



RESEARCH ARTICLE

# Image Segmentation Performance using DeepLabv3+ with Resnet-50 on Autism Facial Classification

Melinda Melinda<sup>1,\*</sup>, Hurriyatul Aqif<sup>2</sup>, Junidar<sup>3</sup>, Maulisa Oktiana<sup>4</sup>, Nurlida Basir<sup>5</sup>, Afdhal Afdhal<sup>6</sup>, and Zulfan Zainal<sup>7</sup>

<sup>1,2,4,6</sup>Dept. of Electrical & Computer Eng., Universitas Syiah Kuala, Banda Aceh 23111, Indonesia

<sup>3</sup>Dept. of Informatics, Universitas Syiah Kuala, Banda Aceh 23111, Indonesia

<sup>5</sup>Fakulti Sains dan Teknologi, Universiti Sains Islam Malaysia, Negeri Sembilan 71800, Malaysia

<sup>7</sup>Computer Eng. Dept., Universitas Serambi Mekkah, Banda Aceh 23245, Indonesia

\*Corresponding email: melinda@usk.ac.id

Received: February 14, 2024; Revised: March 18, 2024; Accepted: May 31, 2024.

**Abstract:** In recent years, significant advancements in facial recognition technology have been marked by the prominent use of convolutional neural networks, particularly in identification applications. This study introduces a novel approach to face recognition by employing ResNet-50 in conjunction with the DeepLabV3 segmentation method. The primary focus of this research lies in the thorough analysis of ResNet-50's performance both without and with the integration of DeepLabV3+ segmentation, specifically in the context of datasets comprising faces of children on the autism spectrum. The utilization of DeepLabV3+ serves a dual purpose: firstly, to mitigate noise within the datasets, and secondly, to eliminate unnecessary features, ultimately enhancing overall accuracy. Initial results obtained from datasets without segmentation demonstrate a commendable accuracy of 83.7%. However, the integration of DeepLabV3+ yields a substantial improvement, with accuracy soaring to 85.9%. The success of DeepLabV3+ in effectively segmenting and reducing noise within the dataset underscores its pivotal role in refining facial recognition accuracy. In essence, this study underscores the pivotal role of DeepLabV3+ in the realm of facial recognition, showcasing its efficacy in reducing noise and eliminating extraneous features from datasets. The tangible outcome of increased accuracy of 85.9% post-segmentation lends credence to the assertion that DeepLabV3+ significantly contributes to refining the precision of facial recognition systems, particularly when dealing with datasets featuring faces of children on the autism spectrum.

**Keywords:** autism spectrum disorder, convolutional neural networks, deeplabv3+, image segmentation, resnet-50.

## 1 Introduction

In recent years, facial recognition technology has advanced significantly, and identification applications convolutional neural networks (CNN) have demonstrated outstanding performance in facial recognition applications [1]. Several studies on facial recognition have been undertaken [2–6]. This study proposes a facial recognition system for autistic children using the Deep Learning and segmentation method to recognize the characteristics of autism through facial images. The faces of autistic and non-autistic children are nearly identical in appearance [4]. However, children diagnosed with autism have an atypically broad upper face, which includes wide-set eyes. They also have a shorter nose and cheekbones in the middle of their faces [7]. Looking at the CDC data year by year over 80 years reveals HC has increased by 11 mm for boys and 7 mm for girls [8]. In the study [7] used MobileNet and two dense layers to perform feature extraction and image classification. In the study [9], there are three different pre-trained deep learning models, namely, Xception, VGG19, and NASNetMobile, that were implemented to detect ASD. The study proposes a practical ASD screening solution using facial images by applying VGG16 transfer learning-based deep learning [10]. However, these studies did not implement any segmentation model. In this study, the autism and non-autism images will be analyzed. This study also explored a CNN architecture using ResNet-50 and a segmentation method called DeepLabV3+. These methods are intended to produce models with good accuracy in the recognition and classification processes and the performance of the architecture. ResNet-50 is one of the popular deep-CNN models to extract features from facial images and it was designed for recognition since it outperforms simple CNN [11, 12]. DeepLabv3+ is one of the best semantic segmentation algorithms currently and has high accuracy in the segmentation [13].

Based on the available research, a benchmark comparison of autistic face image classification using ResNet-50 and ResNet-50 with DeepLabv3+ segmentation can be summarized as follows:

- The accuracy of classifying autistic and non-autistic children using ResNet-50 for facial image classification ranges from 43% to 96% [14–16].
- Deep learning and transfer learning have been applied to the detection of autism through facial images, with various studies reporting different accuracy levels using different models and techniques [14–16].
- The use of DeepLabv3+ segmentation with ResNet-50 for autistic face image classification is not specifically mentioned in the provided search results.

This study analyzes the performance of ResNet-50 and DeepLabV3+ in the recognition of autism and non-autism. The results will be analyzed whether the Deeplabv3+ segmentation method can increase accuracy or not. It is hoped that these methods can be a reference to classify autism and non-autism. The decision to opt for an algorithm other than DeepLabv3+ for segmenting the autistic spectrum face dataset stems from the necessity to identify the most appropriate method, considering various factors such as algorithm compatibility, performance evaluation criteria, feature extraction requirements, and potential for originality. Despite the effectiveness of DeepLabv3+ in semantic segmentation tasks, its specific structure and features might not perfectly match the dataset's characteristics,

prompting the exploration of alternative algorithms that could potentially offer superior performance or novel perspectives. By venturing into uncharted territory with algorithms previously unused in this context, there's an opportunity to discover innovative methodologies and potentially enhance our understanding of autistic spectrum facial attributes, thereby driving progress in the field.

## 2 Research Method

This study will analyze the performance of ResNet-50 in classifying autistic and non-autistic datasets before and after DeepLabv3+ is applied. The following is the flowchart of the study flow. In a comparative benchmark study, this study delved into the realm of autism face dataset classification, employing the formidable ResNet-50 architecture as the baseline model. This stalwart neural network exhibited commendable accuracy and efficiency in discerning facial features associated with autism. However, this study sought to push the boundaries further by integrating the power of segmentation using Deeplabv3+. The augmentation of ResNet-50 with Deeplabv3+ brought a significant enhancement to the classification process. The segmentation capability not only improved the model's understanding of facial structures but also provided a nuanced analysis of spatial relationships within the images. Consequently, the benchmark revealed that the combined approach surpassed the standalone ResNet-50, showcasing the potential of leveraging advanced segmentation techniques for more nuanced and accurate autism face dataset classification.

### 2.1 Facial Recognition System

The primary goal of the face recognition system is to recognize the human identity from the static images. Generally, a face recognition system deals with the input image as a classification problem [2]. Face recognition systems employ technology to recognize faces in an input image. This system will recognize and match the images in the dataset to the supplied images [17]. The face recognition process is carried out using a computerized method, including detecting and verifying a person through an image [18]. Although facial recognition systems have been widely studied, several challenges, such as misalignment, illumination variations, and expression variations, require approaches and tests to improve the accuracy and precision level of face recognition.

### 2.2 Convolutional Neural Network (CNN)

The CNN (or ConvNet) is a popular discriminative deep learning architecture that learns directly from the input without the need for a human feature extraction [19]. CNN is used to construct the majority of computer vision algorithms [20]. CNN represents the most exploited in the image processing field due to its ability to recognize patterns in images. CNN model can contain several types of layers. The most frequently used are convolutional, pooling, and fully connected layers. CNN consists of two layers, namely Convolution and Pooling [21,22].

## 2.3 ResNet-50 Architecture

ResNet-50, with 50 layers is one of the variants of ResNet, a CNN [23]. Fig. 2 shows the architecture of ResNet-50 in detail. ResNet is based on the deep residual learning framework [24]. ResNet50 is previously trained on ImageNet, composed of a large number of around 1.5 million natural scene images [25]. This figure also shows the major layers of ResNet-50, namely:  $3 \times 3$  conv 64,  $3 \times 3$  conv 128,  $3 \times 3$  conv 256,  $3 \times 3$  conv 512, feature vector map, etc [26].

## 2.4 DeepLabV3+

DeepLabv3+ is currently one of the best semantic segmentation algorithms available. Based on DeepLabv3, this technique creates a coding-decoding structure by incorporating a simple and effective decoder. DeepLabV3+ has two components: an encoder and a decoder. The encoder is made up of a backbone feature extraction network and an atrous spatial pyramid pooling (ASPP) structure, whereas the decoder takes low-level features from the backbone feature extraction network and up-samples them to produce pixel-by-pixel classification results the same size as the input image [27]. The framework of DeepLabV3+ is shown in Figure 3.

## 2.5 Dataset

The decision to opt for an algorithm other than DeepLabv3+ for segmenting the autistic spectrum face dataset stems from the necessity to identify the most appropriate method, considering various factors such as algorithm compatibility, performance evaluation criteria, feature extraction requirements, and potential for originality. Despite the effectiveness of DeepLabv3+ in semantic segmentation tasks, its specific structure and features might not perfectly match the dataset's characteristics, prompting the exploration of alternative algorithms that could potentially offer superior performance or novel perspectives. By venturing into uncharted territory with algorithms previously unused in this context, there's an opportunity to discover innovative methodologies and potentially enhance our understanding of autistic spectrum facial attributes, thereby driving progress in the field. The dataset in shown is Figure 3.

### 2.5.1 Backbone network (ResNet-50)

The input will enter the backbone network used, namely ResNet-50, which functions as a feature extractor for the input image. ResNet-50 has deep layers and the ability to transfer information through residual blocks.

### 2.5.2 Atrous spatial pyramid pooling (ASPP)

Atrous Spatial Pyramid Pooling (ASPP) uses atrous convolutions with varying dilation rates to capture information at different scales without sacrificing resolution. Each ASPP branch applies atrous convolution with different dilation rates, followed by Spatial Pyramid Pooling (SPP) to combine information from various receptive field sizes. ASPP output is used as input for subsequent steps in DeepLabv3+, such as up-sampling and merg-

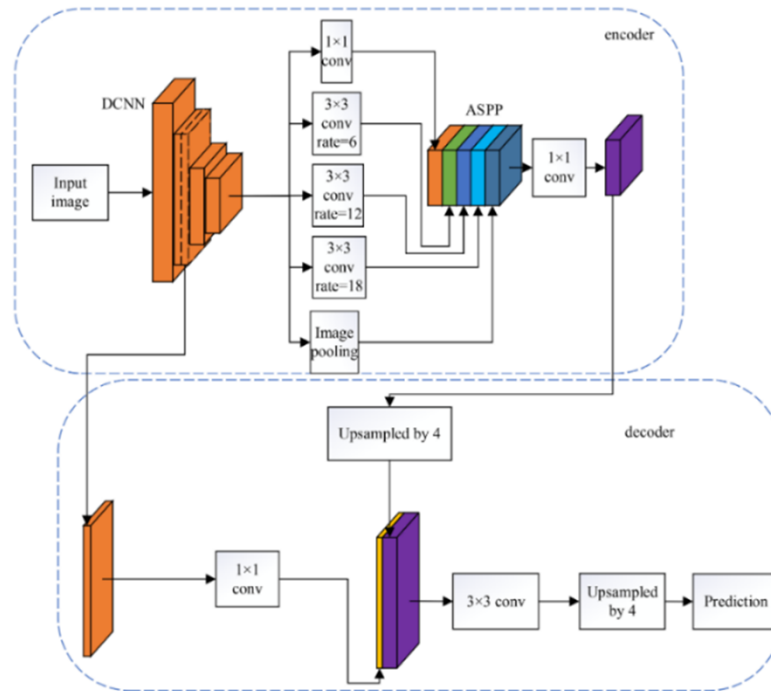


Figure 1: DeepLabV3+ architecture [20].

ing with results from other backbones, assisting the model in semantic segmentation tasks [28], [29], [30].

### 2.5.3 Multi-scale contextual

The multi-scale contextual process in DeepLabv3+ is the model's ability to understand the context at various spatial scales in the image. This is achieved through the use of Atrous Spatial Pyramid Pooling (ASPP), which uses atrous convolutions with different dilation rates to capture information at varying scales. Multiscale features in DeepLabv3+ offer several advantages such as improved object detection, reduced segmentation boundary problems, and robustness to image variations. With this, it can be said that multi-scale features in DeepLabv3+ play an important role in improving the model's ability to detect and segment objects in the input image, reducing the segmentation boundary problem, and increasing the model's robustness to image variations [31,32].

### 2.5.4 Up-sampling

Up-sampling involves the use of bilinear interpolation to increase the resolution of ASPP feature maps. The up-sampling results are used as input for combining with feature maps from the previous backbone, preparing the model to produce more detailed and accurate semantic segmentation predictions [32].

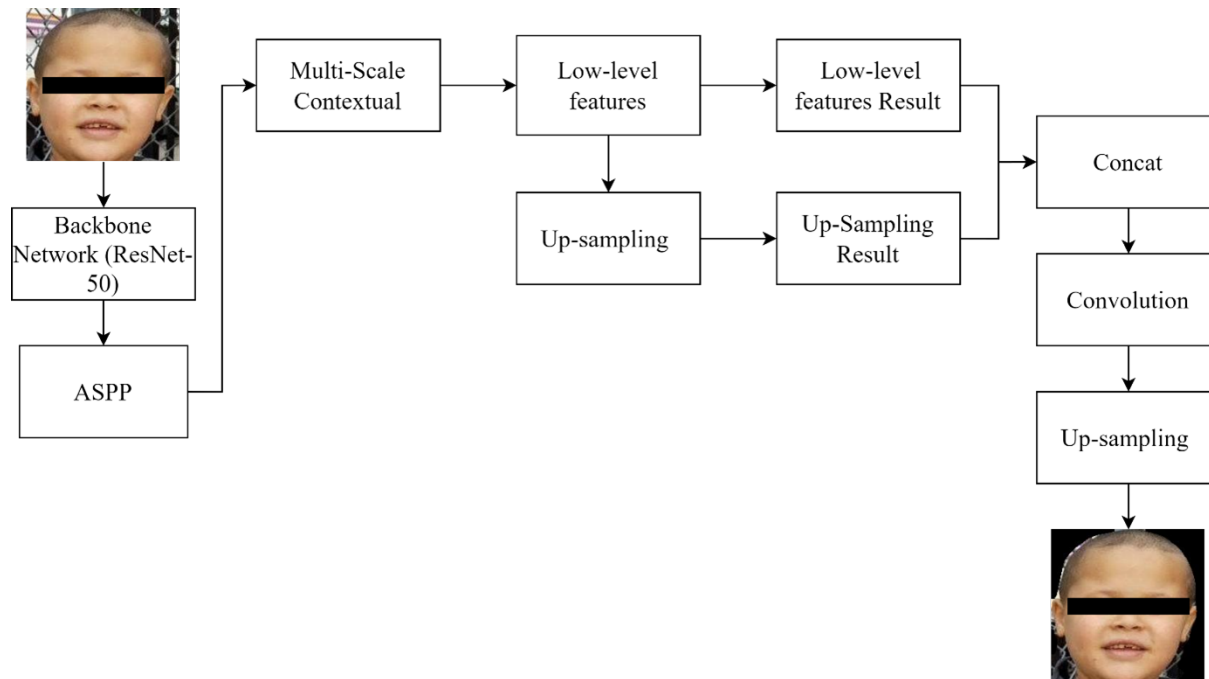


Figure 2: DeepLabV3+ framework.



Figure 3: Used dataset.

### 2.5.5 Low-level features

Low-level features in DeepLabv3+ work by obtaining and storing detailed local and structural information from the initial or lower layers in the network architecture. This is done to preserve spatial information and details that may be lost during deep pooling and convolution processes. Low-level features work collaboratively with high-level features produced by ASPP. They help improve the high resolution of segmentation results and give

the model the ability to capture fine details that are important for semantic segmentation tasks [33–35].

### 2.5.6 Concat (concatenate)

Concatenation is used to combine up-sampling results from the Atrous Spatial Pyramid Pooling (ASPP) module with low-level features originating from the initial or lower layers in the network architecture. Concatenation allows models to combine information from different scales and levels of resolution, improving the model's ability to capture the detail and context necessary for semantic segmentation tasks [36,37].

### 2.5.7 Convolutional

After concatenation, the model can apply additional convolution layers to concatenate and refine the concatenated features. The convolution function in the final stage of Deeplabv3+ is to produce accurate and appropriate semantic segmentation predictions. This allows the model to effectively identify and map each pixel in the input image [38,39].

### 2.5.8 Output/segmentation results

The results obtained are segmentation results from the input image.

## 2.6 Proposed Method

The datasets used in this study have 2168 images which are divided into two classes, autistic and non-autistic. Each class contains 1890 images. In the proposed method (shown in Figure 4) two processes will be carried out:

1. A dataset that has not been segmented will be trained first.
2. Then the segmented dataset will be trained.
3. The classification process can be seen in Figure 4.
4. Both results will be evaluated.

## 2.7 Accuracy

The scikit-learn handles the performance evaluation part. They employ metrics based on the "Confusion Matrix" to assess the binary classification model's performance [39]. Some of the important values that contribute to the performance evaluation were as follows: (i) False positive (FP): when data is negative yet the model predicts positive (ii) False negative (FN): positive data predicted by a model as negative (iii) True positive (TP): when both the data and the model anticipate positive. Accuracy represents the ability of the model to run a system properly, and the trained model will resemble the actual system performance. The system performance can be evaluated using the Accuracy, Recall, Precision, and F-Score metrics [40]. Four combinations of the actual value and the predicted value are used to see the performance of the classification system with the output of two classes [9].

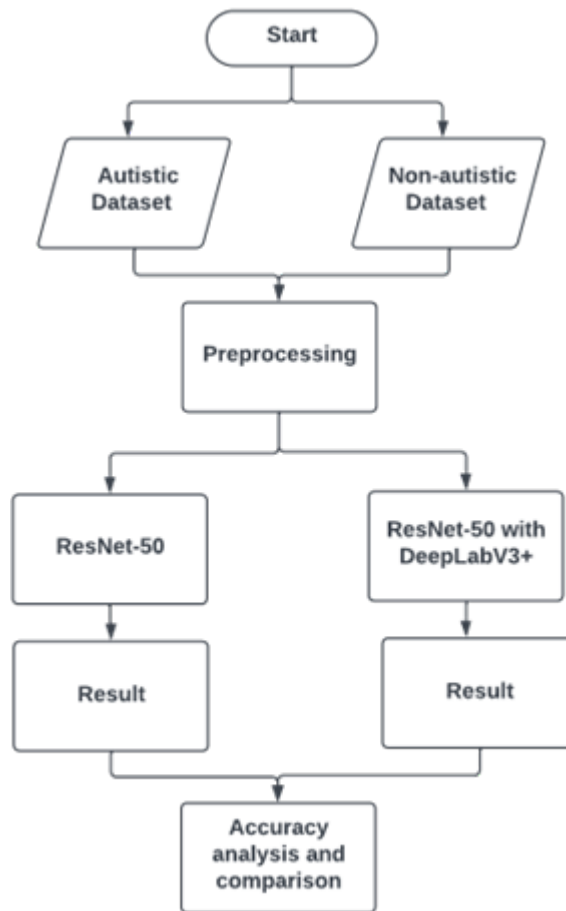


Figure 4: Proposed method.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (1)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (2)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (3)$$

$$\text{F1 - Score} = 2 \times \frac{\text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \quad (4)$$



### 3 Result

#### 3.1 Training And Validation using ResNet-50

The model was developed using the ResNet-50 architecture in the training and validation stages. The dataset is grouped into training and validation data, each of which amounts to 2,168 training data consisting of images of autism and non-autism. At the training and validation stage, where the comparison between training data and validation data is 70:30, there are 1890 images used to be trained. As many as 378 images are used for validation to see the relationship between epochs with the level of accuracy and loss from the training stage. Their respective validations can be observed on the curves shown in Figure 5 and Figure 6.

Figure 5 shows that the blue curve is the accuracy curve at the training stage, and the red curve is the validation accuracy curve. Based on Figure 7 and Figure 8, the resulting curve for the training stage shows the relationship between epochs and losses using the ResNet-50 architecture. It can be observed that the blue curve is the training stage loss curve, and the red curve is the validation loss curve. Based on the curve obtained in Figure 6, it can be seen that the loss resulting from the training and validation stages is still quite large for the implementation of the ResNet-50 architecture.

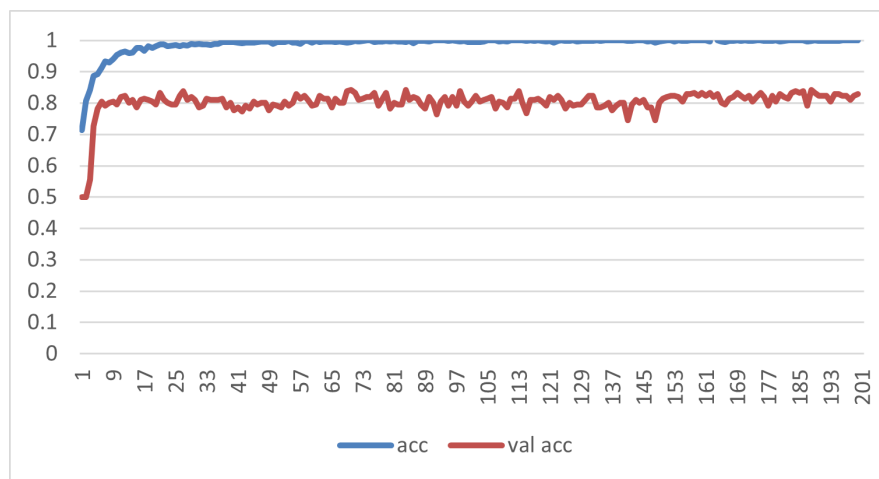


Figure 5: The Relationship of epoch and accuracy using ResNet-50.

#### 3.2 Training And Validation using ResNet-50 with Segmentation using DeepLabV3+

The model was built using the ResNet-50 architecture as the backbone at the training and validation stages. The dataset is grouped into training and validation data, each of which amounts to 2,168 training data consisting of images of autism and non-autism. At the training and validation stage, where the comparison between training data and validation data is 70:30, there are 1890 images used to be trained. As many as 378 images are used for validation to see the relationship between epochs with the level of accuracy and loss from

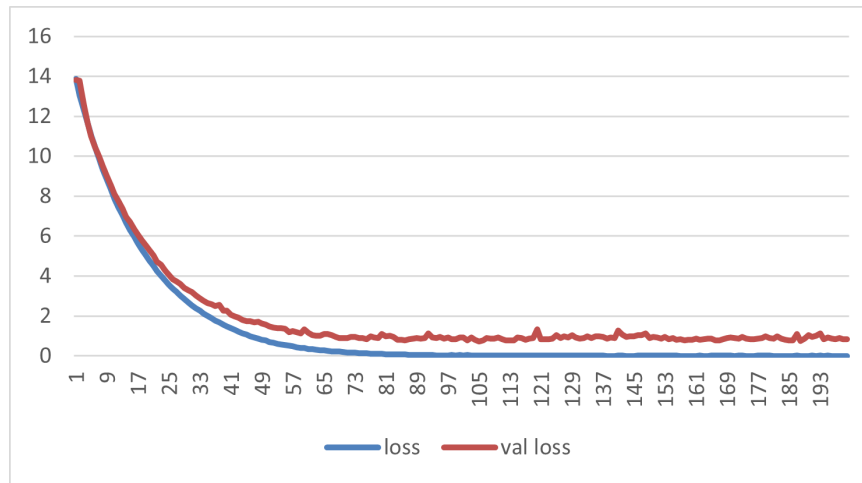


Figure 6: The Relationship of epoch and loss using ResNet-50.

the training stage. The ratio of this dataset is the same as the ratio of the previous dataset. Moreover, the respective validation can be observed on the curves shown in Figure 7 and Figure 8.

Figure 7 shows the epoch relationship curve and accuracy using ResNet-50 with segmentation using DeepLabV3+. Based on these curves, it can be observed that the blue curve is the accuracy curve at the training stage and the red curve is the accuracy curve at the validation stage. Based on Figure 8, it can be observed that the level of accuracy on the resulting curve from the training and validation stages will continue to change as the number of epochs increases so that the number of epochs used during training will be tested to achieve the optimum accuracy value. Based on Figure 9, the epoch-loss relationship curve using ResNet-50 with segmentation using DeepLabV3+. Based on the curve, it can be observed that the blue curve is the loss curve for the training stage, and the red one is the loss curve for the validation stage.

### 3.3 Comparison of Accuracy Results

The accuracy of the ResNet-50 models, both with and without segmentation using DeepLabV3+, has been meticulously evaluated. In assessing the performance of the classification system, the confusion matrix table serves as a robust tool, accompanied by equations (1), (2), (3), and (4) to quantify the accuracy of each method effectively. Initially, without implementing the segmentation method, ResNet-50 achieved an accuracy of 83.7% on the datasets. However, upon segmentation using DeepLabV3+, a notable improvement was observed, with accuracy soaring to 85.9%. This remarkable enhancement underscores the efficacy of DeepLabV3+ in refining classification accuracy, showcasing its potential to augment the performance of existing models. Furthermore, these results not only validate the effectiveness of the applied methodology but also in still confidence in its capability to yield tangible improvements in real-world applications. The comprehensive findings are eloquently presented in Table 1, encapsulating the incremental accuracy achieved through the integration of DeepLabV3+ segmentation into the classification pipeline. Such empir-

ical evidence not only underscores the significance of segmentation techniques but also highlights the promising trajectory of advancements in machine learning methodologies for enhancing classification tasks.

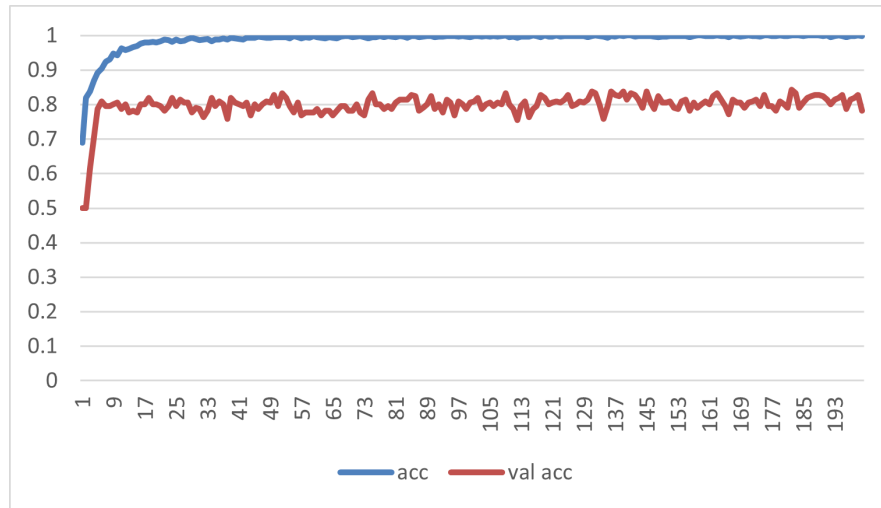


Figure 7: The Relationship of epoch and accuracy using ResNet-50 with segmentation using DeepLabV3+.

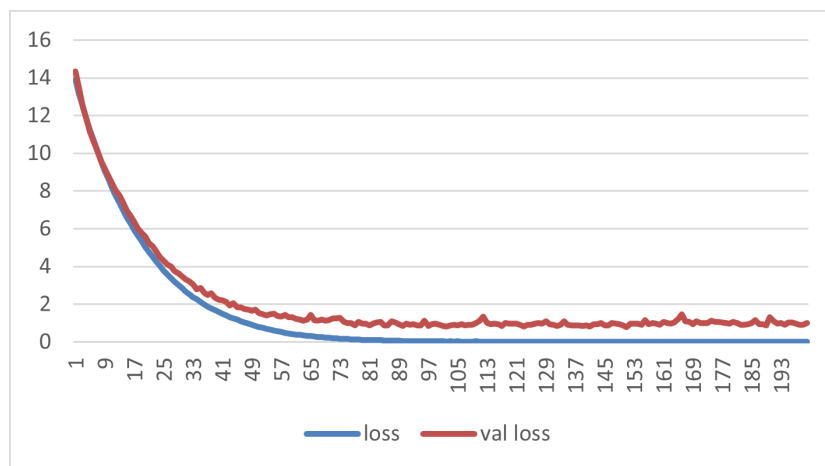


Figure 8: The Relationship of epoch and loss using ResNet-50 with segmentation using DeepLabV3+.

Table 1: Performance comparison

Model	Accuracy %	Recall %	Precision %	F-score %
Without segmentation	83.7%	82.4%	83.7%	83%
With segmentation	85.9%	90%	85.9%	87%

## 4 Discussion

The datasets used has unnecessary features. This is one of the reasons the author decided to use a segmentation method to reduce these features. DeepLabV3+ was chosen because it is one of the best semantic segmentation methods and has good accuracy in segmenting images. The decision to employ DeepLabV3+ was further motivated by its ability to effectively capture intricate details and subtle nuances within the images, providing a comprehensive understanding of the underlying content. This advanced segmentation method not only excels in feature reduction but also ensures precise delineation of objects, enhancing the overall performance and interpretability of the dataset. Moreover, DeepLabV3+ boasts a powerful semantic segmentation architecture that enables the model to discern and classify objects at a pixel level, contributing to a more granular and accurate representation of the data. The model's utilization extends beyond mere feature reduction; it introduces a nuanced comprehension of the contextual relationships between different elements in the images. This ensures that the resulting segmented dataset not only mitigates the impact of unnecessary features but also encapsulates the richness of the visual information, making it a robust choice for intricate image analysis tasks. The author also opted for ResNet-50 as the backbone for DeepLabV3+ due to its proven success in handling complex visual information. ResNet-50's deep architecture facilitates the extraction of hierarchical features from the input images, allowing DeepLabV3+ to capture both low-level and high-level features with remarkable efficiency. This choice not only enhances the model's ability to comprehend intricate patterns but also enables more effective feature integration during the segmentation process. By leveraging the rich representations learned by ResNet-50, DeepLabV3+ can achieve superior accuracy in image segmentation, making it a well-founded selection for the task at hand. Furthermore, ResNet-50's residual learning framework plays a pivotal role in mitigating the vanishing gradient problem, allowing for the successful training of deeper neural networks. This characteristic is particularly advantageous in the context of DeepLabV3+, where intricate segmentation tasks demand a sophisticated and deep architecture. The residual connections in ResNet-50 facilitate the smooth flow of gradients during backpropagation, enabling the model to effectively learn and adapt to the intricacies of the dataset. Additionally, the inherent robustness of ResNet-50 aids in the extraction of meaningful features that are essential for semantic segmentation. The model's ability to capture and retain detailed information across multiple scales contributes to the overall success of DeepLabV3+ in accurately segmenting diverse and complex images. The combination of DeepLabV3+ with ResNet-50 as its backbone thus not only addresses the challenges posed by unnecessary features but also ensures a robust and reliable solution for image segmentation tasks.

## 5 Conclusion

DeepLabV3+ succeeded in increasing accuracy. This also proves that the method applied can work well as expected. ResNet-50 produces 83.7% accuracy with the datasets without segmentation. Accuracy increased after the datasets were segmented using DeepLabV3+ to 85.9%.

## Acknowledgments

The authors express their gratitude to the Universitas Syiah Kuala for supporting this study.

## References

- [1] H. O. Ikromovich and B. B. Mamatkulovich, "Facial recognition using transfer learning in the deep cnn," *Open Access Repository*, vol. 4, p. 502–507, Mar. 2023.
- [2] W. Ali, W. Tian, S. U. Din, D. Iradukunda, and A. A. Khan, *Classical and modern face recognition approaches: a complete review*, vol. 80. Multimedia Tools and Applications, 2021.
- [3] W. Y. Kurniawan and P. H. Gunawan, "Classification of Autism Spectrum Disorder Based on Facial Images Using the VGG19 Algorithm," *Journal of Computing Science and Engineering*, vol. 18, no. 1, pp. 1–9, 2024.
- [4] M. Melinda, M. Oktiana, Y. Nurdin, I. Pujiati, M. Irhamsyah, and N. Basir, "Performance of ShuffleNet and VGG-19 Architectural Classification Models for Face Recognition in Autistic Children," *International Journal on Advanced Science, Engineering and Information Technology*, vol. 13, no. 2, pp. 674–680, 2023.
- [5] T. Debnath, M. M. Reza, A. Rahman, A. Beheshti, S. S. Band, and H. Alinejad-Rokny, "Four-layer ConvNet to facial emotion recognition with minimal epochs and the significance of data diversity," *Scientific Reports*, vol. 12, no. 1, pp. 1–18, 2022.
- [6] Z. Y. Huang, C. C. Chiang, J. H. Chen, Y. C. Chen, H. L. Chung, Y. P. Cai, and H. C. Hsu, "A study on computer vision for facial emotion recognition," *Scientific Reports*, vol. 13, no. 1, pp. 1–13, 2023.
- [7] M.-P. Hosseini, M. Beary, A. Hadsell, R. Messersmith, and H. Soltanian-Zadeh, "Deep learning for autism diagnosis and facial analysis in children," *Frontiers in Computational Neuroscience*, vol. 15, p. 789998, 2021.
- [8] J. Crucitti, C. Hyde, P. G. Enticott, and M. A. Stokes, "Head circumference trends in autism between 0 and 100 months," *Autism*, vol. 24, no. 7, pp. 1726–1739, 2020.
- [9] F. W. Alsaade and M. S. Alzahrani, "Classification and Detection of Autism Spectrum Disorder Based on Deep Learning Algorithms," *Computational Intelligence and Neuroscience*, vol. 2022, 2022.

- [10] A. Lu and M. Perkowski, "Deep learning approach for screening autism spectrum disorder in children with facial images and analysis of ethnoracial factors in model development and application," *Brain Sciences*, vol. 11, no. 11, 2021.
- [11] Y. Pratama, L. M. Ginting, E. H. L. Nainggolan, and A. E. Rismanda, "Face recognition for presence system by using residual networks-50 architecture," *International Journal of Electrical and Computer Engineering*, vol. 11, no. 6, pp. 5488–5496, 2021.
- [12] L. Boussaad and A. Boucetta, "Deep-learning based descriptors in application to aging problem in face recognition," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 6, pp. 2975–2981, 2022.
- [13] H. Fu, D. Meng, W. Li, and Y. Wang, "Bridge crack semantic segmentation based on improved deeplabv3+," *Journal of Marine Science and Engineering*, vol. 9, no. 6, 2021.
- [14] Y. Li, W. C. Huang, and P. H. Song, "A face image classification method of autistic children based on the two-phase transfer learning," *Frontiers in Psychology*, vol. 14, no. August, pp. 1–14, 2023.
- [15] S. Akter, H. Shahriar, and A. Cuzzocrea, "Autism disease detection using transfer learning techniques: Performance comparison between central processing unit vs graphics processing unit functions for neural networks," in *2023 IEEE 47th Annual Computers, Software, and Applications Conference (COMPSAC)*, pp. 1084–1092, 2023.
- [16] K. Ganesh, S. Umapathy, and P. T. Krishnan, "Deep learning techniques for automated detection of autism spectrum disorder based on thermal imaging," *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine*, vol. 235, pp. 1113–1127, October 2021. Epub 2021 Jun 9.
- [17] L. Tidmarsh and F. R. Volkmar, "Diagnosis and Epidemiology of Autism Spectrum Disorders," *Canadian Journal of Psychiatry*, vol. 48, no. 8, pp. 517–525, 2003.
- [18] K. H. Teoh, R. C. Ismail, S. Z. Naziri, R. Hussin, M. N. Isa, and M. S. Basir, "Face Recognition and Identification using Deep Learning Approach," *Journal of Physics: Conference Series*, vol. 1755, no. 1, 2021.
- [19] I. H. Sarker, "Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions," *SN Computer Science*, vol. 2, no. 6, pp. 1–20, 2021.
- [20] D. Bhatt, C. Patel, H. Talsania, J. Patel, R. Vaghela, S. Pandya, K. Modi, and H. Ghayvat, "Cnn variants for computer vision: History, architecture, application, challenges and future scope," *Electronics (Switzerland)*, vol. 10, no. 20, 2021.
- [21] W. Ayadi, W. Elhamzi, I. Charfi, and M. Atri, "Deep cnn for brain tumor classification," *Neural Processing Letters*, vol. 53, pp. 671 – 700, 2021.
- [22] T. Kattenborn, J. Leitloff, F. Schiefer, and S. Hinz, "Review on Convolutional Neural Networks (CNN) in vegetation remote sensing," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 173, pp. 24–49, 2021.



- [23] A. Victor Ikechukwu, S. Murali, R. Deepu, and R. Shivamurthy, "ResNet-50 vs VGG-19 vs training from scratch: A comparative analysis of the segmentation and classification of Pneumonia from chest X-ray images," *Global Transitions Proceedings*, vol. 2, no. 2, pp. 375–381, 2021.
- [24] B. Mandal, A. Okeukwu, and Y. Theis, "Masked Face Recognition using ResNet-50," 2021.
- [25] S. A. ElGhany, M. R. Ibraheem, M. Alruwaili, and M. Elmogy, "Diagnosis of Various Skin Cancer Lesions Based on Fine-Tuned ResNet50 Deep Network," *Computers, Materials and Continua*, vol. 68, no. 1, pp. 117–135, 2021.
- [26] M. K. Yaqoob, S. F. Ali, M. Bilal, M. S. Hanif, and U. M. Al-Saggaf, "Resnet based deep features and random forest classifier for diabetic retinopathy detection†," *Sensors*, vol. 21, no. 11, 2021.
- [27] C. Cai, J. Tan, P. Zhang, Y. Ye, and J. Zhang, "Determining strawberries' varying maturity levels by utilizing image segmentation methods of improved deeplabv3+," *Agronomy*, vol. 12, no. 8, 2022.
- [28] H. Li, H. Xu, X. Tian, Y. Wang, H. Cai, K. Cui, and X. Chen, "Bridge crack detection based on SSEnets," *Applied Sciences (Switzerland)*, vol. 10, no. 12, 2020.
- [29] D. Varshney, M. Rahnemoonfar, M. Yari, J. Paden, O. Ibikunle, and J. Li, "Deep learning on airborne radar echograms for tracing snow accumulation layers of the greenland ice sheet," *Remote Sensing*, vol. 13, no. 14, pp. 0–21, 2021.
- [30] Y. Lin, D. Xu, N. Wang, Z. Shi, and Q. Chen, "Road extraction from very-high-resolution remote sensing images via a nested SE-Deeplab model," *Remote Sensing*, vol. 12, no. 18, pp. 0–20, 2020.
- [31] Y. Liu, M. Zhu, J. Wang, X. Guo, Y. Yang, and J. Wang, "Multi-Scale Deep Neural Network Based on Dilated Convolution for Spacecraft Image Segmentation," *Sensors*, vol. 22, no. 11, 2022.
- [32] H. Chen, Y. Qin, X. Liu, H. Wang, and J. Zhao, "An improved DeepLabv3+ lightweight network for remote-sensing image semantic segmentation," *Complex and Intelligent Systems*, vol. 10, no. 2, pp. 2839–2849, 2024.
- [33] O. Elharrouss, Y. Akbari, N. Almaadeed, and S. Al-Maadeed, "Backbones-Review: Feature Extraction Networks for Deep Learning and Deep Reinforcement Learning Approaches," 2022.
- [34] A. Rashwan, X. Du, X. Yin, and J. Li, "Dilated SpineNet for Semantic Segmentation," 2021.
- [35] M. M. Memon, M. A. Hashmani, A. Z. Junejo, S. S. Rizvi, and K. Raza, "Unified DeepLabV3+ for Semi-Dark Image Semantic Segmentation," *Sensors*, vol. 22, no. 14, 2022.
- [36] H. Chen, Y. Qin, X. Liu, H. Wang, and J. Zhao, "An improved deeplabv3+lightweight network for remote-sensing image semantic segmentation," *Complex & Intelligent Systems*, vol. 10, pp. 2839–2849, April 2024.

- [37] K. Jakhar and N. Hooda, "Big data deep learning framework using keras: A case study of pneumonia prediction," *2018 4th International Conference on Computing Communication and Automation (ICCCA)*, pp. 1–5, 2018.
- [38] M. M. Memon, M. A. Hashmani, A. Z. Junejo, S. S. Rizvi, and K. Raza, "Unified deeplabv3+ for semi-dark image semantic segmentation," *Sensors*, vol. 22, no. 14, 2022.
- [39] A. A. Alnuaim, M. Zakariah, C. Shashidhar, W. A. Hatamleh, H. Tarazi, P. K. Shukla, and R. Ratna, "Speaker Gender Recognition Based on Deep Neural Networks and ResNet50," *Wireless Communications and Mobile Computing*, vol. 2022, 2022.
- [40] C. He, M. Ma, and P. Wang, "Extract interpretability-accuracy balanced rules from artificial neural networks: A review," *Neurocomputing*, vol. 387, pp. 346–358, 2020.

