Dean&Francis
ISSN 2959-6157

Image Segmentation with Different Scale Convolutional Neural Networks

Yuming Ran [1, *]

[1] Ulink College, GuangZhou, China

*Corresponding author:
yumran2660@ulinkcollege.com

Abstract:

Image segmentation is a crucial task in computer vision and image processing. It is widely used in many necessary fields, such as scene understanding, medical image analysis, robot perception, video surveillance, augmented reality, and image compression. Numerous algorithms for image segmentation have been proposed, demonstrating their unique advantages and limitations in their respective application scenarios. In image processing and pattern recognition, the importance and criticality of image segmentation are self-evident. Its core task is to divide the entire image into several regions with specific meaning and define a category label for each area. In recent years, convolutional neural networks (CNN) have performed well in image segmentation and have become one of the most popular and widely used models. This paper focuses on changing the model scale, which significantly impacts the segmentation results by changing the size of the data set used to train the model. This paper aims to explore the impact of data volume on model performance. For example, will the segmentation results become more accurate as the model scale increases? This paper first created and trained a CNN model using different scales. In each training, this paper trains the model for 50 epochs, which can significantly improve the reliability and accuracy of the experimental results. Next, this paper segments the test image, analyzes the segmentation effect, and further explores the relationship between parameters scale and model performance. This research will provide new ideas and references for optimizing image segmentation.

Keywords: Image segmentation; Convolutional Neural Network; Deep Learning; Training Strategy.

ISSN 2959-6157

1. Introduction

As a core task in computer vision, image segmentation is widely used in various vital fields, such as autonomous driving and medical diagnosis. Nevertheless, the research and practice results of image segmentation could be better, and there remain many challenges. Image segmentation is involved in extracting meaningful information from scenarios with a large amount of data or scenes where complex decision-making is required, and image segmentation is indispensable in these aspects. For example, in autonomous driving, precise image segmentation can enable vehicles to identify better the category of road signs, pedestrians, and other obstacles, thus improving driving safety and reliability. Similarly, in the medical field, the progress of image segmentation technology can help increase the accuracy of finding lesions and assist doctors in making a more accurate diagnosis. Therefore, further improving the efficiency and accuracy of image segmentation brings more significant benefits to the existing technological aspects, and the support will also be more powerful [1-3].

Despite the significant advancements, there remains a pressing need to enhance the processing speed and accuracy of Convolutional Neural Network (CNN) models in image segmentation tasks. By optimizing CNNs based on existing technologies, we can achieve greater efficiency and precision, leading to transformative improvements across related fields. Such advancements will propel industry development and bring substantial societal benefits [4, 5].

The research began with pre-processing the Oxford Pet Collection dataset and standardizing image format and size to ensure consistency across the dataset. This step was crucial for maintaining the integrity of the data used in the study. The Convolutional Neural Network (CNN) was then defined using TensorFlow and Keras libraries, which provided the foundational structure for the neural network. The network was composed using the layers module, allowing for a precise definition of each layer within the CNN. In applying the CNN, various layers were employed to extract features from the images progressively, culminating in a final output layer that generated pixel-level category predictions. The model was trained using this structure, effectively integrating the different components to achieve the image segmentation task with high efficiency.

The experimental results demonstrated that, with the epoch count fixed at 50, an increase in the parameters scale led to higher accuracy and greater efficiency in image segmentation. The accuracy improvements were directly correlated with the parameters scale, while the performance of the GPU influenced efficiency. On our device, each epoch took approximately 60-65 seconds to complete; however, with a more powerful GPU, the time per epoch could be reduced to 30-35 seconds. This highlights the critical role of GPU capabilities in determining the overall efficiency of the experiment. The research not only simplifies the understanding of the model but also enhances the accuracy and efficiency of the algorithm. These improvements can potentially advance both the academic and practical applications of image segmentation.

The rest of this paper is organized as follows. Section 2 provides an overview of the proposed CNN model, including the layers of the model used. Section 3 presents the experimental results and compares the results obtained with different data amounts. Section 4 discusses the limitations of this study and outlines future research directions.

2. Convolutional Neural Network

Conv2D: The first process within this layer is the feature extraction from the input image. It involves applying a set of learnable filters that slide through the image and detect local features such as edges, textures, and other patterns. Detecting these multiple features across the input image is paramount in the initial development and understanding of visual content.

MaxPooling2D: Following the extraction of features in the process above, the MaxPooling2D layer runs a downsampling operation on the feature maps for others. It selects the maximum value within a specified window and maintains the most crucial characters while reducing the spatial dimensions of the feature map to minimize computational demand. It also enhances the robustness and computing efficiency of the feature representation, thus preventing overfitting, thanks to the feature abstraction. It accomplishes this by replicating each maximum entry service window.

UpSampling2D: upSampling2D is a two-arrow function used to increase the spatial resolution of the feature layer generated by the previous layer and find it mainly in the encoder-decoder's decoder part, such as the U-Net. Due to the convolutional layer strides values, it is used to expand the peak layer to the original input image's load size or any other spatially structured layer. The re-creation of the broad structure allows some pixel-based predictions and improves accuracy in distinguishing structure results.

Concatenate: This layer combines two arrow layers of the same size along the axis channel- this rank. It combines multi-level future maps during the combined structure, allowing developers to use matching features other layers provide in the further structure design. The combination improves the result's quality and accuracy, enhancing the

layer's application performance by using collected and combined layers' achievement.

Activation: The activation layer introduces non-linearity into the system since non-linearity enables the system to handle beyond linearly distributed data. This measure is mainly applied in decision and classification instances to amplify the output results for better judgment and analysis. It provides designers with a recording function.

BatchNormalization: BatchNormalization layers were introduced to stabilize and speed up the learning process; they normalized the inputs to each layer so the model would converge faster. As a result, the training data distributions became more balanced across the network, increasing the convergence rate. BatchNormalization decreases the network sensitivity to initialization. Neural nets were also affected by vanishing and exploding gradients, which BatchNormalization helped mitigate. Typically, BatchNormalization layers are present after convolutional layers.

Dropout: Dropout is a regularization technique to prevent overfitting in neural networks. During training, the Dropout layer randomly "drops" or turns off a fraction of neurons, which forces the network to develop redundant representations. This random exclusion of neurons during training helps the network generalize better by ensuring it does not rely too heavily on specific features, thereby improving the model's overall robustness and performance on unseen data.

3. Experimental Results and Analysis

The pet set used in this analysis is the Oxford Pet Set, which includes images of 37 different breeds of pets. Tables 1, 2, and 3 show CNN models of three different scales used in this paper.

Table 1. CNNs with 181,027 parameters

Layer (Type)	Output Shape
input_layer	(200,200, 3)
rescaling	(200, 200, 3)
conv2d	(100,100, 16)
conv2d_1	(100,100, 16)
conv2d_2	(50,50, 32)
conv2d_3	(50,50, 32)
conv2d_4	(25,25, 64)
conv2d_5	(25,25, 64)
conv2d_transpose	(25,25, 64)
conv2d_transpose_1	(50,50, 64)
conv2d_transpose_2	(50,50, 32)
conv2d_transpose_3	(100,100, 32)
conv2d_transpose_4	(100,100, 16)
conv2d_transpose_5	(200,200, 16)
conv2d_6	(200,200, 3)

Table 2. CNNs with 721,475 parameters

Layer (Type)	Output Shape
input_layer_1	(200,200, 3)
rescaling_1	(200,200, 3)
conv2d_7	(100,100, 32)
conv2d_8	(100,100, 32)
conv2d_9	(50,50, 64)
conv2d_10	(50,50, 64)
conv2d_11	(25,25, 128)

ISSN 2959-6157

conv2d_12	(25,25, 128)
conv2d_transpose_6	(25,25, 128)
conv2d_transpose_7	(50,50, 128)
conv2d_transpose_8	(50,50, 64)
conv2d_transpose_9	(100,100, 64)
conv2d_transpose_10	(100,100, 32)
conv2d_transpose_11	(200,200, 32)
conv2d_13	(200,200, 3)

As shown in Figure 1, both the training loss and validation size follow very similar trends, thus indicating no overfitting of the model during post-training. Nevertheless, as Figure 2 shows for the test images, this model has more significant deviations, inferring some inconsistency. This misalignment indicates that, even though the model performs well on both the training and validation datasets, there are glaring issues with generalizing it to unseen data, resulting in highly likely inconsistent predictions. This leads to the need for better training of these models, or they can also explore more kinds of augmented data so that their robustness increases and performance on test images improves. Tackling such issues helps close the gap between training and real-world application, allowing for the alignment of model accuracy, reliability, and suitability across all data sets.

Table 3. CNNs with 2,880,643 parameters

Layer (Type)	Output Shape
input_layer	(200,200, 3)
rescaling	(200,200, 3)
conv2d	(100,100, 64)
conv2d_1	(100,100, 64)
conv2d_2	(50,50, 128)
conv2d_3	(50,50, 128)
conv2d_4	(25,25, 256)
conv2d_5	(25,25, 256)
conv2d_transpose	(25,25, 256)
conv2d_transpose_1	(50,50, 256)
conv2d_transpose_2	(50,50, 128)
conv2d_transpose_3	(100,100, 128)
conv2d_transpose_4	(100,100, 64)
conv2d_transpose_5	(200,200, 64)
conv2d_6	(200,200, 3)

Besides, As show in Figure 2, overall image segmentation, as compared with the smaller-scale CNN mode, is more extensive and accurate. Although there is still a little section on the left that isn't wholly segmented, the accuracy has also greatly improved. This improvement in segmentation quality indicates that the larger model can better identify and delineate features within the image, leading to a more thorough analysis.

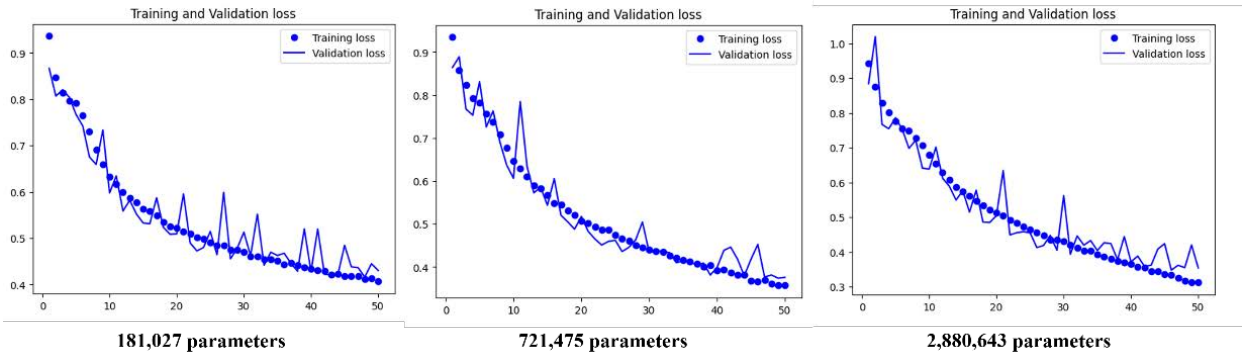

Fig. 1 The training and validation loss

The improvement can be easily seen in the fitting curves and test images. While the final segmentation results, such as this test image, are imperfect, there is an improvement over these overall network accuracies on some foregrounds (in particular). No misaligned points can be found in the regions around some of them. As we scale up the parameters this way and keep improving on refining it, both the accuracy & stability of the model will improve. With more data, the model learns better and generalizes well (in most cases), which results in improved performance across metrics. It indicates that larger model are required for better performance and accuracy with models in such domains as image segmentation.

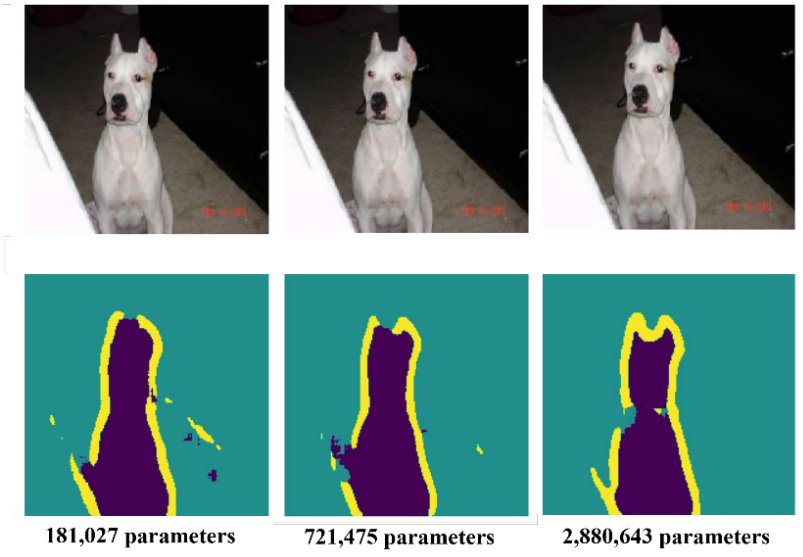

Fig. 2 The case study of model performance

4. Discussion

Although this study has shown that increasing the model scale can dramatically improve the model's accuracy, stability, and efficiency, several limitations need to be noted. First, the dataset is still relatively small. The results indicate that broader datasets result in more accurate education and overall model performance, but the field of study still needs to be improved. In real-life scenarios, the model that deals with a more complex and varied environment may show different results. It would be necessary to use broader and more varied datasets to verify that the proposed model will achieve the same high levels of accuracy under any circumstances. By exposing the model to a greater variety of examples, the expanded dataset will allow for more effective generalisation and create a solution that can be equally good in any real-life situation.

Ultimately, this paper expects to diversify the datasets and expand them significantly. Once the model is trained in many more situations and conditions, it will be a lot more versatile than it is. As a result, this enhancement will make generalising various situations more manageable and accessible. This way, the model will be applied to different environments and used reliably and accurately. The much more comprehensive and diverse dataset will benefit the

ISSN 2959-6157

fields that work with image segmentation. Specifically, medical diagnostics, autonomous driving, agriculture, and other fields in need of precision will significantly benefit from this generalisation. In the future, and with a broadened and diversified dataset, the model may help in many more fields.

5. Conclusion

This study particularly highlights the impact of model scale on the accuracy of image segmentation using the CNN model. Experimental results show that the more data there is, and the number of epochs remains at 50, the more the accuracy of the test image and the stability of the model will be significantly improved. As a core task in computer vision, image segmentation is widely used in various necessary fields, such as autonomous driving and medical diagnosis. However, the current research results and practical effects of image segmentation could be better, and there are still many challenges. Therefore, under the work of these two critical fields, further improving the efficiency and accuracy of image segmentation will bring more significant benefits to existing technologies and strengthen support.

References

[1] Cheng Qiyun, Sun Caixin, Zhang Xiaoxing, et al. Short-Term load forecasting model and method for power system based on complementation of neural network and fuzzy logic. Transactions of China Electrotechnical Society, 2004, 19(10): 53-58.

[2] Kai Han, Victor Sheng, Yuqing Song, et al. Deep semi-supervised learning for medical image segmentation: A review. Expert Systems with Applications, 2024: 123052.

[3] Reza Azad,Ehsan Khodapanah Aghdam, Amelie Rauland, et al. Medical image segmentation review: The success of u-net. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024.

[4] Mayuri Gupta, Ashish Mishra. A systematic review of deep learning based image segmentation to detect polyp. Artificial Intelligence Review, 2024, 57(1): 7.

[5] Xudong Wang, Shufan Li, Konstantinos Kallidromitis, et al. Hierarchical open-vocabulary universal image segmentation. Advances in Neural Information Processing Systems, 2024, 36.