Ensemble Attention Model and Data Augmentation for Generating Segmentation Models with Limited Medical Images

Jin yeong Chae
Department of Intelligence
University of Dongguk
Seoul, South of Korea
jiny419@dgu.ac.kr

Jong ick Hwang
Department of Computer engineering
University of Dongguk
Seoul, South of Korea
puppyjih@gmail.com

Ki yong Hong
Department of Plastic and Reconstructive Surgery
Dongguk University Medical center
Seoul, South of Korea
pskyhong@gmail.com

Jihie Kim
Department of Artificial Intelligence
University of Dongguk
Seoul, South of Korea
jihie.kim@dgu.edu

***Abstract*— In computer vision field, many deep learning models are used for semantic image segmentation. However, often as in medical fields, there are not enough data to generate accurate models. In this paper, we present an approach for mitigating the problem and making the most out of the available dataset, and show results with a problem of pressure ulcer image segmentation. We combine several data processing approaches and a pre-trained model from a relevant domain. First of all, an image pre-processing method using bilinear interpolation is used to resize images and normalize the images. Second, for data augmentation, we make use of rotation, reflection and watershed algorithm that are appropriate for the given skin lesion. For learning, we use a pre-trained deep learning model generated from skin wound images that are similar in nature with pressure ulcer images. An attention block is added that can provide hints on the pressure ulcer image features. Finally, an ensemble of varying combinations of the original and the augmented data is used in order to reduce overfitting. The resulting model provides Intersection of Union of 99.98%, Dice similarity coefficient of 85.26% and Recall of 92.78% for pressure ulcer segmentation which is better than existing approaches.**

Keywords*— *pressure ulcer; pre-training; data augmentation; ensemble; U-Net with Attention

I. INTRODUCTION

Semantic image segmentation is an important field in computer vision and is a high-level problem of not only for classifying photos, but also for understanding the scenes of the overall photograph [1]. The purpose of semantic image segmentation is to classify all the pixels into a specified number of classes and it is also called pixel-level classification as a prediction for all the pixels. Image segmentation is used in various fields such as land cover classification and road signs detection. It is also used in medical fields such as medical device detection during surgery, brain tumors and bedsore detection. However, often times, due to lack of relevant patient data, it is necessary to study efficient image segmentation and image processing techniques with limited data. Our study focuses on image segmentation for medical images where data is scarce, and this paper presents the results in localizing bedsore or pressure ulcer (PU). PU is a skin disease that occurs mainly in elderly patients with sensory and mobility disorders, as shown in Fig 1. Due to the nature of the disease, elderly patients' visits to medical institutions are very expensive, and research on remote management and diagnosis of PU has been increasing. For example, when the patient sends images of his/her sore parts, automated processing of the images can assist clinicians for efficient analyses. However, often times, due to lack of relevant patient data, automated analyses with machine learned models have been limited. There have been several deep learning approaches proposed for handling such limited image dataset for segmenting or classifying wound images. For example, [2] studied end-to-end method based on convolutional neural networks (CNN) using the 2,700 wound images. The Intersection of Union of the study was 47.3%. [3] also makes use of CNN model but performs data augmentation using color variation given less image datasets than [2]. Their method achieved up to 53% in Intersection of Union. In our work, we use U-Net an attention block rather than a simple CNN model and apply an ensemble to reduce overfitting with augmented dataset. We also make use of an image pre-processing and additional data augmentation techniques including rotation, reflection and watershed algorithm for given small datasets.

The presented study combines several approaches. We first perform image pre-processing techniques using the visual method of a medical expert observing the size and color of the ulcer. We also apply data augmentation techniques that are appropriate for medical images, such as rotation, reflection and watershed algorithm. In addition, we make use of pre-training using similar wound images. For generating segmentation models, an attention block is added to help the system focus on relevant visual features. The resulting model provides Intersection of Union of 99.98%, Dice similarity coefficient of 85.26% and Recall of 92.78% for pressure ulcer segmentation which is better than the existing approaches. The contribution of this work is two folds: First, we propose combining approaches to image pre-processing, data augmentation and an ensemble attention model to handle insufficient image data. Second, we applied them for PU image datasets and show that it is more effective than other current approaches. The following section describes related works in Section II. Description of data used in experiments is provided in Section III. And the details on our method are discussed in Section IV, while experiments and results are given in Section V. Finally, we discuss conclusions and future works in Section VI.

II. RELATED WORKS

In computer vision, image segmentation refers to dividing an image into pixels and identifying a region of interest. Image segmentation methods include fully convolutional networks (FCN) [4] and U-Net [5]. These methods can also be useful for medical image analyses. The quality of the analysis can be improved by conveying important information about the shape and the size of the lesion or organ. Much research has been done on the segmentation of lesions. Segmentation can not only identify the scope and the shape of the region of interest, but also help manage and label the area [6]. The most important characteristic of U-Net is the connection between the same resolutions layers in the analysis path for the extension path, and these characteristics provide the high resolutions characteristics required for the deconvolutional layer [6]. Because of these characteristics, it can be said that the U-Net structure is more suitable for medical image segmentation than other CNN model [7]. [8] proposed an improved model to solve the gradient vanishing problem that occurs when the layer becomes deeper and added a skip connection between layers so that the gradient is delivered well. In [9], densely connected convolutional networks (DenseNet) was proposed, which is inspired by the residual neural network (ResNet) idea. It applied a skip connection to all the layers of the entire network. Our work is inspired by the feature extraction in that the decoder concatenates the encoder feature and used it for down-sampling and up-sampling. The difference from the previous study is to design a model that segments to focus only the wound region by adding an attention block to the encoder.

For automated PU analyses, several image processing approaches have been applied, including traditional computer vision technologies [10][11][12]. In [10], PU image segmentation was performed based on contrast changes computed using synthetic frequencies. the authors applied various morphological operations for decomposition the images. example operations include erosion and dilation. From the original image, once the lines were extracted from the entire image, including the wound then the blurring boundary was cleaned using a dilation technique. Then, by applying erosion, unnecessary pixels outside the wound were eliminated and converted into a background image. As a test image, 51 sheets were used and showed an average correlation of 0.89. In this study, inspired by the boundary processing method [10], we use a marker-based watershed algorithm that can distinguish the boundary between a wound region and a general skin area and the algorithm was applied to augment data [13][14][15]. In [16], a reproducible chronic wound image was segmented using fuzzy divergence based thresholding technique. During the pre-processing of the image, the [red, green, blue] (RGB) image was converted to [hue, saturation, intensity] (HIS). The authors said that by converting the RGB image to the HIS values, the boundaries of the wound can be extracted more accurately. The resulting segmentation accuracy was 87.61%. In our work, PU pictures are taken with different devices (private cell phones) and the image can vary such as in brightness. Therefore, we used original RGB images rather than changing the color space HSI.

A lot of image augmentation techniques also exist due to lack of image data. According to [17], there are flipping, cropping, rotation and noise injection etc. There is also a study of segmenting a wound image using image data augmentation. In [18], the number of images was increased using five methods of rotation, shifting, zooming, shearing and flipping. In [3], data augmentation using color variation was applied to the wound image to increase the size of the training data three times. In [19], the wound image was augmented using DCGANs as a deep learning model approach. In our study, since we handle medical images, data augmentation techniques such as flipping and rotation which are appropriate to image nature are used. we applied reflection based on x-axis and y-axis and rotated at an angle between -90 and 90 on the image. In addition, a watershed algorithm which is not introduced in the previous study was applied to augment the data by making the boundary between the wound and the normal skin clear.

U-Net was recently proposed in [5] for segmentation of biomedical images and is currently being actively used as a state-of-the-art technique. In this paper, a baseline model for the segmentation of PU images was created with U-Net. We also apply an attention mechanism in the decoder. This mechanism mitigates the problem of information loss when compressing information into a fixed-sized vector. The basic idea of the attention mechanism starts by referring to the encoder information once again every time the decoder predicts the output information. The approach was initiated for natural language processing problems and now it is applied for image classification and segmentation as well. In [20], attention mechanism with skip-connection was used to decoder, that is up-sampling process. In this study, we apply a similar attention mechanism and when an up-sampling is performed in the decoder, attention blocks are used to extract critical features for PU images. However, unlike the attention method that existed in [20], we used an attention block that can extract both the channel and spatial features of the image to focus on the wound region, not just a simple attention block.

III. DATA

In this study, we used PU images collected by a hospital. The PU image dataset consists of a total of 101 RGB images as shown in Fig 1. Medetec Wound Database (MWD) was also used [21]. The MWD is image datasets that record various other wound types such as venous leg ulcer, arterial leg ulcer, pressure ulcer, malignant wound, and surgical wound infection, and consists of 264 images with a resolutions of 1024x731 as shown as in Fig 2. MWD has many wound images including pressure ulcer. Therefore, for the problem of insufficient data, pre-training was performed with the wound image of the MWD. In addition, training and testing were performed on the PU images.

Fig 1. Images of pressure ulcer patients.

Fig 2. Medetec Wound Data Base images.

IV. METHOD

A. *Image pre-processing*

All the original images are transformed with 224 x 224 size for training since the size of the original image was different respectively. In addition, we used a bilinear interpolation method to resize images as image pre-processing. The equations are listed below as “(1), (2), (3)”. The distance between the x -axis coordinate is corrected as shown in “(1), (2)”. And then about the y-axis coordinate, the distance is corrected as shown in (3). Also, the pixel brightness was adjusted during the image resizing process. For the pixels that are lumped together, the anti-aliasing technique was used to pre-processing to find the boundary of the wound [22].

$$f(x,\ y_1) \approx \frac{x_2 - x}{x_2 - x_1} f(Q_{11}) + \frac{x - x_1}{x_2 - x_1} f(Q_{21}) \tag{1}$$

$$f(x,\ y_2) \approx \frac{x_2 - x}{x_2 - x_1} f(Q_{12}) + \frac{x - x_1}{x_2 - x_1} f(Q_{22}) \tag{2}$$

$$f(P) = f(x,\ y) \approx \frac{y_2 - y}{y_2 - y_1} f(x,\ y_1) + \frac{y - y_1}{y_2 - y_1} f(x,\ y_2) \tag{3}$$

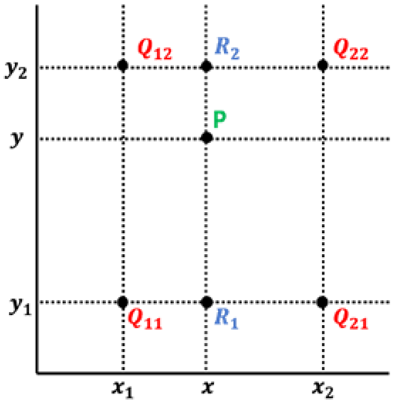

Fig 3. Bilinear interpolation coordinate

B. *Data augmentation*

After image pre-processing, data augmentation techniques such as image rotation, reflection and watershed algorithm are used to solve the problem of data shortage.

1) *Rotation, Reflection*

A random rotation method is applied to augment data. The random value between -90 and 90 is selected and used to increase images. And this process is repeated three times. We also used x-axis reflection and y-axis reflection. Therefore, a total of five images is augmented for each image.

2) *Watershed algorithm*

Fig 4. Left: Original Pressure ulcer image, Right: Pressure ulcer image with watershed algorithm

The watershed algorithm is a method judging the pixel value as the height and increasing the area from the minimum point to the boundary where the different minimum points meet each other [15]. It is important to find and use an appropriate threshold because the area varies depending on the reference threshold value judged to the same [13][14]. A high threshold value is used when segmenting the entire area of the image, and a low threshold value is used when the region characteristics are important in the image. In the case of PU images, since we have a goal to segment the PU region, not classifying the wound type, we used a high threshold value. Therefore, a watershed algorithm with a high threshold is applied and used for data augmentation.

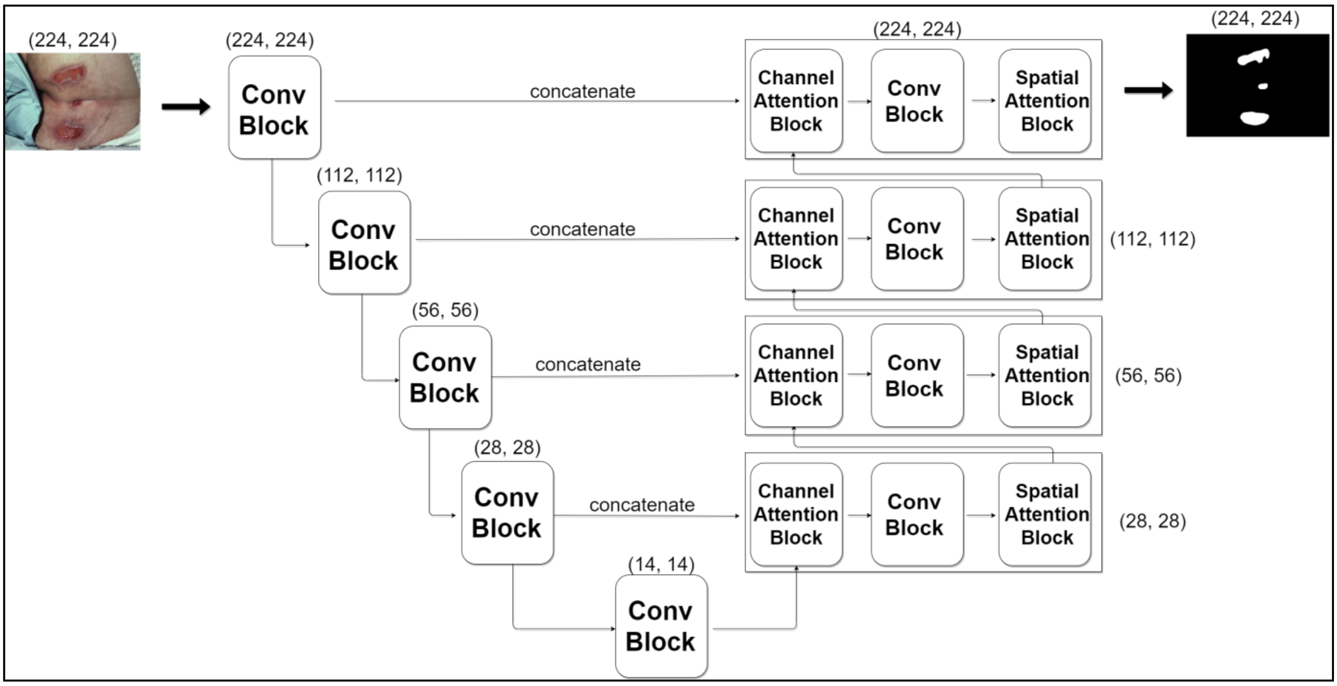

Fig 5. Attention U-Net architecture for Pressure ulcer segmentation

C. *Proposed model*

In this paper, we propose an Ensemble U-Net model with Attention in Fig 3. The images resized become the input of the model and the input data are passed through the convolutional block and attention block. The resulting output is image segmented for the PU region. Our model is divided up to encoder and decoder based on the center of Fig 5. The encoder in the model is the process of down-sampling the image. The down-sampling has a convolution block of 2-D to extract features. The decoder has an attention block and convolutional block of 2-D. In the decoder, spatial attention block and channel attention block are used to focus features and channel directions of the image so that only PU regions are segmented. In addition, when up-sampling is implemented in the decoder, the features of the encoder were concatenated to the decoder for the better up-sampling as shown in Fig5. Therefore, our model architecture can focus on the important area of the image and segment of the PU region.

1) *Pre-training*

Based on the fact the weights learned from similar images can improve the performance of the network, this study proposes a pre-training method. Since the number of training data required for segmentation of the PU area is insufficient, similar wound images are first learned. The weights learned from the wound image are used to segment the PU image. In this study, after pre-training using MWD images, PU images were retrained.

2) *Ensemble*

The fact that the distribution of data with data augmentation techniques differs from the original data distribution occurs. Therefore, an ensemble model is proposed to solve this. The ensemble model can solve the overfitting problem that can be caused by data augmentation. We ensembled model in Fig3 with three different weights for training data to which image pre-processing and data augmentation techniques were applied. And we set a threshold value used to determine a pixel as a PU area. The pixel is predicted to be a PU area if the value is above the threshold value and normal skin if it is below the threshold value. An Otsu algorithm was used to set this threshold [23]. The Otsu algorithm is an algorithm that finds the intensity value that can be best divided into two classes using the histogram of the image and sets the value as a threshold value.

V. RESULTS

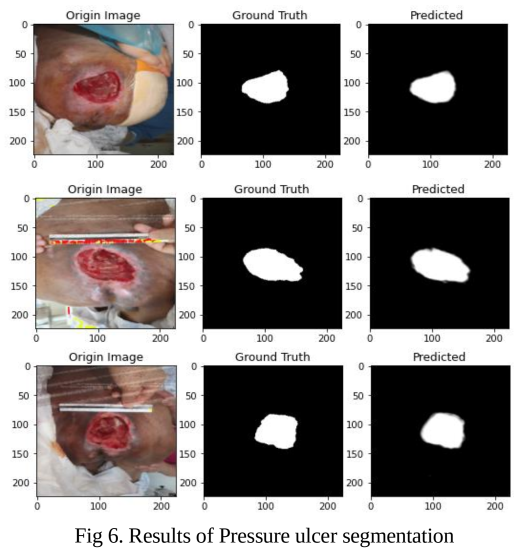

Fig 6. Results of Pressure ulcer segmentation

The metrics of evaluation are Intersection over Union (IoU), Dice similarity coefficient (DSC) and Recall. The following equations are metrics.

$$IoU = \frac{Area\ of\ Overlap}{Area\ of\ Union} \tag{4}$$

$$DSC = \frac{2 * Area\ of\ Overlap}{Total\ pixels\ combined} \tag{5}$$

$$Recall = \frac{TP}{TP + FN} \tag{6}$$

TABLE I.
MEDETEC WOUND DATA BASE RESULTS.

	IoU(%)	DSC(%)	Recall(%)
Wang et al. [2]	47.3	-	-
Pholberdee et al. [3]	53	-	-
Ensemble (Attention block + U-Net) x 3	99.98	89.71	96.67

Table 1. shows performance results that compare with the prior work with ours with MWD images only. Our model shows a better performance than the results from previous studies.

We also evaluated our model with the PU images described in Section III. The Attention U-Net and the ensemble model are trained using 81 train datasets. The test set consisted of the rest 20 PU images. The PU regions segmented are shown in Fig 5 and the results of models are shown in the following Table 2.

TABLE II. COMPARISON MODEL RESULTS.

	IoU(%)	DSC(%)	Recall(%)
Wang et al. [2]	47.3	-	-
Pholberdee et al. [3]	53	-	-
Attention block + U-Net	99.98	78.85	91.60
Ensemble (Attention block + U-Net) x 3	99.98	85.26	92.78

As shown in Table II. Approaches had IoU of 47.3% and 53%, respectively. As a single model, ‘Attention block + U-Net’ showed a better performance than [2] and [3] in IoU. With the ensemble model, DSC was increased by about 6.41%. That is, our approach that combines image pre-processing and data augmentation showed significant better performance.

VI. CONCLUSION & FUTURE WORKS

In this study, we proposed an approach for and making the most out of the available dataset, when the given image dataset is limited. We combined image pre-processing, data augmentation and ensemble attention model. By applying this method to PU image segmentation, we were able to achieve performance improvement over existing approaches. We expect that the approach can be applicable to similar image processing problems when the given image dataset is limited. As future works, we plan to study a model that can further improve the pressure ulcer analysis by making use of temporal changes. We are also interested in estimating the depth of the wound from limited image dataset.

REFERENCES

[1] Liu, Xiaolong, Zhidong Deng, and Yuhan Yang. “Recent progress in semantic image segmentation.”, Artificial Intelligence Review (2018): 1-18

[2] C. Wang et al., “A unified framework for automatic wound segmentation and analysis with deep convolutional neural networks”, 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Milan, pp. 2415-2418, 2015.

[3] N. Pholberdee, C. Pathompatai and P. Taeprasartsit, “Study of Chronic Wound Image Segmentation: Impact of Tissue Type and Color Data Augmentation”, 2018 15th International Joint Conference on Computer Science and Software Engineering (JCSSE), Nakhonpathom, 2018, pp. 1-6, doi: 10.1109/JCSSE.2018.8457392.

[4] Evan S, Jonathan L, Trevor D “Fully convolutional networks for semantic segmentation”, Proceedings of CVPR(Computer Vision and Pattern Recognition) , 2016

[5] Olaf R, Philipp F, Thomas B “U-Net: Convolutional Networks for Biomedical Image Segmentation”, Proceedings of CVPR (Computer Vision and Pattern Recognition), 2015

[6] Mohammad HH, Wenjing J, Xiangjian Hm, Paul K “Deep Learning Techniques for Medical Image Segmentation: Achievements and Challenges” , Springer, 2019r

[7] Yeong CL, Sang-Hyuk J, Hong-Hee W “WonDerM: Skin Lesion Classification with Fine-tuned Neural Networks”, Proceedings of CVPR (Computer Vision and Pattern Recognition), 2018

[8] K. He, X. Zhang, S. Ren and J. Sun, “Deep Residual Learning for Image Recognition,” 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016, pp. 770-778, doi: 10.1109/CVPR.2016.90.

[9] G. Huang, Z. Liu, L. Van Der Maaten and K. Q. Weinberger, “Densely Connected Convolutional Networks,” 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, 2017, pp. 2261-2269, doi: 10.1109/CVPR.2017.243.

[10] Ortiz, D.P., Sierra-Sosa, D. & Zapirain, B.G. “Pressure ulcer image segmentation technique through synthetic frequencies generation and contrast variation using toroidal geometry”, BioMed Eng OnLine 16, 4, 2017 doi: 10.1186/s12938-016-0298-3

[11] M. Elmogy, B. García-Zapirain, C. Burns, A. Elmaghraby and A. Ei-Baz, “Tissues Classification for Pressure Ulcer Images Based on 3D Convolutional Neural Network,” 2018 25th IEEE International Conference on Image Processing (ICIP), Athens, 2018, pp. 3139-3143, doi: 10.1109/ICIP.2018.8451119.

[12] David, Ortiz P et al. “Pressure ulcer image segmentation technique through synthetic frequencies generation and contrast variation using toroidal geometry.” Biomedical engineering online vol. 16,1 4. 6 Jan. 2017, doi:10.1186/s12938-016-0298-3

[13] K V, Lalitha & R, Amrutha & Michahial, Stafford & M, Dr. “Implementation of Watershed Segmentation”. IJARCCE. 5. 196-199, 2016, 10.17148/IJARCCE.2016.51243.

[14] M. Bai and R. Urtasun, "Deep Watershed Transform for Instance Segmentation", 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, 2017, pp. 2858-2866, doi: 10.1109/CVPR.2017.305.

[15] Anna Fabijanska "Normalized cuts and watersheds for image segmentation, IET Conference on Image Processing", mage Processing (IPR 2012), IET Conference on, 2012

[16] Mukherjee et al. "Automated Tissue Classification Framework for Reproducible Chronic Wound Assessment. BioMed Research International", Vol. 2014 (2014). 9. 10.1155/2014/851582.

[17] Shorten, C., Khoshgoftaar, T.M. "A survey on Image Data Augmentation for Deep Learning", J Big Data 6, 60 , 2019, doi: 10.1186/s40537-019-0197-0

[18] Shenoy, Varun & Foster, Elizabeth & Aalami, Lauren & Majeed, Bakar & Aalami, Oliver. "Deepwound: Automated Postoperative Wound Assessment and Surgical Site Surveillance through Convolutional Neural Networks", 2018

[19] Zhang, Junnan & Zhu, En & Guo, Xifeng & Chen, Haohao & Yin, Jianping."Chronic Wounds Image Generator Based on Deep Convolutional Generative Adversarial Networks" 36th National Conference, NCTCS 2018, Shanghai, China, October 13–14, 2018, Proceedings, doi: 10.1007/978-981-13-2712-4_11.

[20] Oktay et al. "Attention U-Net: Learning Where to Look for the Pancreas.", 2018

[21] S. Thomas, "Wound dressings," in The Epidermis in Wound Healing, D. T. Rovee and H. I. Maibach, Eds., CRC Press, Boca Raton, Fla, USA, 2004,Website:http://www.medetec.co.uk/files/medetec/imagedatabases.html

[22] Anton Trusov and Elena Limonova "The analysis of projective transformation algorithms for image recognition on mobile devices", Proc. SPIE 11433, Twelfth International Conference on Machine Vision , January 2020

[23] D. Liu and J. Yu, "Otsu Method and K-means", 2009 Ninth International Conference on Hybrid Intelligent Systems, Shenyang, 2009, pp. 344-349, doi: 10.1109/HIS.2009.74.