in Table 3. As Table 3 demonstrates, it is observable that our model, trained with $L_{TK-BUB-v2}$ could still gain much effective and accurate segmentation results (approximately 0.89 in DSC).

4.3.4. Evaluation on the MRI cardiac ACDC dataset

Lastly, our proposed Modified EffcientNet-encoder U-Net Joint Residual Refinement Module is validated with several state-of-the-arts.

Table 1
Comparison between different methods on the 2018 Data Science Bowl Cell Nucleus Segmentation dataset.

Methods	Params	DSC	IoU	Precision	Recall	F-score
Double U-Net* [57]	29.3M	0.9133	0.8407	**0.9596**	0.6407	0.7684
U-Net++* [11]	10.2M	0.8970	0.8440	*0.9480*	0.8840	0.9110
FANet [54]	–	0.9176	0.8569	0.9194	0.9222	0.9208
TransUNet [59]	66.9M	0.9210	0.8560	–	–	
MFRS-Net [55]	–	0.9224	0.8534	0.9022	*0.9402*	0.9208
SSFormer-L [60]	–	*0.9230*	*0.8614*	–	–	
Ahmed et al. [61]	–	0.8632	0.7715	0.8843	0.8719	0.8632
Xie et al. [56]	–	0.9046	–	–	–	
Attention U-Net* [62]	31.9M	0.8750	0.7782	0.9384	0.8927	0.9150
FCN* [46]	10.9M	0.8939	0.8087	0.9118	0.9201	0.9159
Eff-UNet (B2)* [63]	21.3M	0.9096	0.8345	0.9304	0.9256	**0.9280**
ResU-Net* [64]	17.6M	0.8931	0.8053	0.9278	0.9257	*0.9267*
PraNet* [65]	30.3M	0.8334	0.7149	0.8978	0.8719	0.8847
Swin U-Net* [66]	2.9M	0.8441	0.7306	0.8550	0.9280	0.8900
Trinh et al.* [58]	32.4M	0.9152	0.8446	0.9283	0.9087	0.9184
Our proposed method*	10.3M	**0.9257**	**0.8619**	0.9014	**0.9408**	0.9207

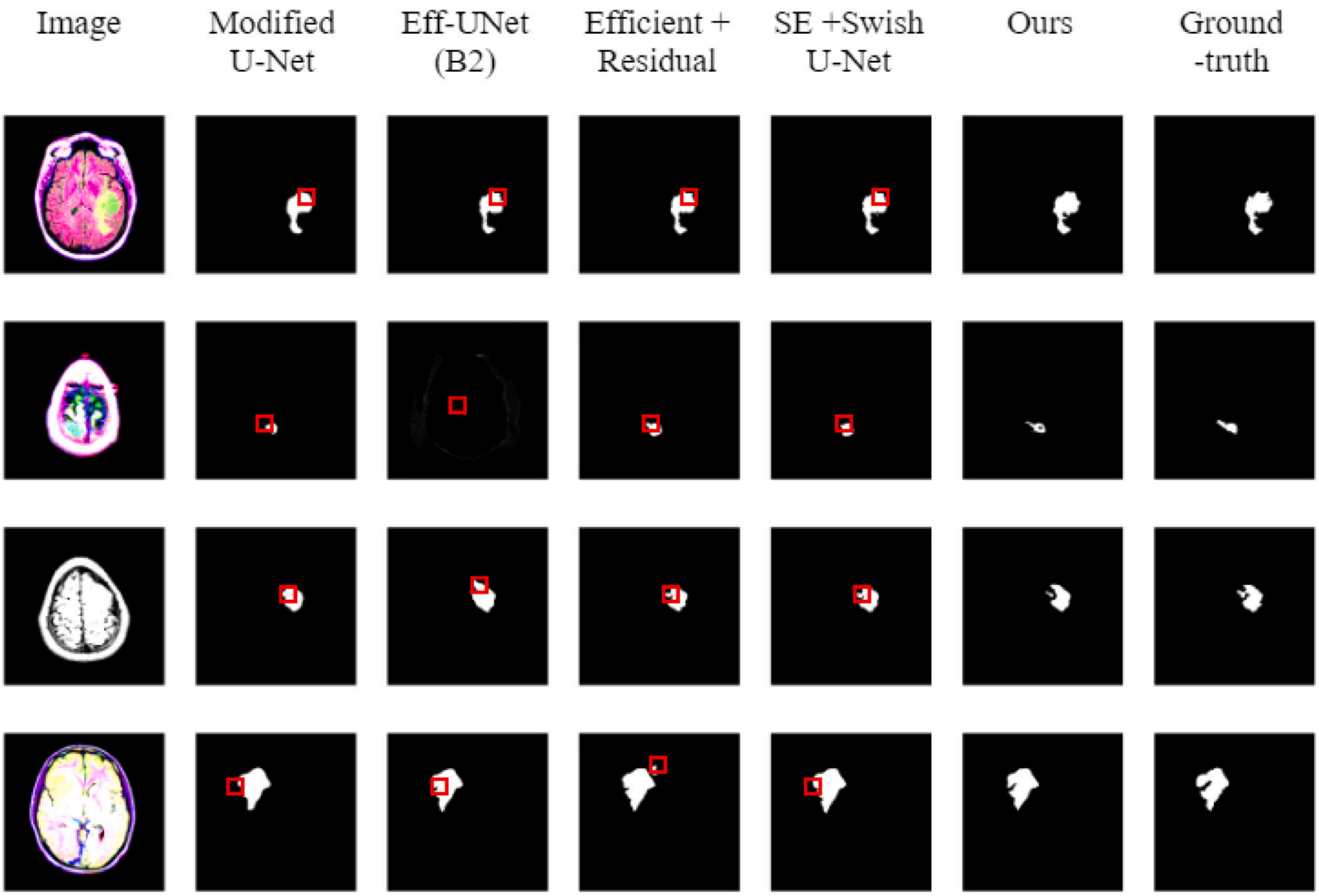


Fig. 4. Qualitative results on top-5 best segmentation performances on the Brain Tumor LGG Segmentation dataset. Important regions are marked with appropriate boxes to easier visualize the difference between predictive results and ground truths.

Qualitative visualizations are displayed in Fig. 6 and segmentation results are observed in Table 4. It could be easily seen that our proposed approach has obtained on par DSC results compared with other state-of-the-arts; except for the LV dice score in comparison with Swin-UNet [66] and TransUNet [59]. Though having noticeably low number of parameters, our DSC scores have still outpaced the second best ones by 0.45% to 0.67%. Furthermore, we have also compared the Hausdorff Distance (HD) between all approaches. It is observable that our HDs on several parts are reasonably low in comparison with other state-of-the-arts; which confirms that our propose approach is robust to noise. The qualitative visualization, which is demonstrated in Fig. 6, has confirmed our effectiveness of our proposed approach.

4.4. Ablation study

In Table 5, performance metrics for different loss functions used in training our proposed model are compared in all the four datasets. The column named “Peaking” denotes the column range for the DSC to reach its peak in each method. In the 2018 Data Science Bowl Cell Nucleus Segmentation dataset, as expected, in Table 5a, it is observed that $L_{TK-BUB-v2}$ produces the top scores on DSC (0.9257), IoU (0.8619) and Recall (0.9408). It is noticeable that when our network is trained with the Focal-Tversky loss function, though this focal loss function is proved with good convergence speed, results generated are more modest than other loss functions whose convergence speeds are mediocre. Table 5b

Table 2
Comparison between different methods on the Brain Tumor LGG Segmentation dataset.

Methods	Params	DSC	IoU	Precision	Recall	F-score
Buda et al. [67]	–	0.82	–	–	–	
Pattabiraman et al. [68]	–	0.87	–	–	–	
Attention U-Net* [62]	31.9M	0.8988	0.8173	0.8940	0.8930	**0.8935**
FCN* [46]	10.9M	0.8782	0.7850	0.8780	0.8855	0.8817
Eff-UNet (B2)* [63]	21.3M	0.9100	0.8310	0.8753	*0.8980*	*0.8865*
ResU-Net* [64]	17.6M	0.8963	0.8127	*0.9150*	0.8329	0.8720
PraNet* [65]	30.3M	0.8878	0.7989	**0.9204**	0.8263	0.8708
Our proposed method w/o RRM*	9.9M	**0.9151**	**0.8458**	0.8173	**0.9057**	0.8592
Our proposed method with RRM*	10.3M	*0.9127*	*0.8411*	0.8555	0.8855	0.8702

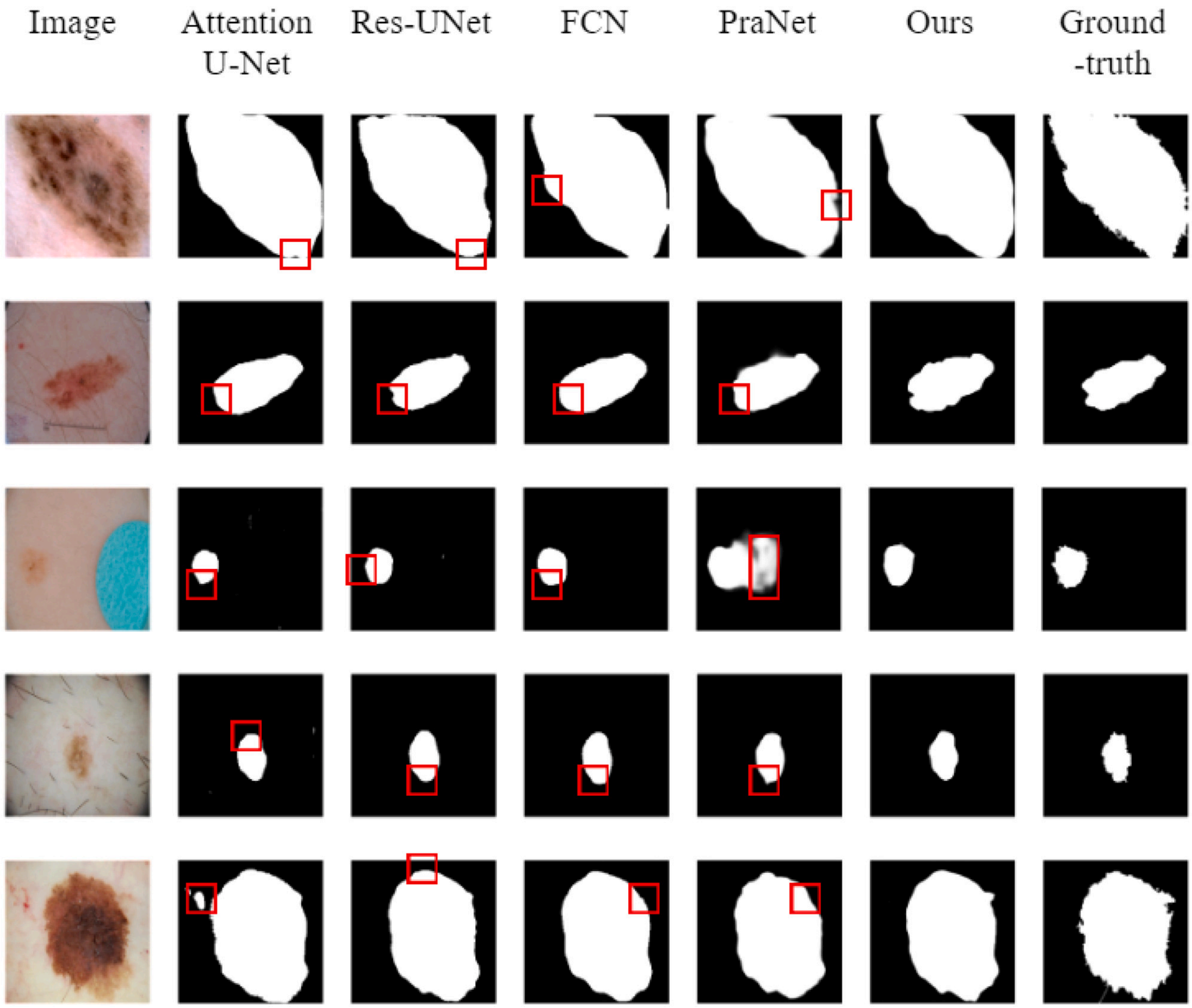


Fig. 5. Qualitative results on some segmentation performances on the ISIC 2018 dataset. Important regions are marked with appropriate boxes to easier visualize the difference between predictive results and ground truths.

further validates the reliability of our proposed $L_{TK-BUB-v2}$ as it exceeds other loss functions on both DSC and IoU indices on the Brain Tumor LGG Segmentation dataset. Table 5c illustrates the superiority of our TK-BUB loss function in the ISIC 2018 dataset despite the fact that the respective scores of all the evaluation metrics are relatively close, that our $L_{TK-BUB-v2}$ highest DSC just defeats the second highest of the one from Accuracy loss function by about 0.15%. In Table 5d, though the DSC results in all columns are not deviant from each other; however, the HD information has proved that our proposed loss helps our proposed overall approach much robust to noise.

To assess the influence of some special components in our proposed approach, we manipulate further comprehensive ablation experiments by removing the elements successively before the experimental outcomes are displayed in Table 6. To be more detailed, "BN" denotes

Table 3
Comparison between different methods on the ISIC 2018 dataset.

Methods	Params	DSC	IoU	Precision	Recall	F-score
FANet [54]	–	–	0.8023	0.9235	0.8650	0.8731
Double-UNet [57]	–	–	**0.8212**	**0.9459**	**0.8780**	*0.8962*
Swin-UNet [69]	–	*0.8846*	0.7962	0.9085	–	
Attention U-Net* [62]	31.9M	0.8593	0.7602	0.9071	0.8458	0.8754
FCN* [46]	10.9M	0.8690	0.7743	0.8992	0.8729	0.8859
ResU-Net* [64]	17.6M	0.8640	0.7673	0.8907	*0.8762*	0.8834
PraNet* [65]	30.3M	0.8734	0.7818	*0.9313*	0.8477	0.8875
Our proposed method*	10.3M	**0.8893**	*0.8046*	0.9280	0.8756	**0.9010**

Table 4
Comparison between different methods on the ACDC dataset. Note that the up arrow symbol after the DSC metric means that the higher the DSC the better; while the down arrow symbol after the HD implies that the lower HD is, the better the segmentation result is.

Methods	Params	RV		Myo		LV		Average	
		DSC↑	HD↓	DSC↑	HD↓	DSC↑	HD↓	DSC↑	HD↓
UNet++* [11]	10.2M	0.3681	12.8581	*0.8720*	*10.2186*	0.9300	9.8913	0.7234	10.9893
VIT-CUP [70]	–	0.8146	–	0.7071	–	0.9218	–	0.8145	–
FCN* [46]	10.9M	0.8055	27.4510	0.8147	48.5603	0.8872	49.4026	0.8358	41.8046
R50-VIT-CUP [70]	–	0.8607	–	0.8188	–	0.9475	–	0.8757	–
Trinh et al.* [58]	32.4M	0.8500	12.8181	0.8678	10.7519	0.9369	9.3127	0.8849	*10.9609*
UNETR [71]	92.58M	0.8529	–	0.8652	–	0.9402	–	0.8861	–
ResU-Net* [64]	17.6M	0.8788	14.3444	0.8557	10.8518	0.9239	*9.0380*	0.8861	11.4114
Eff-UNet (B2)* [63]	21.3M	0.8585	14.6111	0.8680	11.8785	0.9320	10.0479	0.8862	12.1792
Attention U-Net* [62]	31.9M	0.8626	*12.6425*	0.8689	15.2662	0.9321	9.0392	0.8879	12.3160
TransUNet [59]	66.9M	*0.8886*	–	0.8454	–	*0.9573*	–	0.8971	–
Swin-Unet [66]	41.5M	0.8855	–	0.8562	–	**0.9583**	–	*0.9000*	–
Our proposed method*	10.3M	**0.8953**	**11.9263**	**0.8767**	**8.2304**	0.9415	**8.1508**	**0.9045**	**9.4358**

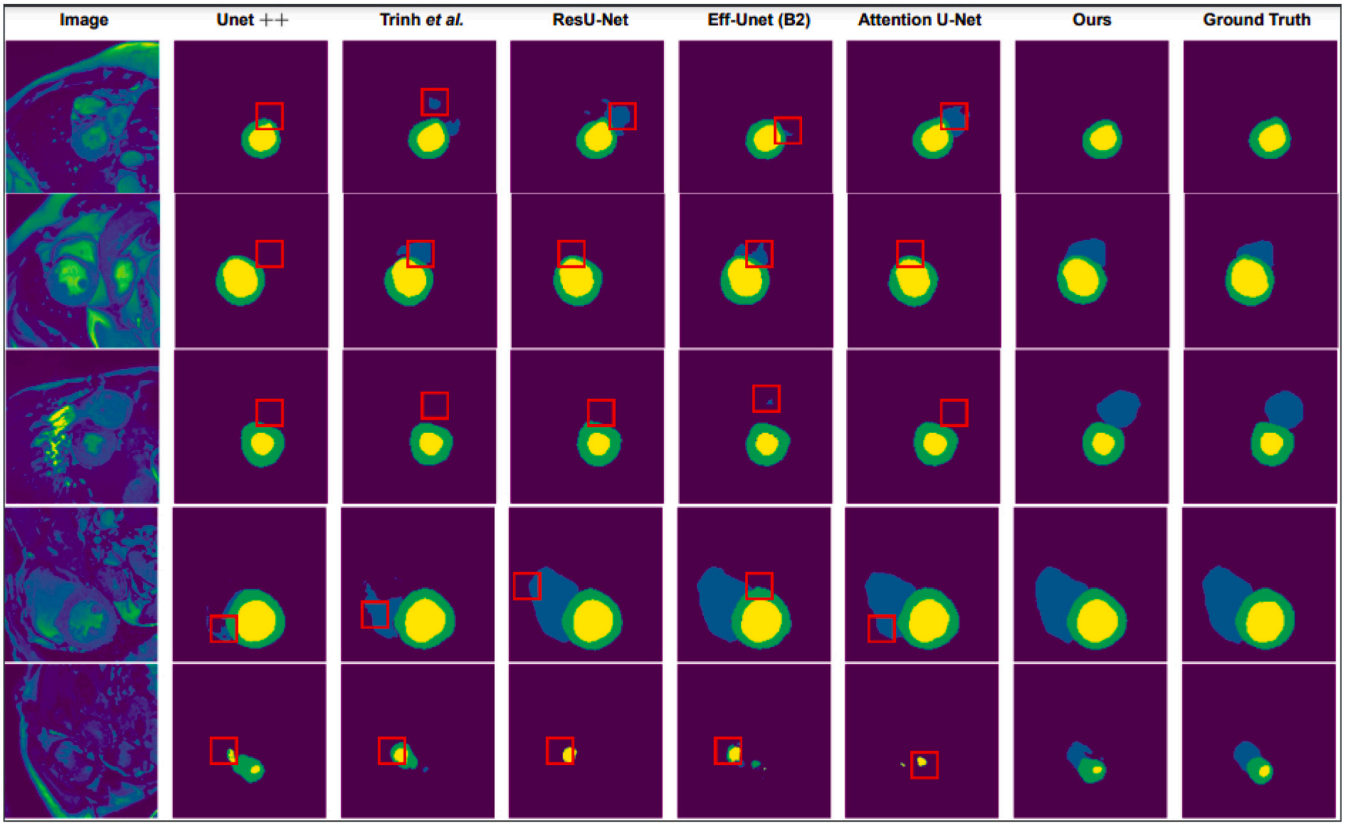


Fig. 6. Qualitative results on some segmentation performances on the ACDC dataset. Important regions are marked with appropriate boxes to easier visualize the difference between predictive results and ground truths.

the appearance of Batch Normalization layer, while "MVN" denotes the Mean-Variance Normalization layer. "RRM" stands for the Residual Refinement Module. "Ours" stands for our proposed method but without the Normalization layers and the Residual Refinement Module.

Table 6 has indicated the importance of MVN layers to the overall model, as without them segmentation performances could be seriously deteriorated. Moreover, when RRM is removed, the overall model would suffer performance degradation, except for the case on

Table 5
Comparison between different loss functions on: (a) - The 2018 Data Science Bowl Cell Nucleus Segmentation dataset (b) - The Brain Tumor LGG Segmentation dataset (c) - The ISIC 2018 dataset (d) - The ACDC dataset. Note that in (d), the up arrow symbol after the DSC metric means that the higher the DSC the better; while the down arrow symbol after the HD implies that the lower HD is, the better the segmentation result is.

(a) - The 2018 Data Science Bowl Cell Nucleus Segmentation dataset

Loss functions	Peaking	DSC	IoU	Precision	Recall	F-score
$L_Tversky$	45–55	0.9198	0.8519	0.8938	**0.9504**	0.9212
$L_Focal-Tversky$	37–42	0.9154	0.8446	0.8886	*0.9476*	0.9172
L_Dice	50–60	*0.9237*	*0.8585*	0.9209	0.9290	*0.9249*
L_BCE	55-63	0.9018	0.8215	**0.9225**	0.9225	0.9225
$L_Accuracy$	42–50	0.9227	0.8570	*0.9220*	0.9280	**0.9250**
$L_TK-BUB-v2$	28–32	**0.9257**	**0.8619**	0.9014	0.9408	0.9207

(b) - The Brain Tumor LGG Segmentation dataset

Loss functions	Peaking	DSC	IoU	Precision	Recall	F-score
$L_Tversky$	48–55	0.9065	0.8302	*0.8894*	0.8701	**0.8796**
$L_Focal-Tversky$	40–48	*0.9142*	*0.8430*	0.8416	**0.9103**	0.8746
L_Dice	45–50	0.9117	0.8383	0.8475	0.9026	0.8742
L_BCE	45–52	0.8947	0.8102	**0.9140**	0.8476	*0.8795*
$L_Accuracy$	55–62	0.9130	0.8418	0.8474	0.9005	0.8702
$L_TK-BUB-v2$	35–40	**0.9151**	**0.8458**	0.8173	*0.9057*	0.8592

(c) - The ISIC 2018 dataset

Loss functions	Peaking	DSC	IoU	Precision	Recall	F-score
$L_Tversky$	40–50	0.8806	0.7905	0.8907	0.8997	0.8952
$L_Focal-Tversky$	35–42	0.8835	0.7983	0.8921	*0.9085*	0.9002
L_Dice	40–48	0.8853	0.8002	0.8915	**0.9125**	*0.9019*
L_BCE	40–50	0.8607	0.7612	0.9222	0.8851	**0.9033**
$L_Accuracy$	38–46	*0.8878*	*0.8031*	**0.9303**	0.8630	0.8954
$L_TK-BUB-v2$	25–35	**0.8893**	**0.8046**	*0.9280*	0.8756	0.9010

(d) - The ACDC dataset

Loss functions	RV		Myo		LV		Average	
	DSC↑	HD↓	DSC↑	HD↓	DSC↑	HD↓	DSC↑	HD↓
$L_Tversky$	0.8868	13.7708	0.8668	17.8619	0.9364	8.8539	0.8967	13.4955
$L_Focal-Tversky$	*0.8879*	15.7588	0.8665	11.0391	0.9358	7.9912	0.8967	11.5964
L_Dice	0.8845	12.7856	0.8690	*10.2654*	*0.9411*	*7.0782*	*0.8982*	*10.0431*
L_CE	0.8759	**10.1739**	*0.8745*	16.5614	0.9321	**6.4137**	0.8942	11.0497
$L_Accuracy$	0.8738	16.4022	0.8608	10.2695	0.9336	7.7862	0.8894	11.4860
$L_TK-BUB-v2$	**0.8953**	*11.9263*	**0.8767**	**8.2304**	**0.9415**	8.1508	**0.9045**	**9.4358**

Table 6
Comparison of mean DSC and mean IoU for Ablation Studies of all the four datasets with different experimental approaches.

Datasets			Ours+BN (11.3M)	Ours+MVN (9.9M)	Ours+BN+RRM (11.7M)	Ours+MVN+RRM (10.3M)
Cell Nuclei	DSC		0.9124	*0.9253*	0.9130	**0.9257**
	IOU		0.8390	*0.8610*	0.8402	**0.8619**
Brain	DSC		0.9095	**0.9151**	0.9081	*0.9127*
	IOU		0.8340	**0.8458**	0.8325	*0.8411*
Skin	DSC		0.8832	0.8871	*0.8877*	**0.8893**
	IOU		0.7981	0.8001	*0.8002*	**0.8046**
ACDC	DSC	RV	0.8583	0.8797	*0.8804*	**0.8953**
		Myo	0.8673	0.8675	*0.8695*	**0.8767**
		LV	0.9292	*0.9323*	0.9315	**0.9415**
		Average	0.8849	0.8932	*0.8938*	**0.9045**
	IOU	RV	0.7811	0.8025	*0.8032*	**0.8181**
		Myo	0.7732	0.7734	*0.7754*	**0.7826**
		LV	0.8793	*0.8824*	0.8816	**0.8916**
		Average	0.8112	0.8194	*0.8201*	**0.8308**

the Brain Tumor LGG Segmentation dataset, where the DSC without RRM is reported to be better than the mean DSC and the mean IoU with RRM (0.9151 compared with 0.9127 and 0.8458 compared with 0.8411). Obviously, all the elements could complement and support each other, which additionally confirms the integrated impacts on general segmentation outcome.

To better understand the efficiency of our $L_{TK-BUB-v2}$, next we analyze the impact of our proposed TK-BUB loss function (version 2) by comparing its performances with other proposed versions trained on the best model. The comparison values in mean DSC and mean IoU evaluation metrics is written down in the Table 7. We confidently see that our $L_{TK-BUB-v2}$ coming on top of the leaderboard, that without this loss function version, the segmentation performances would be unfavorably affected, as our best performances indisputably exceed the second best performances by around 0.08% to 0.17% in term of the mean DSC.

5. Discussion

In an optimization algorithm, the function used for evaluating a candidate solution is defined as the objective function. Our necessity is maximizing or minimizing the objective function. In the context of deep learning problems, a loss function needs to be seeked as the objective function to minimize the error between the predicted masks and labels. In the current study, we propose a novel loss function for training the neural network. This stems from the fact that common

Table 7
Comparison of mean DSC and mean IoU for Ablation Studies with different experimental loss functions.

Datasets			l_1	l_2	$L_{TK-BUB-v1}$	$L_{TK-BUB-v2}$
Cell Nuclei	DSC		0.9220	0.9230	*0.9240*	**0.9257**
	IOU		0.8560	0.8574	*0.8608*	**0.8619**
Brain	DSC	with RRM	0.9106	0.9113	*0.9126*	**0.9127**
		w/o RRM	0.9139	*0.9143*	0.9140	**0.9151**
	IOU	with RRM	0.8400	0.8401	**0.8413**	*0.8411*
		w/o RRM	0.8436	0.8434	*0.8440*	**0.8458**
Skin	DSC		0.8880	*0.8883*	0.8870	**0.8893**
	IOU		0.8041	**0.8050**	0.8020	***0.8046***
ACDC	DSC	RV	0.8747	*0.8797*	0.8755	**0.8953**
		Myo	0.8749	0.8752	**0.8809**	*0.8767*
		LV	*0.9400*	0.9390	0.9394	**0.9415**
		Average	0.8944	0.8964	*0.8986*	**0.9045**
	IOU	RV	0.7940	0.7942	*0.7948*	**0.8181**
		Myo	0.7744	0.7811	**0.7891**	*0.7826*
		LV	*0.8893*	0.8883	0.8887	**0.8916**
		Average	0.8192	0.8212	*0.8242*	**0.8308**

loss functions used for deep learning-based image segmentation such as Cross-Entropy loss, the Dice loss, and the IoU loss (Jaccard loss) have some shortcomings as they are not good enough in handling class-imbalanced problems. Based on Baroni–Urbani–Buser coefficient, we introduce a novel loss function, namely Tversky–Kahneman Baroni–Urbani–Buser loss for the image segmentation task. The proposed loss has advantages over the above mentioned losses and the recently proposed Tversky loss. To the best of our knowledge, the Baroni–Urbani–Buser coefficient and Tversky–Kahneman Baroni–Urbani–Buser loss has not been investigated in segmentation studies before. The detailed explanation on advantage of the proposed loss function is described as the following.

Why we use the Baroni–Urbani–Buser coefficient ($\frac{TN+\sqrt{TP\times TN}}{TN+\sqrt{TP\times TN}+FP+FN}$) for the loss but not Jaccard ($\frac{TP}{TP+FP+FN}$), Dice ($\frac{2\times TP}{2\times TP+FP+FN}$) or Accuracy ($\frac{TP+TN}{TP+TN+FP+FN}$)? Obviously, difficult segmentation cases lying on images which have a very small injury area/white-masked area compared to the safe area/black-masked area (classes imbalance). We have not chosen the Dice and Jaccard coefficients because if we remove the TN quantity (which contributes to a large value), we will ignore a crucial part for updating the loss value and validating our model. On the Accuracy-basedloss, because TN value is larger than the TP, FP and FN for several times, hence after backpropagation stage, the training step size might be not as good as the Baroni–Urbani–Buser-based loss providing, as regards the model convergence speed. Thus we tune the quantities for more importance on TP and less importance on TN quantity; or Baroni–Urbani–Buser coefficient and Jaccard coefficient could be better choices to evaluate the model. Results provided by Table 5 have realized the differences.

The formula of Tversky coefficient for Tversky loss function is $\frac{\alpha\times FP+\beta\times FN}{TP+\alpha\times FP+\beta\times FN}$ where α and β control the magnitude of penalties for FPs and FNs, respectively. However, printing out the pixel number classified with false-positive and false-negative allows us recognize that using any loss function, these values are both small, after several epochs, compared to true-positive and their difference is not large. Hence we decide to tune the quantity of TP and TN. As could be observed from Table 5, results provided by Tversky loss and Focal Tversky loss are not favorable enough. We have not also applied the Cross-Entropy loss, the Dice loss and the Jaccard loss because they are witnessed to be not satisfactory enough in handling class-imbalanced problems [72]. Results provided by Table 5 have pointed out.

Though our proposed method works efficiently on these four datasets with class-imbalanced problem; it might not work well on datasets which have got severe class-imbalanced problem. If injury area/white-masked area is critically small, our loss might converge too fast, which leads to vanish gradient and further training failure.

6. Conclusion

We have presented a new network architecture and a novel loss for biomedical image segmentation in this paper. The network has been proposed with pretrained encoder–decoder framework with refinement module following called Modified EffcientNet-encoder U-Net Joint Residual Refinement Module. For the loss function, encouraged by the similarity matching Baroni–Urbani–Buser coefficient, we further build up several versions of the Baroni–Urbani–Buser loss function for biomedical image segmentation, with the second version of Tversky–Kahneman Baroni–Urbani–Buser loss function ($L_{TK-BUB-v2}$) is the most applicable. Extensive experiments have been performed to prove the dominance of both proposed architecture and proposed loss function. In the future, we will put more consideration on increasing the complexity of the network and proposing new loss functions to apply on training more sophisticated datasets, thus escalating the overall accuracy of the deep-learning segmentation field.

CRediT authorship contribution statement

Do-Hai-Ninh Nham: Conceptualization, Methodology / Study design, Software, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing. **Minh-Nhat Trinh:** Methodology / Study design, Software, Investigation, Data curation, Writing – review & editing, Visualization, Funding acquisition. **Viet-Dung Nguyen:** Software, Investigation, Data curation, Writing – review & editing, Visualization. **Van-Truong Pham:** Conceptualization, Methodology / Study design, Validation, Formal analysis, Investigation, Resources, Writing – original draft, Writing – review & editing, Supervision, Project administration, Funding acquisition. **Thi-Thao Tran:** Conceptualization, Methodology / Study design, Validation, Formal analysis, Investigation, Resources, Writing – original draft, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

There is no interest conflict

Data availability

The authors do not have permission to share data.

Acknowledgments

This research is funded by Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 102.05–2021.34.

References

[1] K. O'Shea, R. Nash, An introduction to convolutional neural networks, 2015.

[2] V.-T. Pham, T.-T. Tran, P.-C. Wang, P.-Y. Chen, M.-T. Lo, EAR-UNet: A deep learning-based approach for segmentation of tympanic membranes from otoscopic images, Artif. Intell. Med. 115 (2021) 102065.

[3] J. Ramya, H. Vijaylakshmi, H. Mirza Saifuddin, Segmentation of skin lesion images using discrete wavelet transform, Biomed. Signal Process. Control 69 (2021) 102839.

[4] N. Otsu, A threshold selection method from gray-level histograms, IEEE Trans. Syst. Man Cybern. 9 (1) (1979) 62–66.

[5] N. Zhang, Y.-X. Cai, Y.-Y. Wang, Y.-T. Tian, X.-L. Wang, B. Badami, Skin cancer diagnosis based on optimized convolutional neural network, Artif. Intell. Med. 102 (2019) 101756.

[6] E. Türetken, X. Wang, C.J. Becker, C. Haubold, P. Fua, Network flow integer programming to track elliptical cells in time-lapse sequences, IEEE Trans. Med. Imaging 36 (4) (2017) 942–951.

[7] E.E. Nithila, S. Kumar, Segmentation of lung from CT using various active contour models, Biomed. Signal Process. Control (ISSN: 1746-8094) 47 (2019) 57–62.

[8] O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional networks for biomedical image segmentation, in: Medical Image Computing and Computer-Assisted Intervention, Springer International Publishing, 2015, pp. 234–241.

[9] K. He, G. Gkioxari, P. Dollár, R. Girshick, Mask R-CNN, in: 2017 IEEE International Conference on Computer Vision, ICCV, 2017, pp. 2980–2988.

[10] A.O. Vuola, S.U. Akram, J. Kannala, Mask-RCNN and U-net ensembled for nuclei segmentation, in: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), 2019, pp. 208–212.

[11] Q. Xu, W. Duan, An automatic nuclei image segmentation based on multi-scale split-attention U-net, in: Proceedings of the MICCAI Workshop on Computational Pathology, in: Proceedings of Machine Learning Research, vol. 156, PMLR, 2021, pp. 236–245.

[12] K. Zhang, Y. Shi, C. Hu, H. Yu, Nucleus image segmentation method based on GAN network and FCN model, 2021.

[13] M. Buda, A. Saha, M.A. Mazurowski, Association of genomic subtypes of lower-grade gliomas with shape features automatically extracted by a deep learning algorithm, Comput. Biol. Med. (ISSN: 0010-4825) 109 (2019) 218–225.

[14] H. Dong, G. Yang, F. Liu, Y. Mo, Y. Guo, Automatic brain tumor detection and segmentation using U-net based fully convolutional networks, in: Medical Image Understanding and Analysis, Springer International Publishing, 2017, pp. 506–517.

[15] T. Brosch, L. Tang, Y. Yoo, D. Li, A. Traboulsee, R. Tam, Deep 3D convolutional encoder networks with shortcuts for multiscale feature integration applied to multiple sclerosis lesion segmentation, IEEE Trans. Med. Imaging 35 (2016) 1.

[16] F. Isensee, P. Kickingereder, W. Wick, M. Bendszus, K.H. Maier-Hein, Brain tumor segmentation and radiomics survival prediction: Contribution to the BRATS 2017 challenge, in: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries, Springer International Publishing, ISBN: 978-3-319-75238-9, 2018, pp. 287–297.

[17] S. Cui, L. Mao, J. Jiang, C. Liu, S. Xiong, Automatic semantic segmentation of brain gliomas from MRI images using a deep cascaded neural network, J. Healthc. Eng. 2018 (2018) 1–14.

[18] M.M.K. Sarker, H.A. Rashwan, F. Akram, S.F. Banu, A. Saleh, V.K. Singh, F.U.H. Chowdhury, S. Abdulwahab, S. Romani, P. Radeva, D. Puig, Slsdeep: Skin lesion segmentation based on dilated residual and pyramid pooling networks, in: Medical Image Computing and Computer Assisted Intervention – MICCAI 2018, Springer International Publishing, ISBN: 978-3-030-00934-2, 2018, pp. 21–29.

[19] H. Li, X. He, F. Zhou, Z. Yu, D. Ni, S. Chen, T. Wang, B. Lei, Dense deconvolutional network for skin lesion segmentation, IEEE J. Biomed. Health Inf. 23 (2) (2019) 527–537.

[20] M. Asadi-Aghbolaghi, R. Azad, M. Fathy, S. Escalera, Multi-level context gating of embedded collective knowledge for medical image segmentation, 2020.

[21] H. Song, W. Wang, S. Zhao, J. Shen, K.-M. Lam, Pyramid dilated deeper ConvLSTM for video salient object detection: 15th European conference, Munich, Germany, september 8-14, 2018, proceedings, part XI, ISBN: 978-3-030-01251-9, 2018, pp. 744–760.

[22] Z. Al Nazi, T.A. Abir, Automatic skin lesion segmentation and melanoma detection: Transfer learning approach with U-net and DCNN-SVM, ISBN: 978-981-13-7563-7, 2018.

[23] H. Zhang, W. Zhang, W. Shen, N. Li, Y. Chen, S. Li, B. Chen, S. Guo, Y. Wang, Automatic segmentation of the cardiac MR images based on nested fully convolutional dense network with dilated convolution, Biomed. Signal Process. Control (ISSN: 1746-8094) 68 (2021) 102684.

[24] W. Shen, W. Xu, H. Zhang, Z. Sun, J. Ma, X. Ma, S. Zhou, S. Guo, Y. Wang, Automatic segmentation of the femur and tibia bones from X-ray images based on pure dilated residual U-Net, Inverse Prob. Imaging (ISSN: 1930-8337) 15 (6) (2021) 1333–1346.

[25] B. Kim, J.C. Ye, Mumford–Shah loss functional for image segmentation with deep learning, IEEE Trans. Image Process. (ISSN: 1941-0042) 29 (2020) 1856–1866.

[26] J. Fang, H. Liu, L. Zhang, J. Liu, H. Liu, Region-edge-based active contours driven by hybrid and local fuzzy region-based energy for image segmentation, Inform. Sci. 546 (2020).

[27] Y. Yang, X. Hou, H. Ren, Accurate and efficient image segmentation and bias correction model based on entropy function and level sets, Inform. Sci. 577 (2021) 638–662.

[28] L. Xiao, T. Zhang, A proximal stochastic gradient method with progressive variance reduction, SIAM J. Optim. 24 (4) (2014) 2057–2075.

[29] C.H. Sudre, W. Li, T. Vercauteren, S. Ourselin, M. Jorge Cardoso, Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations, Lect. Notes Comput. Sci. (ISSN: 1611-3349) (2017) 240–248.

[30] S.S.M. Salehi, D. Erdogmus, A. Gholipour, Tversky loss function for image segmentation using 3D fully convolutional deep networks, in: Machine Learning in Medical Imaging, Springer International Publishing, ISBN: 978-3-319-67389-9, 2017, pp. 379–387.

[31] N. Abraham, N.M. Khan, A novel focal tversky loss function with improved attention U-net for lesion segmentation, in: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), 2019, pp. 683–687.

[32] M. Brusco, D. Cradit, D. Steinley, A comparison of 71 binary similarity coefficients: The effect of base rates, PLoS One 16 (2021) e0247751.

[33] N. Chung, B. Miasojedow, M. Startek, A. Gambin, Jaccard/Tanimoto similarity test and estimation methods for biological presence-absence data, BMC Bioinformatics 20 (2019).

[34] L. Dragomirescu, Clustering program by methods dedicated to biomedical thinking., in: The 1st MEDINF International Conference on Medical Informatics and Engineering, Craiova, Romania, 2003, pp. 146–147.

[35] J. Ingersoll, Non-monotonicity of the tversky-kahneman probability-weighting function: A cautionary note, Eur. Financial Manag. 14 (2008) 385–390.

[36] M. Tan, Q. Le, EfficientNet: Rethinking model scaling for convolutional neural networks, in: Proceedings of the 36th International Conference on Machine Learning, in: Proceedings of Machine Learning Research, vol. 97, PMLR, 2019, pp. 6105–6114.

[37] S. Ioffe, C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift, in: Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37, ICML '15, JMLR.org, 2015, pp. 448–456.

[38] T. Salimans, D.P. Kingma, Weight normalization: A simple reparameterization to accelerate training of deep neural networks, in: Proceedings of the 30th International Conference on Neural Information Processing Systems, NIPS '16, Curran Associates Inc., Red Hook, NY, USA, ISBN: 9781510838819, 2016, pp. 901–909.

[39] J.L. Ba, J.R. Kiros, G.E. Hinton, Layer normalization, 2016.

[40] D. Ulyanov, A. Vedaldi, V. Lempitsky, Instance normalization: The missing ingredient for fast stylization, 2017.

[41] Y. Wu, K. He, Group normalization, in: Computer Vision – ECCV 2018, Springer International Publishing, Cham, ISBN: 978-3-030-01261-8, 2018, pp. 3–19.

[42] G. Lin, A. Milan, C. Shen, I. Reid, RefineNet: Multi-path refinement networks for high-resolution semantic segmentation, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2017, pp. 5168–5177.

[43] Q. Chen, V. Koltun, Photographic image synthesis with cascaded refinement networks, in: 2017 IEEE International Conference on Computer Vision (ICCV), 2017, pp. 1520–1529.

[44] T. Kitrungrotsakul, I. Yutaro, L. Lin, R. Tong, J. Li, Y.-W. Chen, Interactive deep refinement network for medical image segmentation, 2020.

[45] X. Qin, Z. Zhang, C. Huang, C. Gao, M. Dehghan, M. Jagersand, BASNet: Boundary-aware salient object detection, in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR, 2019, pp. 7471–7481.

[46] P.V. Tran, A fully convolutional neural network for cardiac segmentation in short-axis MRI, 2017.

[47] A. Rácz, D. Bajusz, K. Héberger, Life beyond the Tanimoto coefficient: Similarity measures for interaction fingerprints, J. Cheminform. 10 (2018) 48.

[48] D. Karimi, S.E. Salcudean, Reducing the hausdorff distance in medical image segmentation with convolutional neural networks, IEEE Trans. Med. Imaging 39 (2) (2020) 499–513.

[49] B.A. Hamilton, Data science bowl - find the nuclei in divergent images to advance medical discovery, 2018.

[50] N. Codella, V. Rotemberg, P. Tschandl, M.E. Celebi, S. Dusza, D. Gutman, B. Helba, A. Kalloo, K. Liopyris, M. Marchetti, et al., Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (isic), 2019.

[51] P. Tschandl, C. Rosendahl, H. Kittler, The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions, Sci. Data 5 (1) (2018) 1–9.

[52] O. Bernard, A. Lalande, C. Zotti, F. Cervenansky, X. Yang, P.-A. Heng, I. Cetin, K. Lekadir, O. Camara, M.A. Gonzalez Ballester, G. Sanroma, S. Napel, S. Petersen, G. Tziritas, E. Grinias, M. Khened, V.A. Kollerathu, G. Krishnamurthi, M.-M. Rohé, X. Pennec, M. Sermesant, F. Isensee, P. Jäger, K.H. Maier-Hein, P.M. Full, I. Wolf, S. Engelhardt, C.F. Baumgartner, L.M. Koch, J.M. Wolterink, I. Išgum, Y. Jang, Y. Hong, J. Patravali, S. Jain, O. Humbert, P.-M. Jodoin, Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: Is the problem solved? IEEE Trans. Med. Imaging 37 (11) (2018) 2514–2525.

[53] T. Dozat, Incorporating nesterov momentum into adam, in: Proceedings of the 4th International Conference on Learning Representations, 2016.

[54] N.K. Tomar, D. Jha, M.A. Riegler, H.D. Johansen, D. Johansen, J. Rittscher, P. Halvorsen, S. Ali, FANet: A feedback attention network for improved biomedical image segmentation, IEEE Trans. Neural Netw. Learn. Syst. (2022) 1–14.

[55] A. Srivastava, D. Jha, S. Chanda, U. Pal, H.D. Johansen, D. Johansen, M. Riegler, S. Ali, P. Halvorsen, MSRF-net: A multi-scale residual fusion network for biomedical image segmentation, IEEE J. Biomed. Health Inf. 26 (2022) 2252–2263.

[56] X. Xie, Y. Li, Z. Menglu, L. Shen, Robust segmentation of nucleus in histopathology images via mask R-CNN: 4th international workshop, BrainLes 2018, held in conjunction with MICCAI 2018, granada, Spain, september 16, 2018, revised selected papers, part i, ISBN: 978-3-030-11722-1, 2019, pp. 428–436.

[57] D. Jha, M.A. Riegler, D. Johansen, P. Halvorsen, H.D. Johansen, Doubleu-net: A deep convolutional neural network for medical image segmentation, in: 2020 IEEE 33rd International Symposium on Computer-Based Medical Systems, CBMS, 2020, pp. 558–564.

[58] M. Trinh, N. Nguyen, T. Tran, V. Pham, A deep learning-based approach with image-driven active contour loss for medical image segmentation, in: The 2nd International Conference on Data Science and Applications ICDSA 2021, 2021.

[59] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, A.L. Yuille, Y. Zhou, TransUNet: Transformers make strong encoders for medical image segmentation, 2021.

[60] J. Wang, Q. Huang, F. Tang, J. Meng, J. Su, S. Song, Stepwise feature fusion: Local guides global, in: Medical Image Computing and Computer Assisted Intervention – MICCAI 2022, Springer Nature Switzerland, ISBN: 978-3-031-16437-8, 2022, pp. 110–120.

[61] T.S. Ahmed, Zarbega, Y. Gültepe, Nuclei cells detection and segmentation in histological images with semantic deep neural network, 2021.

[62] O. Oktay, J. Schlemper, L.L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N.Y. Hammerla, B. Kainz, B. Glocker, D. Rueckert, Attention U-net: Learning where to look for the pancreas, 2018.

[63] B. Baheti, S. Innani, S. Gajre, S. Talbar, Eff-UNet: A novel architecture for semantic segmentation in unstructured environment, 2020, pp. 1473–1481.

[64] F.I. Diakogiannis, F. Waldner, P. Caccetta, C. Wu, ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data, ISPRS J. Photogram. Remote Sens. (ISSN: 0924-2716) 162 (2020) 94–114.

[65] D.-P. Fan, G.-P. Ji, T. Zhou, G. Chen, H. Fu, J. Shen, L. Shao, PraNet: Parallel reverse attention network for polyp segmentation, in: Medical Image Computing and Computer Assisted Intervention – MICCAI 2020, Springer International Publishing, Cham, ISBN: 978-3-030-59725-2, 2020, pp. 263–273.

[66] H. Cao, Y. Wang, J. Chen, D. Jiang, X. Zhang, Q. Tian, M. Wang, Swin-unet: Unet-like pure transformer for medical image segmentation, 2021.

[67] M. Buda, A. Saha, M.A. Mazurowski, Association of genomic subtypes of lower-grade gliomas with shape features automatically extracted by a deep learning algorithm, Comput. Biol. Med. (ISSN: 0010-4825) 109 (2019) 218–225.

[68] V. Pattabiraman, H. Singh, Deep learning based brain tumour segmentation, WSEAS Trans. Comput. 19 (2021) 234–241.

[69] Y. Dong, L. Wang, Y. Li, TC-net: Dual coding network of transformer and CNN for skin lesion segmentation, PLoS One 17 (2022) e0277578.

[70] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, N. Houlsby, An image is worth 16x16 words: Transformers for image recognition at scale, in: International Conference on Learning Representations, 2021.

[71] A. Hatamizadeh, Y. Tang, V. Nath, D. Yang, A. Myronenko, B. Landman, H.R. Roth, D. Xu, UNETR: Transformers for 3D medical image segmentation, 2022, pp. 1748–1758.

[72] Y. Cui, M. Jia, T.-Y. Lin, Y. Song, S. Belongie, Class-balanced loss based on effective number of samples, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR, 2019.