Automated cervical spinal cord segmentation in Real World MRI of Multiple Sclerosis patients by optimized hybrid residual attention-aware Convolutional Neural Networks

Abstract

Magnetic resonance (MR) imaging is the most sensitive clinical tool in the diagnosis and monitoring of multiple sclerosis (MS) alterations. Spinal cord evaluation has gained interest in this clinical scenario in recent years but, unlike the brain, there is a more limited choice of algorithms to assist spinal cord segmentation. Our goal was to investigate and develop an automatic MR cervical cord segmentation method, enabling automated and seamless spinal cord atrophy assessment and setting the stage for the development of an aggregated algorithm for the extraction of lesion-related imaging biomarkers. The algorithm was developed using a real-world MR imaging dataset of 121 MS patients (96 cases used as a training dataset and 25 cases as a validation dataset). Transversal, 3D-T1 weighted gradient echo MR images (TE/TR/FA=1.7-2.7ms/5.6-8.2ms/12°) were acquired in a 3T system (SignaHD, GEHC) as standard of care in our clinical practice. Experienced radiologists supervised the manual labelling, which was considered the ground-truth. The 2D convolutional neural network consisted of a hybrid residual attention-aware segmentation method trained to delineate the cervical spinal cord. The training was conducted using a focal loss function, based on the Tversky index to address label imbalance, and an automatic optimal learning rate finder. Our automated model provided an accurate segmentation, achieving a validation DICE coefficient of 0.904±0.101 compared with the manual delineation. An automatic method for cervical spinal cord segmentation on T1-weighted MR images was successfully implemented. It will have direct implications serving as the first step for accelerating the process for MS staging and follow-up through imaging biomarkers.

Introduction

Multiple Sclerosis (MS) is an inflammatory and autoimmune disease of the central nervous system (CNS) which generally results in the demyelination and axonal loss over time [1], due to the presence of large focal lesions in the white and grey matter of the brain and spinal cord. Because of its high sensitivity for the evaluation of inflammatory and neurodegenerative processes, Magnetic Resonance (MR) has become the most important tool in confirming diagnosis and in monitoring MS treatment trials. MR images quantify the lesion load, spatial dissemination and longitudinal evolution over time [2]. Spinal cord evaluation by MR has become an increasingly relevant topic in the diagnosis and management of the disease [3].

Initially, MS evaluation focused on demyelinated plaques in the white matter. Later, investigators realized that lesions are also present in the grey matter, including the cortex, the basal ganglia, brain stem, as well as the grey matter of the spinal cord. Moreover, neurodegeneration gives rise to axonal loss in the white matter and diffuse changes in the entire grey matter, affecting the brain and spinal cord in a global sense. These changes will finally result in profound brain tissue loss and atrophy, most pronounced in the progressive stages of the disease [4]. Several neuropathological and MR imaging studies have proved the involvement of the spinal cord in MS [5][6][7].

Furthermore, neurodegeneration of the spinal cord is considered to be the main pathological cause of irreversible locomotor disability [8],[9]. In particular, MR imaging of the spinal cord has shown consequential signs of axonal loss by quantifying cord atrophy, which is defined as the reduction of the cord cross-sectional area correlated with physical disability. In addition, MR images also show cervical cord T2-weighted hyperintense areas related to parenchyma injuries in more than 90% of MS patients [10].

An effective method in the assessment of MS patients is the MR evaluation of spinal cord atrophy and lesions. The evaluation of the cross-sectional area of the spinal cord volume and lesions act as imaging biomarkers, evaluating its structural integrity [11]. However, MR imaging standardization and subsequent analysis is particularly difficult [12] due to the presence of varying MR acquisition protocols, which leads to different images (signal-to-noise ratio, contrast-to-noise ratio, temporal and spatial resolutions) and the presence of artefacts, such as susceptibility, motion, chemical shift, ghosting, blurring and Gibbs artifact. These factors result in special MR images features that limit their use without pre-processing steps and also would affect to the generalization of a deep learning model that can work with a huge variety of MR images, one of our main goals**.** Also, there is a wide range of spinal cord shapes, lengths, and pathologies. All these aspects introduce complexity into spinal cord evaluation and segmentation.

Spinal cord localization, atrophy measurement and focal areas of inflammatory demyelination detection can be depicted by segmentation techniques [13]. Several automatic segmentation methods have been proposed [13-23]. Whereas these methods have shown good performance [13], they often require initial specifications or landmarks and are restricted to specific MR image acquisition protocols as mentioned previously [14]. Some of their limitations relate to traditional intensity based deterministic algorithms, which are limited by factors relating to changing sensitivity between MR images [15]. Lesion intensities can be confused with normal structures due to non-optimal image quality [15]. Moreover, the lack of validation against multicenter data or cases with spinal cord damage has limited their application in large clinical studies [16].

In recent years, outperformance of Convolutional Neural Networks (CNNs) as compared to traditional deterministic algorithms has been proven in many domains, such as computer vision [17]. Furthermore, CNNs are known as the state-of-the-art in biomedical image segmentation, demonstrating to be highly robust to the variation of image quality characteristics [18]. The main innovation of CNN with respect to traditional machine learning models is their ability to learn which filters to apply to the input images to extract the most appropriate features to solve a specific problem, such as image classification.

In the image segmentation domain, the U-Net architecture [19] has meant an important improvement in performance. The success is built on a down-sampling path, to encode input images and get context, and an up-sampling path, to decode feature maps for a segmentation map and recover spatial information. Recent efforts have attempted to develop fully automated methods with deep learning architectures to obtain an accurate spinal cord segmentation, such as the Spinal Cord Toolbox (SCT), which is comprehensive open-source software, dedicated to the processing of spinal cord MR images [20]. Another existing spinal cord automatic segmentation method (BASICSeg) from axial T2-weighted MR imaging is based on U-Net architecture and uses pre-processed data obtained from SCT [21]. These methods use pre-processing techniques such as data transformation, region cropping, feature extraction and selection, which usually results in time consuming steps through the need for user interaction, hindering their acceptance of use as a medical device and therefore, their clinical application.

Different pre-processing techniques are often applied to also deal with the unbalanced spinal cord images problem. In the proposed methodology, we handle this limitation through the CNN design and parameterization, so that we automatically obtain the whole cervical spinal cord segmentation from real world data, i.e., no pre-processed MR images, thus avoiding time consuming techniques which entails the need of human interaction and enabling the use of common data acquired in MR systems used in usual hospital practice. Nowadays, it is easy to find real-world data sets which have irregular class distributions, such as those encountered in the MS spinal cord segmentation task, i.e., most of the data instances belong to one class and only a few instances belong to others. That is why training CNNs on very unbalanced data sets is an important research challenge [22][23]. Data with <1% of spinal cord voxels remains a continued issue of concern and needs to be addressed [24].

To overcome this issue a Residual Attention-Aware mechanism [25] was used, which has become renowned in the fields of computer vision [26] and image classification [27]. This technique generates attention-aware features that change adaptively as layers going deeper and are based on spatial features to solve the imbalanced data problem. Furthermore, the residual units reduce the vanishing gradient problem which take place in networks with too many layers, decreasing their performance and stability as the value of gradients approaches zero in early layers.

Further to a proof of concept study developed with a preliminary methodology in a reduced cohort [28], in this extended work we aim to present a novel approach for automated cervical spinal cord segmentation from MR, based on hybrid residual attention-aware mechanisms that, together with a focal loss function [29] with the Tversky-index [30] as main metric, addresses the problems of unbalanced annotation and MR imaging preprocessing. We evaluate the performance of this methodology by its application to a dataset of real-world MR images (3D-T1 weighted images acquired in a 3T system) acquired in patients suffering from MS.

Materials and methods

All models and methods were implemented in Python 3.7 and Anaconda framework using Keras (v2.2.4) and TensorFlow (v1.14.0) libraries. Model training was carried out by a scientific computer with an Intel i7 processor, working at 2.2 GHz and 16 Gb of RAM memory. The model training was carried out on a Nvidia GeForce GTX 2080 with 8 Gb memory.

Early Stopping callback of Keras was used to monitor the validation loss at each epoch during the training process and impose a stop criterion (if the validation loss had not improved in the last ten epochs, training was interrupted).

The proposed method was divided into 4 main steps (Fig. 1). After image acquisition, data was resized and normalized to avoid heterogeneous image intensities. Then, all the MR images were manually segmented through the delineation of the cervical spinal cord, leading to the ground truth dataset. Finally, part of the acquired data was used to train the designed CNN and validation metrics were obtained in the rest of the dataset (validation dataset).

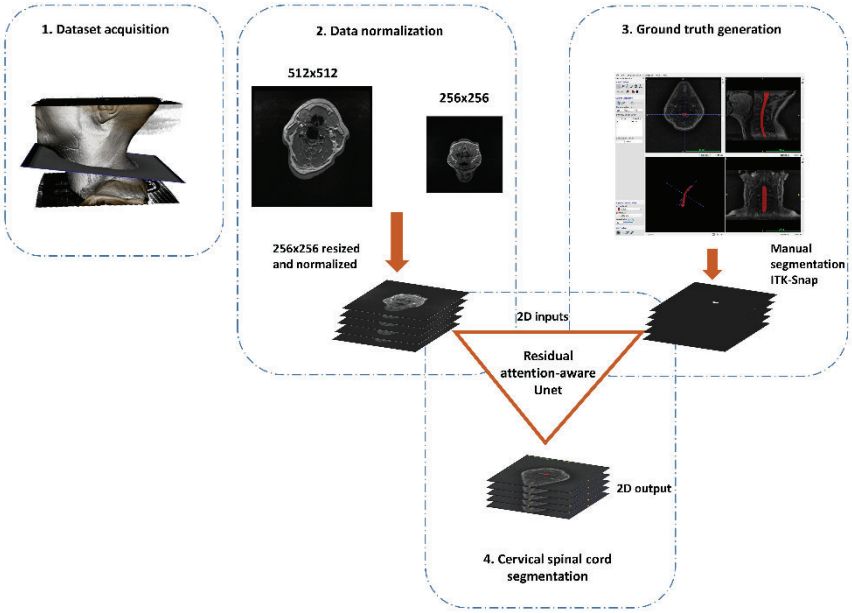


Fig. 1 Pipeline designed to build the proposed automatic cervical spinal cord segmentation. Going from the dataset generation (step 1), going through the data normalization (step 2) and expert ground truth generation (step 3), to finally train and validate a CNN based architecture (step 4).

A. Datasets

The Ethics Committee approved the observational non-interventional study for which the dataset was collected. The informed consent was waived. Anonymised MR exams of 121 MS patients from the hospital Picture Archiving and Communication System (PACS) were collected and used for the development and validation of the proposed algorithm.

The population series consisted of 34 men and 87 women, with a mean age of 45 years old (the range being 23 to 74). In terms of MS clinical course, 70% of the patients were diagnosed as Relapsing-Remiting (RR), 16% as Secondary-Progressive (SP), 8% as Primary- Progressive (PP), 5% as Clinically Isolated Syndrome (CIS) and 1% as Progressive-Relapsing (1%). We don't know a priori the number of patients with presence or not of lesions.

3D T1 weighted gradient echo, axial MR images (TE/TR/FA=1.7- 2.7ms/5.6-8.2ms/12°) were acquired on various 3T scanners (Signa HDxt and Optima MR360, GE Healthcare). All MR images had 1 mm slice thickness and an in-plane isotropic resolution from 0.43x0.43 mm to 1.02x1.02 mm.

This dataset was randomly divided into two different groups, one for training (80%, 96 cases, 15,668 images) and the other one for validation (20%, 25 cases, 4,260 images), following the traditional distribution based on the Pareto principle [31].

We selected MR images acquired in routine clinical practice for reproducibility and feasibility considerations. Cervical spinal cord region is selected for our analysis since accumulating evidence supports routine measurement of cervical spine (C-spine) area loss in MS [32] We are aware that better cervical spinal cord segmentations can be obtained through very high-resolution images, movement synchronism and higher acquisition times not used in most clinical centers.

B. Data normalization

All MR 2D images were resized to a common shape of 256 x 256 pixels using a bicubic algorithm interpolation over 4 x 4 pixels. To avoid inconsistent tissue intensities in the acquisition process, image normalization at slice level was used to set the pixel values between 0 and 1, as in (1):

$$z(i,j) = (x(i,j) - x_{min})/(x_{max} - x_{min}) \quad (1)$$

where z(i, j) was the normalized pixel value, x(i, j) the original pixel intensity and x_{min}, x_{max} the minimum and maximum image intensity values, respectively.

C. Cervical spinal cord ground truth generation

To harmonize the image size, ground truth images were also resized to a shape of 256 x 256 pixels using a nearest-neighbour interpolation. Manual labelling was performed slice by slice by two radiologists (with 12 and 30 years of experience, respectively) using the ITK-SNAP tool [33]. Discrepancies were solved by consensus. The resulting cervical spinal cord mask was considered as being a ground truth.

D. Cervical spinal cord segmentation network training and hyperparameters tunning

To reach our goal of obtaining the desired cervical spinal cord segmentation, a 2D residual attention-aware U-Net, based on a residual attention mechanism and U-Net connections, was used as architecture. The residual block enables a deeper network by having hundreds of layers, whereas the attention mechanism learns to focus on relevant locations. Regarding training design, it was made with an automatic optimal learning rate finder, a focal loss function based on the Tversky-index (FTL) [29] to address the problem of label imbalance and the Matthews correlation coefficient (MCC) as metrics during training [34].

Finally, to evaluate the performance of both the reproducibility of manual segmentations and the spatial overlap accuracy of automatic segmentation, the DICE coefficient [35] was used as statistical validation metric.

Residual Attention-Aware U-net architecture

To solve the high-class imbalance problem (proportion of spinal cord compared to the rest of the image), and due to its ability to improve the model performance, a “very deep” architecture was selected for the cervical spinal cord segmentation.

In the selected network, residual blocks were used in the different U-net layers, except the first and the last layers, to solve the gradient vanishing problem by using identity mappings as the skip connections (Fig. 2). Because the residual units directly propagate features from early to late convolution, an improvement was noted in the model performance.

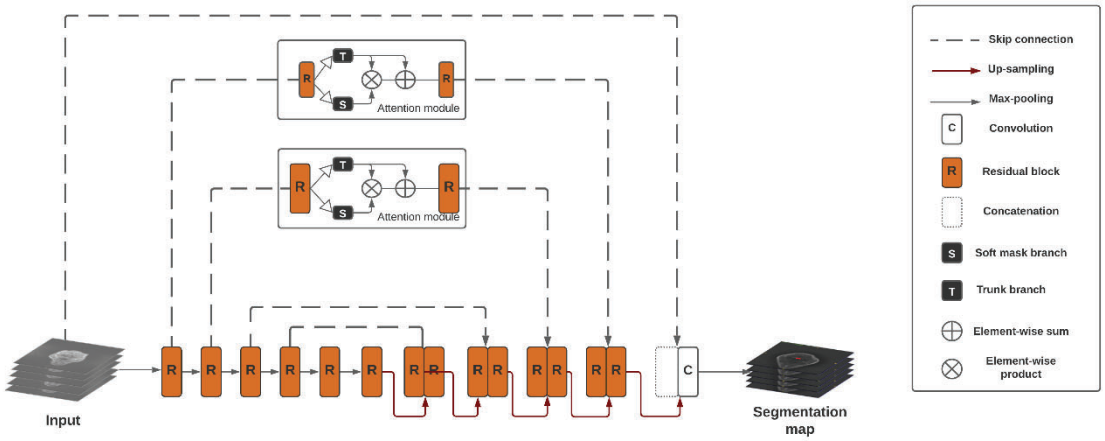


Fig. 2 Overall network architecture based on residual blocks and attention gates. Residual blocks allow a deeper network by solving the vanishing gradient problem. Attention mechanisms solve the unbalanced data problem by learning to focus on locations that are relevant.

The stacked residual blocks were introduced to mitigate the effect of network degradation on model learning by using identity mappings as the skip connections.

The residual unit was made up of a block that contains a batch normalization layer (BN), an activation (ReLU) layer, and a convolutional layer, repeated three times. The detailed residual unit is shown in Fig. 3.

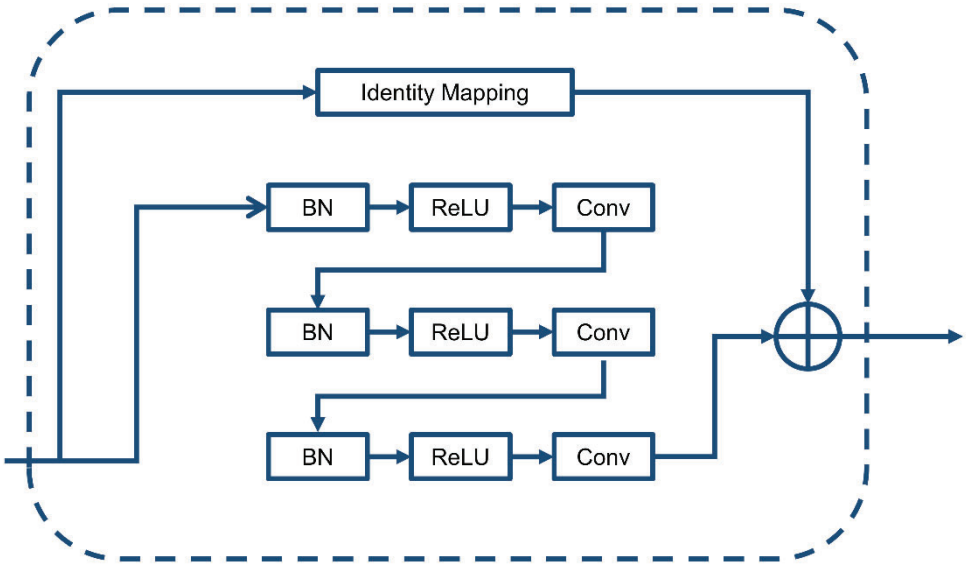


Fig.3 Residual block design. The identity mapping is performed as the skip connection, the output of the identity mapping will be added to the output of the stacked layers.

The batch normalization was employed for each convolutional layer and yields the mean and standard deviation to normalize the activation of current layer before advancing to the next layer. A convolutional identity mapping connection was used to ensure the accuracy as the network goes “deeper” [36].

Residual attention learning divided the attention module into a trunk branch, used to process the original features, and a soft branch, used to construct the identity mapping, enhancing the good features, and reducing the noise from the trunk branch.

Applied to our domain, by using this attention residual mechanism, the original feature information was preserved through the trunk branch and focused on cervical spinal cord relevant locations by the soft mask branch.

Focal Tversky Loss

The classic segmentation losses (such as DICE loss) tend to be less accurate when the training dataset is unbalanced. This is because the DICE loss function equally weights false positives (FP) and false negatives (FN), leading to segmentation maps with high precision but low recall [36]. With highly unbalanced data and small Regions of Interest (ROIs) such as the spinal cord, FN detections need to be weighted higher than FPs to improve recall rate. This was addressed by the Focal Tversky Loss index, which controls the FP and FN and controls the loss contribution in training examples, with difficult ROI detection.

$$FTL_c = \sum_c (1 - TI_c)^{\frac{1}{\gamma}}$$

Where γ [1,3] exponentiates the loss to focus on hard classes detected with lower probability and TI_c is the Tversky Similarity index

Matthews Correlation Coefficient

The Matthews correlation coefficient (MCC), is a measure unaffected by the unbalanced datasets issue which produces a high score only if the prediction provided good results in all of the four confusion matrix categories (true positives (TP), false negatives (FN), true negatives (TN), and false positives (FP)), proportionally both to the size of positive elements and the size of negative elements in the dataset. MCC reads as follows:

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP) \cdot (TP + FN) \cdot (TN + FP) \cdot (TN + FN)}}$$

Learning rate finder

A cyclical learning rate (CLR) method [37] was used, which aims to train neural network with a learning rate that changes in a cyclical way for each batch, instead of in a monotonic way. The minimum and maximum boundaries were set to $base_{lr}$ and max_{lr} (Fig. 4). Keras learning rate finder technique was used to automatically find a suitable max_{lr} and $base_{lr}$ for cyclical learning rate scheduling and apply it to our experiments (Fig. 5).

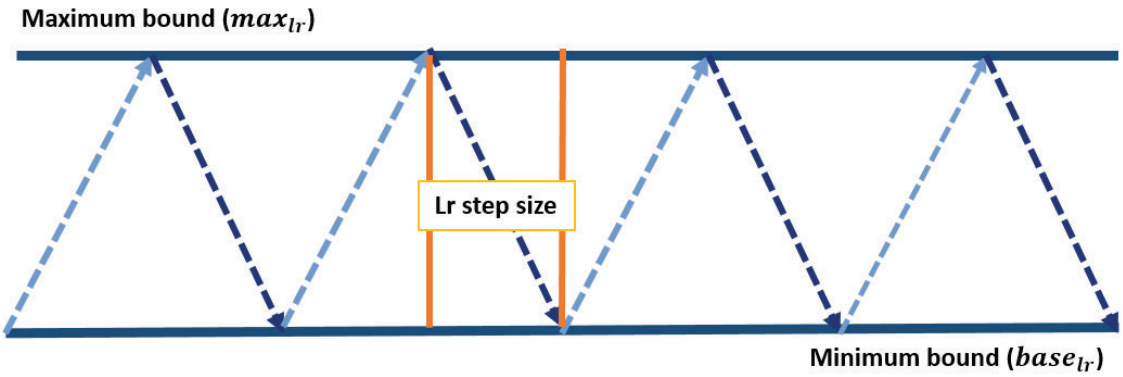


Fig. 4 Triangular learning rate policy. Orange line: learning rate value changing between bounds. Step size is the number of iterations in half a cycle.

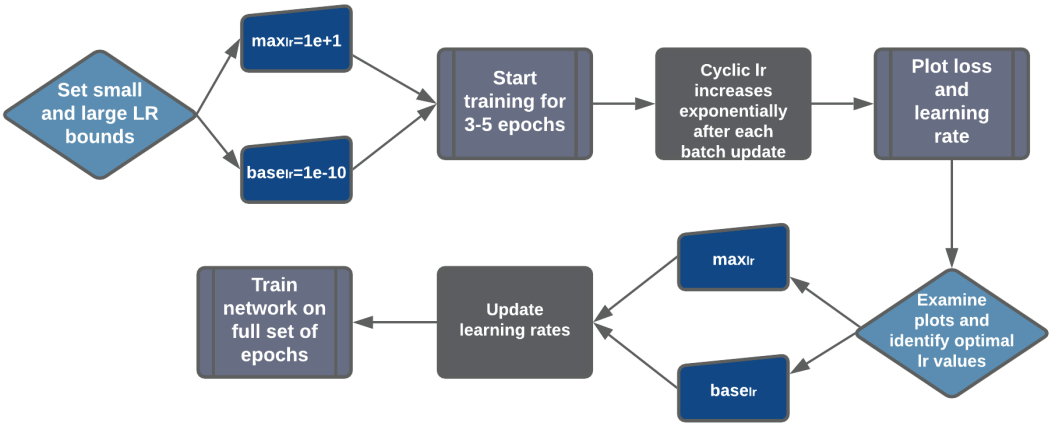


Fig. 5 Learning rate finder, basic procedure.

We started by training during 2 to 5 epochs to see the overall behaviour of our model loss function over time. The lower boundary was set to a small value ($1e^{-10}$) to let the network learn while the upper boundary was defined with a too large value ($1e^{1}$) for our model to learn. After training completion, a smoothed loss over time was plotted (Fig. 5), enabling us to define when the learning rate was just large enough for loss to decrease (optimal lower bound) and then too large, to the point where loss starts to increase (optimal upper bound). Finally, the network was fully trained along all the desired epochs.

In Fig. 6 we can see the plot of the overall behaviour of the loss function during 2 epochs, which led us to determine the optimal CLR bounds for the whole training. In our case, $max_{lr} = 1e^{-3}$ (loss started to increase) and $base_{lr} = 1e^{-5}$ (loss started to decrease).

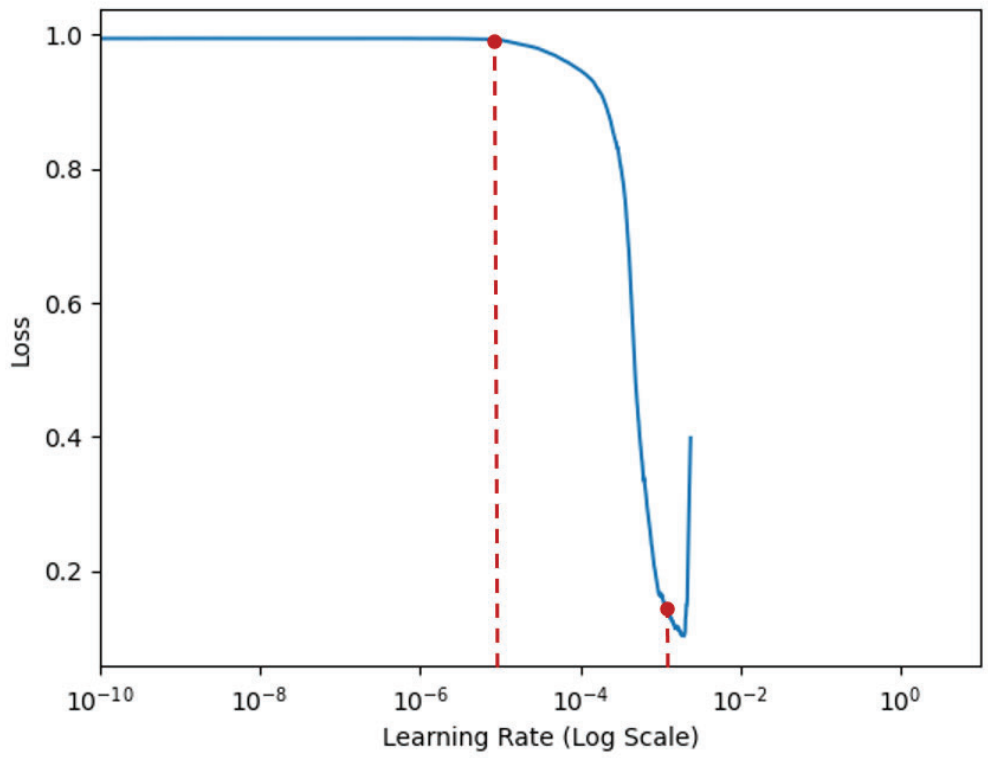


Fig. 6 Loss over learning rate to identify optimal CLR bounds, max_{lr}:loss starts to increase, $base_{lr}$: loss starts to decrease.

Once the final CLR bounds were determined, the network was trained from scratch varying the learning rate in a linear manner between the minimum ($base_{lr}$) and the maximum (max_{lr}) (Fig. 7). In this way, we were able to reduce the work in setting learning rates and improve the network performance.

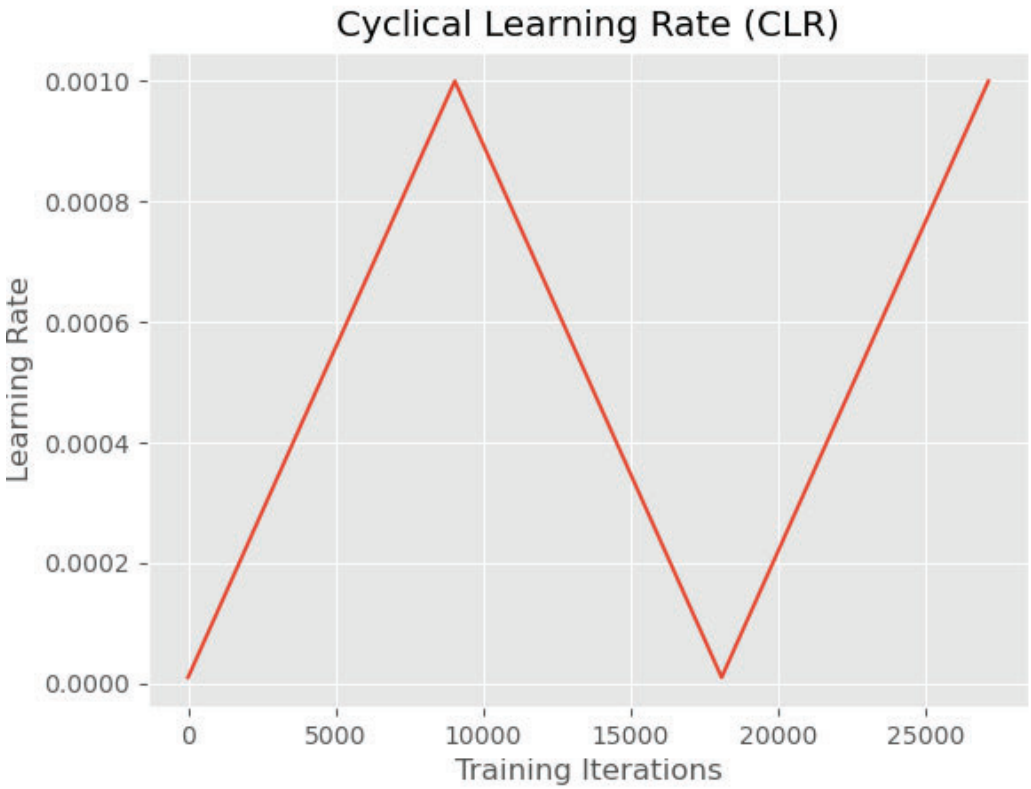


Fig. 7 Learning rate varying between CLR bounds. 'Training iterations' label is the number of weight updates per epoch, i.e., the number of total training examples divided by the batch size.

Results

An initial cervical spinal cord segmentation was performed, applying Residual Attention-Aware U-net architecture with Focal Tversky as loss function and setting the learning rate with the CLR method while training. Our model provided an automated and accurate method, achieving a MCC of 0.95 as training metric and a DICE coefficient of 0.9 as statistical validation metric. The model was trained using a batch size of 15 along 30 epochs, defined with an Early Stopping callback (the training stops after 10 epochs of not training loss improvement). Fig. 8 shows how the MCC and loss obtained evolved over the different training epochs on the validation set of our final model.

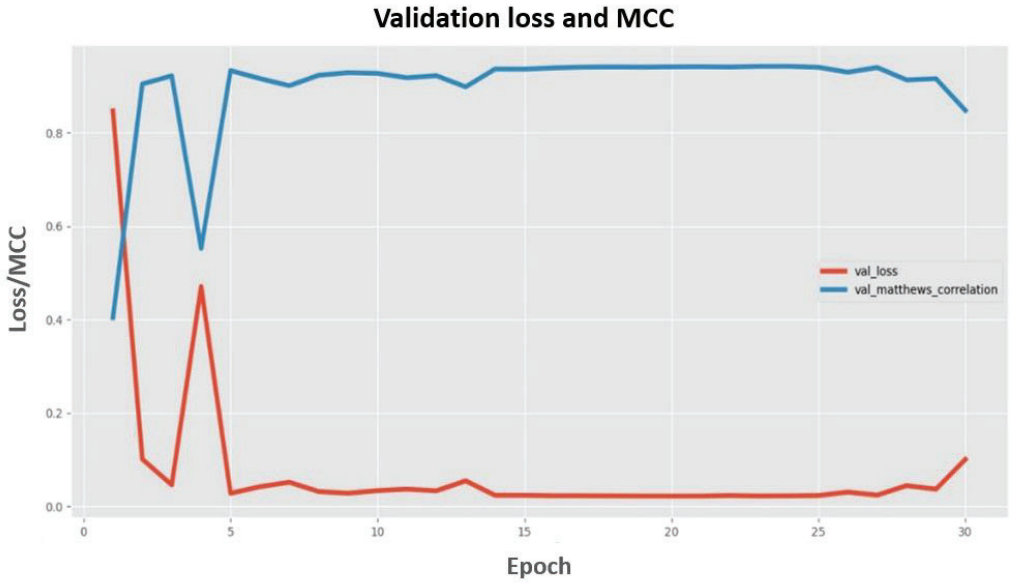


Fig. 8 Validation loss and accuracy evolution over the whole training process.

The trained method was validated on the 20% of the dataset described above (validation dataset), where the target region class is significantly smaller than the background class.

To reach the final model we did several tests by studying a large amount of variation within U-net, different learning rate methods and multiple loss functions. We decided to apply the cyclic learning rate method in those models that showed a better performance with a fixed learning rate and thus try to achieve an improvement on them.

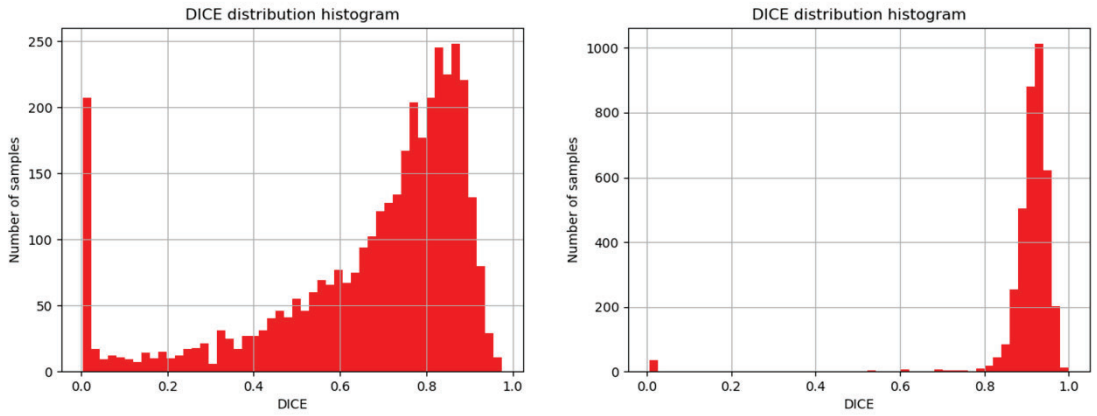
Table 1 shows the comparison of our proposed method with other state-of-the-art CNN configurations trained with our training dataset, in terms of the median DICE coefficient and their statistical parameters obtained with the data dedicated to validation.

Model	Loss function	Parameters	Learning rate	Mean DICE ± STD	MCC	P25	Median (P50)	P75
U-net *LH	TL	α=0.4, β=0.6	$1e^{-5}$	0.66±0.24	0.68	0.57	0.75	0.84
U-net	TL	α=0.3, β=0.7	$1e^{-5}$	0.68±0.24	0.71	0.61	0.77	0.85
U-net	FTL	γ=1.33, α=0.7, β=0.3	$1e^{-5}$	0.62±0.25	0.64	0.53	0.71	0.81
U-net	FTL	γ=1.33, α=0.6, β=0.4	$1e^{-5}$	0.68±0.27	0.70	0.62	0.80	0.87
U-net	FTL	γ=1.6, α=0.6, β=0.4	$1e^{-5}$	0.74±0.23	0.77	0.70	0.82	0.88
RA-U-net	FTL	γ=1.33, α=0.6, β=0.4	$1e^{-5}$	0.86±0.11	0.89	0.85	0.89	0.92
RA-U-net	FTL	γ=1.6, α=0.6, β=0.4	$1e^{-4}$	0.80±0.22	0.84	0.82	0.87	0.90
Deep U-net	FTL	γ=1.33, α=0.6, β=0.4	CLR ($max_{lr} = 1e^{-2}, base_{lr} = 1e^{-4}$	0.79±0.25	0.83	0.82	0.89	0.92
RA-U-net *RH	FTL	γ=1.33, α=0.6, β=0.4	CLR ($max_{lr} = 1e^{-3}, base_{lr} = 1e^{-5}$	0.904±0.101	0.95	0.90	0.92	0.94

Table 1. Results and comparison between model and loss function schemes on validation dataset. Column "Loss function": Tversky Loss (TL), Focal Tversky Loss (FTL). Column "parameters" refers to loss function data. Column

"models": U-net, RA-U-net, Deep U-net [38]. Column "MCC": Matthews Correlation Coefficient, final training metric. Column "P25": 25th Percentile. Column "Median(P50)": Median DICE and 50th Percentile. Column "P75": 75th Percentile. *LH: Left Histogram. *RH: Right Histogram.

Models trained with the U-net algorithm generally performed poorly, even though Tversky Loss and Focal Tversky Loss were used as loss functions. Using RA-U-net as CNN architecture offered an improved performance. The model obtained with the Deep U-net algorithm also showed a slight improvement over those trained with U-net, since deep networks have a higher learning capability. Finally, the results show that Residual Attention-aware U-net architecture with a CLR finder has the best performance, i.e., high DICE coefficient, low standard deviation (STD) and high, similar percentile values. In Fig. 9 the validation dataset DICE distribution histogram of the first (*LH) and the final (*RH) trained models are shown. Fig.9 shows that with a traditional U-Net architecture a wider DICE histogram is generated, which means that lower DICE values are obtained in a high number of samples. However, the DICE values obtained with the proposed architecture and configuration are associated with a narrower histogram with almost all the values concentrated over a DICE value of 0.8.



(a) (b)

Fig. 9 DICE distribution histogram. a) First model distribution, scattered distribution. b) Final model distribution, narrow distribution.

DICE values obtained along the different training epochs are represented in Fig. 10 comparing the results obtained using a CLR or a fixed learning rate of $1e^{-4}$ (a common value chosen in the literature). It is possible to realize how by using a CLR, higher accuracy was achieved with fewer experiments and limited hyperparameter tuning. In our case, by training with CLR a higher and earlier (around 20 epochs before) DICE were obtained when compared to the fixed learning rate.

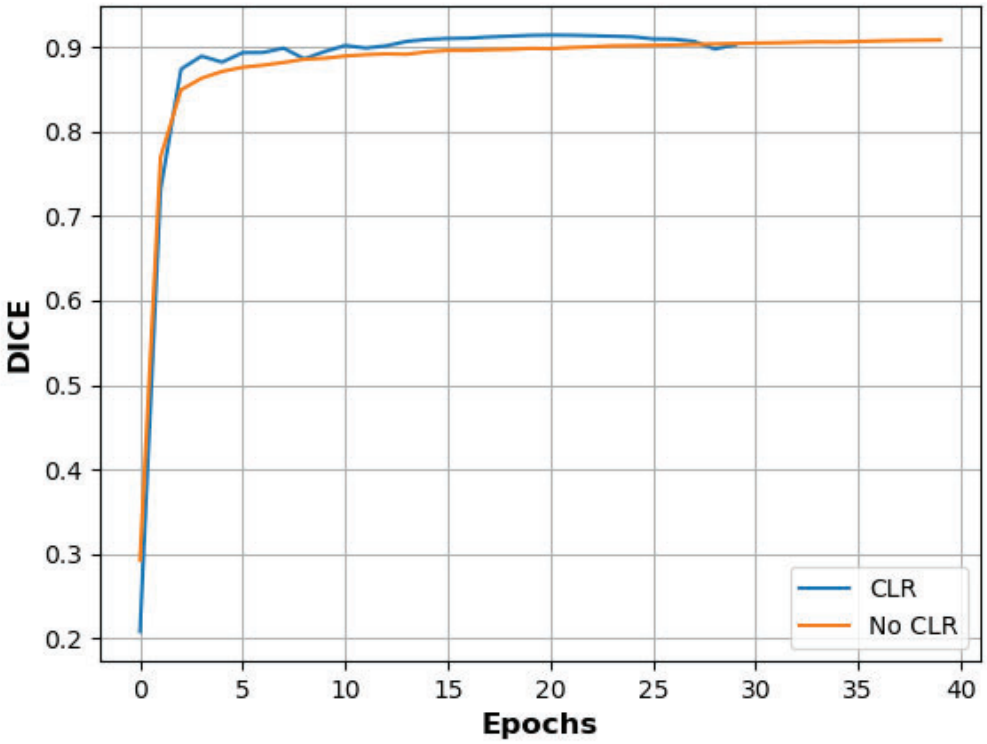


Fig. 10 DICE evolution during training epochs using a CLR and a fixed learning rate of 1^{e-4}. With CLR, a higher DICE is achieved in a lower number of epochs.

The performance of the developed model over a randomly selected axial slice is shown in Fig. 11. Fig. 11a shows the real-world MR image that is given as an input to the network. The red box indicates the part of the 2D axial image that is zoomed in the remaining images, to observe each detail clearly. Ground truth (manually labelled mask) over real world data can be depicted in Fig. 11b, whereas Fig. 11c presents the mask automatically predicted by the proposed model. To better compare these two masks, the overlap of both is shown in Fig. 11d. The red voxels identified the mismatched area (pixels that are only present in one of the masks) and the green voxels indicate the matching area (pixels that belong to both the manual and predicted mask).

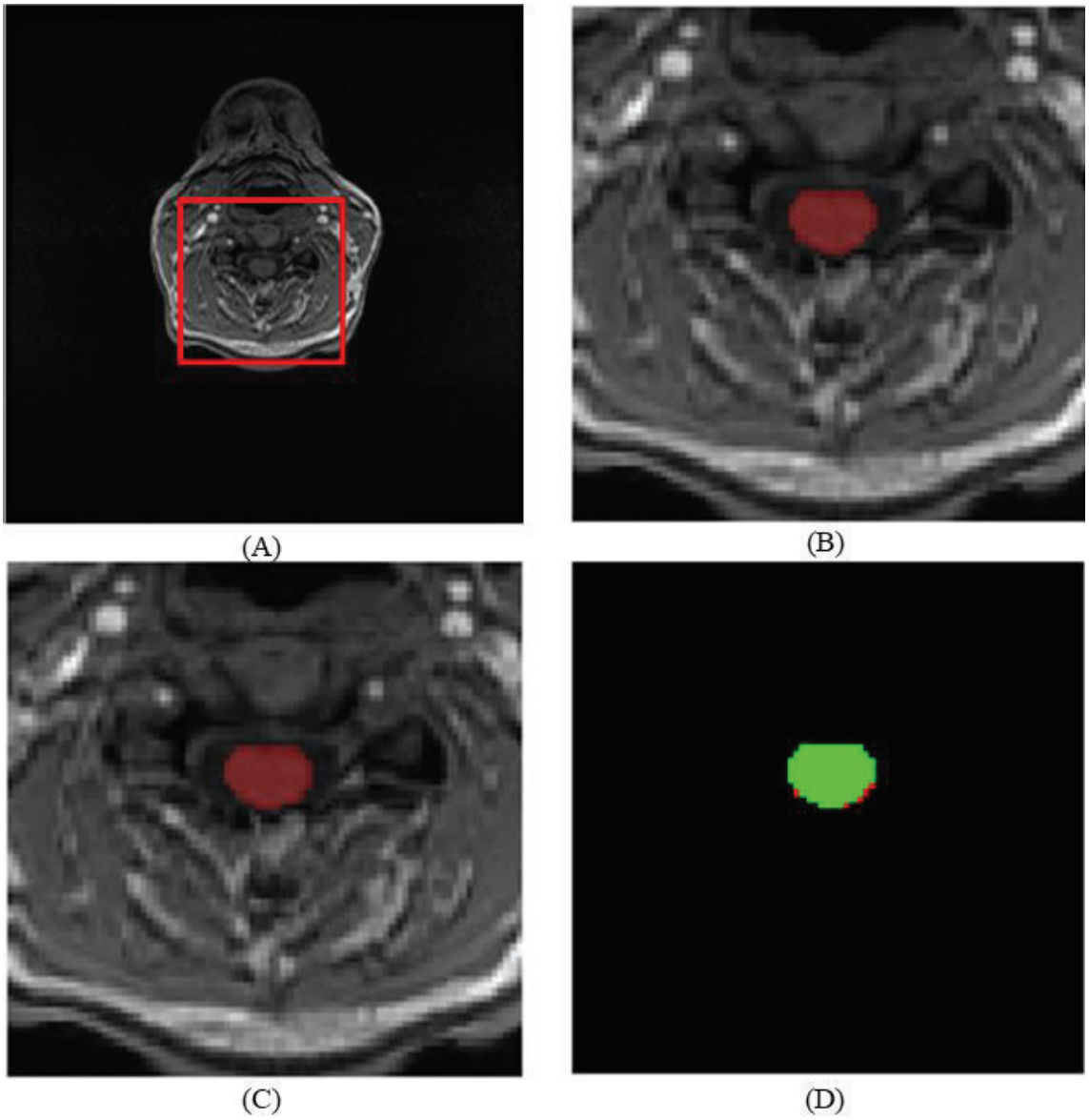


Fig. 11 a) Real world MR data, b) ground truth, c) CNN predicted mask, d) ground truth over predicted mask (green-matched area, red-mismatched area).

According to the Fig. 11, only one random example from more than 2,000 axial images that make up the validation dataset is shown. It is observed that our model achieves a successful automatic prediction which hardly differs from that performed by the expert radiologist. Despite good overall performance, there are some situations that show that our automatic segmentation tool can make some prediction mistakes. One example of that can be seen on Fig. 12 where a segmentation error example is shown. In 3D and sagittal views, we can see two inhomogeneities that are surrounded in yellow. These segmentation failures, that appear as two spikes toward the end of the cervical spinal cord, are caused by the settings in slices 35, 36, and 37. In these images, we can see that the cervical spinal cord portion is very close to another anatomical structure of similar intensity that seem to be practically connected. That is why the prediction mask is bigger than those obtained for slices 34, 38, and 39 (where the cervical spinal cord region is more separated from that anatomical structure) and, therefore, the segmentation mask is fitted to the cervical spinal cord region and there is no confusion.

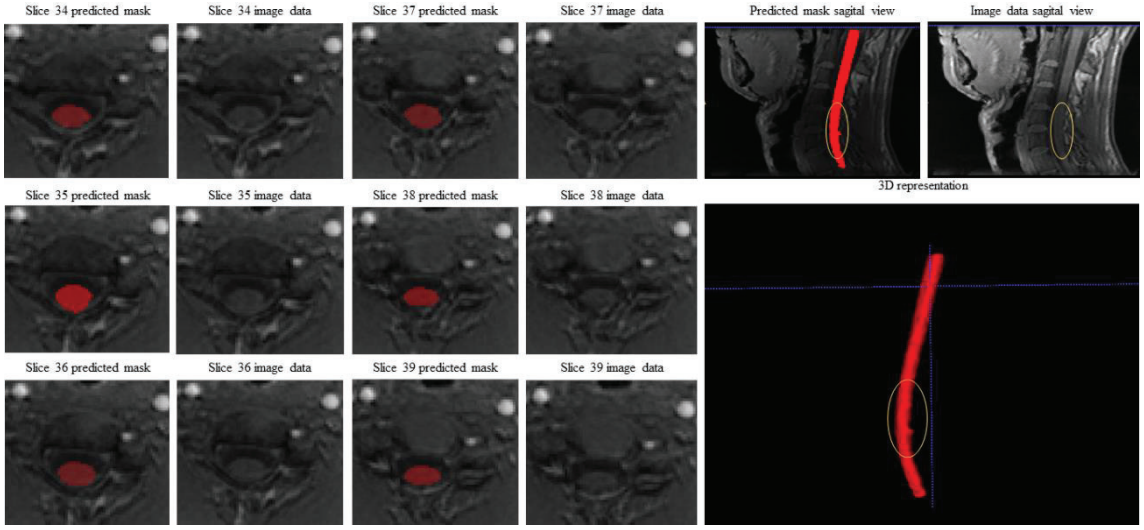


Fig. 12 Poor segmentation performance setting. Slices 35, 36, and 37 show the cervical spinal cord very close to another anatomical structure that could cause oversized prediction masks and is the last inhomogeneity that is surrounded in 3D and sagittal views. Slices 34, 38, and 39 show a well-delimited cervical spinal cord region which leads us to a correct segmentation mask, and these correspond to the adjacent peak areas.

Discussion

In the present study, we obtained an accurate, automatic cervical spinal cord segmentation from 3D-T1 weighted gradient echo MR axial images. Due to the huge variety of MR scanners and the absence of standard acquisition MR images in the real-world data, MR images of MS patients differ as they cover different spinal cord regions with different tissue intensities, different fields of view, different pixel spacing and signal intensity. The major proposal of this work is to handle the cervical spinal cord segmentation by the CNN design and parameterization, avoiding pre-processing MR images techniques that require human interaction, so that an automatic segmentation can be obtained from common data acquired in MR systems used in usual hospital practice. The cervical spinal cord manual segmentation involves a time consuming task for radiologists, so it will be of useful assistance for them and it will be an important prior step for medical analysis issues like the cervical spinal cord atrophy measurement by calculating the volumetry, as it's proved that a volume reduction has direct implications with the atrophy.

This tool was developed thinking about solving some deficiencies that other models present, such as the SCT. This method is based on a sequence of two CNNs and achieved a median DICE coefficient of 0.95. The first CNN with 2D-dilated convolutions detects the spinal cord centreline being necessary human interaction to initialize the spinal cord location (e.g., three points in the spinal cord) for an optimal performance and also carrying out the next CNN for spinal cord segmentation in a framework that is more sensitive to the quality of that detection module.

There is another method that depends on the previous one, for automatic segmentation of the spinal cord after traumatic contusion injury from axial T2-weighted MR imaging and whose DICE coefficient is 0.93. Firstly, the optic algorithm [39] is used from the SCT to first detect the spinal cord centerline and create a square mask around it. Secondly, three different novel 2D-CNN architectures are applied to achieve the segmentation. Moving away from our method generalization purpose, this study cohort is derived from a single institution with all the imaging being performed on a single MR imaging scanner using similar parameters, potentially biasing the results. Furthermore, the pre-processing step of cropping the image around the spinal cord centerline depends on a third-party tool.

In our study, overcoming what we understand as other methods limitations and with the aim of avoiding any type of pre-processing techniques and reducing human interaction, a 2D residual attention-aware U-Net, based on a residual attention mechanism and U-Net connections, was implemented to obtain the desired spinal cord segmentation. We selected a deep architecture since it is proved that deeper networks have a higher learning capacity, although they show some drawbacks like the gradient vanishing problem

(when carrying out back-propagation) and the degradation issue, which results in poor training results. The residual blocks were stacked in the U-net layers, except the first and last layers, to solve the vanishing gradient problem by allowing the network to have hundreds of layers. The attention mechanism learns to focus on locations that are relevant by allowing the retention of the original feature information through the trunk branch and paying attention to those spinal cord regions by the soft mask branch. The training was designed with a focal loss function, based on the Tversky-index, to address the problem of label imbalance in medical image segmentation and an automatic optimal learning rate finder. Finally, a DICE coefficient of 0.904 was achieved.

As we mentioned during the paper, the usefulness of this model, and as future work in development, this segmentation algorithm will allow the quantification of the cervical spinal cord volume and, therefore, atrophy assessment through the comparison of the obtained volume against that associated to a healthy population. In addition, this model will help to develop an aggregated lesion detection model from STIR image sequences which will be focused on the lesion localization on the cervical spine region, since it would be used as first step by obtaining the cervical spinal cord location in the corresponding STIR space after applying image registration algorithms between both sequences.

Limitations detected during our study will be investigated in future works. Some of them primarily relate to the relatively small sample size of patients. Consequently, to proceed with a K-fold cross-validation scheme would have been reasonable as it ensures that every observation from our limited dataset has the chance of appearing in training and validation set. A performance comparison between existing tools would better clarify the quality of our proposed model compared to others although this can sometimes be complicated since not all of them work with the same MR image types. The need for external validation is one of them; our method should be externally validated in other centers using different MR scanners from different vendors, magnetic fields, and acquisition parameters. In this way, we will be able to evaluate the interoperability of our method. Increasing patient numbers will improve our database and lead us to a greater number of training samples, therefore, achieving fuller training and better performance.

Conclusion

An optimal automated method for cervical spinal cord segmentation in real world MRI of MS patients was implemented by optimized hybrid residual attention-aware CNN with Focal Tversky as loss function and CLR method to set, change and tweak the learning rate during training. A dataset consisting of 3D T1-weighted images from 121 MS patients was used to train and validate the algorithm. Our method can address the class imbalance problem without pre-processing techniques, so by using directly real-world data, a better clinical integration can be achieved. Lastly, a comparison of our architecture in combination with several loss functions and learning rate methods and values is presented. The main contributions of this proposed method are the lack of human interaction, the use of real-world data to approach implementable solutions in usual clinical practice, the minimization of the unbalanced data problem without pre-processing techniques, and the time optimisation due to the CLR method.

The proposed automatic segmentation model will have direct implications for accelerating the MS phenotyping, follow-up, and extraction of imaging biomarkers, since we achieve a time optimisation by subtracting effort from the segmentation task and, moreover, increasing resources in the disease's evolution monitoring in order to define the factors behind the increase of irreversible disability, the principal MS symptom. These implications are addressed in subsequent studies, since the purpose of this paper is to reflect the first segmentation method necessary to develop them.

REFERENCES AND FOOTNOTES

1. Matthews PM, De Stefano N, Narayanan S, et al (1998) Putting magnetic resonance spectroscopy studies in context: Axonal damage and disability in multiple sclerosis. Semin. Neurol. 18:327–336

2. Magraner MJ, Bosca I, Simó-Castelló M, et al (2012) Brain atrophy and lesion load are related to CSF lipid-specific IgM oligoclonal bands in clinically isolated syndromes. In: Neuroradiology. pp 5–12

3. Cordovez M J, Gálvez G M, Rojas C G, et al (2013) Uso de volumetría y carga lesional en el seguimiento de pacientes con esclerosis múltiple. Experiencia local y revisión de la literature. Rev Chil Radiol 19:156–164. https://doi.org/10.4067/S0717-93082013000400004

4. Lassmann H (2018) Multiple sclerosis pathology. Cold Spring Harb. Perspect. Med. 8:a028936

5. Bakshi R, Dandamudi VSR, Neema M, et al (2005) Measurement of brain and spinal cord atrophy by magnetic resonance imaging as a tool to monitor multiple sclerosis. J. Neuroimaging 15

6. Bjartmar C, Kidd G, Mörk S, et al (2000) Neurological disability correlates with spinal cord axonal loss and reduced N-acetyl aspartate in chronic multiple sclerosis patients. Ann Neurol 48:893–901. https://doi.org/10.1002/1531-8249(200012)48:6<893::AID-ANA10>3.0.CO;2-B

7. Trapp BD, Ransohoff RM, Fisher E, Rudick RA (1999) Neurodegeneration in Multiple Sclerosis: Relationship to Neurological Disability. Neurosci 5:48–57. https://doi.org/10.1177/107385849900500107

8. Cohen AB, Neema M, Arora A, et al (2012) The Relationships among MRI-Defined Spinal Cord Involvement, Brain Involvement, and Disability in Multiple Sclerosis. J Neuroimaging 22:122–128. https://doi.org/10.1111/j.1552-6569.2011.00589.x

9. Lundell H, Svolgaard O, Dogonowski AM, et al (2017) Spinal cord atrophy in anterior-posterior direction reflects impairment in multiple sclerosis. Acta Neurol Scand 136:330–337. https://doi.org/10.1111/ane.12729

10. Valsasina P, Rocca MA, Horsfield MA, et al (2013) Regional cervical cord atrophy and disability in multiple sclerosis: A voxel-based analysis. Radiology 266:853–861. https://doi.org/10.1148/radiol.12120813

11. Filippi M, Agosta F (2010) Imaging biomarkers in multiple sclerosis. J. Magn. Reson. Imaging 31:770–788

12. Yiannakas MC, Mustafa AM, De Leener B, et al (2016) Fully automated segmentation of the cervical cord from T1-weighted MRI using PropSeg: Application to multiple sclerosis. NeuroImage Clin 10:71–77. https://doi.org/10.1016/j.nicl.2015.11.001

13. De Leener B, Taso M, Cohen-Adad J, Callot V (2016) Segmentation of the human spinal cord. Magn. Reson. Mater. Physics, Biol. Med. 29:125–153

14. Horsfield MA, Sala S, Neema M, et al (2010) Rapid semi-automatic segmentation of the spinal cord from magnetic resonance images: Application in multiple sclerosis. Neuroimage 50:446–455. https://doi.org/10.1016/j.neuroimage.2009.12.121

15. Chen M, Carass A, Oh J, et al (2013) Automatic magnetic resonance spinal cord segmentation with topology constraints for variable fields of view. Neuroimage 83:1051–1062. https://doi.org/10.1016/j.neuroimage.2013.07.060

16. Jovicich J, Czanner S, Greve D, et al (2006) Reliability in multi-site structural MRI studies: Effects of gradient non-linearity correction on phantom and human data. Neuroimage 30:436–443. https://doi.org/10.1016/j.neuroimage.2005.09.046

17. Szlávik Z, Szirányi T (2004) Face Analysis Using CNN-UM

18. Litjens G, Kooi T, Bejnordi BE, et al (2017) A survey on deep learning in medical image analysis. Med. Image Anal. 42:60–88

19. Ronneberger O, Fischer P, Brox T (2015) U-net: Convolutional networks for biomedical image segmentation. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Springer Verlag, pp 234–241

20. De Leener B, Lévy S, Dupont SM, et al (2017) SCT: Spinal Cord Toolbox, an open-source software for processing spinal cord MRI data. Neuroimage 145:24–43. https://doi.org/10.1016/j.neuroimage.2016.10.009

21. McCoy DB, Dupont SM, Gros C, et al (2019) Convolutional neural network–based automated

segmentation of the spinal cord and contusion injury: Deep learning biomarker correlates of motor impairment in acute spinal cord injury. Am J Neuroradiol 40:737–744. https://doi.org/10.3174/ajnr.A6020

22. Lin T-Y, Goyal P, Girshick R, et al (2017) Focal Loss for Dense Object Detection

23. Montahaei E, Ghorbani M, Baghshah MS, Rabiee HR (2018) Adversarial Classifier for Imbalanced Problems

24. Buda M, Maki A, Mazurowski MA (2017) A systematic study of the class imbalance problem in convolutional neural networks. https://doi.org/10.1016/j.neunet.2018.07.011

25. Wang F, Jiang M, Qian C, et al (2017) Residual attention network for image classification. In: Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017. Institute of Electrical and Electronics Engineers Inc., pp 6450–6458

26. Gao L, Li Y, Ning J (2019) Residual attention convolutional network for online visual tracking. IEEE Access 7:94097–94105. https://doi.org/10.1109/ACCESS.2019.2927791

27. Huang P, Wang J, Zhang J, et al (2020) Attention-aware Residual Network based Manifold Learning for White Blood Cells Classification. IEEE J Biomed Heal Informatics 1–1. https://doi.org/10.1109/jbhi.2020.3012711

28. Bueno Gómez A, Alberich-Bayarri A, Bosch I, Carreres Polo J (2021) Automatic MR Spinal Cord Segmentation by Hybrid Residual Attention-Aware Convolutional Neural Networks and Learning Rate Optimization on Real World Data. In: IFMBE Proceedings. Springer Science and Business Media Deutschland GmbH, pp 158–168

29. Abraham N, Khan NM (2018) A Novel Focal Tversky loss function with improved Attention U-Net for lesion segmentation. Proc - Int Symp Biomed Imaging 2019-April:683–687

30. Salehi SSM, Erdogmus D, Gholipour A (2017) Tversky loss function for image segmentation using 3D fully convolutional deep networks. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Springer Verlag, pp 379–387

31. The Pareto Principle | Dunford | The Plymouth Student Scientist. https://bcur.org/journals/index.php/TPSS/article/view/408. Accessed 9 Dec 2019

32. Mina Y, Azodi S, Dubuche T, et al (2021) Cervical and thoracic cord atrophy in multiple sclerosis phenotypes: Quantification and correlation with clinical disability. NeuroImage Clin 30:102680. https://doi.org/10.1016/J.NICL.2021.102680

33. Yushkevich PA, Gerig G (2017) ITK-SNAP: An Intractive Medical Image Segmentation Tool to Meet the Need for Expert-Guided Segmentation of Complex Medical Images. IEEE Pulse 8:54–57. https://doi.org/10.1109/MPUL.2017.2701493

34. Chicco D, Jurman G The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. https://doi.org/10.1186/s12864-019-6413-7

35. Eelbode T, Bertels J, Berman M, et al (2020) Optimization for Medical Image Segmentation: Theory and Practice When Evaluating With Dice Score or Jaccard Index. IEEE Trans Med Imaging 39:3679–3690. https://doi.org/10.1109/TMI.2020.3002417

36. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, pp 770–778

37. Smith LN (2017) Cyclical learning rates for training neural networks. In: Proceedings - 2017 IEEE Winter Conference on Applications of Computer Vision, WACV 2017. Institute of Electrical and Electronics Engineers Inc., pp 464–472

38. Li R, Liu W, Yang L, et al (2017) DeepUNet: A Deep Fully Convolutional Network for Pixel-level Sea-Land Segmentation. IEEE J Sel Top Appl Earth Obs Remote Sens 11:3954–3962

39. Gros C, De Leener B, Dupont SM, et al (2018) Automatic spinal cord localization, robust to MRI

contrasts using global curve optimization. Med Image Anal 44:215–227.
https://doi.org/10.1016/j.media.2017.12.001

Appendix: Supplementary material

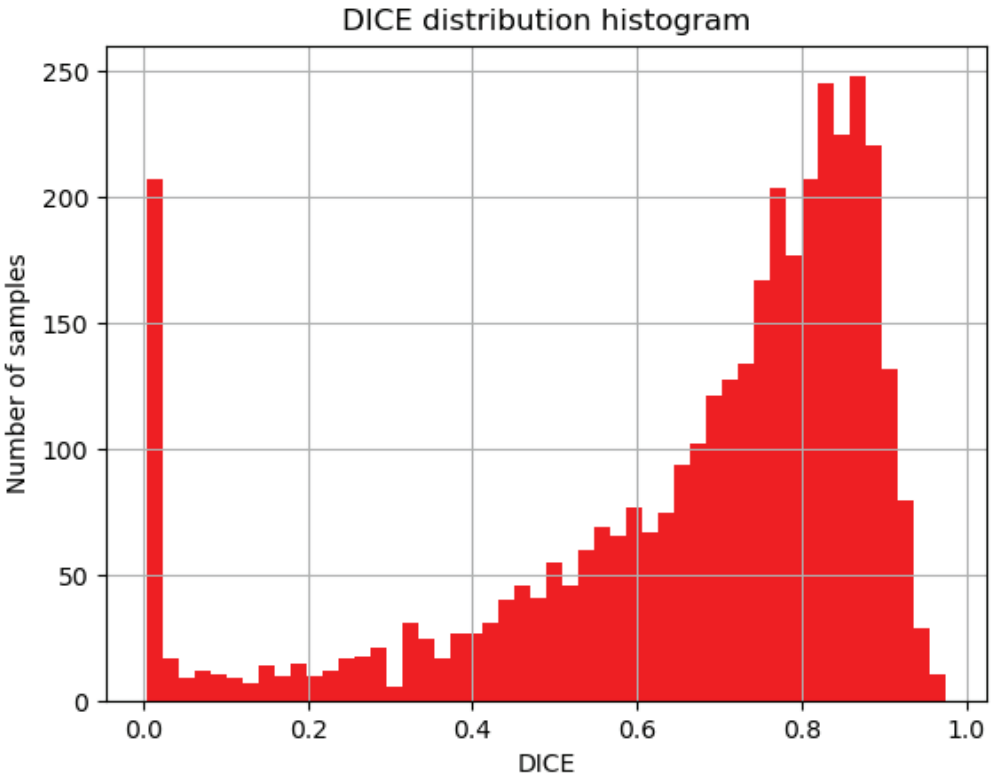


Fig. 13 DICE distribution histogram. Model: U-net. Loss function: TL. Parameters: α=0.4, β=0.6. Learning rate: $1e^{-5}$. Median DICE ±STD: 0.66±0.24. MCC: 0.68. P25: 0.57. P50: 0.75. P75: 0.84.

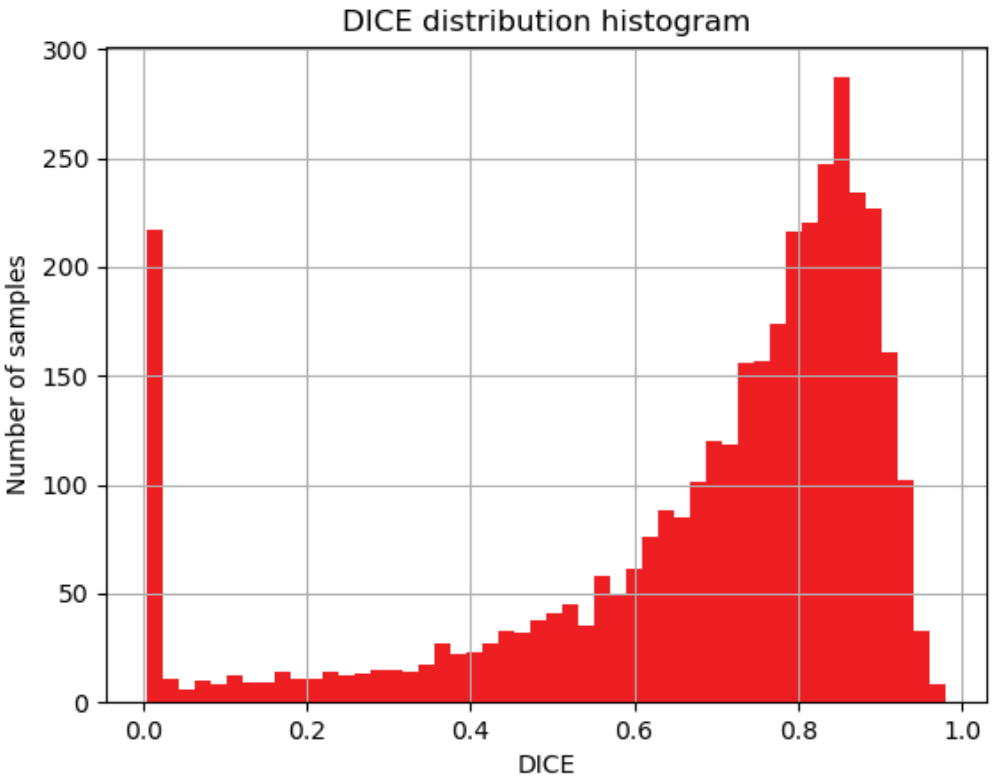


Fig. 14 DICE distribution histogram. Model: U-net. Loss function: TL. Parameters: α=0.3, β=0.7. Learning rate: $1e^{-5}$. Median DICE ±STD: 0.68±0.24. MCC: 0.71. P25: 0.61. P50: 0.77. P75: 0.85.

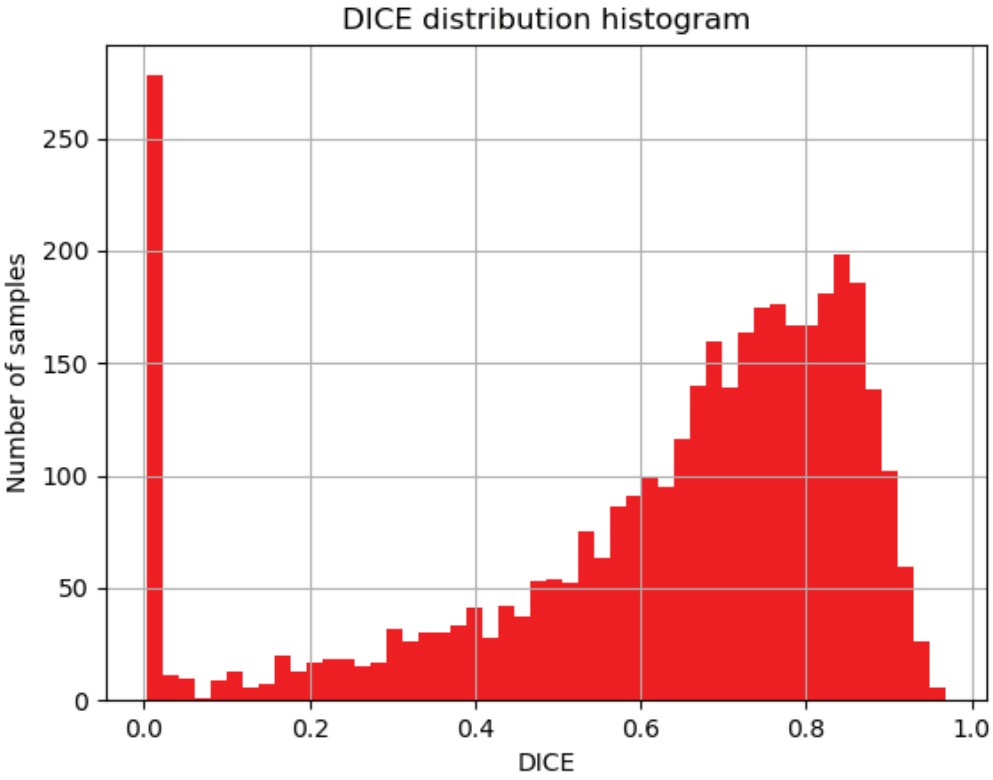


Fig. 15 DICE distribution histogram. Model: U-net. Loss function: FTL. Parameters: γ=1.33, α=0.7, β=0.3. Learning rate: $1e^{-5}$. Median DICE ±STD: 0.62±0.25. MCC: 0.64. P25: 0.53. P50: 0.71. P75: 0.81.

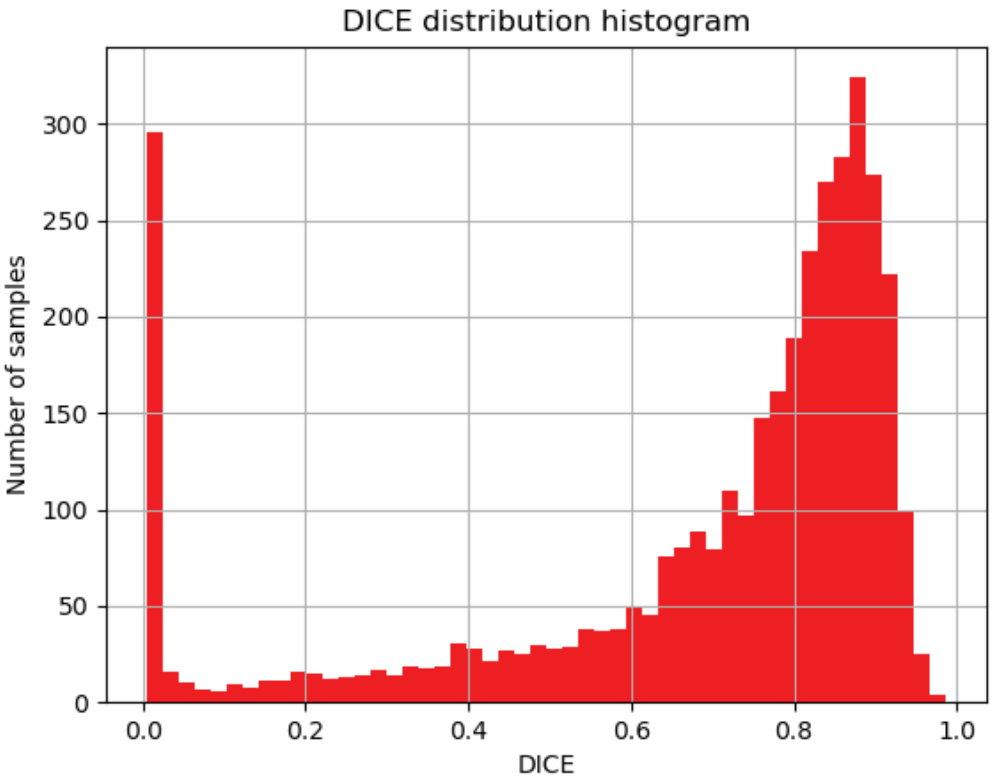


Fig. 16 DICE distribution histogram. Model: U-net. Loss function: FTL. Parameters: γ=1.33, α=0.6, β=0.4. Learning rate: $1e^{-5}$. Median DICE ±STD: 0.68±0.27. MCC: 0.70. P25: 0.62. P50: 0.80. P75: 0.87.

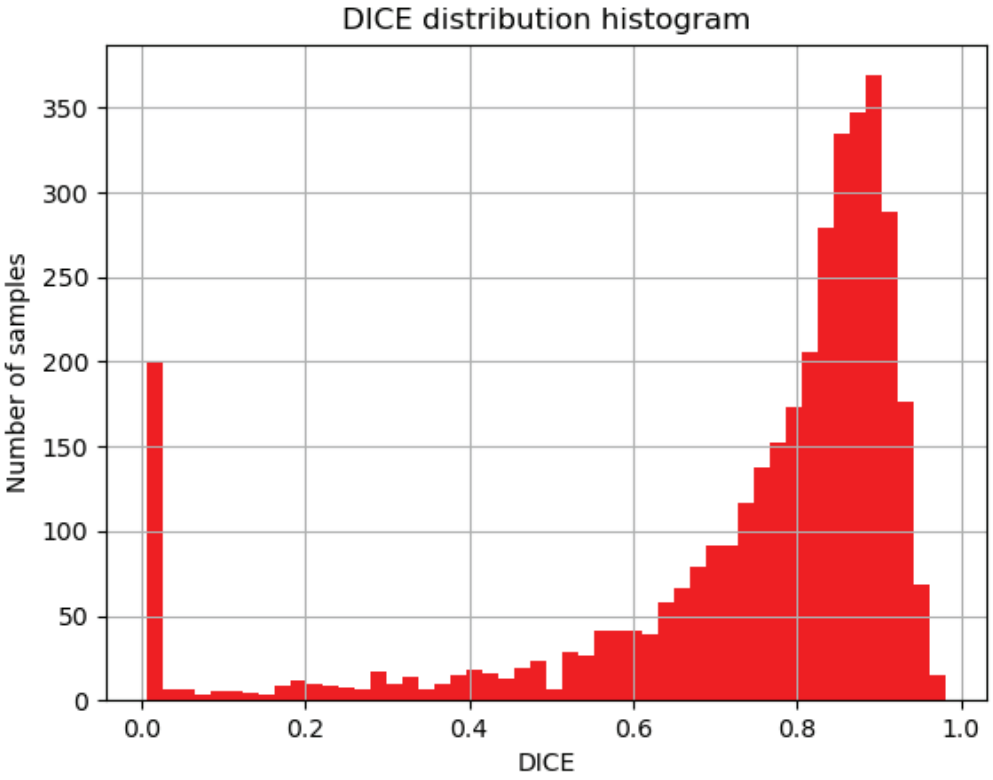


Fig. 17 DICE distribution histogram. Model: U-net. Loss function: FTL. Parameters: γ=1.6, α=0.6, β=0.4. Learning rate: $1e^{-5}$. Median DICE ±STD: 0.74±0.23. MCC: 0.77. P25: 0.70. P50: 0.82. P75: 0.88.

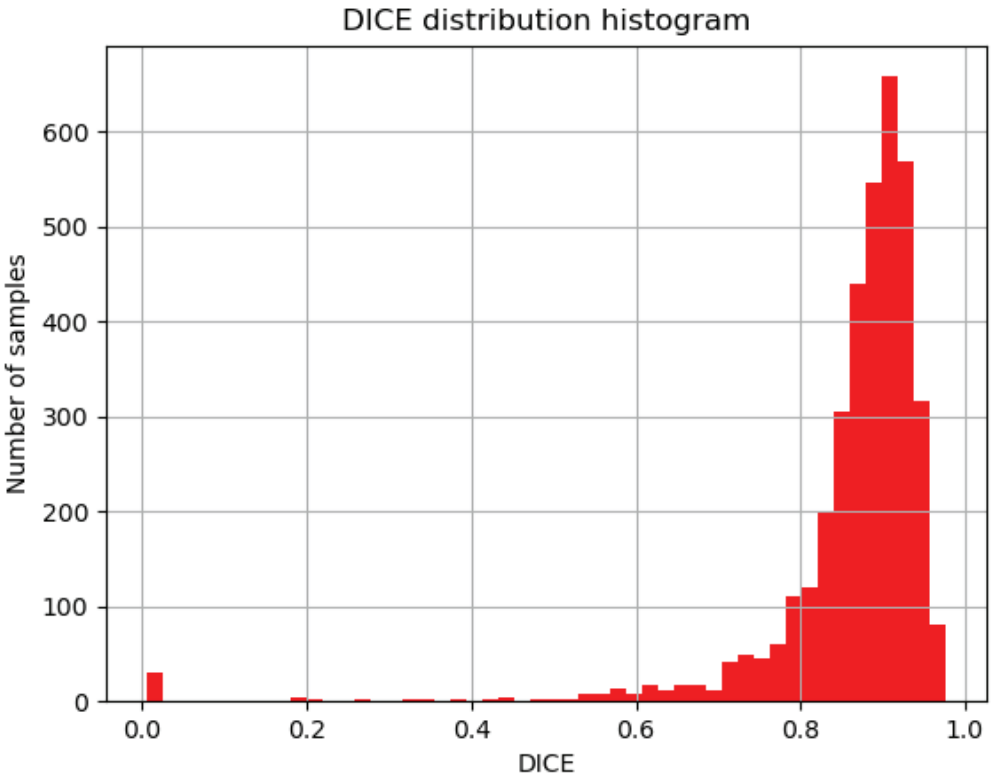


Fig. 18 DICE distribution histogram. Model: RA-U-net. Loss function: FTL. Parameters: γ=1.33, α=0.6, β=0.4. Learning rate: $1e^{-5}$. Median DICE ±STD: 0.86±0.11. MCC: 0.89. P25: 0.85. P50: 0.89. P75: 0.92.

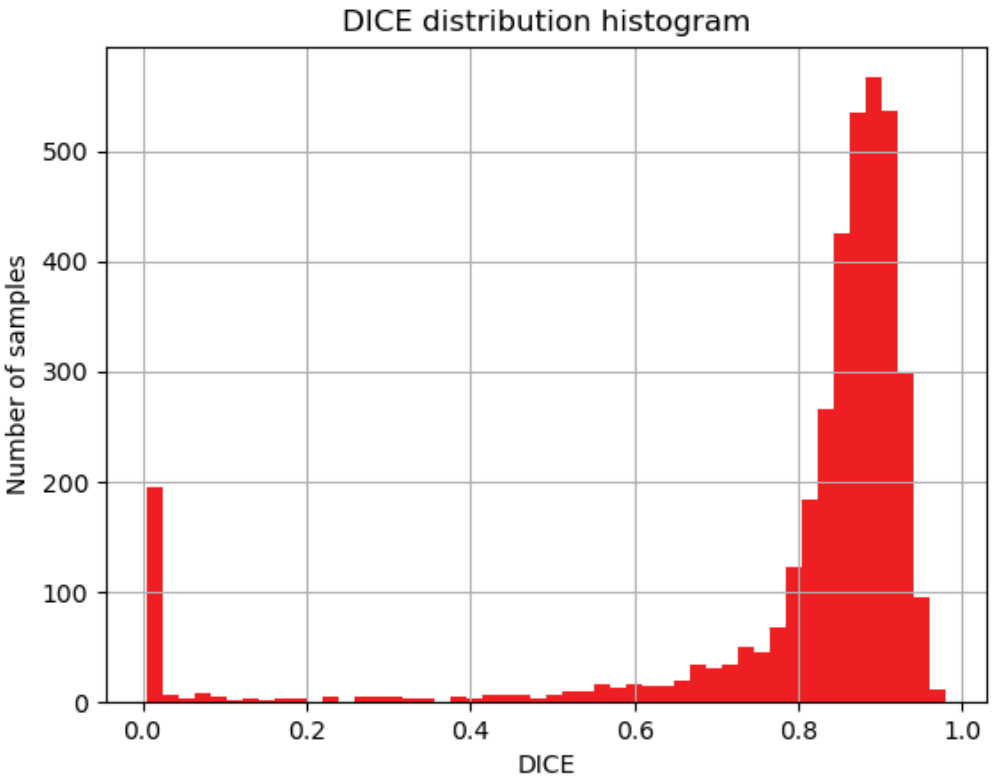


Fig. 19 DICE distribution histogram. Model: RA-U-net. Loss function: FTL. Parameters: γ=1.6, α=0.6, β=0.4. Learning rate: $1e^{-4}$. Median DICE ±STD: 0.80±0.22. MCC: 0.84. P25: 0.82. P50: 0.87. P75: 0.90.

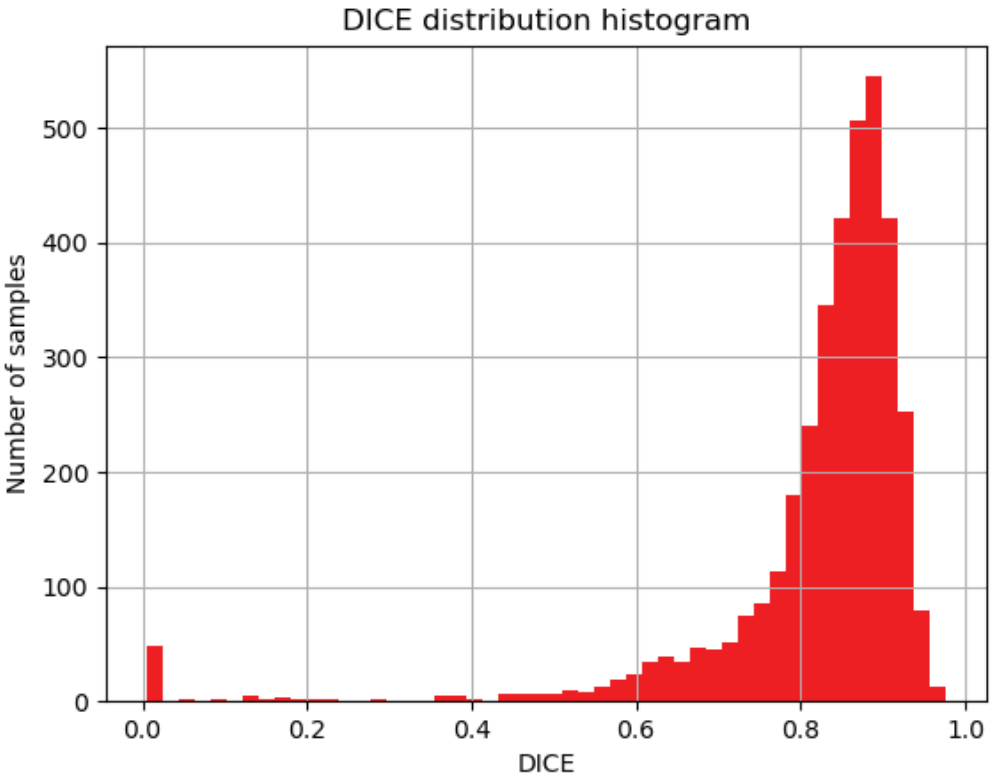


Fig. 20 DICE distribution histogram. Model: Deep U-net. Loss function: FTL. Parameters: γ=1.33, α=0.6, β=0.4. Learning rate: CLR ($max_{lr} = 1e^{-2}, base_{lr} = 1e^{-4}$. Median DICE ±STD: 0.79±0.25. MCC: 0.83. P25:

0.82. P50: 0.89. P75: 0.92.

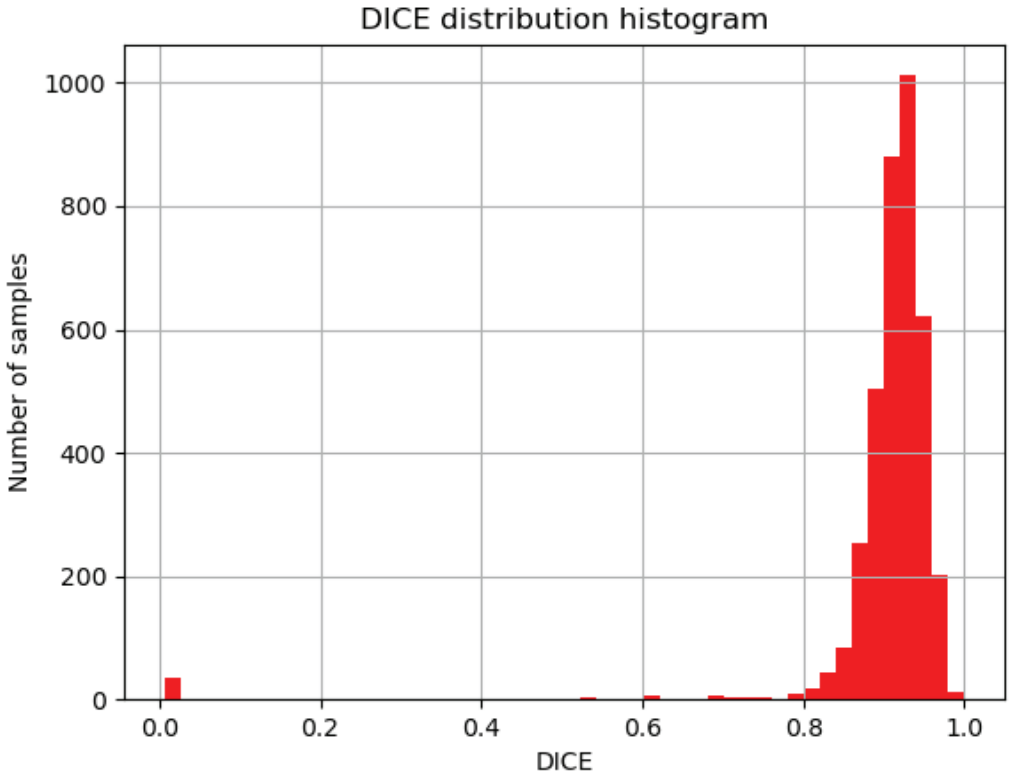
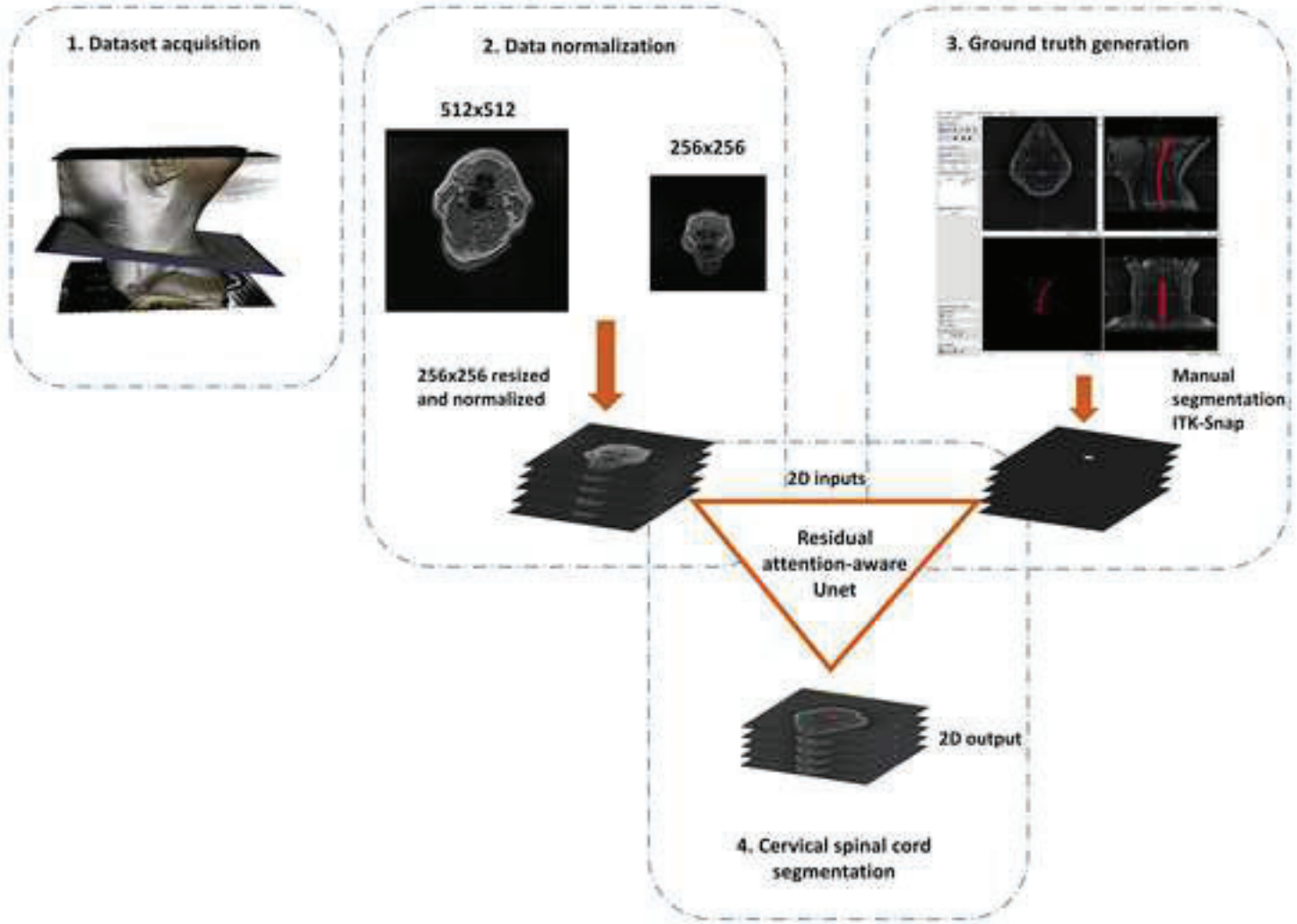


Fig. 21 DICE distribution histogram. Model: RA-U-net. Loss function: FTL. Parameters: γ=1.33, α=0.6, β=0.4. Learning rate: CLR ($max_{lr} = 1e^{-3}, base_{lr} = 1e^{-5}$. Median DICE ±STD: 0.904±0.101. MCC: 0.95. P25: 0.90. P50: 0.92. P75:94



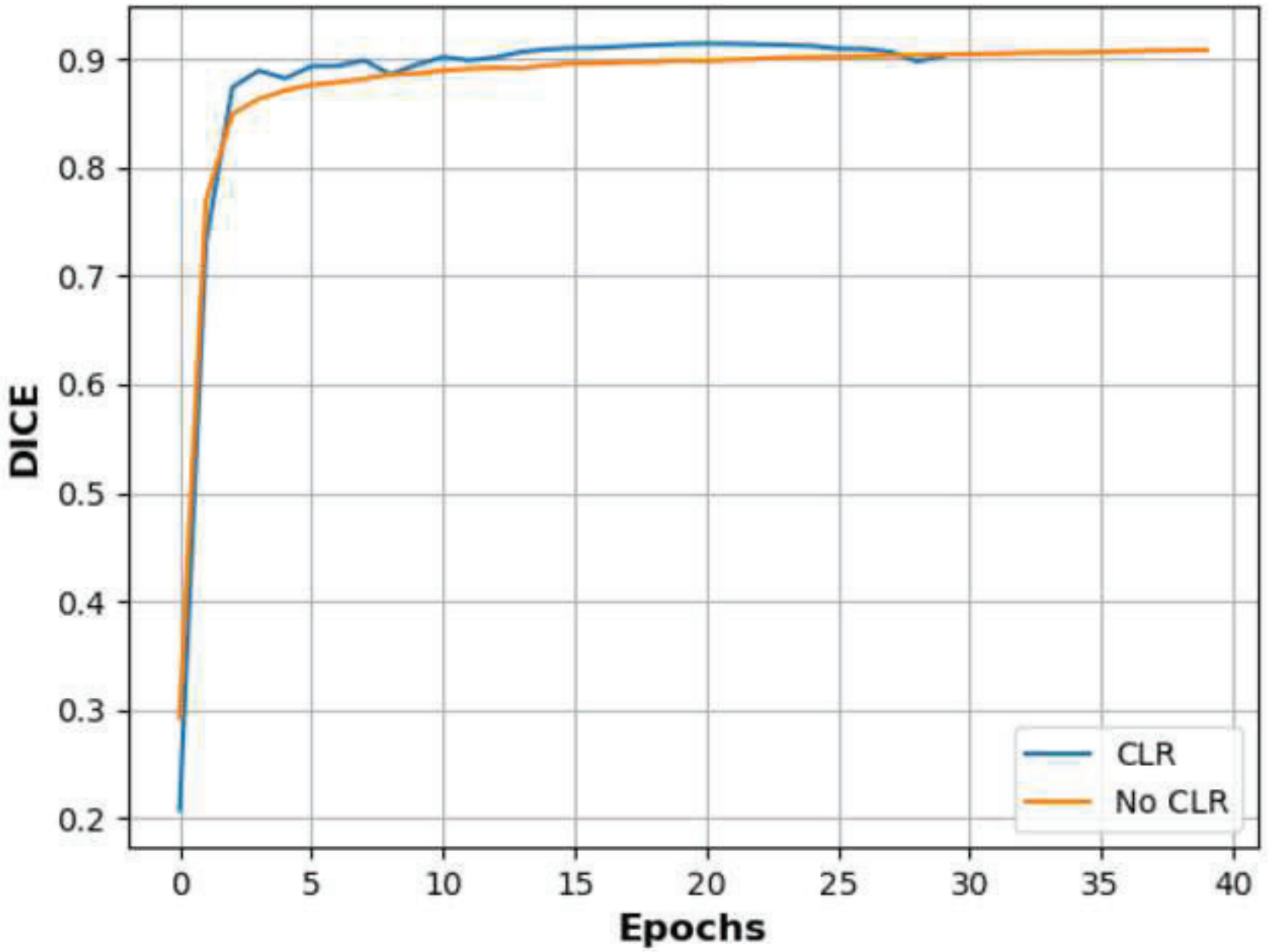
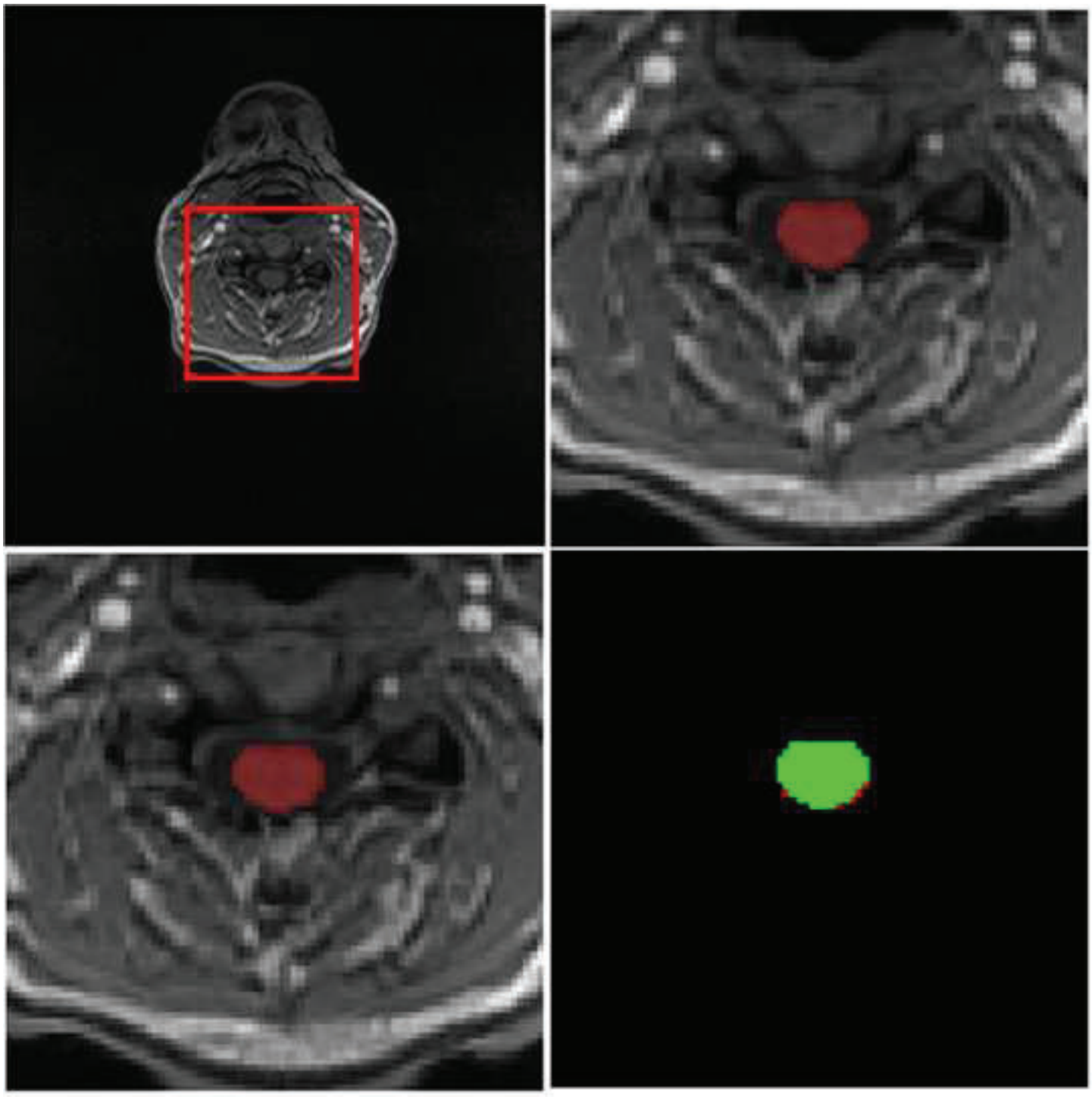
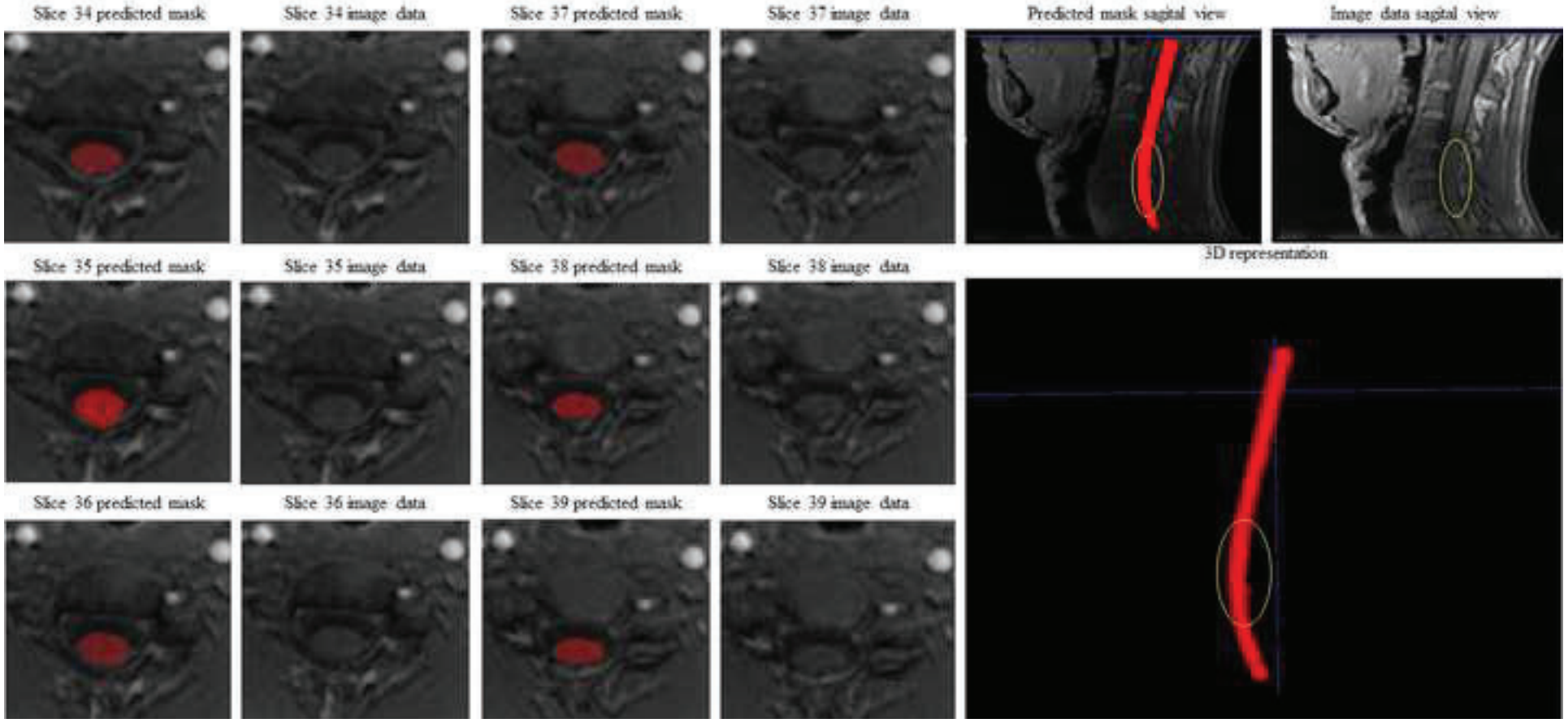
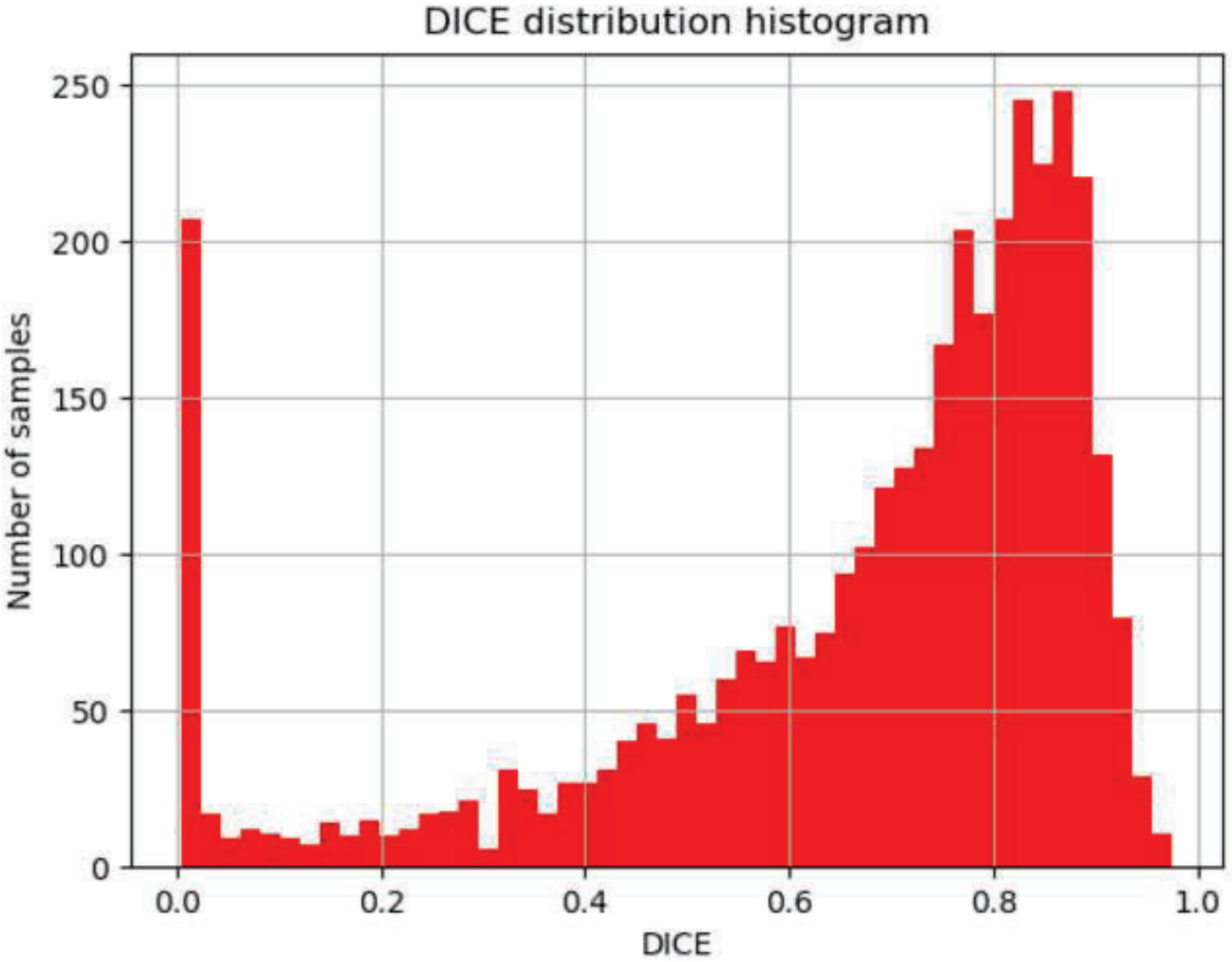


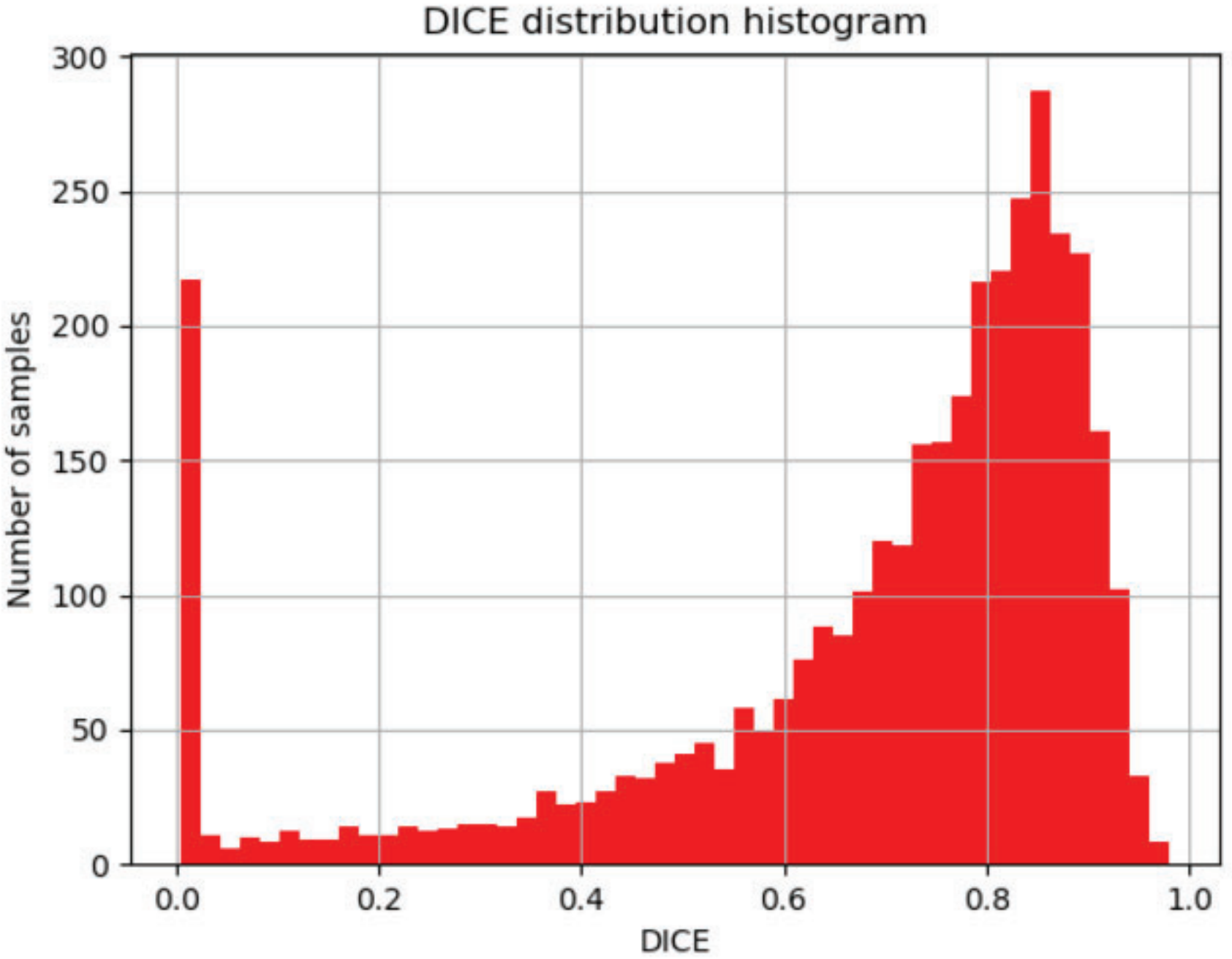
Fig11

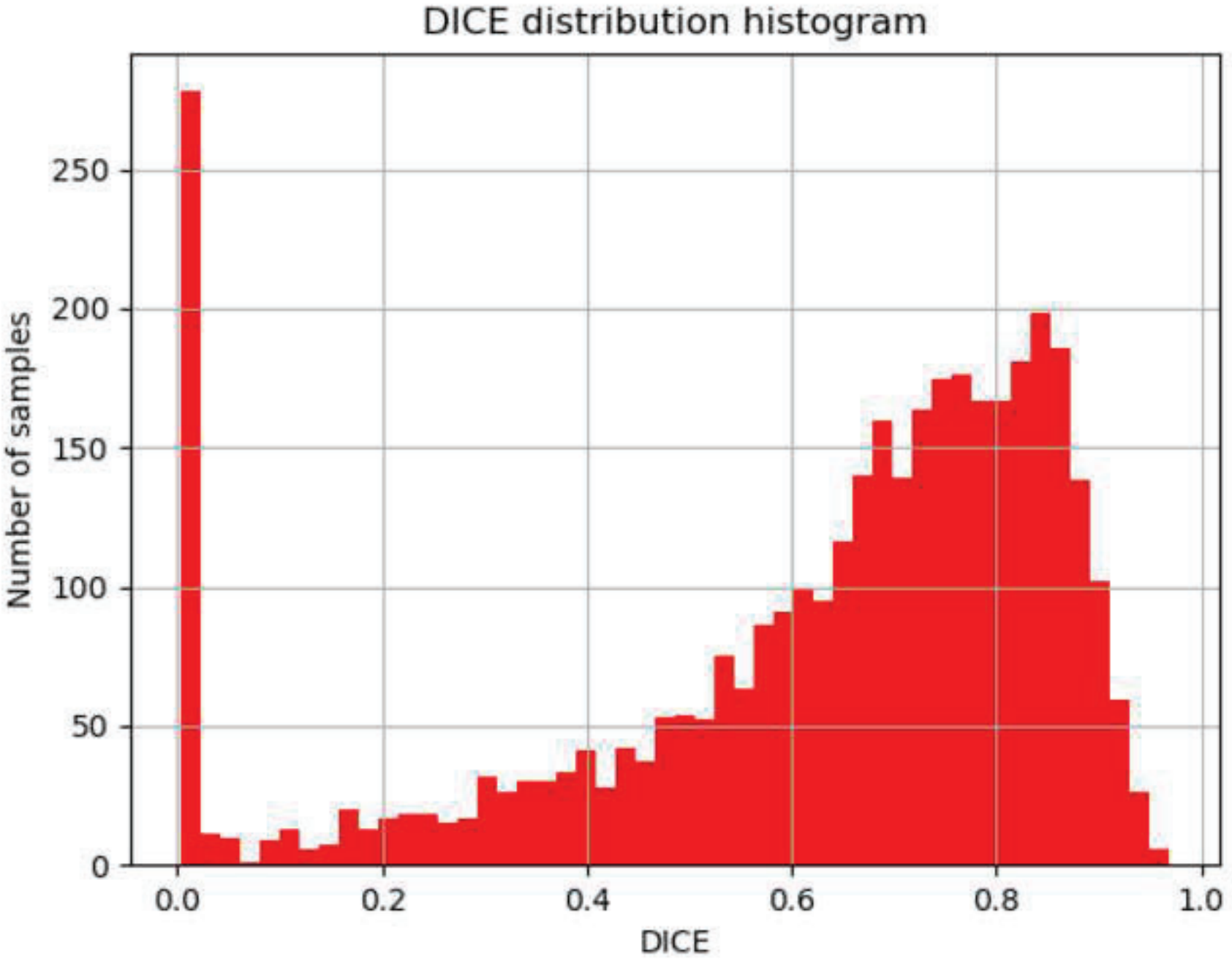
Click here to access/download;Figure;Fig11.jpg

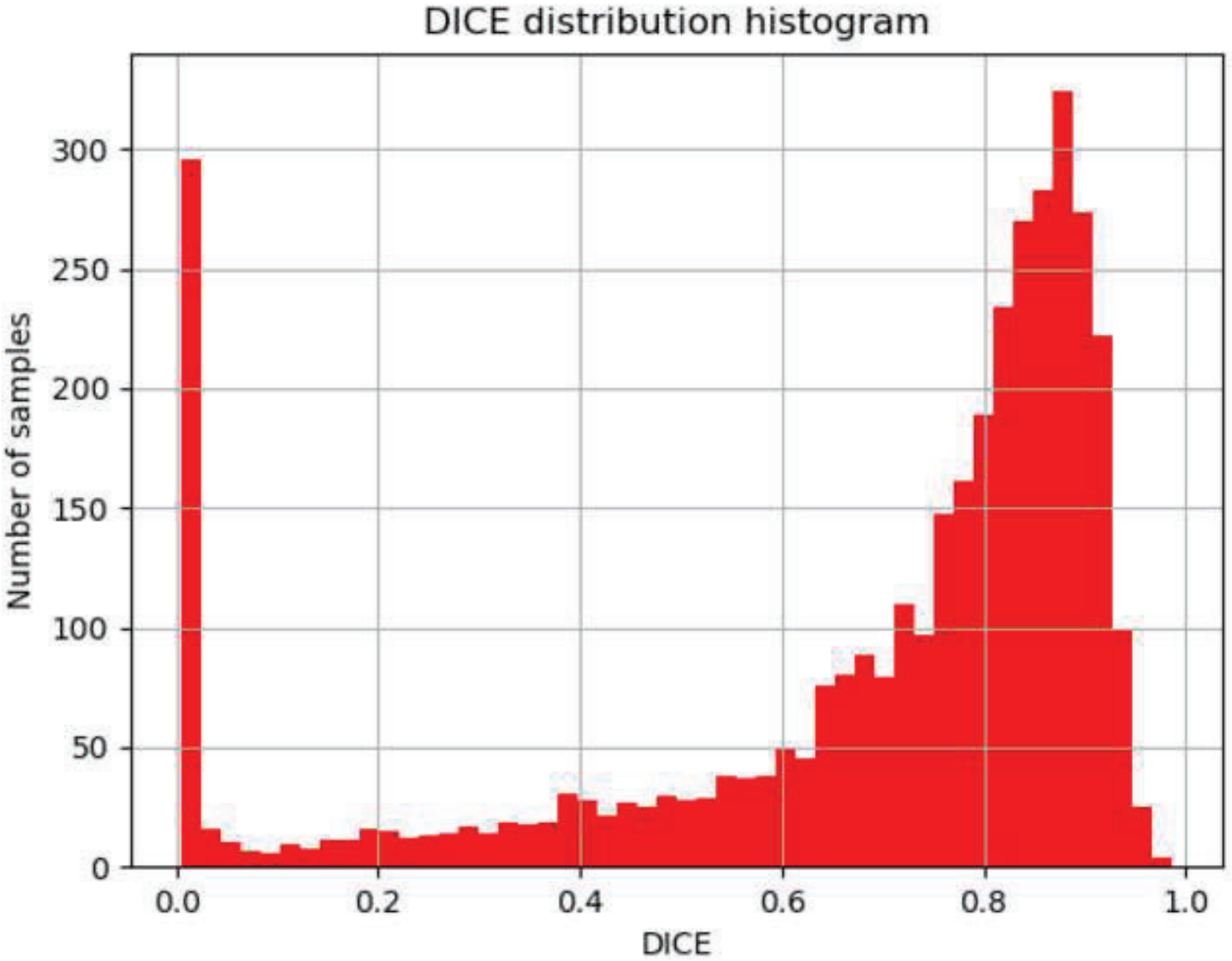


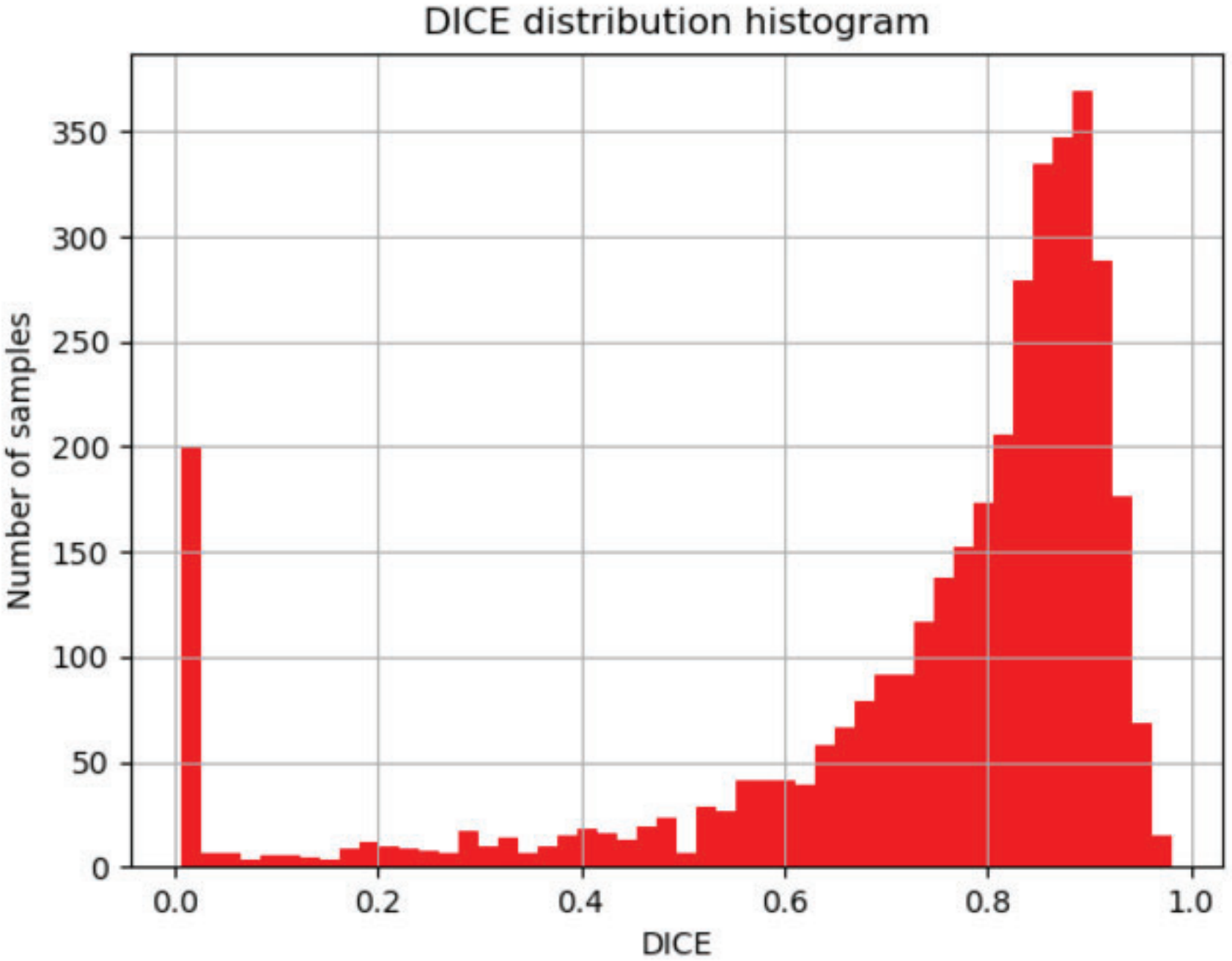


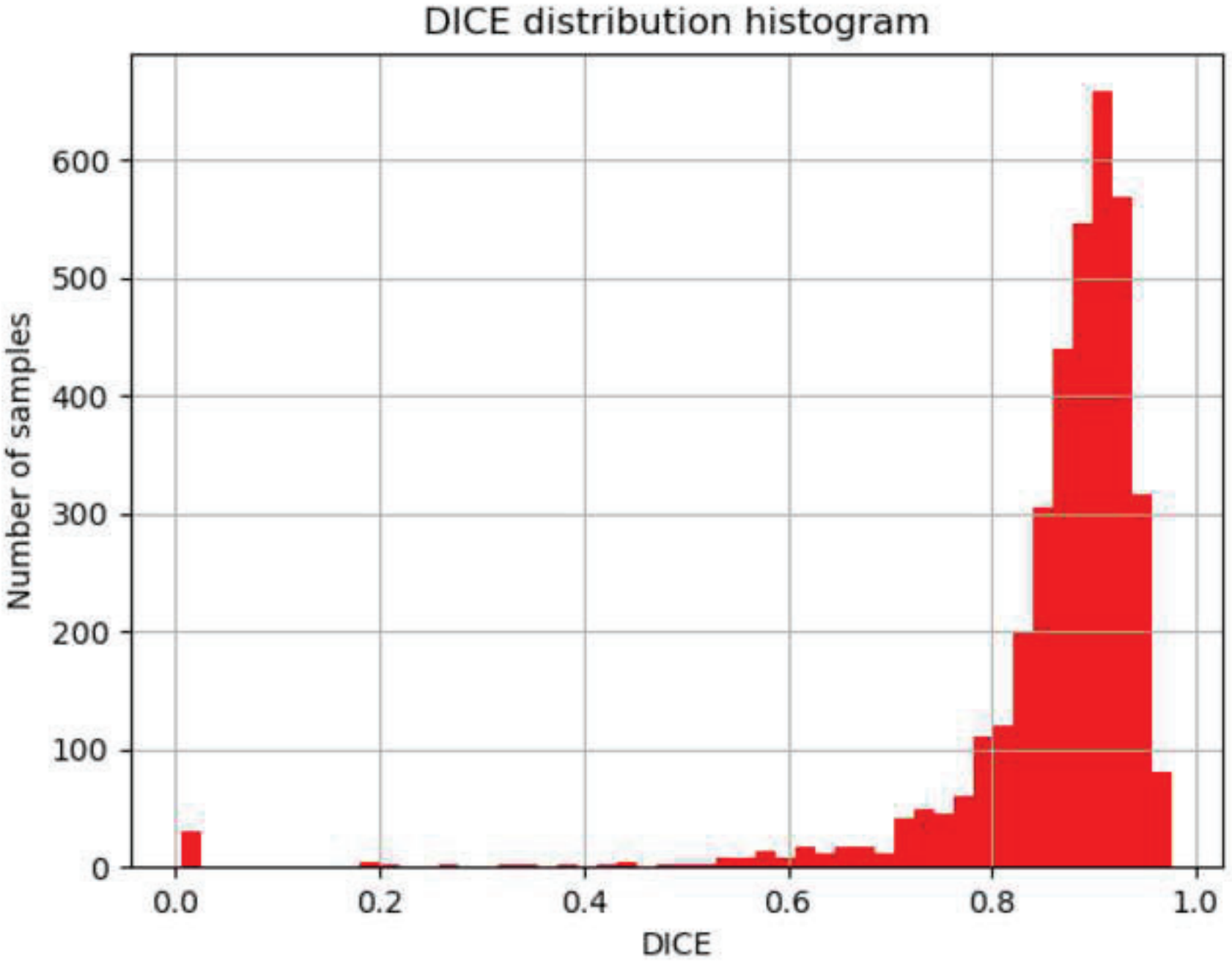












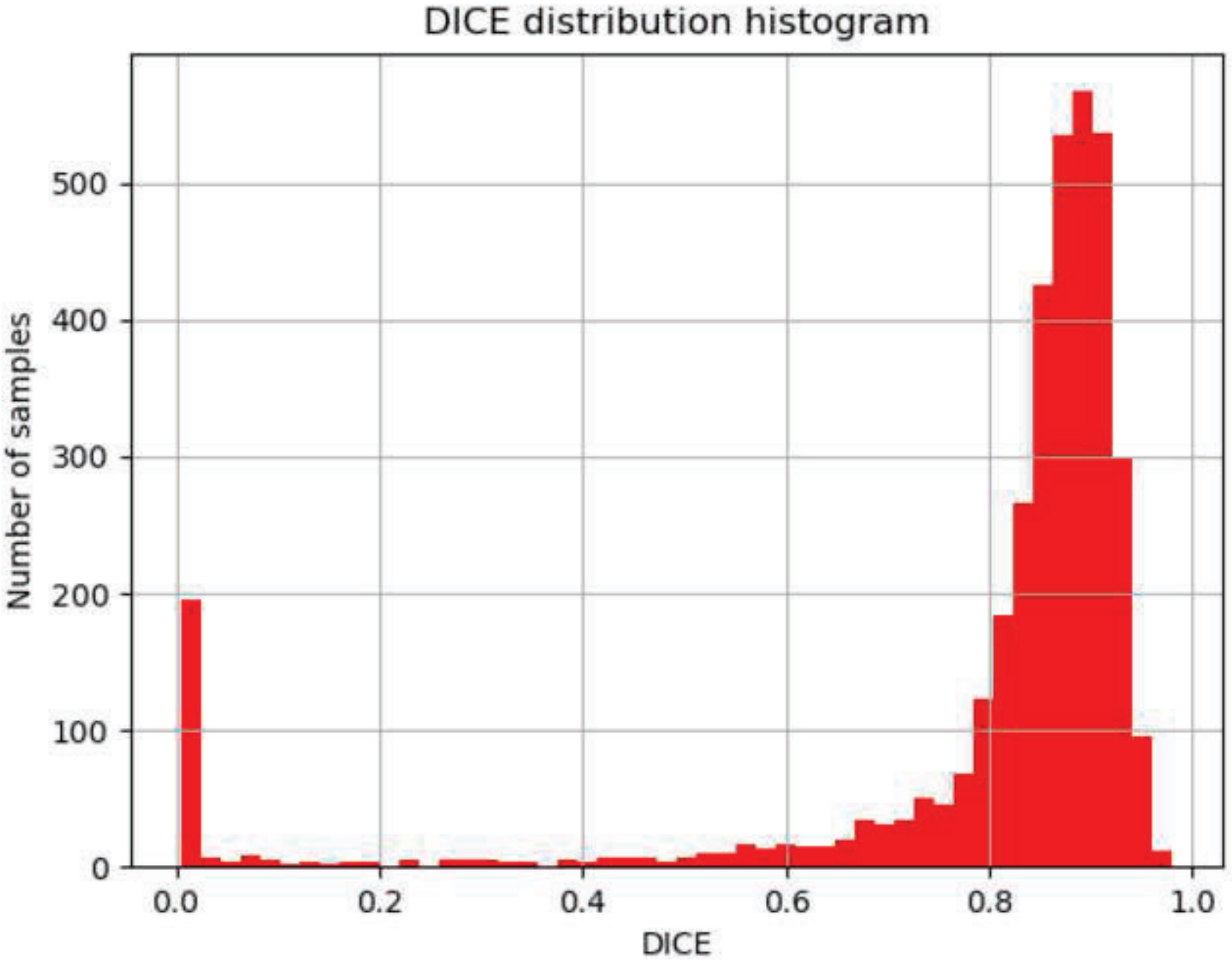
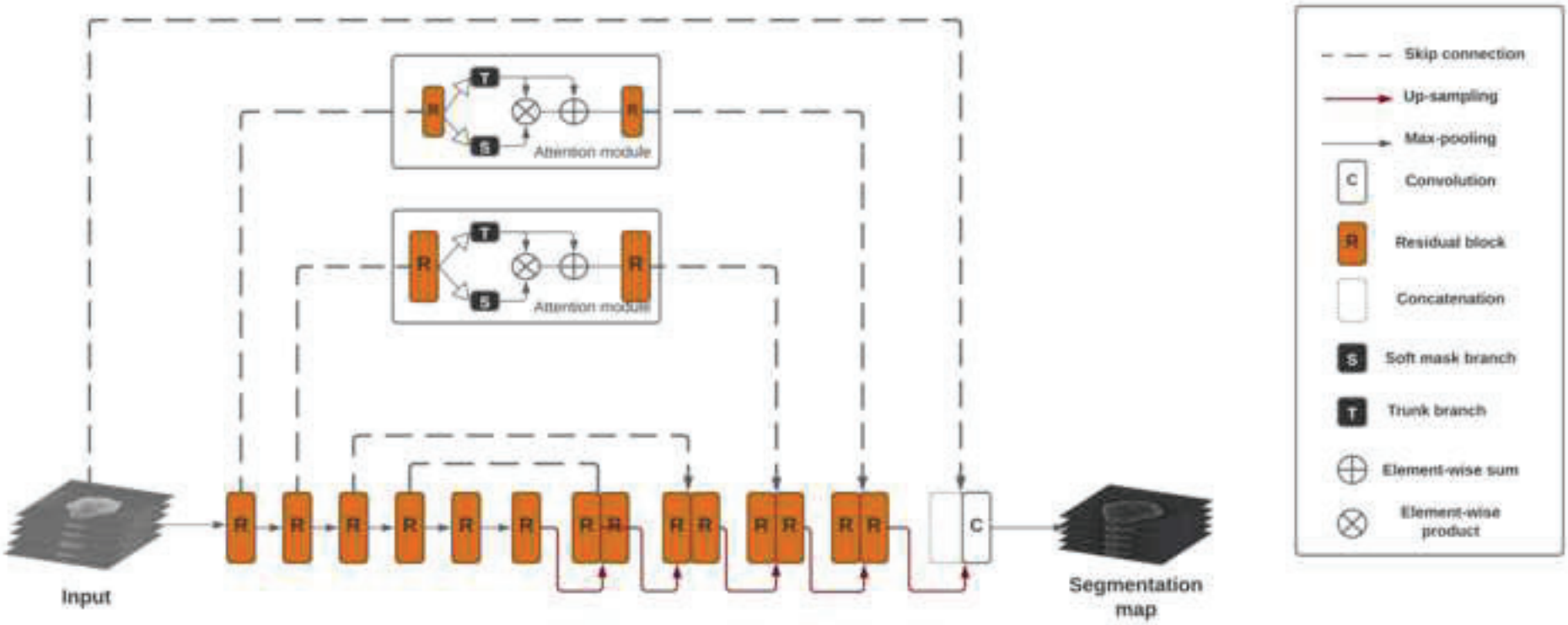
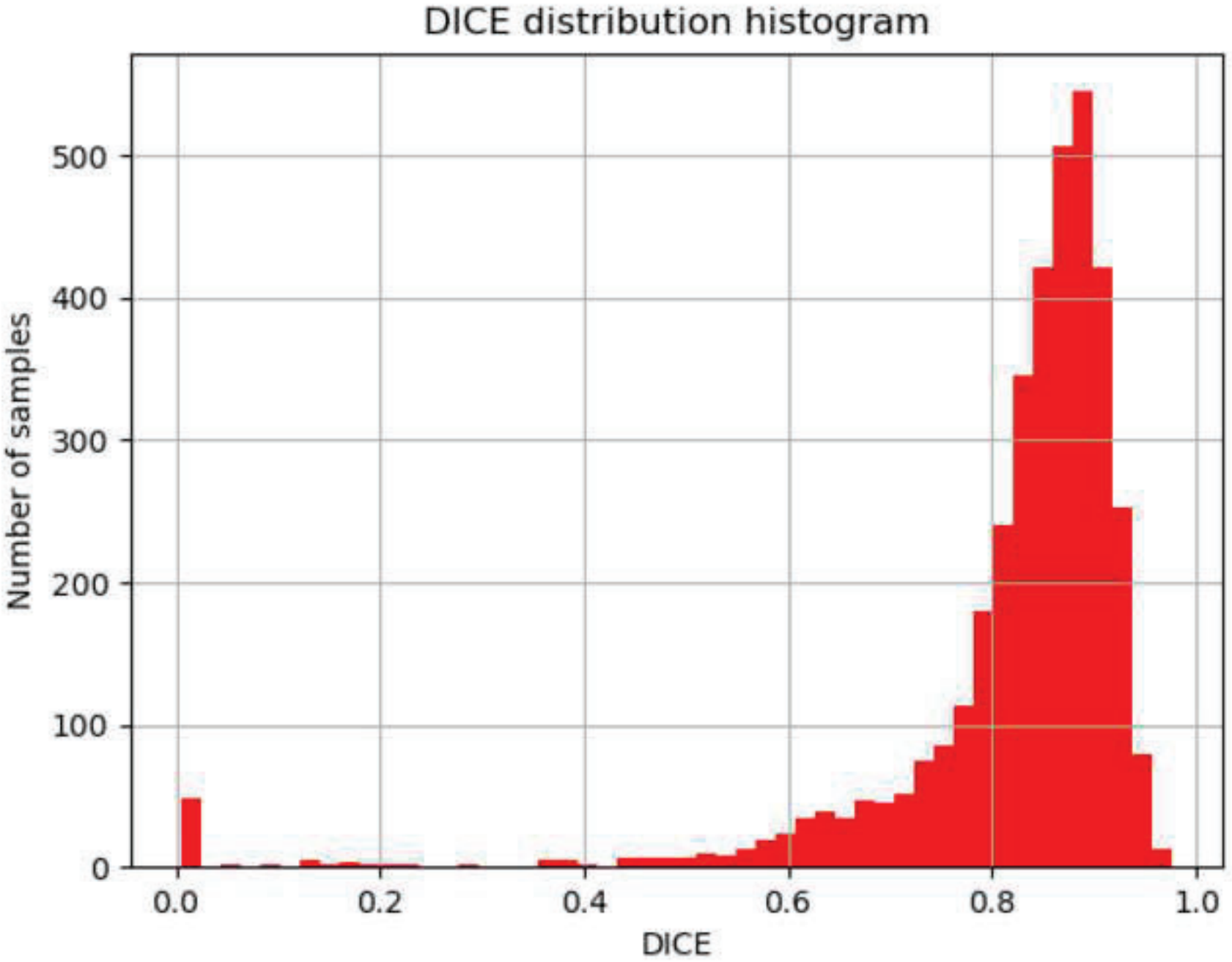
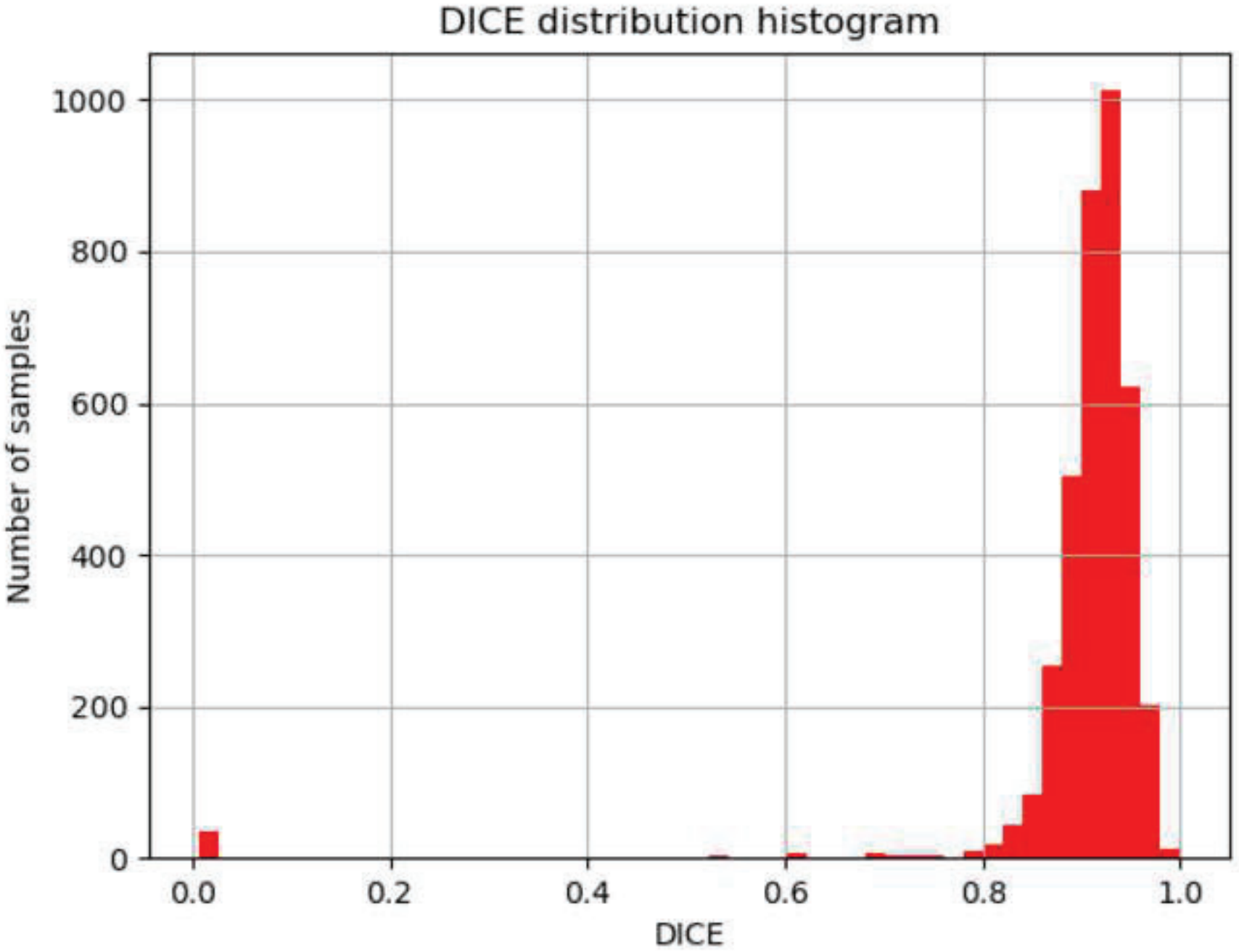
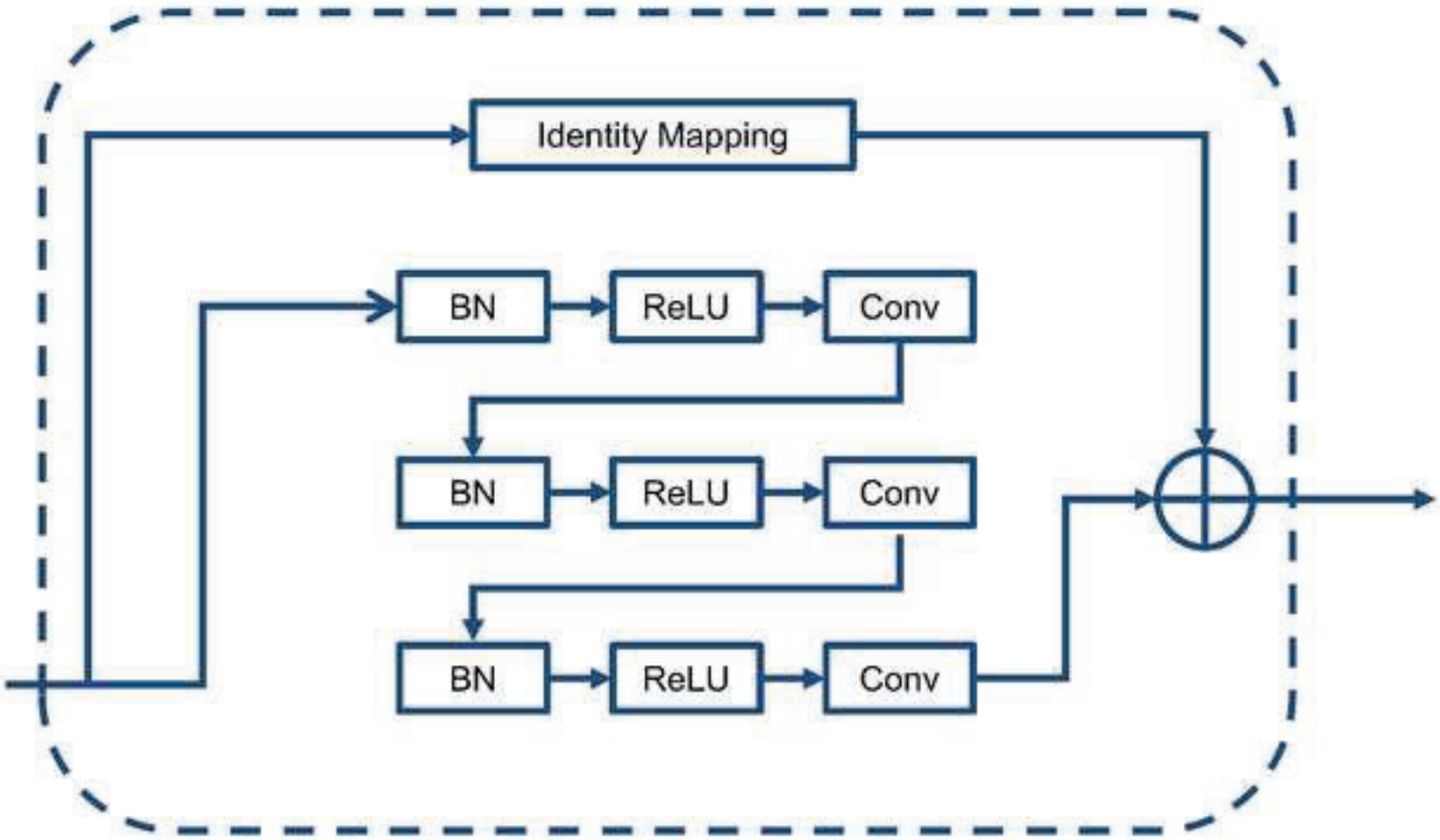


Fig2









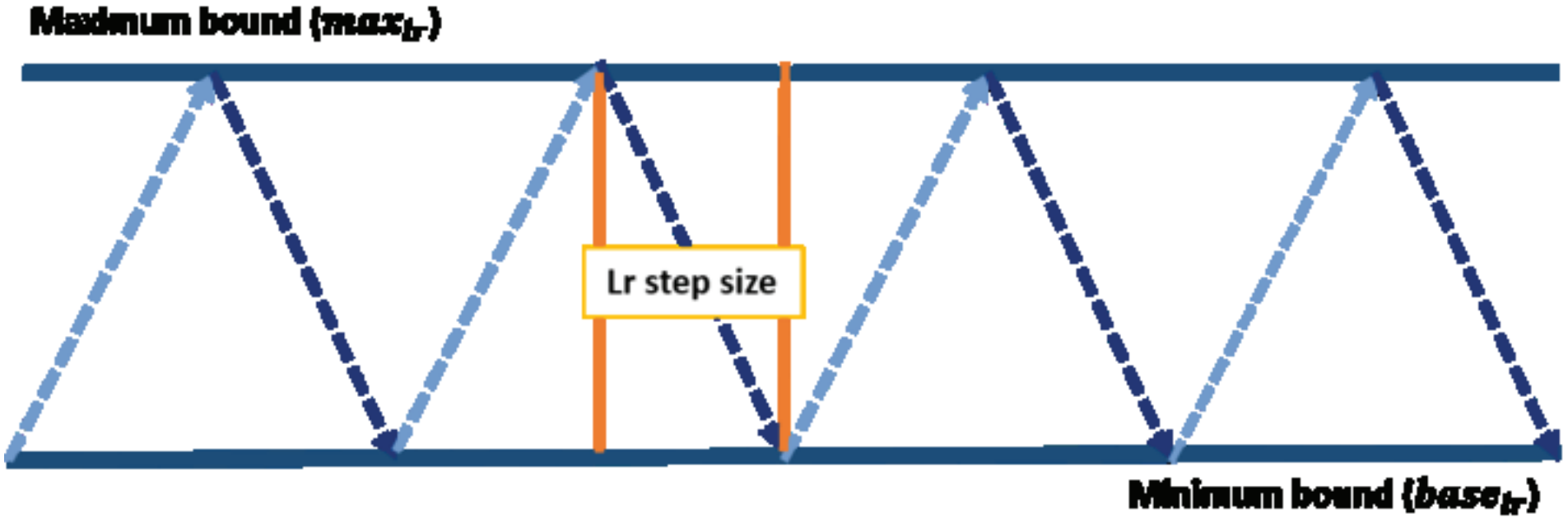


Fig5

Click here to access/download;Figure;Fig5.png

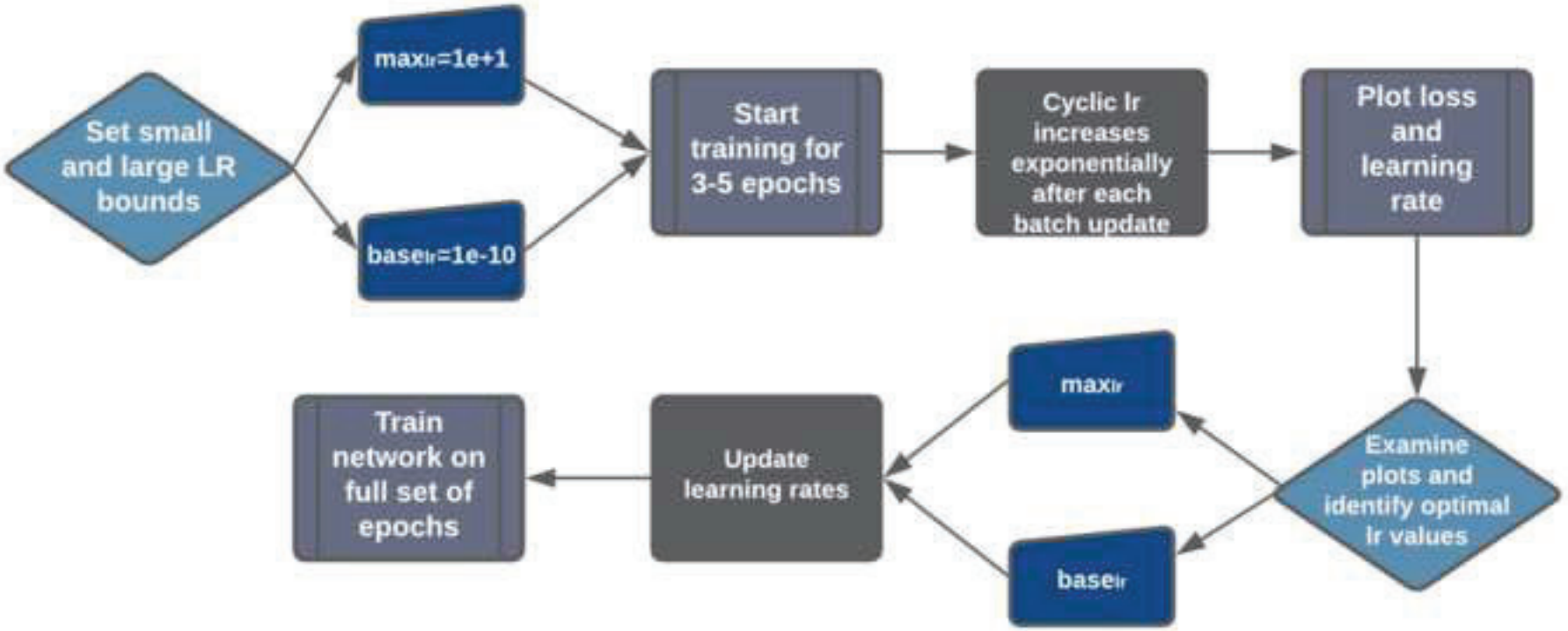
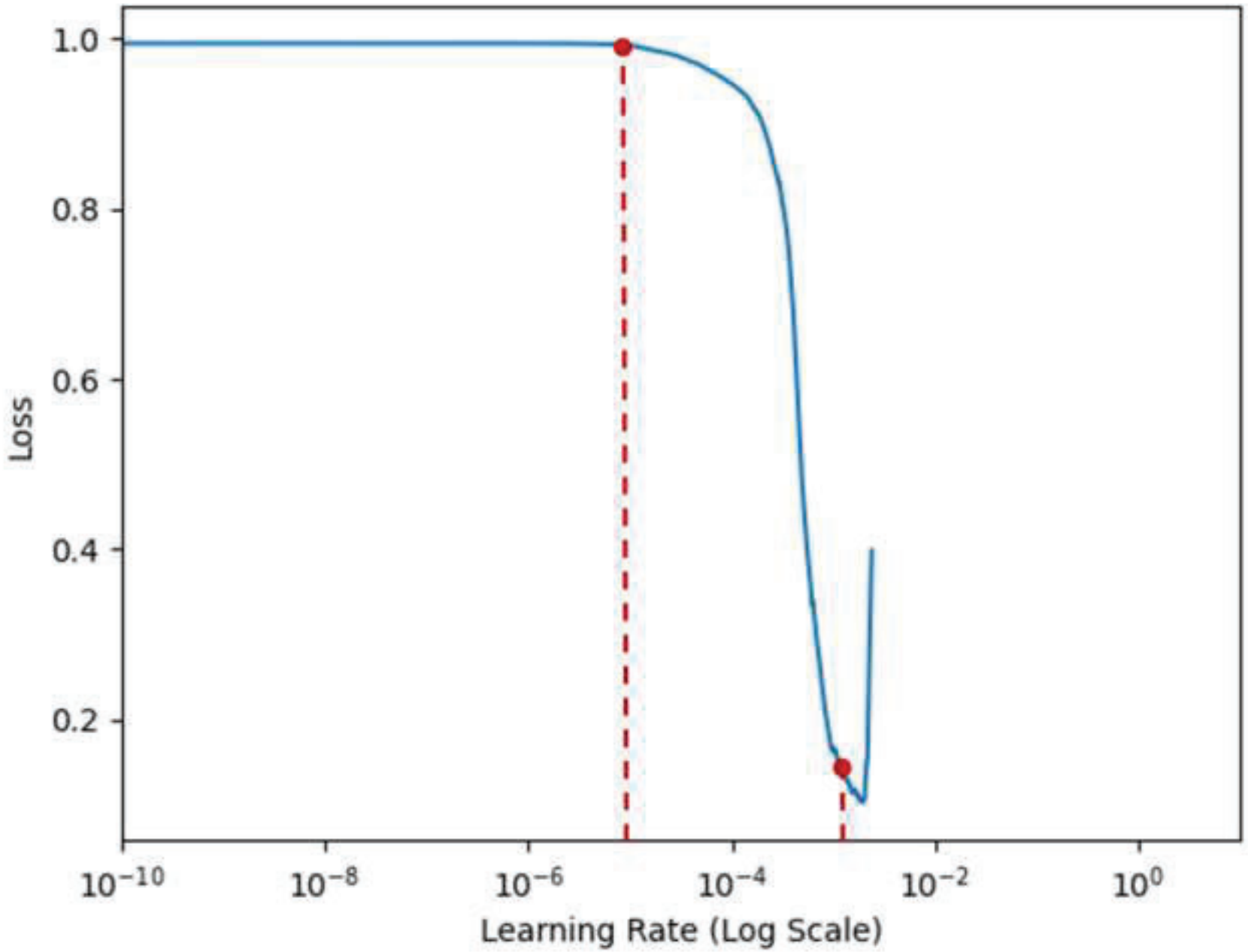


Fig6

Click here to access/download;Figure;Fig6.png



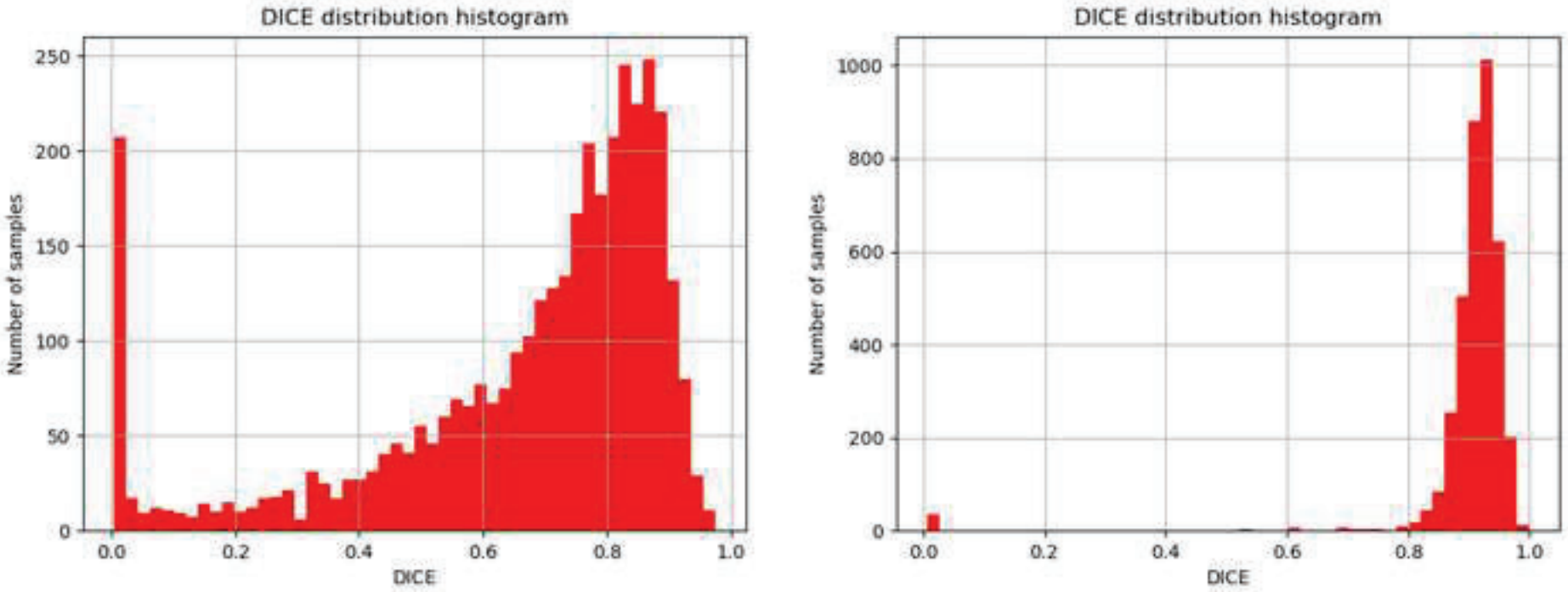
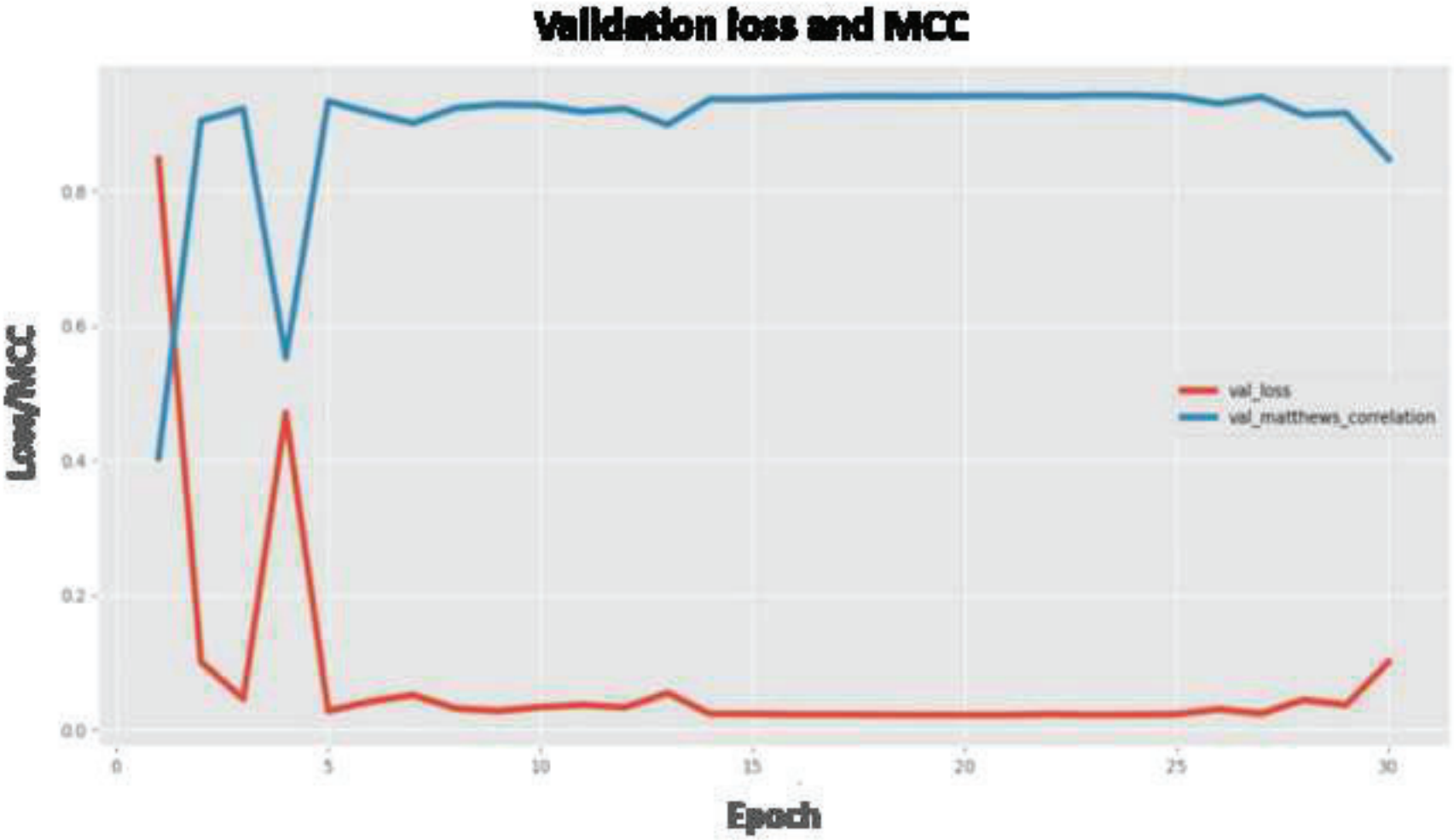
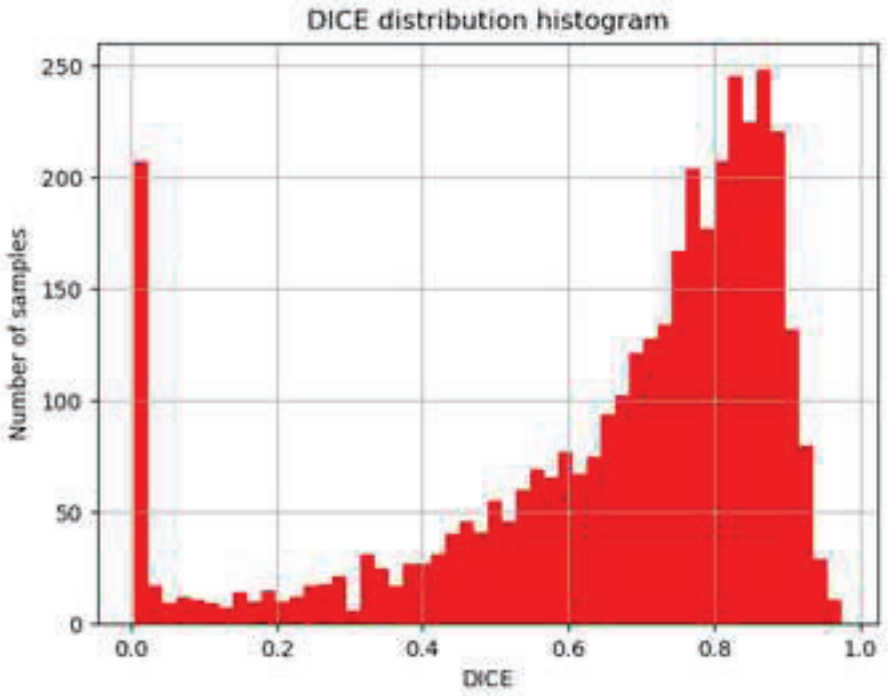
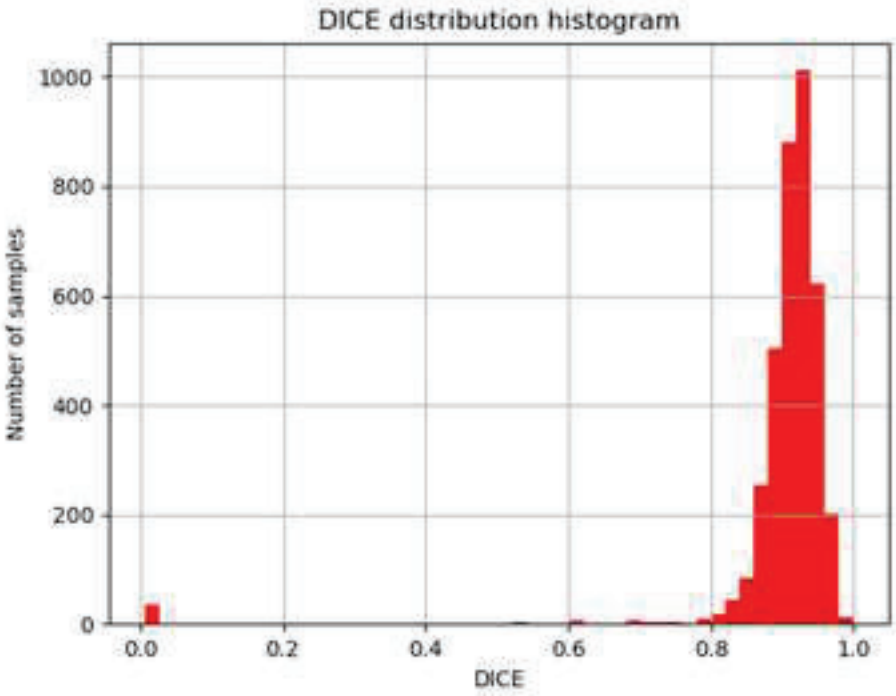


Fig8

Click here to access/download;Figure;Fig8.png







Thank you very much for your comments and time dedicated. We fully agree with that. We have tried to improve the English of all the text added during the first revision as you recommended. We have also changed the figures in the appendix. We hope to have improved both concerns. Thank you one more time.

Title page

Title

Automated cervical spinal cord segmentation in Real World MRI of Multiple Sclerosis patients by optimized hybrid residual attention-aware Convolutional Neural Networks

Author information

América Bueno [1], Ignacio Bosch [1], Alejandro Rodríguez [2], Ana Jiménez [3], Joan Carreres [4], Matías Fernández [2], Luis Marti-Bonmati [2,5], Angel Alberich-Bayarri [2,3]

1. Instituto de Tecnologías y Aplicaciones Multimedia, Universitat Politècnica de València, España.
2. Biomedical Imaging Research Group (GIBI230), Hospital Universitario y Politécnico e Instituto de Investigación Sanitaria La Fe, Valencia, Spain.
3. Quantitative Imaging Biomarkers in Medicine, QUIBIM S.L., Valencia, Spain.
4. Radiology Department, Hospital Universitario y Politécnico La Fe, Valencia, Spain.
5. Imaging La Fe node at Distributed Network for Biomedical Imaging (ReDIB) Unique Scientific and Technical Infrastructures (ICTS), Valencia, Spain.

América Bueno (corresponding author) ambuego@iteam.upv.es ORCID: 0000-0003-1735-3593

Ignacio Bosch ORCID: 0000-0003-3190-3635

Alejandro Rodríguez ORCID: 0000-0003-4781-6368

Ana Jiménez ORCID: 0000-0002-0978-9429

Joan Carreres ORCID: 0000-0003-2667-2019

Matías fernández ORCID: 0000-0001-9374-1411

Luis Marti-Bonmati ORCID: 0000-0002-8234-010X

Angel Alberich-Bayarri ORCID: 0000-0002-5932-2392

Abstract

Magnetic resonance (MR) imaging is the most sensitive clinical tool in the diagnosis and monitoring of multiple sclerosis (MS) alterations. Spinal cord evaluation has gained interest in this clinical scenario in recent years but, unlike the brain, there is a more limited choice of algorithms to assist spinal cord segmentation. Our goal was to investigate and develop an automatic MR cervical cord segmentation method, enabling automated and seamless spinal cord atrophy assessment and setting the stage for the development of an aggregated algorithm for the extraction of lesion-related imaging biomarkers. The algorithm was developed using a real-world MR imaging dataset of 121 MS patients (96 cases used as a training dataset and 25 cases as a validation dataset). Transversal, 3D-T1 weighted gradient echo MR images (TE/TR/FA=1.7-2.7ms/5.6-8.2ms/12°) were acquired in a 3T system (SignaHD, GEHC) as standard of care in our clinical practice. Experienced radiologists supervised the manual labelling, which was considered the ground-truth. The 2D convolutional neural network consisted of a hybrid residual attention-aware segmentation method trained to delineate the cervical spinal cord. The training was conducted using a focal loss function, based on the Tversky index to address label imbalance, and an automatic optimal learning rate finder. Our automated model provided an accurate segmentation, achieving a validation DICE coefficient of 0.904±0.101 compared with the manual delineation. An automatic method for cervical spinal cord segmentation on T1-weighted MR images was successfully implemented. It will have direct implications like a previous stage for accelerating the process for MS staging and follow-up through imaging biomarkers.

Keywords

Segmentation, MRI, Multiple Sclerosis, Deep Learning, Residual attention-aware, CNN.

Declarations

Funding

This work was funded by a Generalitat Valenciana PhD fellowship (grant number ACIF/2017/057) and the Universitat Politècnica de València and Polytechnic La Fe Hospital research project DeepMedul (grant number 2018/0274) (Deep Learning for spinal cord segmentation in Multiple Sclerosis).

Conflicts of interest/Competing interests

The authors declare that they have no conflict of interest.

Availability of data and material (data transparency) Not applicable

Code availability (software application or custom code) Not applicable

Authors' contributions (optional: please review the submission guidelines from the journal whether statements are mandatory) Not applicable

Additional declarations for articles in life science journals that report the results of studies involving humans and/or animals. Not applicable

Ethics approval (include appropriate approvals or waivers)

This research study was conducted retrospectively from data obtained for observational purposes. An official waiver of ethical approval was granted from $GIBI2^{30} - IISLAFE$ of the Hospital Universitari I Politècnic of Valencia.

Consent to participate (include appropriate statements) Not applicable

It is not necessary to obtain consent because of MR images without identifying information in this study.

Consent for publication (include appropriate statements) Not applicable

It is not necessary to obtain consent because of MR images without identifying information in this study.