BRAHMANA: Jurnal Penerapan Kecerdasan Buatan
Terakreditasi Nomor 204/E/KPT/2022 | Vol. 5, No. 1, Desember (2023), pp. 85-95

The Comparison of U-net and Deeplab V3 as Semantic Segmentation Models for Food Images

Natanael Richie Irwanto[1], Gede Putra Kusuma[2]
[1]Informatic Engineering Department, Graduate Program, Bina Nusantara University, Jakarta, Indonesia
[2]Computer Science Department, Graduate Program, Bina Nusantara University, Jakarta, Indonesia
E-mail: [1]natanael.irwanto@binus.ac.id, [2]i.negara@binus.ac.id

Abstract

The Semantic segmentation models have been used for many things, namely image classification, image detection, and other activities, including outdoor and object segmentation. Those models can either work with having a good result with a custom dataset and give up an excellent accurate process or a bad one. This research aimed to compare two models of semantic segmentation model, namely Unet and Deepvlab, for food images. The research procedure is to create an original food image dataset, process the dataset with two models, analyze the IoU of two models, and compare the mIoU between the models. The research results show that U-net has a higher mIoU value of 0.01 than Deeplab V3 but has less processing time and some parameters. The research results also show that the completeness of performance details and the prediction segmentation results in the Deeplab v3 segmentation model are superior to this research. This research supports previous research findingsregarding the use of U-net and Deeplab v3 in semantic segmentation models. It enriches research on using these models in food image recognition. Further research is needed to evaluate other models in semantic segmentation for food images.

Keywords: *Semantic segmentation, food images, U-net Model, Deeplab V3*

1. Introduction

Pandemic covid-19 become the main problem for many countries. Older people affected by COVID-19 have a death rate 23 times greater than young people [1]. It primarily affects older people who have congenital diseases such as diabetes, high cholesterol, heart disease, stroke, cancer, and many other diseases. The disease is generally caused by a lifestyle and unhealthy food consumption [2]. Technology can play a role in creating healthier lifestyles and eating patterns in humans to avoid various diseases that exacerbate the impact of COVID-19 [3]. Currently, interest in using technology to control food intake and health behavior is increasing [4]. One of the technological models that can be applied to help process food control through image recognition is the semantic segmentation model. Semantic segmentation is the computer vision method to perform a specific task that includes object labeling in a bounding box and makes those labels perform as input data [5]. The input data will be defined as segmentation images with label data. Semantic image segmentation is popular for describing, categorizing, and visualizing the object as an image [6].

Computer vision and image processing include semantic segmentation. Those images will be processed either by photos or pictures from the internet. It will processed from each edge using semantic segmentation methods to give the best image segmentation performance that defines the category from the images. Semantic segmentation models are also widely used to recognize food images. The research by [7] has researched the method of identifying the type of food and its volume to improve public health and awareness regarding the food they consume. Kong et al. have developed a system to calculate estimated food calories using a semantic segmentation model for diabetes

ISSN: 2715-9906

patients [8]. Semantic segmentation models can be used to identify and classify input food data. Healthy food comprises proportionate carbohydrates, protein, and fat as macronutrients that can affect health [9]. Segmentation semantics for food image recognition is needed to explore the content of the food so people can get information about the portion and composition of the food.

Several semantic segmentation models are widely used to recognize images [10]. One of the popular models in the semantic segmentation area is U-net, an FCN-based model with many layers of information. U-net has a structure resembling the letter "U" consisting of symmetrical downsampling and upsampling processes [11]. U-net is made up of two pathway: a contract and an expansion. The contract pathway is designed in the manner of a standard convolutional network. Another model that is also widely used in the semantic segmentation model is deeplab v3. Deeplab v3 is the third version of the deepvlab series proposed by Liang-Chieh Chen and the Google team [12]. For dense feature extraction, this model employs atrous convolution with upsampled filters. [13]. Both deeplab and U-net have been used to analyze The semantic segmentation model includes food imagery. A study by [14] used the model to detect almonds and green onions in the flake food process. Meanwhile, deeplab v3 has been used as a semantic segmentation model for Apple image recognition [15].

Unet and Deeplab are semantic segmentation models that can analyze food images based on custom datasets. As a good model, they must have performance and Intersection of Union (IoU) insufficient and exceed the average to show segmentation and detection of food. They can also be applied to AI (Artificial intelligence) and other machinesThe Intersection of Union (IoU) measure is used to assess the efficacy of a model in recognizing and categorizing images or objects. IoU is often referred to as the Jaccard Index. This metric is the most popular one to compare the similarity between two arbitrary shapes [16]. Research by [17] have measured the IoU score of the Unet model which shows a score of 0.66 to 0.71. This shows that the performance and evaluation of segmentation are excellent and accurate according to food segmentation performance data. Meanwhile, Deepvlab v3 can provide an excellent IoU of 0.83 [13]. The two models have good and precise mIoU scores, but no research has compared the MIoU scores in the two models for food image semantic segmentation. This study compares the MIoU scores of the Deeplab v3 and U-net models to determine which model can best recognize food images.

2. Related Works

2.1. Food images semantic segmentation

Grouping image components task that correspond to the same item class is known as semantic image segmentation or pixel-level categorization[18]. There are multiple applications of semantic image segmentation, such as detecting road signs, colon crypts segmentation, land use, and land cover classification, detecting brains and tumors, detecting and tracking medical instruments in operations, and in self-driving car areas [19]. Compared to semantic segmentation in general, food segmentation is a relatively new area and still requires a lot of exploration. Food image segmentation is generally used to measure calories and nutrition, to assess food intake for diets, to measure food portions, to evaluate food quality, to recognize food items, and for many other purposes [20]. Food image segmentation is rarely explored due to a scarcity of high-quality food picture datasets with fine-grained ingredient labeling and complex food displays causes difficulty in recognizing ingredients in food images [21].

2.2. DeepLab v3

DeepLabv3 is an architecture for semantic segmentation that builds on previous version by making various changes. Modules are designed to catch multiple scales context by employing a number of enticing rates in cascade or parallel to tackle the issue of segmenting objects at various scales. Furthermore, DeepLabv2's Module for Atrous Spatial Pyramid Pooling has been enhanced with characteristics at the picture level that encode worldwide context and improve efficiency. The final feature map made of one 11 convolution and three 3 3 convolutions with 256 filters, normalization of batches, and characteristics at the picture level [13].

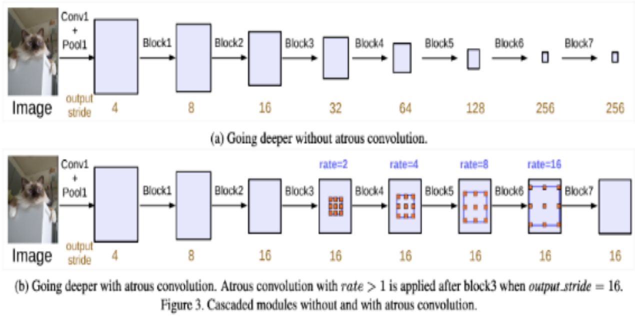


Figure 1. Deeplab V3 architecture

Deeplab v3 has a mIoU performance of 0.85, the same as the PSPNet and PSANet models, but more significant than the ResNet-38 model, which is 0.84 [22]. This research was conducted using the PASCAL VOC 2012 dataset. The value in this study shows a relatively accurate level of performance in each model because it is adjusted to the existing dataset. A study by [23] in the ResNet 101 segmentation model and also the deep lab v2 model showed that the mIoU results were 0.75 for the ResNet 101 model and 0.79 for the deep lab v2 model using the same dataset, namely PASCAL VOC. The performance value between 0.75 and 0.79 shows a relatively good level of performance using older versions of models such as ResNet 101 compared to ResNet 38 and deeplab v2 compared to the deeplab v3 model. In research by [24], the same segmentation model used the DeepLab CRF model, with the same dataset, namely PASCAL VOC 2012, found a mIoU performance value of 0.66 and proved the same performance value as previous research.

2.3. U-net

U-net is a model that frequently used in image segmentation [25]. U-net is made up of two paths: contract and expansive. The contract path is built on a standard convolutional network design. For downsampling, iteratively perform two unpadded convolutions with 2x2 pooling operation and a rectified linear unit (ReLU). The feature map is unsampled on the expansive path by a 2x2 up-convolution and two convolutions of 3x3 that combined with a ReLU. The last layer used 1x1 convolution, so U-net has 23 convolutional layers in total [26].

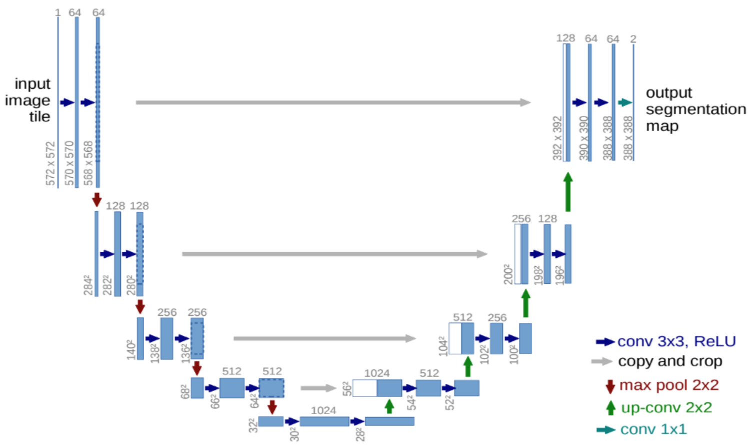


Figure 2. The Architecture of U-net

Semantic segmentation research using deep learning methods [17] explained that the semantic segmentation process uses deep learning methods to optimize remote sensing image performance and can improve U-net performance by applying CNN (Convolutional Neural Network). This research enhances performance for object extraction by leveraging the encoder and decoder structures in CNN. The model of that research was developed into three parts, all of which are developments of the model. The research obtained that before U-net was developed, mIoU (Mean Intersection of Union) was carried out, which was obtained in the method used in the form of a value of 0.66, which was then expanded with CNN to 0.69 to 0.71 in mIoU. Meanwhile, research [27] investigate the used of the model in transportation facility construction. According to the findings of [27] the models can semantically segment multiple target attributes, and the accuracy get more than 80%.

3. Research Methods

This research aims to compare two semantic segmentation models in food images. The research stages are as follows:

3.1. Preparation

In the preparation stage, the researcher collected various data by studying the literature and collecting food image datasets. Researchers compiled the dataset personally labeled with three categories, namely carbohydrates, meat, and vegetables. The dataset consists of 130 food images stored on Google Drive.

3.2. Implementation

The implementation stage in this research was using Google Research Colab to process segmentation modeling. The researcher uploads the Python base program code in the Google research collab. After that, each code will be executed to process the semantic segmentation of the food dataset images. The following is the process of operating the Deepvlab v3 and U-net programs.

3.2.1. U-net model operation

The model operation begins with connecting the U-net program with the data set on Google Drive. Figure 3 shows the process of connecting U-net with a data set on Google Drive. Then, the dataset will be loaded into the segmentation model with program code and val labels with an image pixel size of 256 pixels from all the dataset images that have been prepared.

```
[ ]  from google.colab import drive
     drive.mount('/content/drive')

     Mounted at /content/drive

[ ]  drive_dir = '/content/drive/MyDrive/Semantic'
```

Figure 3. U-net connecting with Google Drive

```
[ ]  framObjTrain = LoadData(framObjTrain,
                             imgPath = drive_dir+os.sep+'Datasets'+os.sep+'Done'+os.sep+'train',
                             maskPath = drive_dir+os.sep+'Datasets'+os.sep+'Done'+os.sep+'train_label',
                             shape = 256)

[ ]  framObjValidation  = LoadData(framObjValidation,
                                   imgPath = drive_dir+os.sep+'Datasets'+os.sep+'Done'+os.sep+'val',
                                   maskPath = drive_dir+os.sep+'Datasets'+os.sep+'Done'+os.sep+'val_label',
                                   shape = 256)
```

Figure 4. Load dataset for a segmentation model

The next step is to run model segmentation training by providing the number of training epochs needed to produce the level of performance at the time the training is executed so that the model can create a semantic segmentation process. Figure 5 shows that 70 epoch training has been run on this U-net model.

```
[ ]  ## trainign our model
     retVal = myTransformer.fit(np.array(framObjTrain['img']), np.array(framObjTrain['mask']),
                                epochs = 70, verbose = 1)
     3/3 [==============================] - 1s 273ms/step - loss: 163.0345 - accuracy: 0.3352
     Epoch 22/70
     3/3 [==============================] - 1s 274ms/step - loss: 173.0443 - accuracy: 0.3391
     Epoch 23/70
```

Figure 5. Epoch training model U-net

3.2.2. Deepvlab v3 model operation

The deepvlab v3 segmentation model also operates with the same number of datasets as the U-net segmentation model. The initial step in the process of running the deeplab v3 model is also carried out in the same steps as the previous model experiment. After training is complete, the process will continue measuring how the two models work on image detection. The implementation step of the segmentation model ends after training for both segmentation models. The hyperparameters that are optimized in the segmentation process are as follows:

Table 1. Hyperparameter

Parameter	Value
Epoch	70
Training Batch	50
Batch Size	8
Learning Rate	0.001
Loss	Binary cross-entropy

3.3. Evaluation

The final stage of this research is to evaluate the results of operating the model to determine the quality and adequacy of the process of the two segmentation models that have been carried out. Evaluation is carried out on the intersection over onion (IoU), time, and number of epochs. The IoU measurement formula is as follows:

$$Intersection\ over\ Union\left(IoU\right) = \frac{|\mathrm{A} \cap \mathrm{B}|}{|\mathrm{A}|\cup|\mathrm{B}|} \quad (1)$$

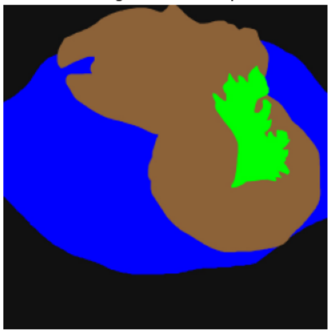
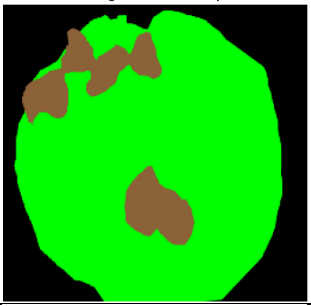
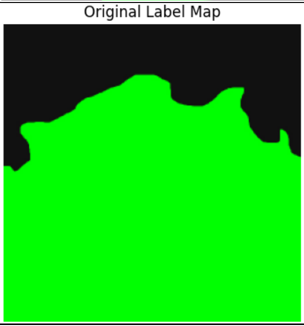
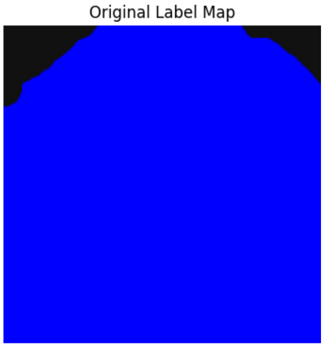
The IoU performance level shows the overlap or intersection area between the bounding boxes, as well as the overall area of the two bounding boxes. Area A in the IoU measurement formula is the true facts bounding box, while the B area is the bounding box prediction. The IoU formula shows that area A, which intersects area B, will be compared with the total area A and area B so that the results of measuring IoU performance from the segmentation model will be obtained. The average IoU (mIoU) performance measurement can be seen if the IoU performance value appears on both objects. The IoU value of the two objects will be given an average value by adding up all the IoU values, which will then be divided by the number of existing objects.

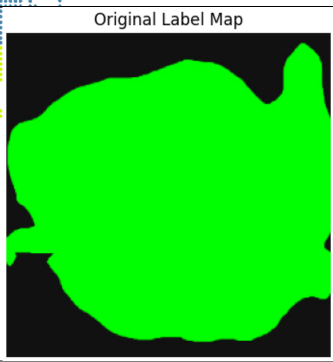
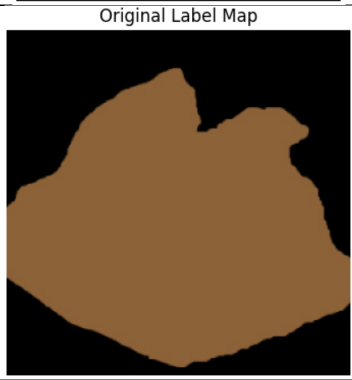
4. Results and Discussion

4.1. Results

The IoU value, processing time, and total of epochs indicate the performance results of two segmentation models. The IoU value is obtained after the food labels have been identified from the food images in the data set. The category labels in the image consist of 3 sections that show the segmentation of carbohydrates, proteins, and vegetables. The IoU performance results in the segmentation model are as follows:

Table 2. Evaluation of segmentation model performance

No	Image Input	Image Output	IoU (0<0.5<1)
1		Original Label Map 	0.81
2	Original Image 	Original Label Map 	1.0
3	Original Image 	Original Label Map 	0.52
4		Original Label Map 	0.37
5	Original Image 	Original Label Map 	0.61

6			1.0
7			0.61

The results of the IoU performance figures are then processed to determine the average figure (mIoU). The mIoU value will show the final performance level of a segmentation model. In the U-Net segmentation model, 16 IoU values are processed from 130 dataset images. Meanwhile, in the deeplab v3 model, each segmentation process that is carried out produces an IoU performance value following the number of training epochs carried out so that the IoU obtained is 70 IoU. Each IoU will be processed to take an average value, showing the segmentation model's final performance results.

Table 3. Results of mIoU evaluation in both segmentation models

No	Model	mIoU	Epoch	Total IoU
1	U-Net Model	0.64	70	7
2	Deeplab v3 Model	0.63	70	70

It is clear from Table 3 that the mIoU score in the two models is almost the same. The U-Net model has a higher mIoU value of 0.01 points than the deeplab v3 model. However, when viewed from the implementation of the workings of the model, there are significant differences between U-Net and Deeplab v3. The deeplab model can calculate 70 IoU of IoU from 70 segmented epoch images. Meanwhile, in U-Net model, It is critical to assess the IoU results in segmented images, and the greater the desired IoU number, the longer it will take. This causes the U-net model to display only 7 IoU values in 7 food images out of 130 images in the dataset. In addition, the more IoU values can also affect the calculation results at a smaller mIoU score.

4.2. Discussion

The research results show that U-net has a greater mIoU score than Deeplab V3. This follows the research results [28] that the IoU score on the U-net is 0.667, while deeplab V3 has an IoU score of 651. U-net also provided the best recall in the segmentation of unmanned aerial vehicle pictures. In addition, research results by [29] also show that U-net has the highest mIoU score among other models, namely Deeplab, ICNet, PSPNet, and Fast-SCNN. The results of this study are also under the results of the study by [30] that the mIoU score on the U-net model is always higher than Deeplab V3 in several baseline models tested. The high IoU score on Unet is due to the network upsampling stage. DeepLabV3 has a poor IoU score since the final picture of segmentation is created by a single-step upsampling method with the outputs of the feature extractor magnified, which may result in some information being lost. U-Net is upsampled in four steps, resulting in a higher IoU On the validation set, the IoU was higher, but on the test set, it was lower [31].

However, when viewed from the number of parameters processed, deeplab has an advantage compared to Unet. In this research, Deeplab V3 was able to produce 70 parameters, while U-net was only able to create seven parameters from the 130 available datasets. The results of this study support the research results [32], which obtained the result that the U-net can only measure a number of 1/5 parameters that can be measured by Deeplab v3. The U-net has the smallest measurements compared to the Deeplab and Swin transformer models examined in the study. The small number of parameters measured causes the U-net to have a shorter time to process data. U-net does not take long to process data because only a few parameters are displayed. Meanwhile, deeplab V3 requires a longer time because the number of parameters processed is almost 10x that of U-net.

The precision of the two models also suggests the performance of the semantic segmentation model. Research result by [32] shows that Deeplab v3 has higher accuracy in rice image object recognition than U-net. U-net tends to be wrong in identifying rice and non-rice objects and has a low concordance between predicted and actual values. The results of other studies also show that the traditional U-Net network can only identify parts with significant color differences, so it is less accurate in identifying almost similar colors [30]. Meanwhile, Deeplab v3 shows more complete results on image recognition. This follows the research results [33], which showed that the DeepLab V3 method had relatively complete object segmentation results but was not detailed enough to process the image's details. Researching food semantic segmentation is challenging. This is because tThe same meal might have various characteristics and appearances depending on the image background, presentation, and proportions of the food [34]. U-net and Deeplab V3 models in food images can still be explored and improved to achieve maximum accuracy and mIoU results.

4. Conclusion

This study aims to compare two popular models in semantic segmentation, namely Deeplab V3 and U-net. The results show that U-net has a higher mIoU value than Deeplab V3 but has less processing time and the number of parameters. In this study, the U-net model only displays 1/10 prediction results from the Deeplab V3 model. The lower mIoU value on Deeplab V3 can be affected by the predicted amount of data that is more from the U-net. Both segmentation models are implemented using the Python-based programming language. The research results also reveal that the Deeplab v3 segmentation model's completeness of performance details and prediction segmentation results outperform this research. This study backs up the findings of prior studies on the usage of two semantic segmentation models. It enriches research on using these models in food image recognition. However, this research still needs to be improved and improved, one of which is using the model segmentation for real-time segmentation process for segmentation data records. The added feature that includes a food database and information on the food data, calories, and nutrition provided at the time also needs to be improved. Future research can also compare other models in the semantic segmentation model for food images.

References

[1] Amber L. Mueller, Maeve S.McNamara, and David A. Sinclair, "Why does COVID-19 disproportionately affect older people?" *Aging (Albany. NY).*, vol. 12, no. 10, pp. 9959–9981, 2020.

[2] I. A. Fajarini and R. A. D. Sartika, "Obesity as a common type-2 diabetes comorbidity: Eating behaviors and other determinants in Jakarta, Indonesia," *Kesmas*, vol. 13, no. 4, pp. 157–163, 2019, doi: 10.21109/kesmas.v13i4.2483.

[3] S. Dixit and G. Nandakumar, "Promoting healthy lifestyles using information technology during the COVID-19 pandemic," *Rev. Cardiovasc. Med.*, vol. 22, no. 1, pp. 115–125, 2021, doi: 10.31083/J.RCM.2021.01.187.

[4] F. Zhu *et al.*, "The use of mobile devices in aiding dietary assessment and evaluation," *IEEE J. Sel. Top. Signal Process.*, vol. 4, no. 4, pp. 756–766, 2010, doi: 10.1109/JSTSP.2010.2051471.

[5] Y. Zhang, S. Mehta, and A. Caspi, "Rethinking Semantic Segmentation Evaluation for Explainability and Model Selection," 2021, [Online]. Available: http://arxiv.org/abs/2101.08418.

[6] M. K. Kar, M. K. Nath, and D. R. Neog, *A Review on Progress in Semantic Image Segmentation and Its Application to Medical Images*, vol. 2, no. 5. Springer Singapore, 2021.

[7] F. S. Konstantakopoulos, E. I. Georga, and D. I. Fotiadis, "A Review of Image-based Food Recognition and Volume Estimation Artificial Intelligence Systems," *IEEE Rev. Biomed. Eng.*, pp. 1–17, 2023, doi: 10.1109/RBME.2023.3283149.

[8] X.-Y. Kong *et al.*, "Food Calorie Estimation System Based on Semantic Segmentation Network," *Sensors Mater.*, vol. 35, no. 6, p. 2013, 2023, doi: 10.18494/sam4061.

[9] Y. J. Kwon, H. S. Lee, J.-Y. Park, and J. W. Lee, "Associating intake proportion of carbohydrate, fat, and protein with all-cause mortality in Korean adults.," *Nutrients*, vol. 12, pp. 1–12, 2020.

[10] S. Wang, Z. Zhou, and W. Zhao, "Semantic Segmentation and Defect Detection of Aerial Insulators of Transmission Lines," *J. Phys. Conf. Ser.*, vol. 2185, no. 1, 2022, doi: 10.1088/1742-6596/2185/1/012086.

[11] X. Xia, Q. Lu, and X. Gu, "Exploring An Easy Way for Imbalanced Data Sets in Semantic Image Segmentation," *J. Phys. Conf. Ser.*, vol. 1213, no. 2, 2019, doi: 10.1088/1742-6596/1213/2/022003.

[12] H. Zeng, S. Peng, and D. Li, "Deeplabv3+ semantic segmentation model based on feature cross attention mechanism," *J. Phys. Conf. Ser.*, vol. 1678, no. 1, 2020, doi: 10.1088/1742-6596/1678/1/012106.

[13] L. C. 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," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, 2018, doi: 10.1109/TPAMI.2017.2699184.

[14] G. J. Son, D. H. Kwak, M. K. Park, Y. D. Kim, and H. C. Jung, "U-Net-based foreign object detection method using effective image acquisition system: A case of almond and green onion flake food process," *Sustain.*, vol. 13, no. 24, 2021, doi: 10.3390/su132413834.

[15] L. Mo, Y. Fan, G. Wang, X. Yi, X. Wu, and P. Wu, "DeepMDSCBA: An Improved Semantic Segmentation Model Based on DeepLabV3+ for Apple Images," *Foods*, vol. 11, no. 24, pp. 1–19, 2022, doi: 10.3390/foods11243999.

[16] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese, "Generalized intersection over union: A metric and a loss for bounding box regression," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2019-June, pp. 658–666, 2019, doi 10.1109/CVPR.2019.00075.

[17] X. Wang, Z. Hu, S. Shi, M. Hou, L. Xu, and X. Zhang, “A deep learning method for optimizing semantic segmentation accuracy of remote sensing images based on improved UNet,” *Sci. Rep.*, vol. 13, no. 1, pp. 1–13, 2023, doi: 10.1038/s41598-023-34379-2.
[18] M. Thoma, “A Survey of Semantic Segmentation,” *ArXiv*, pp. 1–16, 2016, [Online]. Available: http://arxiv.org/abs/1602.06541.
[19] X. Liu, Z. Deng, and Y. Yang, “Recent progress in semantic image segmentation,” *Artif. Intell. Rev.*, vol. 52, no. 2, pp. 1089–1106, 2019, doi: 10.1007/s10462-018-9641-3.
[20] V. Burkapalli and P. Patil, “A Review on Segmentation Techniques for Food Images,” *Int. J. Emerg. Technol. Comput. Sci. Electron.*, vol. 23, no. 6, pp. 250–255, 2016.
[21] X. Wu, X. Fu, Y. Liu, E. P. Lim, S. C. H. Hoi, and Q. Sun, *A Large-Scale Benchmark for Food Image Segmentation*, vol. 1, no. 1. Association for Computing Machinery, 2021.
[22] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, “Pyramid scene parsing network,” *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017*, vol. 2017-Janua, pp. 6230–6239, 2017, doi: 10.1109/CVPR.2017.660.
[23] Z. Wu, C. Shen, and A. van den Hengel, “Wider or Deeper: Revisiting the ResNet Model for Visual Recognition,” *Pattern Recognit.*, vol. 90, no. December 2016, pp. 119–133, 2019, doi 10.1016/j.patcog.2019.01.006.
[24] H. Sharif, F. Rehman, A. Rida, and A. Sharif, “Segmentation of Images Using Deep Learning: A Survey,” *2022 2nd Int. Conf. Digit. Futur. Transform. Technol. ICoDT2 2022*, 2022, doi: 10.1109/ICoDT255437.2022.9787440.
[25] W. Weng and X. Zhu, “INet: Convolutional Networks for Biomedical Image Segmentation,” *IEEE Access*, vol. 9, pp. 16591–16603, 2021, doi: 10.1109/ACCESS.2021.3053408.
[26] N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, “Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015: 18th International Conference Munich, Germany, October 5-9, 2015 proceedings, part III,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9351, no. Cvd, pp. 12–20, 2015, doi: 10.1007/978-3-319-24574-4.
[27] X. Hao, L. Yin, X. Li, L. Zhang, and R. Yang, “A Multi-Objective Semantic Segmentation Algorithm Based on Improved U-Net Networks,” *Remote Sens.*, vol. 15, no. 7, pp. 1–14, 2023, doi: 10.3390/rs15071838.
[28] L. Xia *et al.*, “Evaluation of deep learning segmentation models for detection of pine wilt disease in unmanned aerial vehicle images,” *Remote Sens.*, vol. 13, no. 18, pp. 1–15, 2021, doi: 10.3390/rs13183594.
[29] S. Zhao, G. Hao, Y. Zhang, and S. Wang, “A Real-Time Semantic Segmentation Method of Sheep Carcass Images Based on ICNet,” *J. Robot.*, vol. 2021, 2021, doi: 10.1155/2021/8847984.
[30] S. Gao and Z. Y. Chen, “Semantic Segmentation of Germinated Oil Palm Seeds Based on Deep Convolutional Neural Networks with a Novel Channel Attention Mechanism,” pp. 1–26.
[31] T. Lee, J. H. Kim, S. J. Lee, S. K. Ryu, and B. C. Joo, “Improvement of Concrete Crack Segmentation Performance Using Stacking Ensemble Learning,” *Appl. Sci.*, vol. 13, no. 4, 2023, doi: 10.3390/app13042367.
[32] H. Xu, J. Song, and Y. Zhu, “Evaluation and Comparison of Semantic Segmentation Networks for Rice Identification Based on Sentinel-2 Imagery,” *Remote Sens.*, vol. 15, no. 6, 2023, doi: 10.3390/rs15061499.

[33] D. Yang, Y. Du, H. Yao, and L. Bao, "Image semantic segmentation with hierarchical feature fusion based on deep neural network," *Conn. Sci.*, vol. 34, no. 1, pp. 1772–1784, 2022, doi: 10.1080/09540091.2022.2082384.

[34] S. Aslan, G. Ciocca, D. Mazzini, and R. Schettini, "Benchmarking algorithms for food localization and semantic segmentation," *Int. J. Mach. Learn. Cybern.*, vol. 11, no. 12, pp. 2827–2847, 2020, doi: 10.1007/s13042-020-01153-z.