IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

## IN COMPUTER & COMMUNICATION ENGINEERING

Volume 11, Issue 5, May 2023

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105157 |

# Real-Time Image Segmentation using Android App

Sunny Kumar Yadav, Saurabh Mishra, Manish Kumar Sah, Shivam Kasaudhan

Dept. of Computer Science, Jain University, Bengaluru, India

**ABSTRACT**:Image segmentation is a vital field that has a lot of potential for future computer vision research. It is essential in many real-world applications. Deep learning has recently made incredible strides in image segmentation, allowing computers to recognize, separate, and classify particular items of interest within images. However, frontal view or asymmetric field of view objects have been the main focus of the majority of existing techniques. The unique strategy that is suggested in this paper combines the benefits of the two current approaches. In order to augment DeepLabv3 and achieve better segmentation outcomes, particularly at object borders, we propose a model we call DeepLabv3+. An encoder, also known as a contracting path, captures the image context, and a decoder, sometimes known as an expanding path, allows for precise localisation. As the encoder, we use a trained CNN to compress the feature maps of the input image. The decoder up-samples and reconstructs the output using the crucial data that the encoder collected. We also examine the Xception model to build a faster and more efficient encoder-decoder network by using Depthwise separable convolution in the Atrous Spatial Pyramid Pooling and the decoder modules. On the PASCAL VOC 2012 and Cityscape datasets, respectively, our proposed model achieves test set accuracies of 89.0% and 82.1% without the need of any post-processing procedures. This demonstrates the potency of our model.

**KEYWORDS**: Camera2API, Keras, Atrous Convolution, DeepLabV3+, PASCAL VOC 2012

## I. INTRODUCTION

### A. Overview

Semantic segmentation, which entails putting specific picture pixels into semantic categories, is the main goal of this study. In comparison to manually developed feature-based systems, the efficacy of deep convolutional neural networks, more notably the Fully Convolutional Neural Network (FCN), in successfully executing segmentation is explored.We investigate encoder-decoder structures and spatial pyramid pooling modules as two classes of neural networks. While spatial pyramid pooling modules, which combine data from various resolutions, provide contextual information, encoder-decoder models excel at establishing accurate object boundaries. The research suggests employing Atrous convolutions to extract denser feature maps to overcome the issue of losing accurate boundary information caused by pooling and convolutions in the network backbone. This tactic, however, makes calculations more challenging. As an extension of DeepLabv3, the authors propose DeepLabv3+, which combines the benefits of both approaches for faster computation in the encoder channel and progressive recovery of object boundaries in the decoder path. The encoder module of the encoder-decoder networks is enhanced with multi-scale contextual data , and a decoder module is inserted to recover object boundaries.

DeepLabv3+ makes use of the extensive semantic data stored in the output of DeepLabv3 to alter the density of encoder features using Atrous convolutions. The decoder module aids in obtaining accurate object bounds.

### B. Problem Definition

The Android software under development attempts to process either a single image or a stream of photos continuously taken by the device's camera. The primary objective is to carry out pixel-level segmentation, separating the background from the object of interest. Modern deep learning models will be included into the programme to enable precise and effective image-segmentation, completing the work.

### C. Objectives

i. To create an Android application that makes use of the Camera2API and DeeplabV3+ segmentation model for real time image segmentation and object categorization.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105157 |

### D. Methodology

This study's main objective is the real-time segmentation and classification of objects from video feeds. Modern AI applications, including those used in mobile devices and self-driving automobiles, heavily rely on object segmentation. Deep learning methods have been used to demonstrate the efficacy of this strategy. The model used in this project was trained in Tensorflow, and we stored it in the. tfliteformat to make it easier to use in our Android application.

1. Acquire the DeepLabV3+ model that has been trained and convert it to. tflite format.
2. Take a photo and send it to the simulation.
3. The model attempts to categorise the image's objects into one of 20 categories and separate them into various colour overlays.
4. Our model accepts an image in the form of four-dimensional tensors.
5. The image is initially converted into three dimensions as part of the pre-processing.
6. The image is then trimmed to a 257x257 size before being fed into the model.
7. To extract features from images, our model employs encoder-decoder with Atrous convolution.
8. The next step is to categorise each pixel as either an object or a backdrop.
9. Classified pixels are then positioned in their appropriate locations in a new image that is the same size as the original image.
10. This image is overlaid over the original image.

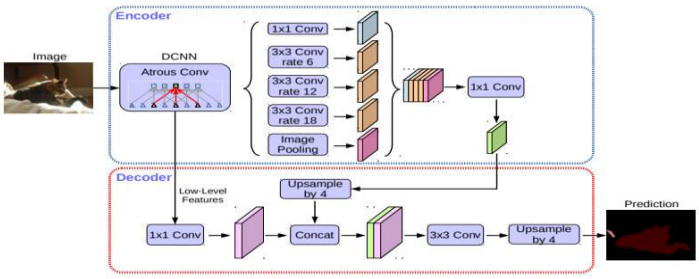



Fig: Flow Chart of the system used in this project

## II. LITERATURE REVIEW

### A. Existing System

Fully convolutional network (FCN)-based models have shown notable improvements in a number of benchmarks for segmentation. Several model variants have been suggested to use contextual information for segmentation. These comprise models that use multi-scale inputs, such as image pyramids, as well as models that use probabilistic graphical models like DenseCRF with efficient inference techniques. The main focus of this paper is on the examination of models with a spatial pyramid pooling and encoder-decoder structure.

**Pyramid pooling in space:**Atrous Spatial Pyramid Pooling (ASPP), also known as multiple parallel Atrous convolutions, is used in models like PSPNet or DeepLab., or combine spatial pyramid pooling, including image-level pooling, at different grid scales. These models make use of multi-scale data to obtain encouraging segmentation benchmark results.

**Encoder-decoder:** Semantic segmentation, object detection, and human position prediction are just a few of the computer vision tasks that these neural networks have proven to be incredibly effective at. An encoder module, a part of an encoder-decoder network, gradually reduces feature maps while recording high-level semantic data. The decoder module eventually recovers spatial data. We suggest adding a straightforward yet efficient decoder module and using DeepLabv3 as the encoder module to enhance the quality of border delineation and segmentation.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105157 |

**Depthwise separable convolution:** Using group convolution, also known as Depthwise separable convolution, is a potent method for lowering processing needs and parameters without sacrificing performance. This method has been significantly included into several contemporary neural network topologies. The Xception model is specifically examined in the context of semantic segmentation, and its gains in terms of accuracy and speed are demonstrated.

**B. Limitation of Existing System**

• Traditional image segmentation methods tend to over-segment and are susceptible to noise.
• The existing system lacks the capability for on-site log file viewing.

Proposed System- In this study, transfer learning was used to convert an image segmentation model's weightsinto. tflite files. The MobileNetV3 architecture, which is recognised for its ability to extract high-level information from pictures, serves as the foundation for DeepLabV3's capability. In order to deploy and integrate the model into an Android app, we ultimately converted the model into TensorFlow Lite format.

## III. METHODOLOGY

Atrous convolution and depth wise separable convolution are introduced briefly in this section. Before describing the suggested decoder module that is added to the encoder's output, we will first examine DeepLabv3, the encoder module that we use. We also provide a modified Xception model that, by permitting faster calculation, improves efficiency.

**A. Encoder-Decoder with Atrous Convolution**

Atrous convolution is a useful technique for extending the capabilities of traditional convolution techniques. Deep convolutional neural networks' feature resolution can be directly adjusted, and the filter's field of view can be adjusted to collect data at various scales. In the setting of two-dimensional signals, Atrous convolution is the process of applying the convolution filter, denoted as w, across the input feature map x at each point i on the output feature map y. $y[i] = \Sigma\, k\, x[i + r \cdot k]w[k]$

The Atrous rate, abbreviated "r," determines the input signal's sampling step. The cited source has more thorough explanations of this concept for readers who are interested. Keep in mind that the Atrous rate for standard convolution is set to 1. However, by altering the rate value, the filter's field of vision can be modified adaptively.

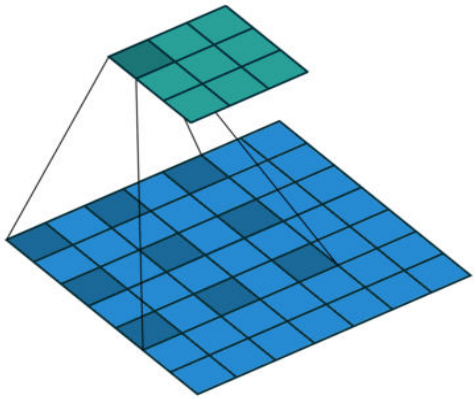


Fig: Atrous Convolution

**Depthwise separable convolution:**In order to simplify processing, the study introduces depth wise separable convolution, which separates a convolution into a depth wise convolution and a pointwise convolution. With Atrous separable convolution, performance is maintained or enhanced while computation complexity is further reduced. This resulting convolution is referred to in this study as an Atrous separable convolution.

**DeepLabv3 as encoder:** To extract features from deep convolutional neural networks at various resolutions, DeepLabv3 uses Atrous convolution. The ratio between the input image resolution and the final output resolution is known as the output stride. By using appropriate Atrous convolutions and lowering striding in the final blocks, which allows for output strides of 16 or 8, it is possible to extract more detailed features for semantic segmentation. The Atrous Spatial Pyramid Pooling module gains image-level capabilities with DeepLabv3, enabling convolutional feature probing at various sizes while utilising Atrous convolutions at varied rates. In the suggested encoder-decoder arrangement, the DeepLabv3 original encoder output is used as the final feature map prior to the logits. The encoder feature map, which contains a wealth of semantic data, has 256 channels. Atrous convolution can be employed at the required resolutions depending on the available computer resources.

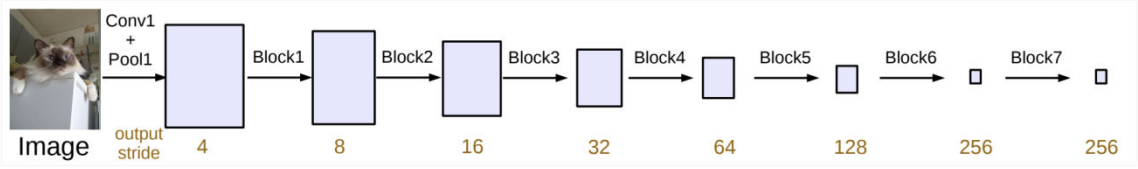
| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

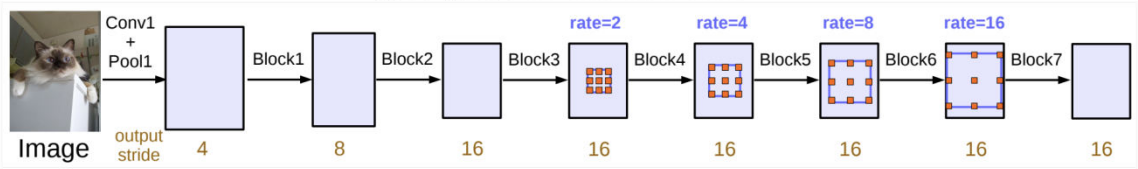
| DOI: 10.15680/IJIRCCE.2023.1105157 |

**Proposed decoder:** DeepLabv3's standard output stride for computing encoder characteristics is 16. However, because the decoder module in this study uses bilinear upsampling of features by a factor of 16, which may not reliably retrieve object segmentation details, it is seen as being overly simplistic. A more effective decoder module is suggested in order to overcome this restriction.

The recommended decoder module integrates similar, low-level, and geographically identical data from the network backbone with the upsampled encoder features. The channel count is decreased using a 1x1 convolution to minimise the importance of the encoder features—which have fewer channels—from being overshadowed by the high channel count of the low-level features. In order to enhance the features after concatenation, additional 3x3 convolutions are performed before a fourth bilinear upsampling.The output stride of the encoder module should be adjusted to 16 to obtain the best balance between speed and precision, as shown in Section 4 of the paper. Although performance is marginally improved when the output stride is set to 8, the computational complexity also increases.




(a) Going deeper without atrous convolution.




(b) Going deeper with atrous convolution. Atrous convolution with $rate > 1$ is applied after block3 when $output\_stride = 16$.

Figure 3. Cascaded modules without and with atrous convolution.

Fig: Decoder model with Atrous Convolution

The mean Intersection-over-Union (IoU) was used as the measurement tool to assess the performance of semantic picture segmentation. A popular evaluation metric in the area of semantic image segmentation is intersection-over-union.

The IoU metric is defined as follows for a specific class:

IoU=TP/(TP+FP+FN)

## IV. TOOL DESCRIPTION

This section provides in-depth information on the hardware and software tools used to create this system.

### A. Software:

Python: One of the most well-known programming languages for machine learning is Python. It provides specialised libraries for studying linear algebra and the kernel technique in machine learning, including SciPy, pandas, Scikit-learn, and NumPy. Python is ideally suited for implementing machine learning algorithms due to its simple syntax.

TensorFlow Lite: The creation of a deep learning framework enables on-device inference, making it possible to train and deploy machine learning models on a variety of gadgets, including Android, iOS, Raspberry Pi, and IoT devices. As a result, there is no longer a requirement for intensive cloud-based processing, and real-time inference and decision-making at the edge are made possible. This architecture enables the deployment of machine learning capabilities directly on the devices.

Java: Popular high-level programming language Java uses a class-based structure and an object-oriented paradigm. It provides a flexible development environment for a range of applications, including games, desktop software, web applications, and mobile apps. Java is made to reduce implementation dependencies, giving developers a powerful and broadly compatible language for a range of software development requirements.

### B. Hardware:

- An Android phone with a camera module and Camera2API functionality.

## V. IMPLEMENTATION

• Jupyter Notebook and Android Studio are two programming tools used in this project.

• To execute the tflite model on Android, the 'org.tensorflow:tensorflow-lite:0.0.0-nightly' requirement is added to the app level gradle file.

The label file and trained model file are stored in the assets folder.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105157 |

• The Camera2API is used to access cameras and take photographs.
• The 20 classes of the Java-based TFLite Object Detection Model are used to categories objects into different groups.

**STEPS TO RUN THE PROGRAM:**

- Installed on an Android phone with a camera module is the Image Segmentation application.
- The Image Segmentation app can be opened by clicking on its icon.
- The programme will ask the user for permission to access the camera on first launch.
- The app needs users' permission to access their cameras.
- An interface for the app will show a camera fragment along with the segmented objects inside the view frame window.

The simulation results demonstrated that the suggested approach outperforms the maximum number of hops metric when using the total transmission energy meter. The suggested method maximizes the lifetime of the entire network and offers an energy-efficient way for data transfer. The performance of the suggested algorithm can be compared with other energy-efficient algorithms as the performance of the method is evaluated between two metrics in the future with minor alterations in design considerations. We've only employed a relatively modest network of five nodes; as the number of nodes rises, so will its complexity. We can add more nodes and assess how well they work.

## VI. RESULTS AND ANALYSIS

### A. Result Discussion

The result of different ML algorithms used are:

| Algorithm | mIOU |
|---|---|
| FCN | 80% |
| U-Net | 82% |
| DeepLabV3 | 84% |
| DeepLabV3+ (Xception) | 89% |

### B. Analysis

The results in the table make it evident that the DeepLabV3+ algorithm achieved the highest mIOU score. Our thorough testing shows that the suggested model outperforms cutting-edge methods, setting a new standard for performance on the Cityscapes and PASCAL VOC 2012 datasets. Mean Intersection-over-Union (mIOU) was employed in the evaluation as the accuracy metric. The "DeepLabV3+" model was used in the final product in order to accurately forecast fresh data.

## VII.CONCLUSION

Our proposed architecture, named "DeepLabv3+," uses an encoder-decoder structure for accurate semantic segmentation. The encoder module is based on DeepLabv3, which gathers rich contextual information, while the decoder module efficiently retrieves object boundaries. Depending on the available processing power, Atrous convolution can obtain the encoder properties at different resolutions. To further boost our model's effectiveness and speed, we compare it to the Xception model and use Atrous separable convolution.

Through extensive testing, we show that our suggested model, DeepLabv3+, offers state-of-the-art performance on well-known benchmark datasets including PASCAL VOC 2012 and Cityscapes. The experimental results confirmed the excellence and effectiveness of the model we propose in this study.

## REFERENCES

1. Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam, Google Inc., “Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation”
2. S. C. Yurtkulu, Y. H. Şahin and G. Unal, "Semantic Segmentation with Extended DeepLabv3 Architecture," 2019

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105157 |

27th Signal Processing and Communications Applications Conference (SIU), 2019.

3. Qashlim, Akhmad & Baba, Basri & Haeruddin, Haeruddin & Ardan, Ardan & Nurtanio, Ingrid & Ilham, Amil A. (2020). Smartphone Technology Applications for Milkfish Image Segmentation Using OpenCV Library. International Journal of Interactive Mobile Technologies (iJIM).

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor: 8.379

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

## IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details