Pattern Recognition Letters 125 (2019) 791–797

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

Multi-projection deep learning network for segmentation of 3D medical images

Rarasmaya Indraswari[a,*], Takio Kurita[b], Agus Zainal Arifin[a], Nanik Suciati[a], Eha Renwi Astuti[c]

[a] Department of Informatics, Faculty of Information and Communication Technology, Institut Teknologi Sepuluh Nopember (ITS), Jl, Raya ITS, Surabaya 60111, Indonesia
[b] Department of Information Engineering, Graduate School of Engineering, Hiroshima University, 1-4-1 Kagamiyama, Higashi-Hiroshima-shi, Hiroshima 739-8527, Japan
[c] Department of Dentomaxillofacial Radiology, Faculty of Dental Medicine, Universitas Airlangga, Kampus A UNAIR, Jl. Mayjen Prof. Dr. Moestopo No.47, Surabaya 60132, Indonesia

ARTICLE INFO

Article history:
Received 9 January 2019
Revised 3 July 2019
Accepted 2 August 2019
Available online 3 August 2019

Keywords:
Deep learning
Image segmentation
Imbalanced dataset
Neural networks
Three-dimensional medical image

ABSTRACT

Segmentation of three-dimensional (3D) medical images using deep learning is a challenging task due to the lack of a 3D medical image dataset and their ground truth, resource memory limitations, and imbalanced dataset problem. In this paper, we propose advanced deep learning network for segmentation of 3D medical images. The proposed Multi-projection Network can preserve resource memory by applying two-dimensional (2D) kernels while still obtaining the 3D information from the image by incorporating slices from different planar projections of the 3D image to achieve good segmentation results. The proposed network uses a weighted cost function to address the imbalanced dataset problem and introduces an adaptive weight that considers the probability of each class in the image. The experimental results showed that the proposed Multi-projection Network can produce the highest sensitivity (true positive rate) compared to other architectures despite the high class imbalance in the dataset and small amount of training data.

© 2019 Elsevier B.V. All rights reserved.

1. Introduction

Due to the advanced development of medical technology, three-dimensional (3D) image acquisition methods, such as computed tomography (CT) scanning and magnetic resonance imaging (MRI), are commonly used. This advancement has led to the need for more accurate and efficient 3D image segmentation methods. By performing image segmentation, important objects in the medical image can be recognized and further analysis can be done to extract relevant information about the object and decisions regarding the patient can be made. Many computer-aided methods for medical image segmentation have been proposed. These methods can be divided into threshold-based, region-based, texture-based, model-based, atlas-based, and artificial neural network-based approaches [1–3]. Recent studies have shown that deep learning techniques, such as convolutional neural network (CNN), are useful in medical image segmentation because they can provide high segmentation accuracy [4–11]. Instead of extracting features manually, deep learning can find informative and distinctive features that represent the data using a learning process [9]. Hence, the burden of finding the right features to perform image segmentation or classification shifts from humans to computers [12].

The lack of medical image datasets and their ground truth is a problem when using deep learning for medical image segmentation because this approach requires a large-scale dataset for training to produce good segmentation results [12–14]. Moreover, segmentation of 3D medical images is challenging due to the relations between the three dimensions: ignoring any of these relations will result in information loss. Hence, the most common deep learning approaches, which use two-dimensional (2D) kernels, are not the most suitable methods for the segmentation of 3D medical images. Deep learning approaches that use 3D kernels directly generate large feature vectors, therefore their implementation has high computational cost (resource memory limitations and long training time) [15]. Moreover, because networks that use 3D kernels use 3D images or 3D image patches as the input data, they are also limited by the number of 3D images in the dataset [13,16].

Another challenge in the segmentation of 3D medical images is the high class imbalance in 3D medical image datasets [8]. Not

* Corresponding author.
E-mail address: rarasmaya16@mhs.if.its.ac.id (R. Indraswari).

https://doi.org/10.1016/j.patrec.2019.08.003
0167-8655/© 2019 Elsevier B.V. All rights reserved.

every image contains a segmentation object and when there is an object in the image, the size of the object region tends to be much smaller than the background region. Deep learning algorithms tends to prefer the majority class and suppress the minority class to minimize the network cost, even though the minority class is the one of interest. If this problem is not addressed, the segmentation result of 3D medical images using deep learning will have low sensitivity (true positive rate). Several researches have proposed solutions for this problem by introducing a weight into the cost function of the network [17–19]. Although these methods are effective for binary classification problems using deep learning, it is difficult to calibrate the proposed functions for multiclass classification problems.

In this paper, we propose an advanced deep learning network for segmentation of 3D medical images. The proposed architecture, called Multi-projection Network, preserve resources' memory by applying 2D kernels while still obtain the 3D information of the image by incorporating slices from different planar projections of the 3D image to achieve a good segmentation result. We also introduce an adaptive weighted cost function for the network to address the imbalanced dataset problem in 3D medical image segmentation. This weight is calculated by considering the probability of each class in the image batch. The proposed cost function improves the network's performance in detecting the minority class and can be also used for multiclass segmentation. Three kinds of datasets were used in an experiment to represent the problems of 3D medical image segmentation, such as the imbalanced dataset problem and a limited amount of available training data. This paper also provides a comparison of the proposed method with several other network architectures to analyze the accuracy of the proposed network for binary segmentation of one-modality 3D medical images.

2. Related works

Image segmentation classifies each pixel in the image individually. CNN can be used for image segmentation by using image patches to determine the class of a pixel. The drawback of this approach is that input patches from neighboring pixels overlap so that the same convolutions are computed many times over [20]. Fully Convolutional Network (FCN) substitutes the fully connected layer in CNN with a convolution layer [21]. Using FCN, the network can take an entire image as the input and produce an entire image as the output, thus speeding up the computation process. However, FCN approaches for medical image segmentation do not have an adequate spatial resolution in their direct output label space [22]. To solve this problem, a modified version of FCN, called U-Net [5] has been proposed. U-Net and its variations are among the most well known deep learning methods for medical image segmentation [20]. It has a similar architecture as Autoencoder, a symmetrical neural network that learns the features of the dataset in an unsupervised manner [23]. It combines an equal number of layers and processes on the encoder and decoder side. The difference between Autoencoder and U-Net is that U-Net uses skip connections that bypass information from the encoder layers to the decoder layers to help recover the full spatial resolution at the network's output [24]. U-Net can take an entire image as input and give the segmentation map directly as output, therefore it can take into account the global information of the image.

Several researches have been conducted for 3D medical image segmentation using a number of different approaches. The first approach is to treat the 3D images as a set of 2D slices and applying a 2D deep learning strategy to segment the images [4,7,8,25,26]. Although this approach generates more training data for the network, it dismisses the connection between consecutive slices in the 3D image. Several researches examined this problem by adding a post-processing step, such as using 3D Conditional Random Field, to obtain the inter-slice information [25]. Another approach for 3D medical image segmentation is by adding the image depth in the deep learning process [6,18,27,28]. Voxels are used instead of pixels, therefore 2D convolution and 2D pooling will be substituted by 3D convolution and 3D pooling. Although this approach is the most obvious way to preserve the inter-slice information, it is limited by the small size of the 3D dataset and the resource memory.

Several researches have attempted to use a 2D deep learning strategy while preserving the inter-slice information by integrating several 2D projections from different points of view of the 3D image in a network. For 3D medical images, this is done by dividing the 3D image into three different planes (x, y, and z) and creating three different sets of 2D slices [29–31]. Prasan, et al. [29] adopted this approach in a method called Triplanar Convolutional Neural Network on knee cartilage segmentation images. The 3D image patches consisting of voxels are extracted from the image and then divided into three different sets of 2D slices. Each set of slices will become the input for a CNN. The outputs of the three CNNs are concatenated to obtain a joint output, which will be fed into the softmax classifier to obtain the classification result of each voxel.

3. Material and methods

3.1. Dataset

We used three datasets in our experiments. The first dataset used was BRATS-2012 [32,33], a dataset of brain images for brain tumor segmentation that was acquired using MRI (Magnetic Resonance Imaging) scanning. There are 80 brain images in the dataset, consisting of multi-contrast MR scans of 10 low-grade and 20 high-grade glioma patients and simulated images of 25 low-grade and 25 high-grade glioma subjects. The size of the 3D images in this dataset vary. The second dataset used was BRATS-2018, the latest version of the BRATS dataset. It consists of the BRATS-2012 and BRATS-2013 datasets, manually annotated and revised by clinical experts. This dataset consists of 285 MRI scans divided into 210 high-grade glioma subjects and 75 low-grade glioma subjects. Each MRI scan consists of 155 2D images with a size of 240×240 pixels.

There are four modalities for each brain image in the BRATS dataset, namely T1 (native), T1C (post-contrast T1-weighted), T2 (T2-weighted), and Flair (T2 Fluid Attenuated Inversion Recovery). The ground truth provided four segmentation labels, namely non-tumor, edema, necrosis, and enhanced structures. We treated the segmentation problem using BRATS dataset as a binary segmentation problem in which the segmentation object was the whole brain tumor. This included edema, necrosis and enhanced structures. Hence in our research there were only two classes in the BRATS-2012 and BRATS-2018 datasets, i.e. whole brain tumor and non-tumor. We used only one modality for the input of the network, i.e. the T1C modality that is better at showing brain tumors than other modalities.

The third dataset used was CBCT (Cone-Beam Computed Tomography) containing scans of human jaws. We acquired this scan from the hospital *Rumah Sakit Gigi dan Mulut, Universitas Airlangga* (RSGM UNAIR), which used an ORTHOPANTOMOGRAPH^{TM} OP300 3D X-ray unit. The field of view (FOV) width and height of the scanner are 79.8 mm and 60 mm, respectively. The dataset consisted of jaw images from 7 patients. The segmentation object was teeth and the manually annotated ground truth was confirmed by radiologist experts. The 3D images had sizes of 266 x 266 x 200 voxels. The 2D images were obtained by slicing the 3D image along the axial plane. To make them uniform, the 2D images in the BRATS-2012, BRATS-2018 and CBCT datasets were resized to 128×128 pixels while only the middle 128 images of each 3D scan

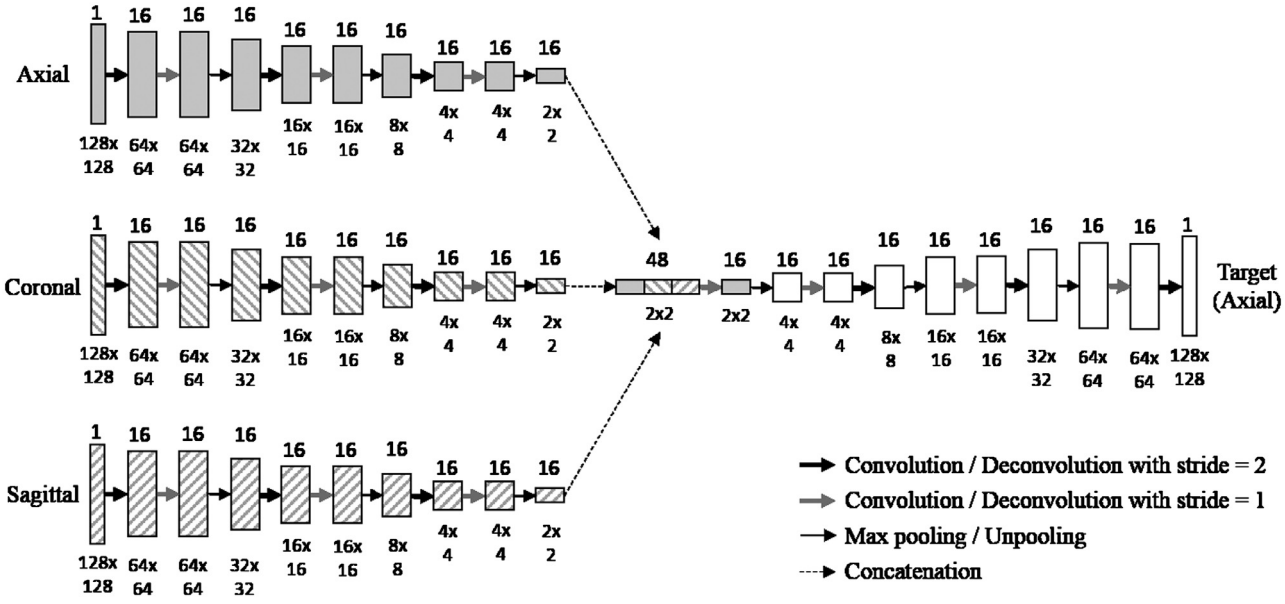

Fig. 1. Architecture of the proposed Multi-projection Network.

were selected. Therefore, the size of each processed 3D images was $128 \times 128 \times 128$ voxels.

3.2. Network architecture

The proposed network architecture, which incorporates three different projections of the 3D input images, is illustrated in Fig 1. The structure of the proposed network architecture is developed from the structure of Convolutional Autoencoder network that consist of encoder and decoder side. Autoencoder has symmetrical structure where the process in decoder side is the reverse of the process in encoder side [34]. The Multi-projection Network uses different encoders to take each of the planar projections of the 3D image as input. The input 3D medical image is sliced into 3 different sets of 2D images according to axial, coronal, and sagittal plane. Axial slices show the 3D image from top to bottom, coronal slices show the 3D image from front to back and sagittal slices show the 3D image from left to right. The proposed Multi-projection Network consists of three sets of two convolution - deconvolution (up-convolution) layers and one pooling - un-pooling (up-sample) layer.

In the convolution layer, convolution with step size or stride s using a number of filters or kernels $\{w_1, w_2, \ldots, w_n\}$ is done to the input image I to produce feature maps or output channels $\{I \times w_1, I \times w_2, \ldots, I \times w_n\}$. Each kernel w_i is a matrix with size $l \times m$. The kernel values are changed during the training process to obtain the objective of the network, usually to minimize the error of the network that was calculated by the network cost function. For all the convolution layers, the kernel size $l \times m$ is set to 3×3 and the number of output channels n is set 16. The stride s of first convolution layer is set to 2, while the stride of second convolution layer is set to 1. Each convolution layer is followed by a transfer function to help the classifier build a non-linear decision boundary. A ReLU (rectified linear unit) activation function was chosen as the transfer function. Therefore, the output of the transfer function will be as in Eq (1),

$$f(I \times w_i) = max(0, I \times w_i). \tag{1}$$

The pooling process using the maximum value (max pooling) is done to reduce the dimensionality and avoid overfitting by downsampling the input image of the process. This is done by applying max filtering with kernel size $l \times m$, which moves through the image with stride s. For all pooling layers, the kernel size $l \times m$ is set to 2×2 and the stride s is set to 1. After the final pooling layer, the outputs of the three encoders are concatenated and a convolution with kernel size = 3, strides = 1, and number of output channels = 16 is performed. The target image of the network is one of the 2D projections of the 3D image, i.e. the axial slices. Adam optimizer [35] is employed for the training process of the network. Adam optimizer, recommended for achieving fast convergence [36], computes adaptive learning rates for each parameter using momentum.

Using this architecture, the 3D information of the image can still be obtained while preserving the resource memory by using 2D kernels. The information from axial, coronal, and sagittal slices are combined after the last pooling process on the encoder side to reduce the dimensionality of the combined input. The concatenated information is then convoluted to produce the most important feature maps from the combined input.

3.3. Weighted cost function

The idea of the proposed cost function is to make the network prefer the occurrence of false positives over false negatives by providing a larger error value when a false negative occurs. False positives occur when the segmented class, which is the minority class, covers more area than the target. False negatives occur when the segmented class covers less area than the target. For example, using root mean squared error (RMSE) as the cost function, the error obtained from a segmented object that has n more pixels or n less pixels than the target will be the same. The formula of RMSE is shown in Eq (2) where N is the number of pixels in the image, $\hat{x}_i$ is the target pixel, and x_i is the corresponding segmentation result.

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{x}_i - x_i)^2} \tag{2}$$

Let the image have K number of classes and C_k, $k = \{1, 2, .., K\}$ representing a class in the image. Let n_{C_k}/N be the probability of each class where n_{C_k} is the number of elements in class C_k. In a dataset with high class imbalance, the probability of the majority class is much bigger than the probability of minority class. A larger

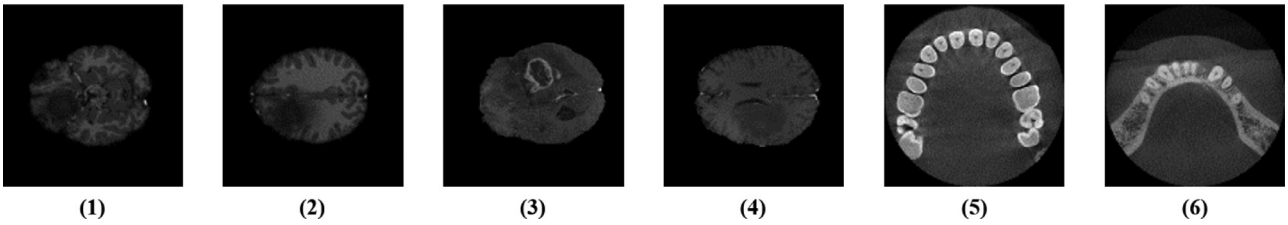

Fig. 2. Input images.

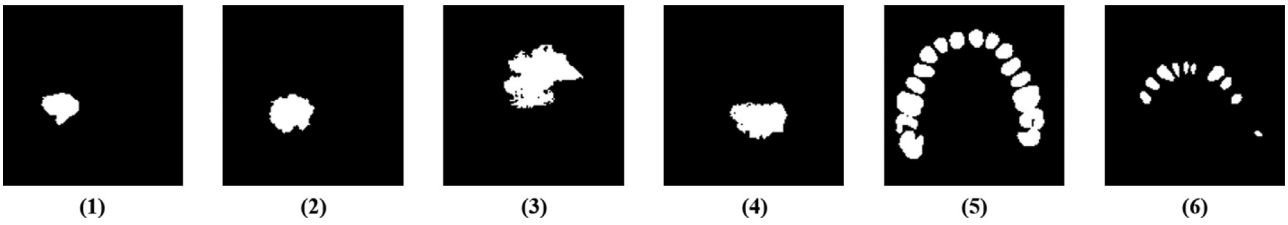

Fig. 3. Target images.

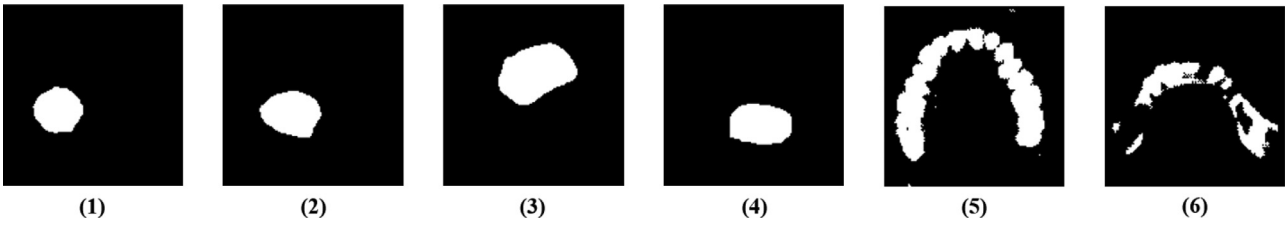

Fig. 4. Segmentation results of Multi-projection network with the proposed weighted cross entropy as the cost function.

weight of the minority class can be achieved by using the probability of majority class as the weight of the minority class. However, using a reversed class probability directly as the weight can make the cost function return zero value when the number of elements of one of the classes is zero, hence each class must have minimal one element. The reverse of the normalized probability function of each class will be $1-((n_{C_k}+1)/(N+K))$ where $K=2$ in a binary segmentation task. This function can be expanded for multiclass segmentation problems as in Eq (3).

$$\text{Weighted RMSE} = \sqrt{\frac{1}{N}\left(\sum_{k=1}^{K}\frac{1-\frac{n_{C_k}+1}{N+K}}{K-1}\sum_{\widehat{x}_i\in C_k}(\widehat{x}_i-x_i)^2\right)} \tag{3}$$

However, while the RMSE cost function can be applied to binary classification tasks by rounding or thresholding its value, it is difficult to apply in multiclass classification tasks. A cross-entropy cost function is better suited for classification tasks. The proposed weight can be assigned to the cross-entropy cost function using a similar approach. Let $\widehat{y}_k^{(i)}$ be the actual probability of pixel $(\widehat{y}^{(i)})$ belonging to class k and $y_k^{(i)}$ be the output probability of pixel $(y^{(i)})$ belonging to class k. The cross-entropy error for binary segmentation is calculated using Eq (4). Using this formula as the cost function will result in the same error value, either when a higher number of false negatives or a higher number of false positives occurs. To make the function prefer the occurrence of false positives over false negatives, a weight w needs to be introduced in the first part of the formula. Using the proposed weight $w=(1-((n_{C_k}+1)/(N+K)))/(K-1)$, the weighted cross-entropy formula will be as in Eq (5).

$$\text{Cross}-\text{entropy} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K}(\widehat{y}_k^{(i)}\ \log(y_k^{(i)})) \tag{4}$$

$$\text{Weighted Cross}-\text{entropy} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K}(w\,\widehat{y}_k^{(i)}\ \log(y_k^{(i)})) \tag{5}$$

If the proposed cost functions are applied to a neural network that takes batches or mini-batches consisting of several images as input, the weight of each class in the cost function will be calculated for each batch. The weight of the cost function adaptively provides a different value based on the class probability of each batch. Moreover, while the original cost function will give an error value in the range of [0,1], the proposed cost function will give an error value in the range of $[0, \infty]$ due to its weight.

4. Experimental results

This research was implemented on Python using the Tensorflow library. The specifications of the machine on which network was run were is GPU GTX 1080, RAM 2 x 8GB 2400 MHz DDR4. Examples of the input images and their corresponding ground truth (target images) are shown in Figs. 2 and 3, respectively. Examples of the output images from Multi-projection Network using the proposed weighted cost-entropy as the cost function are shown in Fig 4. Image (1) and image (2) are from the BRATS-2012 dataset, image (3) and image (4) are from the BRATS-2018 dataset, and image (5) and image (6) are from the CBCT dataset.

Because of the imbalanced dataset problem, accuracy cannot be the only evaluation metric for measuring the performance of deep learning in 3D medical image segmentation. A high accuracy is not enough to demonstrate the goodness of the evaluated method because in dataset with high class imbalance a method may presents a high accuracy value even though it fails to recognize the area of interest (object) [37]. Therefore, sensitivity (true positive rate) and specificity (true negative rate) were also used to evaluate the performance of the proposed network. The segmentation result is

Table 1
Comparison of deep learning methods on several datasets.

Network	Performance (%)								
	BRATS-2012			BRATS-2018			CBCT		
	Acc	Sen	Spe	Acc	Sen	Spe	Acc	Sen	Spe
Autoencoder [34]	98.30	89.49	98.31	97.84	70.24	98.02	94.66	87.39	94.54
FCN [21]	98.31	91.18	98.30	97.44	68.69	97.63	95.12	89.09	94.91
2D U-Net [5]	98.59	91.03	98.58	97.64	71.34	97.81	95.16	88.41	95.07
3D U-Net [27]	97.90	78.33	98.00	97.32	**74.54**	97.54	94.31	85.22	94.19
Multi-projection	98.07	**93.67**	98.06	97.49	74.51	97.62	94.26	**90.09**	94.07

Table 2
Comparison of cost function on the proposed Multi-projection network.

Cost function	Performance (%)								
	BRATS-2012			BRATS-2018			CBCT		
	Acc	Sen	Spe	Acc	Sen	Spe	Acc	Sen	Spe
RMSE	99.59	58.41	99.88	98.58	2.53	99.98	96.77	74.74	97.91
Cross-entropy (CE)	99.55	67.35	99.77	98.58	0.33	100	96.74	74.90	97.75
Globally-weighted RMSE	98.64	91.09	98.64	98.17	70.62	98.37	94.70	85.19	94.74
Globally-weighted CE	98.66	89.05	98.67	97.72	72.53	97.95	94.63	89.92	94.47
Proposed RMSE	98.68	92.16	98.67	98.19	71.14	98.41	95.21	88.52	95.11
Proposed CE	98.07	**93.67**	98.06	97.49	**74.51**	97.62	94.26	**90.09**	94.07

considered good if it has high accuracy, sensitivity, and specificity. High sensitivity means that the method has good ability to detect the object class, while high specificity means that the method has good ability to detect the background class.

4.1. Data splitting

Each of the datasets was split into a training set and a test set. The BRATS-2012 dataset, consisting of 80 MRI images, was divided into 70 training data and 10 testing data. The BRATS-2018 dataset, consisting of 285 MRI images, was divided into 250 training data and 35 testing data. The CBCT dataset, consisting of 7 images, was divided into 5 training data and 2 testing data. The 3D images from the dataset were randomly assigned for training or testing process. The test set is also used for the validation process. A number of experiments was conducted using k-fold cross validation for the data splitting process but the results were not significantly different from using the training and testing split. Therefore we chose not to use the cross validation method so that the computational cost would not be increased.

4.2. Comparison with other networks

Experiments were conducted to evaluate several network architectures for binary segmentation of 3D medical images. The architectures were: Convolutional Autoencoder [34], FCN [21], 2D U-Net [5], 3D U-Net [27], and the proposed Multi-projection Network. The 3D U-Net has the same network architecture as the 2D U-Net but it uses 3D volumetric image as the input and uses 3D kernels for its process. The size of the image batch for Autoencoder, FCN, and 2D U-Net was set to 128 2D images while the size of the image batch for 3D U-Net and Multi-projection Network was set to 1 3D image (128 2D images). For each training set and architecture, we set 50 as the training epoch because the given training error has reached convergence.

We implemented the architecture and the cost function of the compared methods according to their respective paper. However, their hyper-parameters setting, consist of the number of convolution and pooling layers, kernel size, stride, and number of output channels, are made similar with the proposed Multi-projection Network to provide equivalent comparison. In their respective paper, the compared methods usually use cross-entropy as their cost function and it is mentioned that weight should be used in case of imbalanced dataset. However, the weight is also a hyper-parameter. For this research, the weight value for the cost function of compared methods is set according to the ratio between minority and majority class in each dataset, in which we called as globally-weighted cross-entropy. We train all of the networks from scratch and do not use transfer learning.

Table 1 shows the comparison results for the BRATS-2012, BRATS-2018, and CBCT datasets, respectively. The network's performance was measured using accuracy (Acc), sensitivity (Sen), and specificity (Spe). The performance measurements in Table 1 show that the proposed Multi-projection Network had a higher sensitivity value than Autoencoder, FCN, and 2D U-Net for all datasets. This means that Multi-projection Network can handle the imbalance problem better than the other networks that use 2D kernels. However, 3D U-Net which uses 3D kernels had the highest sensitivity value compared to the other networks for BRATS-2018 dataset. It can be concluded that obtaining the 3D information for segmentation of 3D medical images is crucial to produce an accurate segmentation result in terms of the sensitivity metric. However, the 3D U-Net method had the lowest sensitivity value compared to the other networks for the BRATS-2012 and CBCT datasets. This is because for 3D U-Net, which takes 3D images as input, the BRATS-2012 and CBCT training sets do not contain enough images to make the network perform well.

4.3. Evaluation of cost function

The performance of Multi-projection Network using different cost functions was compared, as shown in Table 2. The compared cost functions are root mean squared error (RMSE), cross-entropy (CE), weighted RMSE using the global probability of each class in the dataset (globally-weighted RMSE), weighted cross-entropy using the global probability of each class in the dataset (globally-weighted CE), RMSE using the proposed weight (proposed RMSE), and cross-entropy using the proposed weight (proposed CE). This experiment was conducted on BRATS-2012, BRATS-2018, and CBCT datasets respectively, as shown in Table 2.

The performance evaluation in Table 2 shows that the use of the proposed weighted cost function produced a higher sensitivity

value, especially on datasets with high class imbalance, such as BRATS-2012 and BRATS-2018. The ratio between minority and majority class in the BRATS-2012, BRATS-2018, and CBCT datasets is 15:1000, 13:1000, and 45:1000, respectively. Higher class imbalance will result in lower sensitivity of non-weighted cost functions. The use of a weighted cost function that considers the class probability in each batch gives Multi-projection Network a slightly better accuracy and sensitivity value than the use of a globally-weighted cost function, which uses the class probability in the dataset directly.

5. Discussion

Although deep learning has many advantages in 3D medical image segmentation, this approach is not always the best method for specific datasets because it has limitations, such as requiring sufficient training data and its inability to deal with imbalanced datasets. The common approach of addressing the imbalanced dataset problem is to balance the dataset by reducing the number of majority class instances in a sample subset or by oversampling the minority class, including the creation of synthetic samples [38]. However, in this research we preferred to address the imbalanced dataset problem by modifying the network cost function rather than modifying the samples because of the characteristics of the input data that were used. In this research, 3D medical images were used as the input data that consisted of sequential 2D images. This characteristic will make it difficult to reduce the 2D slice images or add 2D synthetic images to the sample subset because it can damage the data sequence. Moreover, creating additional 3D synthetic images as samples is not an option because it does not solve the imbalanced dataset problem because the class imbalance in one 3D medical image for segmentation purpose is already high.

The proposed method was tested on datasets with class imbalance. High class imbalance makes harmful effect on the classification results. A method can be identified as better than others if it performs better on the data with high class imbalance [39]. Adding weight to the cost function of Multi-projection Network can solve the imbalanced dataset problem. By considering the local probability contained in each batch for the cost function's weight, the network's kernels will be updated according to all of the images that are contained in the input batch. Because the 2D slices are inserted sequentially to form the appropriate 3D image, there are input batches that have higher class imbalance than the others which results in higher error rate in some of the batches. After 50 iteration, the error of the training process using cross-entropy as the cost function (without weight) is about 0.07 on BRATS-2018 dataset, meanwhile the error rate of the training process using weighted cross-entropy is about 0.0075. Evaluation of the network's average error or cost rate in each iteration of training process shows that the proposed cost function gives an uneven cost graph due to its adaptive weight. However, this does not pose a problem because the rate of the training cost is globally decreased.

In this research, Multi-projection Network used max pooling for the pooling layer because it leads to faster convergence by selecting superior invariant features, which improves generalization performance [40]. The kernel sizes are important parameters for the convolution layer in deep learning networks. A large kernel size can capture more information from the image than a small kernel size. However, a larger kernel means more computation therefore it leads to higher computational cost and longer training time. In this research we chose to use 3 as the kernel size for all the convolution processes. The selection of a small kernel size is also done by many state-of-the-art deep learning methods to keep both computation and number of parameters contained [41]. Stride = 1 is usually used for the convolution process. However, in this research there are convolution layers that use stride = 2 to save the resource's space and speed-up the computational time. Multi-projection Network uses Rectifier Linear Unit (ReLU) activation function because it increases the network sparsity and makes the network learn faster [42–45].

The architecture of Multi-projection Network can be further developed for segmentation of 3D objects, where the encoders are used for 2D projections of the 3D object from different sides. However, because the number of encoders depends on the number of projections that are used, it is necessary to consider a training strategy that can reduce the dimensionality of the feature maps. The proposed Multi-projection Network method with it cost function showed promising results for binary segmentation of 3D medical images. However, many aspects can still be investigated and improved in future work. Further research regarding deep learning network parameters, such as effective mini-batch size and number of network layers related to the type of input data and the task of the network, needs to be conducted.

6. Conclusion

In this paper we proposed an advanced network architecture and a cost function for segmentation of 3D medical images. The proposed Multi-projection Network method can preserve resource memory by applying 2D kernels while still obtaining the 3D information from the image by incorporating a number of slices of the 3D image to achieve a good segmentation result. The proposed network's cost function addresses the imbalanced dataset problem by introducing an adaptive weight to the network cost function, which considers the probability of each class in the image. The proposed cost function improves the network's performance in detecting the minority class and can also be used for multiclass segmentation. It had accuracy, sensitivity, and specificity of 97.49%, 74.51%, and 97.62%, respectively, on the BRATS-2018 dataset, which has high class imbalance. Furthermore, the method for assigning adaptive class weight can also be applied to other network cost functions.

The experimental results showed that the proposed Multi-projection Network can reduce the effect of the imbalanced dataset problem and had the highest sensitivity value among the compared network architectures, i.e. Autoencoder, FCN (Fully Convolutional Network), 2D U-Net, and 3D U-Net, on the BRATS-2012 and CBCT datasets. Multi-projection Network had a slightly lower sensitivity value than 3D U-Net on the BRATS-2018 dataset, which confirms the importance of 3D information in segmentation of 3D medical images. Multi-projection Network had accuracy, sensitivity, and specificity of 94.26%, 90.09%, and 94.07%, respectively, on CBCT dataset that consists of 7 3D images. This proves that despite the small amount of training data, the proposed method can have excellent performance on segmentation of 3D medical images. The experimental results showed that the proposed method has the potential to be used and further developed in conducting 3D image analysis and other medical applications, such as brain cancer detection and oral surgery.

Declaration of Competing Interest

We wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

Acknowledgments

This work was supported by the Ministry of Research, Technology and Higher Education, Indonesia with grant number

135/SP2H/LT/DRPM/IV/2017. We are deeply grateful for PMDSU (*Program Magister Menuju Doktor untuk Sarjana Unggul*) and PKPI (*Program Peningkatan Kualitas Publikasi Ilmiah*) sandwich-like program, which enabled this joint research with Hiroshima University.

References

[1] D.L. Pham, C. Xu, J.L. Prince, Current methods in medical image segmentation, Annu. Rev. Biomed. Eng. 2 (1) (2000) 315–337.
[2] D.J. Withey, Z.J. Koles, Medical image segmentation: Methods and software, in: Noninvasive Functional Source Imaging of the Brain and Heart and the International Conference on Functional Biomedical Imaging, 2007. NFSI-ICFBI 2007. Joint Meeting of the 6th International Symposium on, IEEE, 2007, pp. 140–143.
[3] N. Sharma, L.M. Aggarwal, Automated medical image segmentation techniques, J. Med. Phys./Assoc. Med.Phys. India 35 (1) (2010) 3–14.
[4] W. Zhang, R. Li, H. Deng, L. Wang, W. Lin, S. Ji, D. Shen, Deep convolutional neural networks for multi-modality isointense infant brain image segmentation, NeuroImage 108 (2015) 214–224.
[5] O. Ronneberger, P. Fischer, T. Brox, U-net: convolutional networks for biomedical image segmentation, in: International Conference on Medical image computing and computer-assisted intervention, Springer, 2015, pp. 234–241.
[6] J. Kleesiek, G. Urban, A. Hubert, D. Schwarz, K. Maier-Hein, M. Bendszus, A. Biller, Deep mri brain extraction: a 3d convolutional neural network for skull stripping, NeuroImage 129 (2016) 460–469.
[7] S. Pereira, A. Pinto, V. Alves, C.A. Silva, Brain tumor segmentation using convolutional neural networks in mri images, IEEE Trans. Med. Imag. 35 (5) (2016) 1240–1251.
[8] M. Havaei, A. Davy, D. Warde-Farley, A. Biard, A. Courville, Y. Bengio, C. Pal, P.-M. Jodoin, H. Larochelle, Brain tumor segmentation with deep neural networks, Med. Image Anal. 35 (2017) 18–31.
[9] E. Goceri, N. Goceri, Deep learning in medical image analysis: recent advances and future trends, in: International Conference on Computer Science and Engineering (UBMK), 10, 2017, p. 5.
[10] O. Charron, A. Lallement, D. Jarnet, V. Noblet, J.-B. Clavier, P. Meyer, Automatic detection and segmentation of brain metastases on multimodal mr images with a deep convolutional neural network, Comput. Biol. Med. 95 (2018) 43–54.
[11] X. Fu, T. Liu, Z. Xiong, B.H. Smaill, M.K. Stiles, J. Zhao, Segmentation of histological images and fibrosis identification with a convolutional neural network, Comput. Biol. Med. 98 (2018) 147–158.
[12] D. Shen, G. Wu, H.-I. Suk, Deep learning in medical image analysis, Annu. Rev. Biomed. Eng. 19 (2017) 221–248.
[13] Y. Zheng, D. Liu, B. Georgescu, H. Nguyen, D. Comaniciu, 3d deep learning for efficient and robust landmark detection in volumetric data, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2015, pp. 565–572.
[14] P. Meyer, V. Noblet, C. Mazzara, A. Lallement, Survey on deep learning for radiotherapy, Comput. Biol. Med. 98 (2018) 126–146.
[15] J. Chen, L. Yang, Y. Zhang, M. Alber, D.Z. Chen, Combining fully convolutional and recurrent neural networks for 3d biomedical image segmentation, in: Advances in Neural Information Processing Systems, 2016, pp. 3036–3044.
[16] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, J. Garcia-Rodriguez, A review on deep learning techniques applied to semantic segmentation, CoRR abs/1704.06857 (2017) 1–23.
[17] P. Vorraboot, C. Lursinsap, S. Rasmequan, K. Chinnasarn, A modified error function for imbalanced dataset classification problem, in: Computing and Convergence Technology (ICCCT), 2012 7th International Conference on, IEEE, 2012, pp. 854–859.
[18] T. Brosch, Y. Yoo, L.Y. Tang, D.K. Li, A. Traboulsee, R. Tam, Deep convolutional encoder networks for multiple sclerosis lesion segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2015, pp. 3–11.
[19] S. Wang, W. Liu, J. Wu, L. Cao, Q. Meng, P.J. Kennedy, Training deep neural networks on imbalanced data sets, in: Neural Networks (IJCNN), 2016 International Joint Conference on, IEEE, 2016, pp. 4368–4374.
[20] G. Litjens, T. Kooi, B.E. Bejnordi, A.A.A. Setio, F. Ciompi, M. Ghafoorian, J.A. van der Laak, B. van Ginneken, C.I. Sánchez, A survey on deep learning in medical image analysis, Med. Image Anal. 42 (2017) 60–88.
[21] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 3431–3440.
[22] H.-C. Shin, H.R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, R.M. Summers, Deep convolutional neural networks for computer-aided detection: cnn architectures, dataset characteristics and transfer learning, IEEE Trans. Med. Imag. 35 (5) (2016) 1285–1298.
[23] H.-C. Shin, M.R. Orton, D.J. Collins, S.J. Doran, M.O. Leach, Stacked autoencoders for unsupervised feature learning and multiple organ detection in a pilot study using 4d patient data, IEEE Trans. Pattern Anal. Mach.Intell. 35 (8) (2013) 1930–1943.
[24] M. Drozdzal, E. Vorontsov, G. Chartrand, S. Kadoury, C. Pal, The importance of skip connections in biomedical image segmentation, in: Deep Learning and Data Labeling for Medical Applications, Springer, 2016, pp. 179–187.
[25] P.F. Christ, F. Ettlinger, F. Grün, M.E.A. Elshaera, J. Lipkova, S. Schlecht, F. Ahmaddy, S. Tatavarty, M. Bickel, P. Bilic, et al., Automatic liver and tumor segmentation of ct and MRI volumes using cascaded fully convolutional neural networks, CoRR abs/1702.05970 (2017) 1–20.
[26] N. Tomita, Y.Y. Cheung, S. Hassanpour, Deep neural networks for automatic detection of osteoporotic vertebral fractures on ct scans, Comput. Biol. Med. 98 (2018) 8–15.
[27] Ö. Çiçek, A. Abdulkadir, S.S. Lienkamp, T. Brox, O. Ronneberger, 3d u-net: learning dense volumetric segmentation from sparse annotation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2016, pp. 424–432.
[28] F. Milletari, N. Navab, S.-A. Ahmadi, V-net: fully convolutional neural networks for volumetric medical image segmentation, in: 3D Vision (3DV), 2016 Fourth International Conference on, IEEE, 2016, pp. 565–571.
[29] A. Prasoon, K. Petersen, C. Igel, F. Lauze, E. Dam, M. Nielsen, Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network, in: International conference on medical image computing and computer-assisted intervention, Springer, 2013, pp. 246–253.
[30] D. Yang, S. Zhang, Z. Yan, C. Tan, K. Li, D. Metaxas, Automated anatomical landmark detection ondistal femur surface using convolutional neural network, in: Biomedical Imaging (ISBI), 2015 IEEE 12th International Symposium on, IEEE, 2015, pp. 17–21.
[31] H.R. Roth, L. Lu, J. Liu, J. Yao, A. Seff, K. Cherry, L. Kim, R.M. Summers, Improving computer-aided detection using convolutional neural networks and random view aggregation, IEEE Trans. Med. Imag. 35 (5) (2016) 1170–1181.
[32] B.H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest, et al., The multimodal brain tumor image segmentation benchmark (brats), IEEE Trans. Med. Imag. 34 (10) (2015) 1993–2024.
[33] S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J.S. Kirby, J.B. Freymann, K. Farahani, C. Davatzikos, Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features, Scientif. Data 4 (2017) 170117.
[34] X. Guo, X. Liu, E. Zhu, J. Yin, Deep clustering with convolutional autoencoders, in: International Conference on Neural Information Processing, Springer, 2017, pp. 373–382.
[35] D.P. Kingma, J. Ba, Adam: a method for stochastic optimization, CoRR abs/1412.6980 (2014) 1–16.
[36] S. Ruder, An overview of gradient descent optimization algorithms, CoRR abs/1609.04747 (2016) 1–14.
[37] M. Xu, D.P. Papageorgiou, S.Z. Abidi, M. Dao, H. Zhao, G.E. Karniadakis, A deep convolutional neural network for classification of red blood cells in sickle cell anemia, PLoS Comput. Biol. 13 (10) (2017) e1005746, doi:10.1371/journal.pcbi.1005746.
[38] S.C. Wong, A. Gatt, V. Stamatescu, M.D. McDonnell, Understanding data augmentation for classification: when to warp? CoRR abs/1609.08764 (2016) 1–6.
[39] R. Zhu, Z. Wang, Z. Ma, G. Wang, J.-H. Xue, Lrid: a new metric of multi-class imbalance degree based on likelihood-ratio test, Pattern Recognit. Lett. 116 (2018) 36–42.
[40] J. Nagi, F. Ducatelle, G.A. Di Caro, D. Cireşan, U. Meier, A. Giusti, F. Nagi, J. Schmidhuber, L.M. Gambardella, Max-pooling convolutional neural networks for vision-based hand gesture recognition, in: Signal and Image Processing Applications (ICSIPA), 2011 IEEE International Conference on, IEEE, 2011, pp. 342–347.
[41] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A.L. Yuille, Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs, IEEE Trans. Pattern Anal. Mach.Intell. 40 (4) (2018) 834–848.
[42] X. Glorot, A. Bordes, Y. Bengio, Deep sparse rectifier neural networks, in: Proceedings of the fourteenth international conference on artificial intelligence and statistics, 2011, pp. 315–323.
[43] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (7553) (2015) 436.
[44] M.S. Kavitha, T. Kurita, S.-Y. Park, S.-I. Chien, J.-S. Bae, B.-C. Ahn, Deep vector-based convolutional neural network approach for automatic recognition of colonies of induced pluripotent stem cells, PloS one 12 (12) (2017) e0189974, doi:10.1371/journal.pone.0189974.
[45] F. Godin, J. Degrave, J. Dambre, W. De Neve, Dual rectified linear units (drelus): a replacement for tanh activation functions in quasi-recurrent neural networks, Pattern Recognit. Lett. 116 (2018) 8–14.