IEICE TRANSACTIONS

on Fundamentals of Electronics, Communications and Computer Sciences

DOI:10.1587/transfun.2024TAP0010

Publicized:2024/08/16

A PUBLICATION OF THE ENGINEERING SCIENCES SOCIETY

The Institute of Electronics, Information and Communication Engineers

Kikai-Shinko-Kaikan Bldg., 5-8, Shibakoen 3 chome, Minato-ku, TOKYO, 105-0011 JAPAN

LETTER

Application of Adversarial Training in the Detection of Calcification Regions from Dental Panoramic Radiographs

Sei TAKANO[†], *Nonmember*, **Mitsuji MUNEYASU**[†], *Fellow*, **Soh YOSHIDA**[†], **Akira ASANO**[††], *Members*, **Nanae DEWAKE**[†††], **Nobuo YOSHINARI**[†††], *and* **Keiichi UCHIDA**[††††], *Nonmembers*

SUMMARY Calcification regions, which may be observed on dental panoramic radiographs, are a sign of vascular disease. Therefore, automatic detection methods based on semantic segmentation (SS) have been proposed. However, because of the small amount of data in the available dataset, the segmentation accuracy was insufficient. This paper proposes a method that uses adversarial features (AFs) for this problem. We extend AFs, which are an adversarial training method for discriminative problems, to SS. The proposed method can improve performance, even with a small amount of data.

key words: *adversarial training, arteriosclerosis, semantic segmentation*

1. Introduction

Arteriosclerosis is a cause of strokes and cardiac strokes, which are the second most frequent causes of death without subjective symptoms worldwide [1]. Atherosclerosis may become calcified as it progresses [2], and it has been shown that arterial wall stiffness at carotid arteries is associated with arterial stiffness at other vascular sites [3]. Therefore, it is essential to detect signs of calcification as soon as possible. Carotid arteries, a common site of atherosclerosis [4], often appear in dental panoramic radiographs taken during a dental checkup, and calcification may also be observed. It is helpful for dentists to look for the presence of calcification in the carotid artery from dental panoramic radiographs and to encourage patients to consult a medical doctor regarding atherosclerosis.

Therefore, methods for automatically diagnosing the presence of calcification regions from dental panoramic radiographs have been proposed to assist dentists [5]–[8]. In particular, the methods in [6]–[8] achieve good identification accuracy using a detection method based on deep learning, but further improvement is needed. One reason for this insufficient accuracy is that deep learning requires the many parameters of the discriminative network to be optimized on a large dataset. However, collecting and annotating many dental panoramic radiographs is difficult because of cost and privacy issues. Therefore, it takes substantial work to construct a large dataset.

This paper proposes the use of regularization based on adversarial features (AFs) [9] when training a network for the semantic segmentation (SS) of calcification regions in dental panoramic radiographs. AFs introduce adversarial perturbations (APs) [10], [11] during training to increase the data virtually and have performed well on discriminative problems for convolutional neural network (CNN) models when the training datasets are small. We extend the AFs for use in SS and consider the effect on supervised learning (SL) and semi-supervised learning (SSL), which is a learning method using both labeled and unlabeled data. The use of unlabeled data enables the amount of data to be increased without increasing annotation cost.

Our contribution in this paper is summarized as follows.

1. AFs, which were initially designed for discriminative tasks, are extended to SS. This method is not just a combination of existing methods but an extension of the technique to a new field.
2. To design AFs for SS, we carefully considered where and how to effectively introduce APs for generating AFs in a complex hybrid Transformer-CNN structure such as TransFuse [12]. Therefore, in the paper, we experimentally verified the effect of the location at which the AFs are introduced on performance and showed the most effective application method.
3. We show that the AFs, which have previously been shown to be effective for CNNs, are also effective for the Transformer, a modern architecture. This original finding demonstrates the versatility of AFs.

This paper extends our previous work in [13]. The main difference is the addition of the SSL results. Experiments reveal the effectiveness of the proposed method compared with other methods and the effects on datasets that have different numbers of data.

2. Related Works

It has been shown that adding imperceptibly small perturbations called APs to the input data of a machine-learning model can cause misidentification [10], [11]. Such data are

[†]The authors are with the Graduate School of Science and Engineering, Kansai University 3-3-35 Yamate-cho, Suita-shi, Osaka, 564-8680, Japan

[††]The author is with the Faculty of Informatics, Kansai University, 2-1-1 Ryozenji-cho, Takatsuki-shi, Osaka 569-1095, Japan

[†††]The authors are with the School of Dentistry, Matsumoto Dental University, 1780 Hirookagoubara, Shiojiri-shi, Nagano 399-0781, Japan

[††††]The author is with the Division of Comprehensive Oral Treatment, Matsumoto Dental University Hospital, 1780 Hirookagoubara, Shiojiri-shi, Nagano 399-0781, Japan

called adversarial examples (AEs). In [11], an adversarial training method was proposed to construct robust discriminators by introducing AEs into training.

In this section, we introduce virtual adversarial training (VAT) [14], a method that uses APs to perform regularization in SL/SSL. VAT [14] is an adversarial training method that regularizes the loss function by incorporating $\mathcal{R}_{vadv}$, which is calculated by the following equation:

$$\mathrm{LDS}(\boldsymbol{x}_*, \boldsymbol{\theta}) = D[p(y|\boldsymbol{x}_*, \hat{\boldsymbol{\theta}}), p(y|\boldsymbol{x}_* + \boldsymbol{r}_{vadv}, \boldsymbol{\theta})] \quad (1)$$

$$\boldsymbol{r}_{vadv} = \underset{\boldsymbol{r};\|\boldsymbol{r}\|_2 \le \varepsilon}{\mathrm{argmax}} D[p(y|\boldsymbol{x}_*, \hat{\boldsymbol{\theta}}), p(y|\boldsymbol{x}_* + \boldsymbol{r})] \quad (2)$$

$$\mathcal{R}_{vadv}(D_l, D_{ul}, \boldsymbol{\theta}) = \frac{1}{N_l + N_{ul}} \sum_{\boldsymbol{x}_* \in D_l, D_{ul}} \mathrm{LDS}(\boldsymbol{x}_*, \boldsymbol{\theta}), \quad (3)$$

where $D[\cdot]$ is a function calculating the degree of separation between probability distributions, D_l and D_{ul} are the labeled and unlabeled datasets, respectively, N_l and N_{ul} are their sizes, $\boldsymbol{x}_*$ denotes the training sample, $\boldsymbol{r}_{vadv}$ is the AP, ε is an upper bound on the size of the perturbation, $\boldsymbol{\theta}$ is a network parameter, $\hat{\boldsymbol{\theta}}$ is a parameter at a specific iteration, and p is the probability distribution of the network output. This induces the network to maintain the output values for the adversarial samples generated from the input training data.

Using $\mathcal{R}_{vadv}$, the loss function is given by

$$l(D_l, \boldsymbol{\theta}) + \alpha \mathcal{R}_{vadv}(D_l, D_{ul}, \boldsymbol{\theta}), \quad (4)$$

where $l(D_l, \boldsymbol{\theta})$ is the loss with respect to the labeled dataset and α is a hyperparameter that adjusts the weight between the two terms. Since the computation of $\mathcal{R}_{vadv}$ does not require correct labels, Eq. (4) can be employed in SSL by setting $l(D_l, \boldsymbol{\theta})$ to 0 for the unlabeled data. The method that employs VAT by randomly generating $\boldsymbol{r}_{vadv}$ instead of using Eq. (2) is called random perturbation training (RPT) [14].

3. Proposed Method

We modify the AF-based method for the SS task and apply the method to the training of TransFuse [12] to improve performance with less data. TransFuse is employed as an example of an SS network that includes different kinds of network architectures to indicate the effectiveness of the AF-based method for networks with complex structures. For detection, we first perform the preprocessing in [6] and then use TransFuse to estimate the location of calcification regions.

3.1 AF-based method [9]

In the following, a summary of AFs is presented. The AF-based method proposed in [9] involves two types of AFs: constrained and unconstrained. First, features from each layer are obtained through feedforward propagation of the training data, and we choose a target hidden layer to generate the AFs. To obtain unconstrained AFs, a sub-network from the selected layer to the output layer is extracted and used to generate AEs. Additionally, a coefficient layer is added as its input layer to the sub-network, and constrained AFs are obtained similarly. Finally, the network is updated by minimizing the objective function using the original training data and their AFs.

The method for generating AFs is similar to VAT, except that sub-networks are extracted. Here, $\boldsymbol{x}$ and $\boldsymbol{h}^{(i)}$ represent the input and the node values of layer i (feature representation of the middle layer), respectively, and the AF $\boldsymbol{h}^{(i)}_{adv}$ corresponding to $\boldsymbol{h}^{(i)}$ is obtained using the following:

$$\boldsymbol{h}^{(i)}_{adv} = \boldsymbol{h}^{(i)} + \boldsymbol{r}_{adv} \quad (5)$$

$$\boldsymbol{r}_{adv} = \underset{\boldsymbol{r}}{\mathrm{argmax}}[D_i(\boldsymbol{r}, \boldsymbol{h}^{(i)}, \boldsymbol{\theta}); \|\boldsymbol{r}\| \le \varepsilon \|\boldsymbol{h}^{(i)}\|] \quad (6)$$

$$D_i(\boldsymbol{r}, \boldsymbol{h}, \boldsymbol{\theta}) = \mathrm{KL}(g_i(\boldsymbol{h}|\boldsymbol{\theta}), g_i(\boldsymbol{h} + \boldsymbol{r}|\boldsymbol{\theta})) \quad (7)$$

where $\boldsymbol{\theta}$ is a parameter of the network, $g_i(\boldsymbol{h})$ is the output of the extracted sub-network after layer i, ε is a hyperparameter indicating the magnitude of the perturbation, and $\mathrm{KL}(p, q)$ is the Kullback–Leibler (KL) divergence between the probability distributions p and q.

Because this method does not consider whether the generated AFs follow the distribution of the original features, it may create features that are not valid for actual data. Therefore, the conversion of AFs to linear combinations of feature representations in the middle layer is restricted to generate more accurate data.

Let B be the size of the mini-batch, $\boldsymbol{h}^{(i)}_j$ be the j-th node value of layer i in the batch, and $\boldsymbol{Z}^{(i)} = [\boldsymbol{h}^{(i)}_1, \cdots, \boldsymbol{h}^{(i)}_B]$ be the matrix of $\boldsymbol{h}^{(i)}_j$. From its definition, node value $\boldsymbol{h}^{(i)}_j$ can be expressed as follows:

$$\boldsymbol{h}^{(i)}_j = \boldsymbol{Z}^{(i)} \boldsymbol{c}_j, \quad (8)$$

where $\boldsymbol{c}_j$ is a vector of coefficients with the j-th element set to 1 and the others set to 0. Equation (8) can be viewed as a single-layer, fully connected network with $\boldsymbol{c}_j$ as the input and $\boldsymbol{Z}^{(i)}$ as the weight. Therefore, by adding a new input layer under the input layer of the sub-network, in which the input values represent $\boldsymbol{c}$ and the weight is set to $\boldsymbol{Z}^{(i)}$, we can obtain the same output as when $\boldsymbol{h}^{(i)}$ is input to the original sub-network by inputting $\boldsymbol{c}$ instead of $\boldsymbol{h}^{(i)}$. This new input layer is called a coefficient layer. Suppose we input the perturbation to this coefficient layer; a linear combination of the original middle layer outputs represents the perturbation input to the original sub-network and is obtained as follows:

$$\boldsymbol{Z}^{(i)}(\boldsymbol{c}_j + \boldsymbol{r}) - \boldsymbol{Z}^{(i)} \boldsymbol{c}_j = \boldsymbol{Z}^{(i)} \boldsymbol{r}. \quad (9)$$

Hence, the AFs obtained by this method are called constrained AFs.

Using these AFs, the network is regularized by adding a regularization term $R(\boldsymbol{x}, \boldsymbol{\theta})$ to the loss function, expressed as

$$R(\boldsymbol{x}, \boldsymbol{\theta}) = \mathrm{KL}(g_i(\boldsymbol{h}^{(i)} \mid \boldsymbol{\theta}), g_i(\boldsymbol{h}^{(i)}_{adv-u} \mid \boldsymbol{\theta})) + \mathrm{KL}(g_i(\boldsymbol{h}^{(i)} \mid \boldsymbol{\theta}), g_i(\boldsymbol{h}^{(i)}_{adv-c} \mid \boldsymbol{\theta})), \quad (10)$$

where $\boldsymbol{h}^{(i)}_{adv-u}$ and $\boldsymbol{h}^{(i)}_{adv-c}$ represent unconstrained and constrained AFs, respectively. Here, $\boldsymbol{h}^{(i)}_{adv-u}$ is obtained by Eq. (5) and let the j-th $\boldsymbol{h}^{(i)}_{adv-c}$ in the batch be $\boldsymbol{h}^{(i)}_{adv-c(j)}$. It is obtained by

$$\boldsymbol{h}^{(i)}_{adv-c(j)} = \boldsymbol{Z}^{(i)}(\boldsymbol{c}_j + \boldsymbol{r}_{adv}) = \boldsymbol{h}^{(i)}_j + \boldsymbol{Z}^{(i)}\boldsymbol{r}_{adv}. \quad (11)$$

Like VAT, SSL can be performed by computing the loss on unlabeled data using Eq. (10).

3.2 Extension of AFs for SS

AFs were developed for discriminative problems. For application in SS, the AF-based method must be modified to make it suitable for segmentation tasks. We hence propose an extension of AFs for SS.

The output of SS in the calcification region problem is the probability that each pixel is a calcification region. Since AFs were designed for discriminative problems, it is necessary to extend Eq. (7), which compares the probability distribution of perturbed and unperturbed results, such that it compares the distribution for each pixel. For this reason, we use the method applied in the context-aware VAT (CaVAT) method [15] to extend VAT to SS.

Specifically, the following changes are made. The segmentation network's outcome is obtained by applying the sigmoid function to the output score for each pixel. This outcome represents the probability that each pixel belongs to a calcification region. The KL divergence is calculated from the probability distribution of each pixel. Here, $g_i(\cdot)$ in Eq. (7) is replaced by a probability vector $\boldsymbol{f}_i(\cdot)$ consisting of the outputs of the segmentation network from the layer i onward, as is done in AFs. In other words,

$$D_i(\boldsymbol{r}, \boldsymbol{h}, \boldsymbol{\theta}) = \mathrm{KL}(\boldsymbol{f}_i(\boldsymbol{h} \mid \boldsymbol{\theta}), \boldsymbol{f}_i(\boldsymbol{h} + \boldsymbol{r} \mid \boldsymbol{\theta})) \quad (12)$$

is used for segmentation.

3.3 Application of the proposed method to TransFuse [12]

When detecting calcification regions in dental panoramic radiographs, the relative position of the calcification region in the image is often treated as known because the calcification is observed at the location of the carotid artery. Therefore, we employ TransFuse [12], an SS network for the discriminator task with a global feature-handling ability, to detect calcification regions. TransFuse consists of a Transformer-based network and a CNN-based network in parallel, as shown in Fig. 1. This hybrid architecture allows TransFuse to take advantage of both the Transformer's superiority in capturing global features and the CNN's superiority in capturing local features.

Especially in SS, global and local feature extractions are essential and often realized using encoder–decoder models consisting of multiple CNNs, such as PraNet [16]. However, CNN-based methods have been shown to have limitations in global feature extraction [17]. By contrast, the Transformer is superior at capturing global features because it can reference the entire input. Considering this feature, the proposed method is applied here to TransFuse. Moreover, this application indicates that the proposed method could also be applied to other segmentation models with complex network architectures.

To apply AFs, it is important to consider where and how to introduce APs effectively into the TransFuse as well as how to divide the TransFuse into multiple networks and extract features. In this case, the network is divided into three locations corresponding to the red dashed lines in Fig. 1. The Transformer part (Loc. 1), the ResNet part (Loc. 2), and the part immediately after the outputs of the Transformer and ResNet are combined (Loc. 3) are the locations that are split. Furthermore, according to the method in [10], we select only one location for each network update and generate AFs by applying an AP to the features obtained from that location. Splitting the Transformer and ResNet sections produces effective AFs that affect all the global or local feature maps outputted from each part. Furthermore, the split in the combination part enables APs to be added to the part that combines the feature maps from both the Transformer and ResNet parts and is expected to generate AFs that affect the outputs.

4. Experiment

4.1 Experimental conditions

We used 578 dental panoramic radiographs annotated by specialists at Matsumoto Dental University. Each image was preprocessed according to [6] and divided into left and right images to obtain 1156 images 352×352 pixels in size. The images were then converted to 224×224 pixel images for input to the network. From 1156 images, images without calcification regions were removed, and these images were divided into three groups: training data (491 images), validation data (61 images), and test data (141 images). To investigate changes in performance with respect to the size of the dataset, we created training datasets using 3/4, 1/2, and 1/4 of the original training data and performed the same experiment. Here, the original training data refer to the full dataset, and the reduced training data are called the sub-datasets. The number of images in the sub-datasets were 366, 244, and 122, respectively. Sub-datasets of each size were created by randomly selecting the necessary data from the full dataset and fixed. In SSL, a sub-dataset is used as the labeled data, while all the remaining data in the full dataset are treated as unlabeled data.

The Dice score and IoU were used as evaluation measures. The methods used were a method with no adversarial training ("normal"), RPT [11], VAT [11], and AFs ("proposed"). VAT and RPT were extended for application in SS using the same approach as used for the proposed method. Parameter values $\alpha = 1.0$ and $\varepsilon = 0.001$ for RPT and VAT, $\alpha = 0.01$, and $\varepsilon = 0.1$ for AFs were selected through prior experiments to obtain the best results. These param-

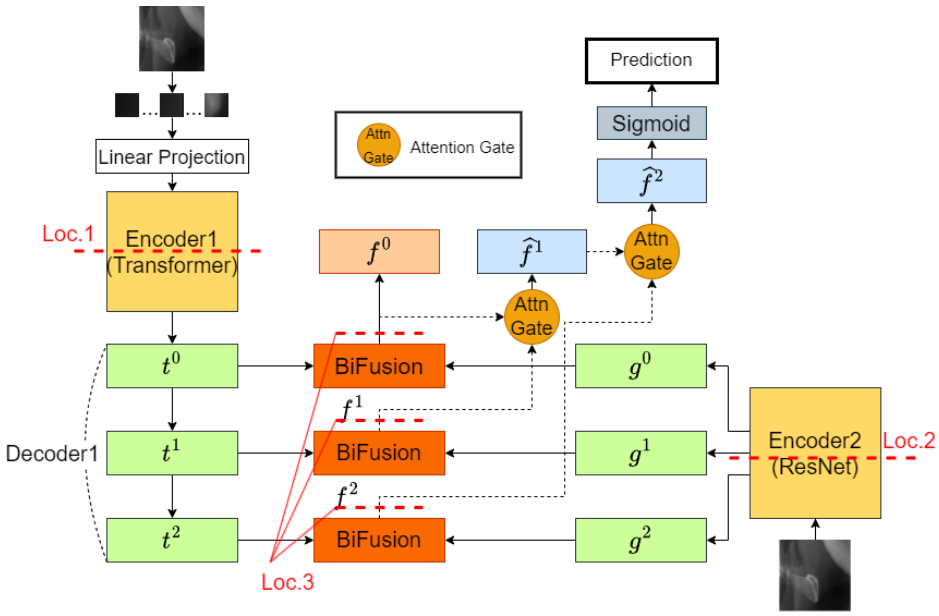

Fig. 1 Network architecture of TransFuse[12] and application of AFs.

eters were used throughout all experiments below. Ten experiments were conducted on the full dataset and each sub-dataset for each method to obtain the mean values and their standard deviation (SD) for each measure.

In the experiments, we randomly shifted, scaled, and rotated the data with a probability of 50%; randomly changed the luminance value with a probability of 50%; and flipped the data horizontally with a 50% probability to augment the data. As parameters for the experiment, the number of training epochs was 60, and the batch size was 20. Adam [18] was adopted as the optimization method; the learning rate was 1.0×10^{-4}, $\beta_1 = 0.5$, and $\beta_2 = 0.999$. In SSL, the batch size was set to 10 for labeled data and 20 for unlabeled data.

4.2 Experimental results

The mean and SD of Dice score and IoU for the test data after training on the full dataset are shown in Table 1. Bolded entries in the table indicate the best values. Table 1 reveals that the value for each measure obtained by the proposed method is superior to those of the other cases. These results show that the proposed method can be used for effective learning.

Table 2 presents the results using the sub-datasets with SL and SSL. The bolded entries indicate the best values, and the numbers in parentheses of the dataset entries indicate the number of images in the labeled/unlabeled dataset. On the sub-dataset, the proposed method outperforms all other methods. The result for the proposed method obtained by training with 366 images is especially comparable to that of the "normal" case obtained using 491 images, indicating that using the proposed method may be equivalent to learning with about 1.3 times as much data.

In the sub-datasets of 122 and 244 data, the proposed method (SSL) outperforms the other methods. In the 336 data case, the proposed method (SL) is superior to the other methods. These results may be because the amount of labeled data was insufficient compared with the amount of unlabeled data in the 122 and 244 data cases. This means that unlabeled data work well when there is little labeled data, but the effect diminishes when more labeled data are used. Therefore, SSL is particularly effective when there is little labeled data and a sufficient amount of unlabeled data.

Table 1 Comparison of each method.

Method	Result(mean±SD)	
	Dice	IoU
normal	0.6173(±0.0088)	0.4867(±0.0082)
RPT	0.6209(±0.0102)	0.4902(±0.0086)
VAT	0.6208(±0.0060)	0.4906(±0.0047)
Proposed	**0.6341(±0.0066)**	**0.5043(±0.0076)**

As an ablation study, we investigated the influence of the choice of the dividing location. The results are summarized in Table 3. The bolded entries indicate the best values. The "dividing locations" are shown in Fig. 1. The experiments were conducted 10 times on the full dataset for each location to obtain each measure. From Table 3, it is clear that the best performance was obtained when all locations were selected (Proposed). When a single location was specified, Loc. 1 and Loc. 2 performed better. This can be explained by the fact that Loc. 1 and Loc. 2 have a relatively larger impact on learning because there are more parameters in the latter part of the network than there are at Loc. 3. When two locations were specified, the performance was not necessarily better than in the case of one, but we can see that combinations with Loc. 3 perform better. Consequently, the proposed method is ultimately effective because Loc. 3 is valid when multiple locations are used.

Table 2 Results for the sub-dataset cases.

Dataset	Method	Result(mean±SD)	
		Dice	IoU
122 (122/369)	Normal	0.4979(±0.0132)	0.3639(±0.0126)
	RPT(SL)	0.4901(±0.0178)	0.3561(±0.0169)
	RPT(SSL)	0.5262(±0.0093)	0.3966(±0.0083)
	VAT(SL)	0.5026(±0.0134)	0.3704(±0.0130)
	VAT(SSL)	0.5264(±0.0146)	0.3976(±0.0137)
	Proposed(SL)	0.5062(±0.0145)	0.3709(±0.0139)
	Proposed(SSL)	**0.5421(±0.0128)**	**0.4056(±0.0138)**
244 (244/247)	Normal	0.5824(±0.0091)	0.4477(±0.0088)
	RPT(SL)	0.5883(±0.0091)	0.4539(±0.0078)
	RPT(SSL)	0.5967(±0.0079)	0.4644(±0.0071)
	VAT(SL)	0.5871(±0.0084)	0.4531(±0.0080)
	VAT(SSL)	0.5905(±0.0070)	0.4567(±0.0077)
	Proposed(SL)	0.5945(±0.0100)	0.4595(±0.0098)
	Proposed(SSL)	**0.6005(±0.0083)**	**0.4695(±0.0086)**
366 (366/125)	Normal	0.6147(±0.0075)	0.4812(±0.0076)
	RPT(SL)	0.6077(±0.0075)	0.4748(±0.0067)
	RPT(SSL)	0.6010(±0.0111)	0.4720(±0.0101)
	VAT(SL)	0.6071(±0.0071)	0.4740(±0.0068)
	VAT(SSL)	0.5995(±0.0086)	0.4704(±0.0079)
	Proposed(SL)	**0.6183(±0.0088)**	**0.4863(±0.0086)**
	Proposed(SSL)	0.6152(±0.0066)	0.4861(±0.0044)

Table 3 Comparison of dividing locations.

Locations	Result(mean±SD)	
	Dice	IoU
Loc. 1	0.6300(±0.0089)	0.5005(±0.0077)
Loc. 2	0.6302(±0.0098)	0.5000(±0.0097)
Loc. 3	0.6243(±0.0078)	0.4949(±0.0078)
Loc. 1&2	0.6288(±0.0060)	0.4984(±0.0059)
Loc. 1&3	0.6295(±0.0087)	0.5022(±0.0081)
Loc. 2&3	0.6300(±0.0089)	0.5005(±0.0077)
Proposed	**0.6341(±0.0066)**	**0.5043(±0.0076)**

5. Conclusions

In this paper, we proposed a method of applying AFs, a regularization method using adversarial examples, to the training method as a solution to the insufficient amount of data in medical image processing, especially in SS for the detection of calcification regions in dental panoramic radiographs. The proposed method provides effective training for various network architectures. We also confirmed that using AFs improves the accuracy when training with smaller datasets.

Acknowledgment

Part of this research was financially supported by a JSPS Grant-in-Aid for Scientific Research (C) (No. 23K03846), and we thank Kimberly Moravec, PhD, from Edanz for editing a draft of this manuscript.

References

[1] R. Lozano, M. Naghavi, K. Foreman, S. Lim, K. Shibuya, V. Aboyans, et al., "Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010," Lancet, Vol.380, pp.2095–128, 2012.

[2] A. Fleckenstein, M. Frey, J. Zorn, and G. Fleckenstein-Grün, "The role of calcium in the pathogenesis of experimental arteriosclerosis," Trends in Pharmacological Sciences, Vol.8, No.12, pp.496–501, 1987.

[3] P. Boutouyrie, P. Lacolley, and B. Laurent, "Arterial stiffness and cardiovascular outcome," Clinical and Experimental Pharmacology and Physiology, Vol.34, no.7, pp.647–651, 2007.

[4] T.Z. Naqvi and M.S. Lee, "Carotid intima-media thickness and plaque in cardiovascular risk assessment," JACC Cardiovasc Imaging, Vol.10, No.10, pp.1025–1038, 2014.

[5] T. Nasu, K. Kawachi, M. Muneyasu, K. Chamnongthai, A. Asano, K. Uchida, Y. Ishioka, N. Yoshinari, and A. Taguchi, "Detection of Calcification Region in Dental Panoramic Radiographs Using Snakes," Proc. 2017 International Workshop on Smart Info-Media Systems in Asia (SISA), Dazaifu, Fukuoka, Sept.6–8, pp.110–113, 2017.

[6] Y. Yamazaki, M. Muneyasu, S. Yoshida, A. Asano, K. Uchida, Y. Ishioka, N, Yoshinari, and A. Taguchi, "Improvement of detection accuracy of calcified area from dental panoramic radiograph using deep learning," IEICE Tech. Report, SIS2019-24, pp. 5–10, 2019.

[7] T. Murano, M. Muneyasu, S. Yoshida, K. Chamnongthai, A. Asano, K. Uchida, N. Dewake, Y. Ishioka, and N. Yoshinari, "Detection of Calcification Regions from Dental Panoramic Radiographs Based on Semantic Segmentation Using Deep Learning," Proc. 2021 International Workshop on Smart Info-Media Systems in Asia (SISA), online, Sept.20–22, pp.122-127, 2021.

[8] T. Murano, M. Muneyasu, S. Yoshida, K. Chamnongthai, A. Asano, and K. Uchida, "New Method of Detecting Calcification Regions in Dental Panoramic Radiographs Based on U-PraNet," Proc. 2021 20th International Symposium on Communication and Information Technologies (ISCIT), Tottori, Japan, Oct.20–22, W1-3, pp.11–14, 2019

[9] M. Ishii and A. Sato "Training Deep Neural Networks with Adversarially Augmented Features for Small-scale Training Datasets", Proc. 2019 International Joint Conference on Neural Networks (IJCNN), Budapest, Hungary, July 14-19, pp.1–8, 2019.

[10] C. Szegedy, W. Zaremba, I. Sutskever, J. Bruna, D. Erhan, I. Goodfellow, and R. Fergus, "Intriguing properties of neural networks," arXiv:1312.6199, 2013.

[11] I.J. Goodfellow, J. Shiens, and C. Szegedy, "Explaining and Harnessing Adversarial Examples", arXiv:1412.6572v3, 2015.

[12] Y. Zhang, H. Liu, and Q. Hu, "TransFuse: Transformers and CNNs for Medical Image Segmentation," Proc. 24-th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI), Strasbourg, France, Sept.27-Oct.1, pp.14-24, 2021.

[13] S. Takano, M. Muneyasu, S. Yoshida, A. Asano, N. Dewake, N. Yoshinari, and K. Uchida, "Application of Adversarial Training in Detection of Calcification Regions from Dental Panoramic Radiographs," Proc. 2023 International Workshop on Smart Info-Media Systems in Asia (SISA), Online, Aug.31-Sept.1, pp.7–12, 2023.

[14] T. Miyato, S. Maeda, M. Koyama, K. Nakae, and S. Ishii, "Distributional smoothing with virtual adversarial training," Proc. 4-th International Conference on Learning Representation (ICLR), San Juan, Puerto Rico, May 2-4, 2016.

[15] P. Wang, J. Peng, M. Pedersoli, Y.Zhou, C, Zhang, and C. Desrosiers "Context-aware virtual adversarial training for anatomically-plausible segmentation," Proc. 24-th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI), Strasbourg, France, Sept.27-Oct.1, pp.304-314, 2021.

[16] D.-P. Fan, G.-P. Ji, T. Zhou, G. Chen, H. Fu, J. Shen, and L. Shao, "PraNet: Parallel Reverse Attention Network for Polyp Segmentation," arXiv:2006.11392, 2020.

[17] N. Baker, H. Lu, G. Erlikhman, and P. Kellman, "Deep convolutional networks do not classify based on global object shape," PLoS Computational Biology, vol.14, No.12, e1006613, 2018.

[18] D.P. Kingma, and J. Ba "Adam: A method for stochastic optimization," arXiv:1412.6980, 2014.