Garbage

by Garbage Garbage

Garbage Image Segmentation Using Combination of Thresholding Algorithms and Pyramid Scene Parsing Network

Daffa Muammar
Computer Science
Universitas Diponegoro
Semarang, Indonesia
daffamuammar@student.undip.ac.id

Sukmawati Nur Endah*
Computer Science
Universitas Diponegoro
Semarang, Indonesia
*corresponding author:
sukmane@lecturer.undip.ac.id

Priyo Sidik Sasongko
Computer Science
Universitas Diponegoro
Semarang, Indonesia
priyoss_undip@yahoo.co.id

Retno Kusumaningrum
Computer Science
Universitas Diponegoro
Semarang, Indonesia
retno@live.undip.ac.id

Khadijah
Computer Science
Universitas Diponegoro
Semarang, Indonesia
khadijah@live.undip.ac.id

Rismiyati
Computer Science
Universitas Diponegoro
Semarang, Indonesia
rismiyati@live.undip.ac.id

Lestari Handayani
PRISME Laboratory, EA 4229,
INSA-CVL,
F18020, Bourges, France;
Informatics Engineering
UIN Suska Riau
Pekanbaru, Indonesia
lestari.handayani@insa-cvl.fr

***Abstract*— Garbage is one of the biggest problems in Indonesia, even most of country in the world. The amount of garbage increases every year, therefore a reliable garbage management system is needed to prevent the contamination of garbage on environment. Garbage sorting (based on its material) is an important part of garbage management system. Since the amount of garbage is really high, a fast and accurate garbage sorting process is needed to improve the overall garbage management system. One of the most important tasks in garbage sorting process is garbage image segmentation. This research proposes garbage image segmentation process by using Pyramid Scene Parsing Network (PSPNet) and three binary images from the combination of thresholding algorithms as the input for the network. Then it is compared with PSPNet that uses a RGB image as an input. The results show that when using PSPNet, RGB image is better than the image from combination of thresholding algorithms as an input for PSPNet with the maximum F1 Score is up to 98%. However, the results of this comparison are competitive where the difference of F1 score between the proposed method and PSPNet with RGB image is less than 0.02.**

Keywords—pspnet, garbage image, image segmentation

I. Introduction

Garbage is one of the biggest problems in Indonesia, even most of country in the world. In accordance with The Indonesian Ministry of Environment and Forestry, Indonesia produces 175.000 tons of garbage per day or equivalent to 64 million tons per year. The garbage composition is dominated by biodegradable garbage which is 60% of total garbage. Plastic garbage is in the second place with 14% of total garbage, followed by paper garbage with 9% of total garbage. The rest of the garbage compositions are metal garbage, textile garbage, glass garbage, et cetera. The amount of garbage increases every year, therefore a reliable garbage management system is needed [1].

Sorting of garbage materials is necessary part of waste management system. Since the amount of garbage is really high, a fast and accurate garbage sorting process is needed to improve the overall garbage management system which can be done by using machine learning.

One of the most important tasks in garbage sorting process is garbage image segmentation. Image segmentation is used to separate the main object from another object in an image. The output is a mask image which separates the main object (foreground) from another object (background). There are some methods in image segmentation like UNet [2], LinkNet [3], MobileNet [4], and PSPNet [5]. PSPNet (Pyramid Scene Parsing Network) is a deep learning method which needs a large image dataset to perform well. PSPNet has some advantages, such as it can capture global context in an image through pyramid pooling module. This research uses PSPNet as a method for garbage image segmentation because PSPNet outperformed other image segmentation method such as FCN, DeconvNet, and CRF-RNN, which is able to achieve mIOU score of 85,4% in PASCAL VOC 2012 benchmark and 80,2% in Cityscapes benchmark [5].

Zeng et al (2019) found that segmentation process using the input from three binary images from different background substraction algorithms (BGS) which is then combined by CNN encoder-decoder, can improve the segmentation result [6]. In accordance with that achievement, we propose the use of binary images from combination of some thresholding algorithms as input for PSPNet. Then we compare such type of input with RGB image.

978-1-7281-9526-1/20/$31.00 ©2020 IEEE

PSPNet as a deep learning architecture needs a large data to perform well. Therefore, we use Trashnet public dataset as the data source for PSPNet because it has large amount of garbage images from various types and it is free to download. This paper will discuss research methodology, parameters, and results.

II. Methodology

This research methodology comprises of data collection, image resolution conversion, ground truth generation, preprocessing, data separation, segmentation, and evaluation. Those steps can be seen at Fig 1.

A. Data Collection

We use garbage images as the data in this study. The garbage images are obtained from public dataset. We use Trashnet dataset which has 2.527 of garbage images. This dataset consists of 501 glass garbage images, 594 paper garbage images, 403 cardboard garbage images, 482 plastic garbage images, 410 metal garbage images, and 137 other garbage images. All images have the resolution of 512 x 384 pixels and were taken by smartphone camera.

B. Image Resolution Conversion

There are some images in Trashnet dataset which do not have background object or the main object is all over the image. In segmentation task, it can be a problem. To solve it, we cut the image then add a solid gray color as the background of the image. Then, to speed up model training time all images are downsampled to 224 x 224 pixels.

C. Ground Truth Generation

Ground truth is made to all images as the actual output of the segmented image. We do it manually because Trashnet dataset does not have ground truth to use for segmentation task. The generated ground truth is then validated by expert in environmental engineering.

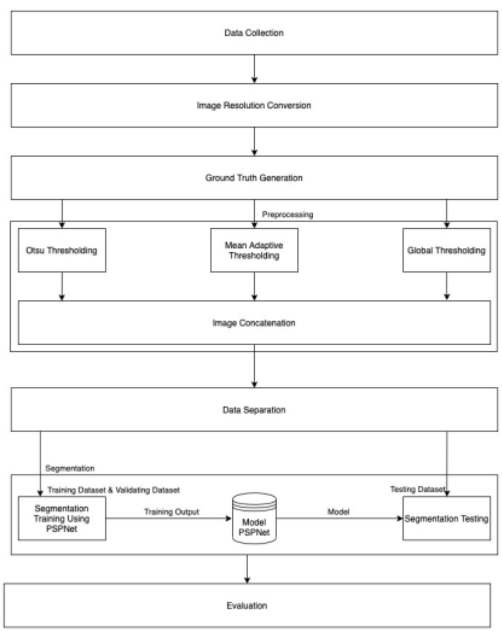

Fig. 1. The research methodology of garbage image segmentation

D. Preprocessing

All images are preprocessed before they are fed to PSPNet. Preprocessing is applied by using three image thresholding algorithms. The image thresholding algorithms used in this research are otsu thresholding, adaptive mean thresholding, and global thresholding. Each thresholding algorithm has an output of binary image. The output of those thresholding algorithms are concatenated as a 3 channel image.

Otsu thresholding is one of image segmentation method which is using a threshold value. Some calculation is needed to get the proper threshold value. The first step is getting the grayscale histogram of the image. The threshold value from a grayscale image is stated as k. Later, pixel with smaller value than k will have the value of 0 and pixel with greater value than k will have the value of 1. The value of k is ranged from 0 to L-1 where L has a total value of 255. The second step is calculating the probability value of all pixels. The third step is calculating the cumulative amount for L=0 to L=255. The fourth step is calculating the mean cumulative for L=0 to L=255 and calculating the global mean intensity. The fifth step is calculating the within class variance. The maximum value from the calculated within class variance is used as threshold value (k) as stated below [7] :

$$\sigma_B^2(k^*) = max_{1\le x\le L}\, \sigma_B^2(k) \quad (1)$$

Mean adaptive thresholding is one of image segmentation method which is using threshold value. The threshold value (k) is calculated through mean of n x n neighboring pixels values from pixel T(x,y). All pixels values are subtracted from a constant C. Pixel with smaller value than k will have the value of 0 and pixel with greater value than k will have the value of 1. The value of k is ranged from 0 to 255 with a is the sum of neighboring pixels values and b is the number of neighboring pixels [8].

$$k = a / b \quad (2)$$

Global thresholding is one of image segmentation method which is using threshold value. Pixel with smaller value than threshold value will have the value of 0 and pixel with greater value than threshold value will have the value of 1. The same threshold value is applied to all pixels in an image [9].

E. Data Separation

Preprocessed images are separated to testing dataset, training dataset, and validating dataset. We use static method to make the data separate. In general, the ratio of training dataset : testing dataset : validating dataset is 70 : 15 : 15 or 80 : 10 : 10.

F. Segmentation

Segmentation process comprises of training process and testing process. The goal of training process is to build a model that can be used to segment garbage image by using training dataset and validating dataset, while the goal of testing process is to test the performance of the model from training process by using testing dataset. We use PSPNet as the architecture in this study. In general, the architecture of PSPNet can be seen at Fig 2.

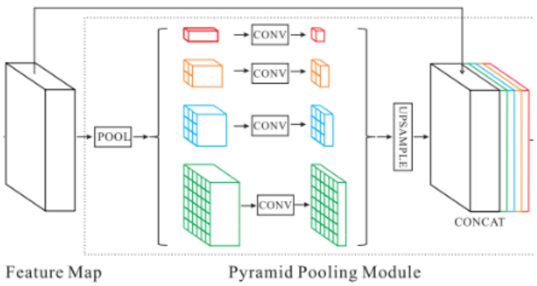

Fig. 2. PSPNet Architecture

PSPNet (Pyramid Scene Parsing Network) is an architecture that can be used in image segmentation. This architecture is known for its accuracy and it is better than other models such as FCN, DeepLab, and UNet in terms of scene parsing [5]. The main component of PSPNet is a Pyramid Pooling Module.

Input image is fed to Convolutional Neural Networks (CNN) architecture. The CNN architecture in PSPNet is Residual Network (ResNet) with dilated network strategy. The dilated network strategy is DeepLab. The output of ResNet architecture is a feature map with 8 times smaller size than the input image. The feature map then enters Pyramid Pooling Module. In Pyramid Pooling Module there are 4 main processes. The processes are sub-region average pooling, 1 x 1 convolution, 3 bilinear interpolation, and concatenation. The output from Pyramid Pooling Module will enter a convolutional layer to generate the output segmented image [5].

In conventional deep learning networks, generally there are convolutional layer then fully connected layer without skip connection. But if the number of layer is too high, vanishing / exploding gradients can become a problem. It affects in increasing error in training process. To overcome this problem, a Residual Network (ResNet) is made. ResNet has skip connection which can avoid vanishing / exploding gradients problem [10].

Skip connection is used to add input x to output after some weight layers. Since the number of layers used is high, the time complexity is also high. To overcome this problem, a bottleneck design is needed. The 1 x 1 convolutional layer is added at the beginning and the end of network to lessen the parameters without a significant performance decrease. The working of skip connection can be seen at Fig 3 while the comparation between network with bottleneck design and without bottleneck design can be seen at Fig 4.

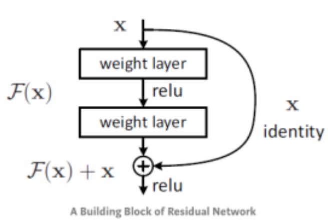

Fig. 3. Skip Connection

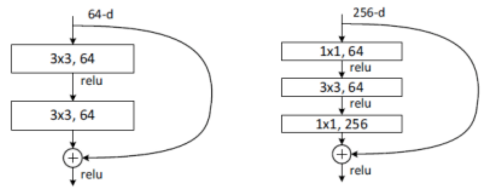

Fig. 4. Network comparation without bottleneck design (left) and with bottleneck design (right).

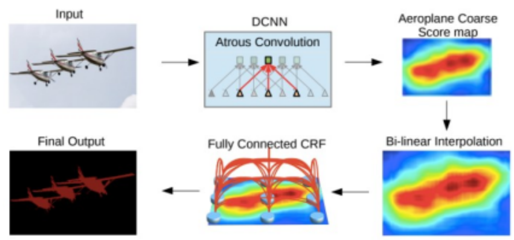

Fig. 5. DeepLab Steps

DeepLab is a feature extraction method in dilated networks. There are 3 main components of DeepLab. The main components are Atrous Convolution, Astrous Spatial Pyramid Pooling, and Fully Connected Conditional Random Field [11]. Atrous convolution is useful in controlling the resolution where feature is taken to Deep Convolutional Neural Networks (DCNN) which can be VGG or ResNet. It makes DCNN can see wider context without adding parameter. Atrous Spatial Pyramid Pooling (ASPP) is useful in segmenting object in various scale. ASPP checks convolutional feature layer before ASPP by using filters in various sampling rates and effective context field of view therefore objects and contexts from various scales can be captured in an image. To improve the localization of the edge of an object, DCNN method and probabilistic graphic models need to be combined. The frequently used method to get invariance but sacrifice localization accuracy is combining max pooling and downsampling in DCNN. To solve this, the response of last layer of DCNN and fully connected Conditional Random Field (CRF) is combined.

In general, DeepLab can be described as a fully convolutional DCNN then using Atrous Convolution to decrease the degree from signal downsampling from 32x to 8x. Then, there is bilinear interpolation to increase the feature map to the same resolution as the input image. After bilinear interpolation, there is fully connected CRF to improve the segmented image. All steps in DeepLab can be seen in Fig 5

Pyramid Pooling Module consists of 4 main process. The main processes are sub-region average pooling, 1 x 1 convolution, 3 bilinear interpolation, and concatenation. Sub-region average pooling consists of some feature map stages. There are average pooling, 1 x 1 convolution, and bilinear interpolation in every stage. In general, there are 4 stages of feature map. At first stage, feature map is used as a whole. At second stage, feature map is divided into 2 x 2 sub-region. At third stage, feature map is divided into 3 x 3 sub-region and at fourth stage, feature map is divided into 6 x 6 sub-region.

At the end of Pyramid Pooling Module, all feature maps from Pyramid Pooling Module is concatenated with feature maps from DeepLab. The combined feature maps then enter a convolutional layer which also play a role as fully connected layer to output predicted image of the segmentation.

G. Evaluation

We use F1 Score as the evaluation metric in this study. F1 Score or also known as Dice Coefficient is an evaluation metric that can be used in image segmentation [12]. This metric compares predicted image with ground truth image. To calculate F1 Score a formula as in (3) is used where AoO is the area of overlap and N is the total number of pixels in predicted image and ground truth image.

$$D = (2\ x\ \text{AoO}) / \text{N} \qquad (3)$$

III. EXPERIMENT RESULTS

We use 2 scenarios in this study. Scenario 1 uses the input image from combination of three thresholding algorithms (otsu, mean adaptive and global thresholding), while scenario 2 uses the RGB input image (without any thresholding algorithm). The parameters in the experiments are the dropout value and the number of ResNet layer. Table 1 shows the scenarios we use in this experiment.

TABLE I. TESTING SCENARIOS

Scenario	**Experiment**	**Preprocessing**		**Dropout**			**ResNet Layer**	
		With	***Without***	**0**	**0.1**	**0.3**	**50**	**38**
1	a	v		v			v	
	b	v			v		v	
	c	v				v	v	
	d	v		v				v
	e	v			v			v
	f	v				v		v
2	a		v	v			v	
	b		v		v		v	
	c		v			v	v	
	d		v	v				v
	e		v		v			v
	f		v			v		v

To evaluate each scenario, we measure F1 Score in training dataset and testing dataset for every experiment. Training process is done in 35 epochs with learning rate of 0.0001. The result for every experiment in every scenario can be seen in Fig. 6, Fig. 7, and Fig. 8.

Based on the experiment results in this study, we can see that the value of parameters (dropout and ResNet layer) used in each scenario do not affect to F1 Score quite much. It can be seen from the F1 Score deviation between every experiment which is not reaching 3%.

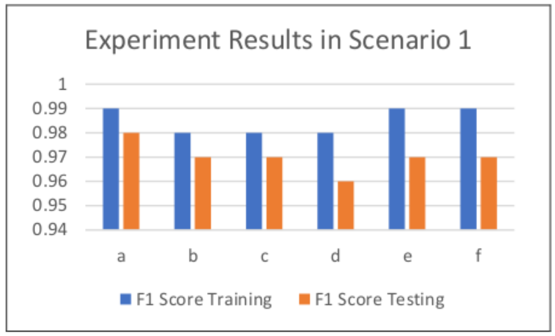

Fig. 6. Experiment results in Scenario 1

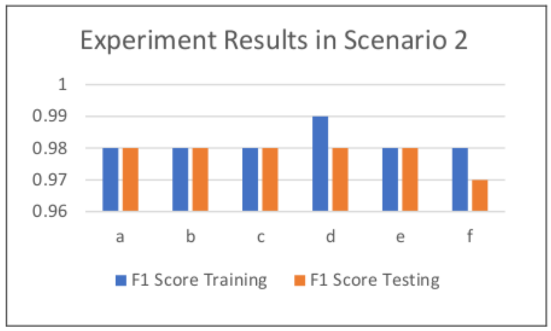

Fig. 7. Experiment results in Scenario 2

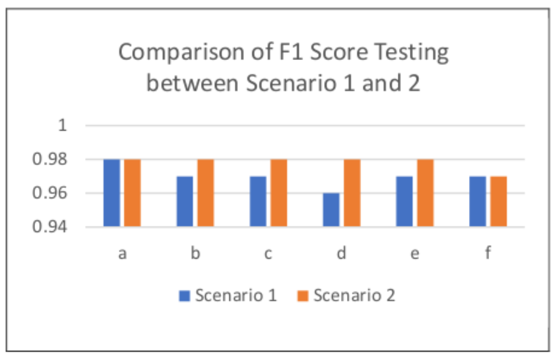

Fig. 8. F1 Score testing comparison between Scenario 1 and Scenario 2

The dropout usage does not affect quite much because there is only 1 convolutional layer that uses dropout and there already is batch normalization in every convolutional layer. The number of ResNet layer does not affect quite much because there is bottleneck design in ResNet architecture which can maintain the performance of ResNet although there are less parameters. Meanwhile the usage of thresholding algorithms does not affect quite much because the output image from thresholding algorithms still has a lot of noises.

Fig 8 shows the F1 Score on testing between scenario 1 and scenario 2. Based on Fig 8, scenario 2 has higher F1 Score compared to scenario 1 with F1 Score in scenario 2 is up to 0.98 in experiment 2b, 2c, 2d, and 2e. The smallest deviation of F1 Score between Scenario 1 and Scenario 2 is 0.01 which is in experiment 1e and 2e. This indicates that using thresholding algorithms does not affect quite much because the output image from thresholding algorithms still has a lot of noises. It can be seen in Fig 9-12. Fig 9 shows the original images of 3 samples. Fig 10 shows the preprocessed image of the respective sample in Fig 9. Fig 11 shows the ground truth. Fig 12 shows some segmentations result by using PSPNet.

However, the results of this comparison are competitive where the difference of F1 score between scenario 1 and scenario 2 is less than 0.02.

IV. Conclusion

Based on the experiment in this study we can conclude that the value of parameters (dropout and number of ResNet layer) and the use of image thresholding algorithms are not quite much affects the resulting image segmentation model in terms of F1 Score. That can be seen in the F1 Score deviation between every experiments in each scenario is not reaching 3%. The original RGB garbage image is better to be used as input image for PSPNet than combined binary images from some thresholding algorithms. However, the results of this comparison are competitive where the difference of F1 score between scenario 1 and scenario 2 is less than 0.02.

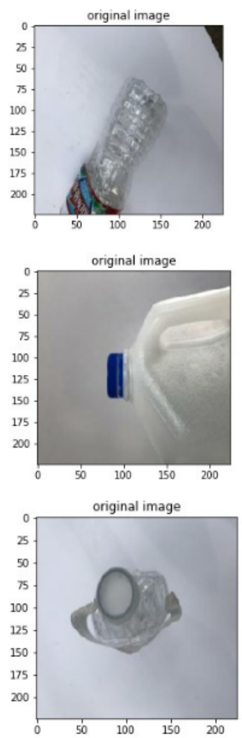

Fig. 9. Samples of original image

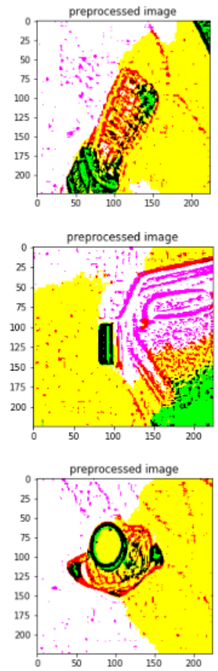

Fig. 10. Samples of preprocessed image

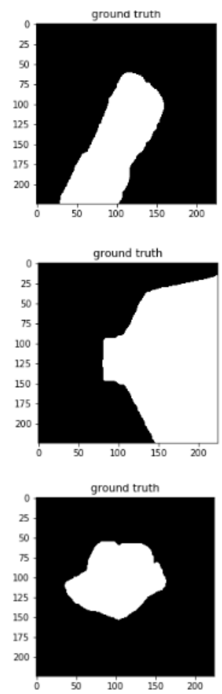

Fig. 11. Samples of ground truth

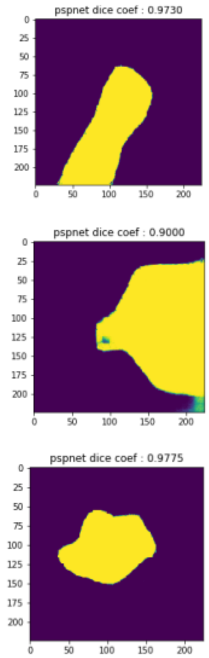

Fig. 12. Samples of segmentation result with its dice coefficient using PSPNet

ACKNOWLEDEMENT

The authors would like to acknowledge the research funding supported by Faculty of Science and Mathematics, Universitas Diponegoro under the Grant of Primary Research – Contract Number 2007/UN7.5.8/PP/2020.

REFERENCES

[1] Kementrian Lingkungan Hidup dan Kehutanan (2017). Retrieved August 1, 2020, from https://databoks.katadata.co.id/datapublish/2019/11/01/komposisi-sampah-di-indonesia-didominasi-sampah-organik

[2] Ronneberger, O., Fischer, P., & Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention* (pp. 234-241). Springer, Cham.

[3] Chaurasia, A., & Culurciello, E. (2017, December). Linknet: Exploiting encoder representations for efficient semantic segmentation. In *2017 IEEE Visual Communications and Image Processing (VCIP)* (pp. 1-4). IEEE.

[4] Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., ... & Adam, H. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861.*

[5] Zhao, H., Shi, J., Qi, X., Wang, X., & Jia, J. (2017). Pyramid scene parsing network. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2881-2890).

[6] Zeng, D. Zhu, M, Kuijiper, A. (2019). Combining Background Subtraction Algorithms with Convolutional Neural Network, J. of Electronic Imaging, 28(1), 013011 (2019). https://doi.org/10.1117/1.JEI.28.1.013011

[7] Vala, H. J., & Baxi, A. (2013). A review on Otsu image segmentation algorithm. *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, *2*(2), 387-389.

[8] Singh, T. R., Roy, S., Singh, O. I., Sinam, T., & Singh, K. (2012). A new local adaptive thresholding technique in binarization. arXiv preprint arXiv:1201.5227.

[9] Lee, S. U., Chung, S. Y., & Park, R. H. (1990). A comparative performance study of several global thresholding techniques for segmentation. *Computer Vision, Graphics, and Image Processing*, *52*(2), 171-190.

[10] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *CVPR*, 2016. 2, 3, 4, 5, 6

[11] L. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *arXiv:1606.00915*, 2016. 5, 7, 8

[12] Zou, K. H., Warfield, S. K., Bharatha, A., Tempany, C. M., Kaus, M. R., Haker, S. J., Wells, W. M., 3rd, Jolesz, F. A., & Kikinis, R. (2004). Statistical validation of image segmentation quality based on a spatial overlap index. *Academic radiology*, *11*(2), 178–189. https://doi.org/10.1016/s1076-6332(03)00671-8

Garbage

ORIGINALITY REPORT

8%	%	8%	%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

MATCH ALL SOURCES (ONLY SELECTED SOURCE PRINTED)

6%

★ Urratul Aqyuni, Sukmawati Nur Endah, Priyo Sidik Sasongko, Retno Kusumaningrum, Khadijah, Rismiyati, Hanif Rasyidi. "Waste Image Segmentation Using Convolutional Neural Network Encoder-Decoder with SegNet Architecture", 2020 4th International Conference on Informatics and Computational Sciences (ICICoS), 2020
Publication

Exclude quotes	On	Exclude matches	< 2%
Exclude bibliography	On		