Video Segmentation Using Deep Learning

Project report submitted in partial fulfilment of the requirement for the degree of Bachelor of Technology

in

Computer Science and Engineering/Information Technology

By

Arjun Seth (191407),
Anshul Jaiswal (191411)

Under the supervision of

Dr. Pardeep Garg & Dr. Rakesh Kanji

to

Department of Computer Science & Engineering and Information Technology

Jaypee University of Information Technology
Waknaghat, Solan-173234, Himachal Pradesh

Candidate's Declaration

We hereby declare that the work presented in this report entitled **"Video Segmentation Using Deep Learning"** in partial fulfilment of the requirements for the award of the degree of **Bachelor of Technology** in **Computer Science and Engineering/Information Technology** submitted in the department of Computer Science & Engineering and Information Technology**,** Jaypee University of Information Technology, Waknaghat is an authentic record of my own work carried out over a period from July 2022 to May 2023 under the supervision of **Dr. Pardeep Garg** (Assistant Professor (Senior Grade), Department of Electronics and Communication Engineering) and **Dr. Rakesh Kanji** (Assistant Professor (Senior Grade), Department of Computer Science and Engineering).

We also authenticate that we have carried out the above-mentioned project work under the proficiency stream **Data Science.**

The matter embodied in the report has not been submitted for the award of any other degree or diploma.

Arjun Seth, 191407

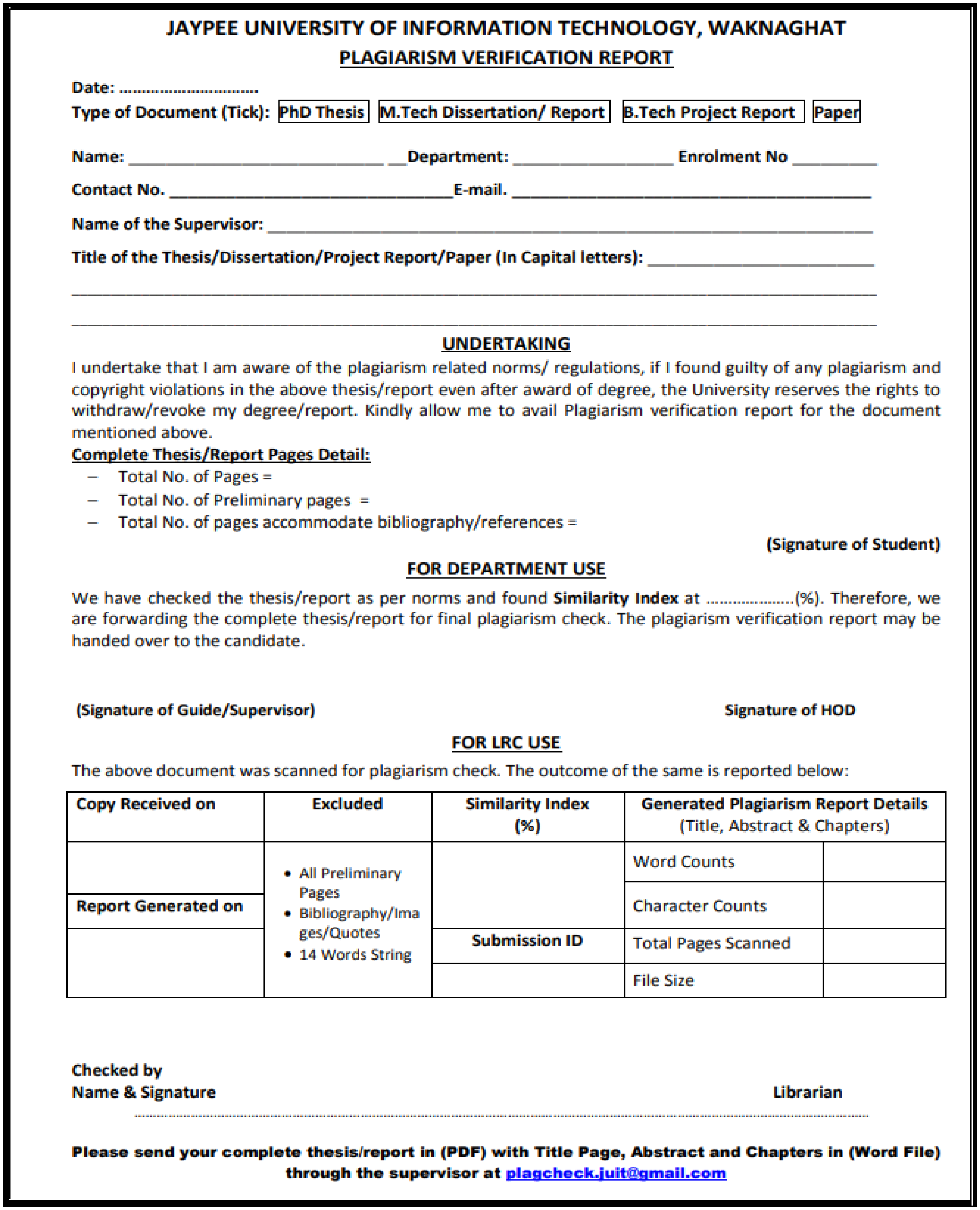
Anshul Jaiswal, 191411

This is to certify that the above statement made by the candidate is true to the best of my knowledge.

Dr. Pardeep Garg
Assistant Professor (Senior Grade)
Department of Electronics and Communication Engineering
Dated:

Dr. Rakesh Kanji
Assistant Professor (Senior Grade)
Department of Computer Science and engineering
Dated:

Plagiarism Certificate



JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

PLAGIARISM VERIFICATION REPORT

Date: ....

Type of Document (Tick): □ **PhD Thesis** □ **M.Tech Dissertation/ Report** □ **B.Tech Project Report** □ **Paper**

Name: ___________________________ __**Department:** _________________ **Enrolment No** _________

Contact No. ______________________________**E-mail.** ______________________________________

Name of the Supervisor: ____

Title of the Thesis/Dissertation/Project Report/Paper (In Capital letters): ________________________

____

____

UNDERTAKING

I undertake that I am aware of the plagiarism related norms/ regulations, if I found guilty of any plagiarism and copyright violations in the above thesis/report even after award of degree, the University reserves the rights to withdraw/revoke my degree/report. Kindly allow me to avail Plagiarism verification report for the document mentioned above.

Complete Thesis/Report Pages Detail:

– Total No. of Pages =
– Total No. of Preliminary pages =
– Total No. of pages accommodate bibliography/references =

(Signature of Student)

FOR DEPARTMENT USE

We have checked the thesis/report as per norms and found **Similarity Index** at ....(%). Therefore, we are forwarding the complete thesis/report for final plagiarism check. The plagiarism verification report may be handed over to the candidate.

(Signature of Guide/Supervisor) **Signature of HOD**

FOR LRC USE

The above document was scanned for plagiarism check. The outcome of the same is reported below:

Copy Received on	**Excluded**	**Similarity Index (%)**	**Generated Plagiarism Report Details** (Title, Abstract & Chapters)	
	• All Preliminary Pages • Bibliography/Images/Quotes • 14 Words String		Word Counts	
Report Generated on			Character Counts	
		Submission ID	Total Pages Scanned	
			File Size	

Checked by
Name & Signature **Librarian**

....

Please send your complete thesis/report in (PDF) with Title Page, Abstract and Chapters in (Word File) through the supervisor at **plagcheck.juit@gmail.com**

Acknowledgement

We are thankful to all the people who joined as part of making this journey of fulfilling this project into a working model. We are grateful to Gujarat Technological University for giving us a wonderful platform for exploring our software developing skills during the making of this project.

We also express our heartfelt gratitude to Dr. Vivek Sehgal, Head of the Department, for rendering all possible help and support during the development, implementation and presentation of the project.

We are also thankful to our mentors Dr Pardeep Garg as well as other staff members of the Computer Science and Engineering department, Jaypee University of Information Technology for their constructive and helpful inputs.

Table of Contents

List of Figures

Abstract

One of the key areas of visual signal processing research is now video segmentation. The goal of Applications of Video Segmentation is to cover the theoretical frameworks, practical applications, and emerging techniques of video analysis and computer vision communities while presenting the most recent developments in methods for segmenting and analysing videos. Pattern recognition in computer vision, and the transformation of digital material, segmentation is crucial. Segmenting images and videos are a job that arises in a variety of application domains, including interpreting images, analysing videos, and comprehension, digital entertainment, video indexing and summarising. The challenge of image/video data has been segmented developed into a basic issue during the past two decades, having a substantial both new pattern recognition, algorithms and implementations. Foreground object recognition and segmentation in videos is a vital task issue in computer vision research and a crucial element for many different uses. encompassing more serious vision issues like semantic scene comprehension and video summary, to basic video editing, the whole range of video-related editing tools, and video segmentation tasks. This varied group of applications achieves various goals and imposes various demands on the level of effectiveness, quality, and physical labor needed.

Chapter 1: Introduction

1.1 Introduction

Every day millions of individuals per day worldwide produce enormous amounts of video data, which are then made available online to the general public. Users typically correlate labels with this vast amount of visual information to denote content and location. Although these noisy labels are a type of weak annotation that is effective for some supervised machine learning tasks like scene categorization and action detection, they do not provide enough context to take use of the rich spatiotemporal data that films represent. Away from the noisy scene categorization labels, dense, per-pixel exact hand annotations of videos are at the opposite extreme of the spectrum (Figure 1.1). For instance, in the context of self-driving, this form of annotation permits a greater level of visual scene interpretation that is necessary automobiles and video surveillance In addition to providing separate processing of various image regions, pixel-wise annotations are widely used in the media content post-production pipeline. Large-scale datasets are essential for enabling rapid advancement in computer vision, as shown by recent success on the tasks of object recognition and detection in still images [Lin et al., 2014; Russakovsky et al., 2014; Torralba and Efros, 2011]. The main distinction between static photos and videos is the motion of the objects, which is a major driver of visual attention. In other words, some things' mobility may make them stand out more than others. Techniques for segmenting videos may be roughly divided into over-segmentation and object segmentation depending on the manner of grouping. The former seeks to gather compact sections of a video that are perceptually similar, whereas the latter seeks to group pixels that are a part of the same object instance. In addition to simple tasks like label propagation and video analysis, video object segmentation is essential for a wide range of sophisticated media content creation applications. In particular, segmenting video objects is crucial for post-production of special effects. Complex editing, such as compositing, needs independent processing of multiple aspects of a scene and video segmentation techniques might allow the artists to speed up their work flow. Video object segmentation, however, continues to be a difficult problem despite significant advancement in recent years, and the majority of approaches still show to many substantial restrictions on effectiveness and quality used inn real-world applications, such as video editing and post-production for the visual effects sector.

1.2 Problem Statement

The main problem face with videos is that to understand the particular scenario when there are multiple objects captured or present in a particular frame. As per the research a trained human mind can only deal with 10 to 15 frames a day but in most of the scenario there are more than that for example a traffic camera video consists of multiple vehicles at the same time some of them follows the traffic rule and some of them not and to understand that a semiautomatic video segmentation system can perform efficiently.

1.3 Objectives

This project's primary goal is to segment that is mark the important attributes of the particular frame of the video using image segmentation with the help of Mark-CNN which leads us to better understanding of the video frame by frame. With the help of image/video segmentation we can perform many different applications like Medical Imaging, Traffic Control System, etc.

1. Self-Driving Cars: Self-driving automobiles can utilize image segmentation to make clear distinctions between various objects. be it traffic lights, billboards, people, or automobiles. It may enable the algorithm that generates driving instructions to more accurately evaluate the environment before producing the next instruction.

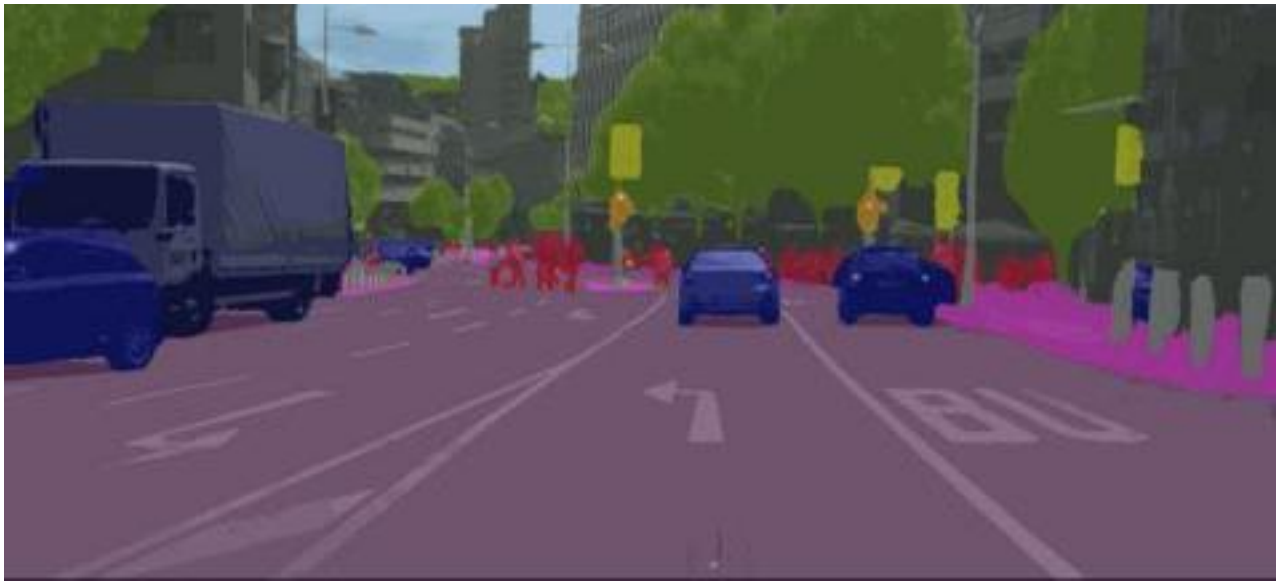


Fig 1.1 Self Driving Cars

2. Face Detection: Today, most smartphone cameras feature portrait mode, as has been noted. Technically speaking, portrait mode is a result of image segmentation. Aside from this, when faces can be distinguished from distracting things, security surveillance will be much more successful.

3. Circuit Board Defect Detection: Defective products must be the responsibility of the manufacturer. A lot of money and time can be saved on mending a defective gadget if a camera supported by an Image Segmentation model continuously scans for problems produced in the finished product.

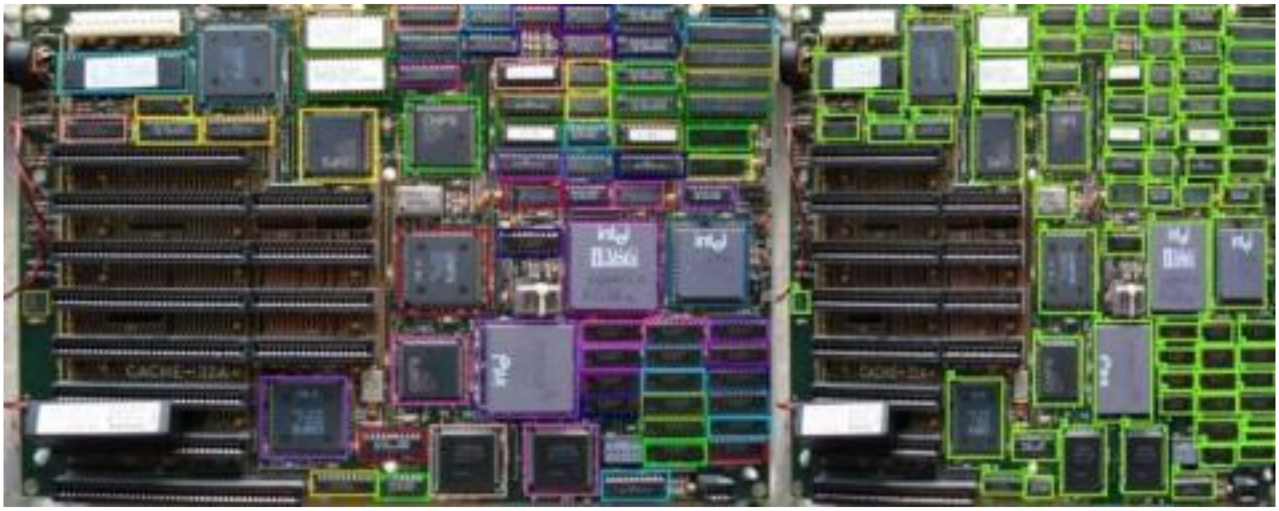


Fig 1.2 Circuit Board Defect Detection

4. **Medical Imaging:** Clinically pertinent information can be extracted from medical reports using image segmentation. For instance, malignancies can be divided up using picture segmentation.

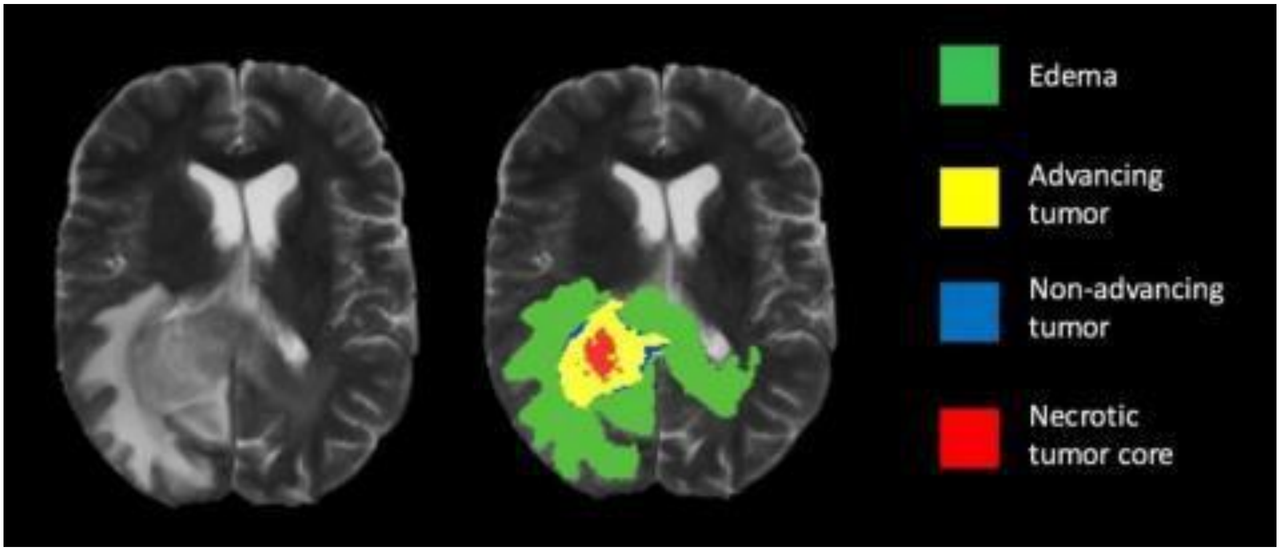


Fig 1.3 Medical Imaging

1.4 Algorithms Used

We're going to use the Mask R-CNN architecture to segment images or frames of any particular video. The Faster-CNN Model, which is favoured for object identification tasks, is extended by this model. In addition to the class label and object bounding box, the Mask R-CNN returns the binary object mask as well. In other words, we can say that Convolutional Neural Networks (CNNs) like the Mask RCNN are state-of-the-art fore segmenting pictures. This particular variant of deep neural networks recognizes produces a better segmentation mask for each of the objects in an image. In terms of segmenting pixels, Mask-CNN is effective. Explained below: We first need to comprehend what a CNN is and how it functions in order to comprehend the distinctions between Faster-CNN and Mask-CNN, and RCNN.

CNN (Convolutional Neural Networks):

An artificial neural network called a convolutional Image identification is performed using neural networks (CNNs), which are created particularly to handle pixel input. and processing, because of this, the essential and fundamental convolutional neural networks (CNN segmentation) serve as the foundation for the computer vision job of picture segmentation.

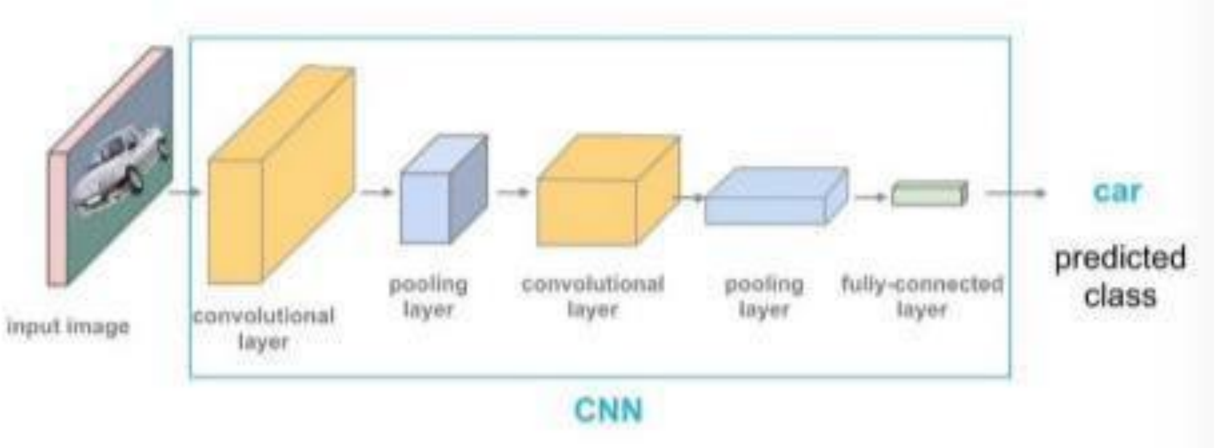


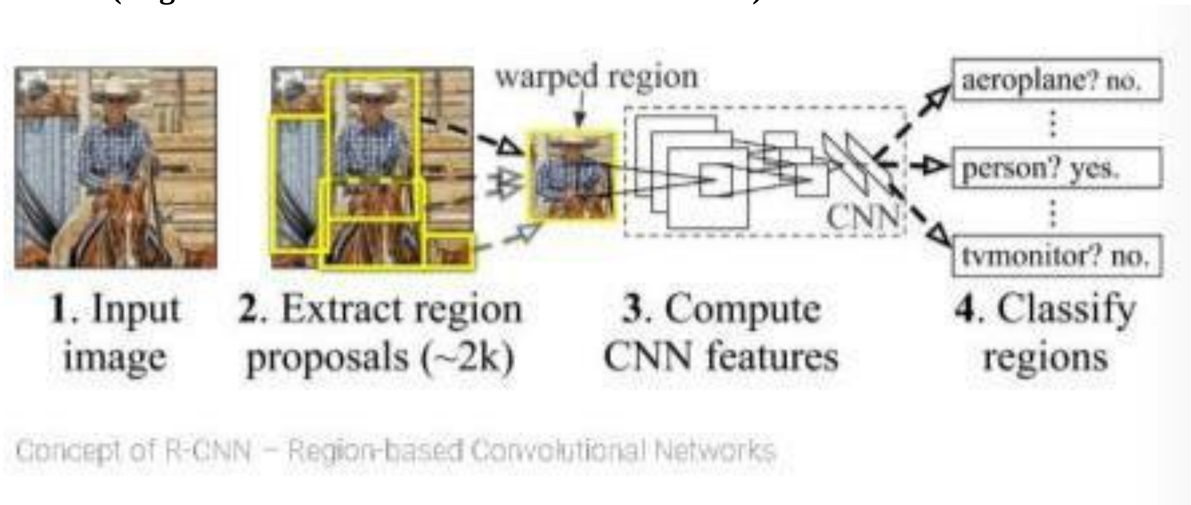
Fig 1.4 CNN

Three primary layers make up the Convolutional Neural Network Architecture:

- Convolutional layer: Using filters and kernels, this layer aids in abstracting the input image as a feature map
- Pooling layer: By summarizing the presence of characteristics in different patch feature maps, this layer aids down sampling in feature maps. Each neuron in a layer is connected to every other neuron in the layer below by completely integrated layers.
- Fully-connected layer: each layer's individual neurons is coupled for each neuron in the other layer by fully connected layers.

A straightforward CNN architecture isn't the best option in situations where there are several objects in the image. MaskR-CNN is the cutting-edge architecture that is based on R-CNN and is designed for certain situations (also referred to as RCNN).

R-CNN (Region-Based Convolutional Neural Networks):



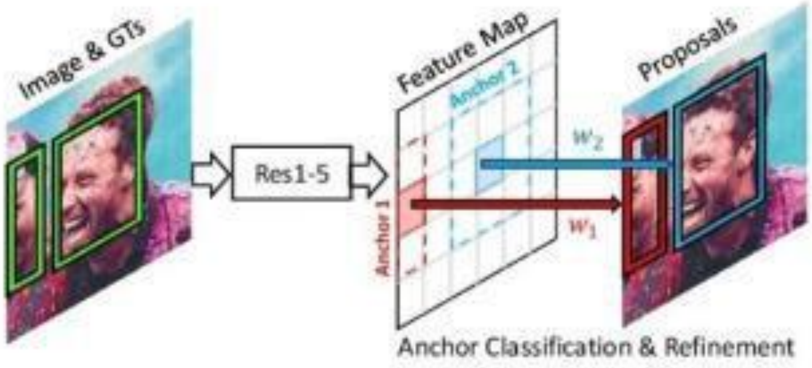
Concept of R-CNN – Region-based Convolutional Networks

Fig 1.5 Concept of R-CNN

A king of machine learning model called R-CNN. Sometimes referred as RCNN, is utilised for Computer vision applications, notably for object detection. Region-Based Convolutional Neural Network is known as R-CNN. Next, in order to understand what the RCNN architecture is , we Look at it.

A regional CNN (R-CNN) is depicted in the image below as an example. Convolutional networks are then assessed independently on each Region of Interest (ROI) to categorise various picture areas into the indicated category after bounding boxes are utilised to create object regions. The RCNN architecture was designed to deal with image detection issues. Also based on the R-CNN architecture that was improved to create Faster R-CNN is Mask R-CNN.

Faster R-CNN:



Concept of Region proposal Networks (RPN)

Fig 1.6 Concept of RPN

An enhanced variant of R-CNN architectures with two stages is called Fast R-CNN. RPN (Region Proposal Network): RPN is only such a neural network suggests various items that are present in a specific picture.

Faster RCNN: Each candidate box is used to extract features using ROI Pool (Region of Interest Pooling), ands then classification and bounding-boxes regressions are carried out. From each ROI in detection, ROI Pool is a procedure that extracts a tiny feature map.

By using a Fast R-CNN with Region Proposal Network design to understand the attention mechanism, Faster RCNN advances this stream. Since there is no need to continuously feed the convolutional neural network 2000 region suggestions, "FastR-CNN" is more rapid than RCNN. An alternate is a feature map produced because of the convolution operation, which is only performed each picture once.

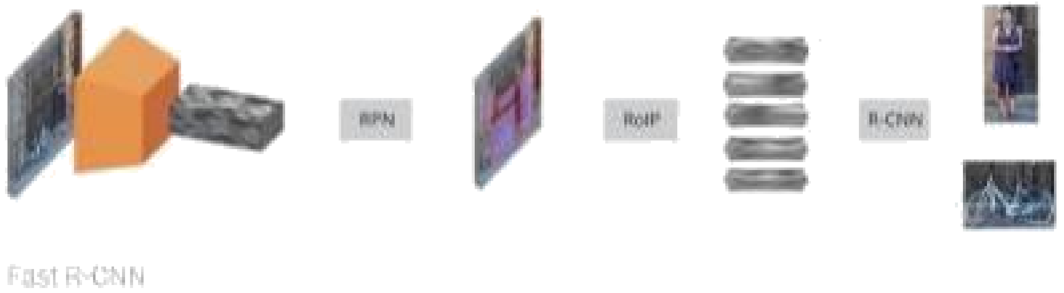


Fig 1.7 Fast R-CNN

Mask R-CNN:

MaskR-CNN is the most sophisticated Convolutional Neural Network (CNN) for image segmentation, for example. Region based convolutional neurals networks, faster R-CNN, laid these starting point for the development of MaskR-CNN.

The initial step in comprehending Working with MaskR-CNN needs comprehension of the idea segmentation of images. The task of computer visions the method of dividing a digital picture into several parts is called image segmentation (Set of pixel, often considered as picture objects). In this segmentation, borders and objects are located (lines, curvatures, etc.).

The two main types of image segmentation covered by MaskR-CNN are:

- Semantic Segmentation
- Instance Segmentation

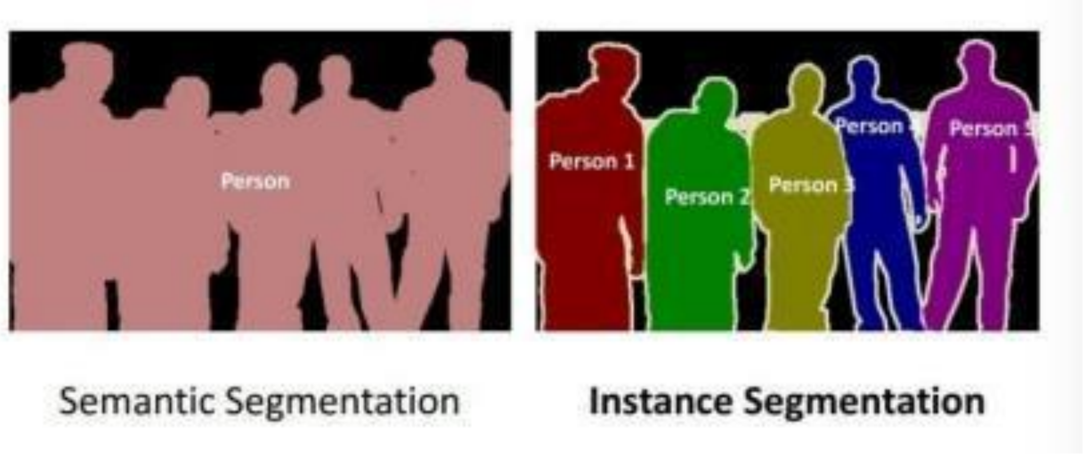


Fig 1.8 Types of Mask R-CNN

Semantic Segmentation: It assigns every single pixels to one of aa predetermined number off categories without making a distinction between different occurrences of the same item. To put it another way, semantic segmentation works at the pixel level to identify and classify comparable objects into a single category. Everything was brought together as seen in the aforementioned photo (person). The process of semantic segmentation, also known as backdrop segmentation, separates the image's themes from its background.

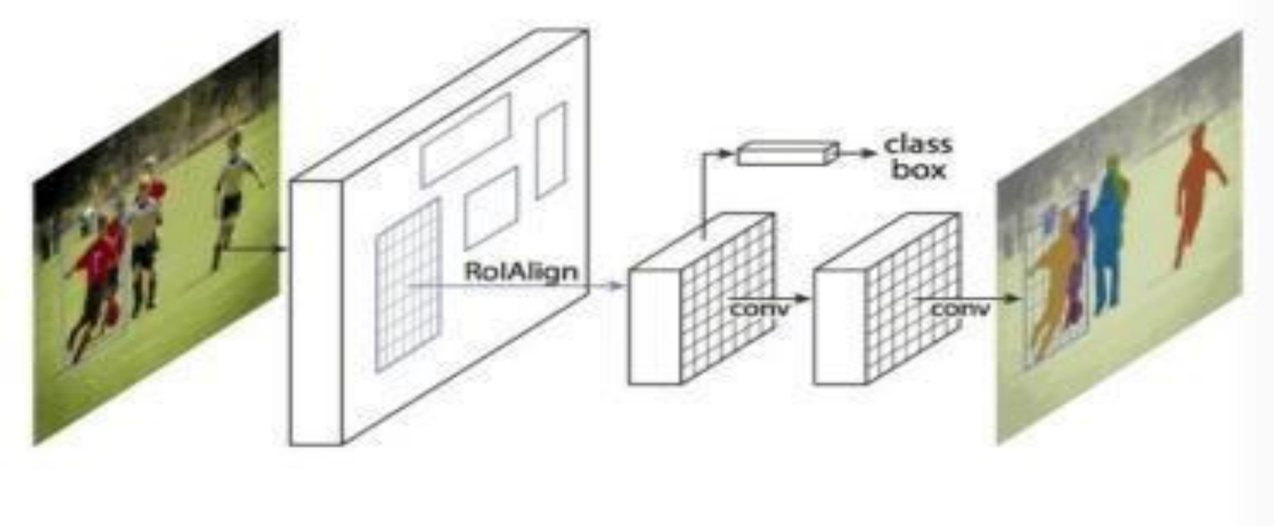
Instance segmentation: It is also known as instance-recognition, is the process of accurately recognising each item in a picture while simultaneously precisely segmenting each instance. As a result, it integrates object localisation, object detection, and object categorization. In other words, this type of segmentation goes a step further to make it evident that each item that is identified as a similar instance is distinct from the others. As seen in the example image above, all objects for the segmentation example are humans; nevertheless, throughout the segmentation process, each person is regarded as a separate entity. Semantic segmentation, also known as foreground segmentation, emphasises the topics of the image rather than the backdrop.

Working of Mask R-CNN:

Mask R-CNN was developed using Faster R-CNN. While Mask R-CNN adds a third branch that outputs the object mask, FasterR-CNN outputs a classes labels and a bounding-box offset for each candidates items. Since it is different from the class and box outputs, the additional mask output necessitates the extraction of a significantly more precise spatial arrangement of an object.

By concurrently adding a branch for predicting an object mask (Region of Interest) and the branch for bounding box detection, CNN's Mask R-CNN improves upon the capabilities of faster R-expansion.

since it differs from the outputs of the class and box, of an object. By concurrently adding a branch for predicting an object mask (Region of Interest) and the branch for bounding box detection, CNN's Mask R-CNN improves upon the capabilities of faster R-expansion.



Mask R-CNN – The Mask R-CNN Framework for Instance Segmentation

Fig 1.9 Framework of Mask R-CNN

Benefits of MaskR-CNN:

- Simplicity: It is straightforward to train a mask R-CNN.
- Performance: On every challenge, Mask R-CNN outperforms all currently available single-model entries.
- Efficiency: The technique is incredibly effective and just slightly increases the overhead of Faster R-CNN.
- Flexibility: It is simple to apply Mask R-CNN to various tasks. For instance, in the same framework, Mask R-CNN can be used to estimate human pose.

The pixel-to-pixel alignment is a crucial aspect of Fast/Faster R-CNN that Mask R-CNN does not have. Mask R-CNN (which is RPN) employs the same two-step procedure with a similar first stage. In addition to class and box offset predictions, Mask R-CNN also generates a binary mask for each ROI in the second stage. On the other hand, the bulk of contemporary systems rely on categorization using mask predictions. Additionally, it is simple to create and train Mask R-CNN thanks to the Faster R-CNN framework, which supports a wide range of customizable architecture designs. The mask branch also has a minimal computational cost, enabling speedy system development and testing.

1.4.1 Tools and Libraries used

Tools Required :

VS Code and **Google Colab** are the two IDEs required for this project.

VS Code : Microsoft created the source-code editor Visual Studio Code, generally known as VS Code, for Windows, Linux, and macOS using the Electron Framework. Debugging support, syntax highlighting, intelligent code completion, snippets, code refactoring, and embedded Git are among the features. Users can modify the theme, keyboard shortcuts, preferences, and add functionality by installing extensions. A source-code editor called Visual Studio Code works with many different programming languages, such as C, C#, C++, Fortran, Go, Java, JavaScript, Node.js, Python, and Rust. The Electron framework, which is used to create Node.js web apps that leverage the Blink layout engine, serves as its foundation. The "Monaco" editor component from Azure DevOps is also used in Visual Studio Code. It enables users to open one or more directories, which may subsequently be saved in workspaces for later usage in place of a project system. As a result, it can function as a language-neutral code editor for any language. It supports a wide variety of programming languages, each with its own set of capabilities.

Google Colab : Colab is a completely cloud-based Jupyter notebook environment that is free to use. The notebooks you create can be simultaneously modified by your team members, exactly like you edit documents in Google Docs, and most significantly, it doesn't require any setup. Many well-known machine learning libraries are supported by Colab and are simple to load in your notebook. A product of Google Research is Colaboratory, or "Colab" for short. Colab is particularly well suited to machine learning, data analysis, and education. It enables anyone to create and execute arbitrary Python code through the browser. Technically speaking, Colab is a hosted Jupyter notebook service that offers free access to computer resources, including GPUs, and requires no setup to use. Some of the impressive features of Google Colab notebook are :

1. Our notebook can be saved to our individual Google Drive.
2. We can import a specific notebook from Google Drive as well.
3. It offers cost-free cloud solutions.
4. Any notebook from github can be directly imported or published.
5. It is simple to combine with TensorFlow, PyTorch, or OpenCV.
6. From a notebook, we can run commands on the terminal.
7. Without any local setup, we can write and run Python3 codes.

Libraries Required :

Numpy, **Pandas, Scipy, sklearn** and **matplotlib** are the various libraries required in this project. All these libraries perform a particular task and are mandatory. Detailed explanations for each of them are provided below :

Numpy : Numerical Python is referred to as Numpy. It is an opensource library for the Python programming language, as implied by the name. Large, multi-dimensional matrices and arrays are now supported by Numpy, along with a vast array of cutting-edge mathematical operations that can be used on them. Its goal is to simplify the transformation of challenging functions for you or the calculation of certain data analysis. The primary benefit of Numpy is its speed. Compared to utilising the built-in Python functions, it is significantly faster. As an illustration, it enables you to quickly determine the mean and median of a dataframe using a single line of code for each. It contains various features including these important ones:

1. A robust multi-dimensional object of array.
2. C/C++ integrating tools for Fortran code.
3. Ability to create random numbers and useful linear algebra.
4. Size is fixed and multi-dimensional.
5. Fast array loops.
6. It uses primitive data types to hold numbers. A primitive data type simply denotes that the information is kept in byte form.

Pandas : Pandas is an open-source Python package that provides you with a very useful collection of tools for conducting data analysis. To improve your machine learning skills, you must learn Pandas. Data science, machine learning, and other applications employ it in addition to data analysis. Simply said, you will require Pandas if it uses data. It can speed up the loading, preparing, merging, joining, reorganising, processing, and changing of data. As previously mentioned, Pandas is a free and open-source package that makes it simple to utilise data structures and data analysis tools for Python. The core of Pandas are Data-Frame objects.

Pandas are highly effective methods of presenting data. This facilitates better understanding and data analysis. Better outcomes for data science initiatives are assisted by simpler data representation. Pandas are extremely powerful animals. They give you access to a wide range of essential tools and guidance that are used to quickly analyse your data. Pandas can be used for a number of tasks, such as filtering your data based on specific criteria or segmenting and splitting the data based on preferences.

Scipy : An SciPy is a Python library that is available for free and open source and is used for technical and scientific computing. SciPy includes modules for a variety of common tasks in science and engineering, including optimisation, linear algebra, integration, interpolation, special functions, FFT, signal and image processing, and ODE solvers.SciPy uses a multidimensional array as its fundamental data structure, which is made available through the NumPy module. There are some functions for linear algebra, Fourier transformations, and random number generation in NumPy, although they lack the generality of their SciPy counterparts. Additionally, NumPy functions well as a multidimensional data container that supports arbitrary datatypes. As a result, NumPy can quickly and easily interact with a wide range of databases.

1. Easy to understand and smooth learning curve.
2. Fast processing.
3. Open-source.
4. Powerful and interactive sessions with python.
5. Classes, and web and database routines for parallel programming.

Despite the fact that SciPy includes all of NumPy's capabilities, we still prefer NumPy when dealing with basic array concepts. Python software called SciPy is used. It functions slowly in spite of having a wide range of capabilities. We utilise SciPy to perform complex numerical calculations.

Sklearn : For the Python programming language, scikit-learn (formerly known as scikits.learn and also referred to as sklearn) is a free machine learning package. Support-vector machines, random forests, gradient boosting, k-means, and DBSCAN are just a few of the classification, regression, and clustering algorithms it offers. It is also built to work with Python's NumPy and SciPy scientific and numerical libraries. Scikit-learn is mostly designed in Python and heavily utilises NumPy for high-performance array and linear algebra operations.
Here we will discuss some of them.

1. Unlabeled data can be grouped using the sklearn clustering tool.
2. Several supervised learning results can be combined into one prediction using the Ensemble feature.
3. Scikit-learn can be used to test the accuracy and the validity of supervised models Unobserved data.
4. You can extract features from text and images using scikit-learn.
5. This algorithm collection includes factoring, cluster analysis, principal component analysis, and unsupervised neural networks.

Matplotlib : For the For the Python programming language and its NumPy numerical mathematics extension, Matplotlib is a plotting library. For integrating plots into programmes utilising all-purpose GUI toolkits like Tkinter, wxPython, Qt, or GTK, it offers an object-oriented API. It is not recommended to use the procedural "pylab" interface, which is based on a state machine (similar to OpenGL) and was created to closely mimic the MATLAB interface. Matplotlib is used by SciPy.

1. Create zoomable, pannable, and updateable interactive plots and figures.
2. Alter the layout, visuals, and diagrams.
3. Make use of the many third-party packages based on Matplotlib.

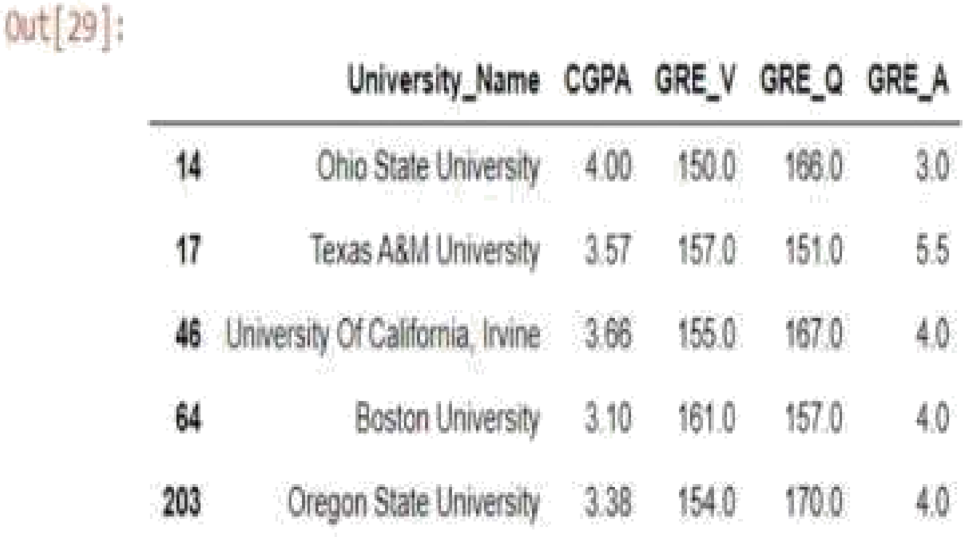
1.4.2 Steps performed :

1. Data collection is the first and most important phase in training our model. We are all aware that a machine first gains knowledge from the information we feed it with before making predictions based on its training. The caliber of the data you give the computer will determine how precise your model is. Inaccurate or outdated data will produce unreliable outcomes or useless predictions. In our instance, both undergraduate and graduate student data were scraped. Web scraping produced about 271807 rows of data, which are now being processed to create the final dataset..

2. Importing Libraries: The primary project file is then imported with libraries like Numpy, Pandas, Scipy, sklearn , and matplotlib. Each of these libraries carries out a certain function that is very important to the project.

3. Preparing the data : After you have your data, you have to prepare it. You can do this by :

- You should combine and randomize all of the available data. By doing this, it is ensured that the distribution of the data is balanced and that the ordering will not have an impact on the learning process.
- Removing extraneous information, duplicate and missing values, rows and columns, data type conversion, etc. from the data. It might even be necessary to modify the dataset's rows, columns, and index of rows and columns.
- To understand the structure of the data and the relationships between the numerous variables and classes it contains, visualise the data.
- Separating the cleaned data into two sets: a training set and a test set. The training set is the set from which your model learns. A testing set is used to evaluate the accuracy of your model after training.
- For the feature weighted algorithm to work in our situation, the data also needed to be normalised



Out[29]:

	University_Name	CGPA	GRE_V	GRE_Q	GRE_A
14	Ohio State University	4.00	150.0	166.0	3.0
17	Texas A&M University	3.57	157.0	151.0	5.5
46	University Of California, Irvine	3.66	155.0	167.0	4.0
64	Boston University	3.10	161.0	157.0	4.0
203	Oregon State University	3.38	154.0	170.0	4.0

Fig 1.10 Example of dataset

4. Selecting a model: A good machine learning model is what defines the results you get after running a machine learning algorithm on the collected data. It's essential to pick a model that works for the current task. Over the years, scientists and engineers have created numerous models that are appropriate for a variety of tasks, including speech recognition, image recognition, prediction, etc. Additionally, you must select whether your model works best with numerical or categorical data and then make the proper decision. He output you get after applying a machine learning algorithm to the gathered data is determined by a machine learning model. Selecting a model that is applicable to the current task is crucial.

5. Over the years, scientists and engineers have created numerous models that are appropriate for a variety of tasks, including speech recognition, image recognition, prediction, etc. Additionally, you must decide whether your model works best with numerical or categorical data and then make the appropriate decision. Assessing the model As soon as your model has been trained, you must evaluate its performance. To do this, the effectiveness of the model is assessed using novel data. The unseen data used is the testing set that you previously divided our data into.

6. After creating and evaluating your model, perform parameter tuning to improve its accuracy. By tweaking the model's parameter parameters, this is achieved. The variables that are normally selected by the programmer are the model's parameters. When a certain parameter value is reached, accuracy will be at its highest. Tuning parameters is the process of determining these values.

7. **Normalizing :** If we take a practical approach, our dataset has approximately fifteen columns, and not all of them can be equally important to a student attempting to get into a university. We have given these criteria greater weights because any university would likely place more emphasis (weight) on a student's GPA or CGPA. This will guarantee that, in comparison to other criteria, these factors will have a greater influence on the final list of colleges. Assume that F1 and F2 are the features and W1 and W2 are the weights. The formula can be used to determine the similarity score. : **W1* F1+W2*(1-F2)**.

8. Dividing dataset into training and testing data : To evaluate how well machine learning algorithms work in conjunction with prediction-based techniques and applications, the train-test split is used. Using this quick and easy method, we may compare the output of our own machine learning model to that of other machines. By default, 70% of the real data is included in the training set, while only 30% is included in the test set.

We must separate a dataset into train and test sets in order to evaluate how effectively our machine learning model performs. Since the train set's statistics are available, they are used to adjust the model. The second set of data, the test data set, is only used to make predictions. The scikit-learn library, also referred to as sklearn, is the most efficient and trustworthy Python machine learning package. The model selection module of the Scikit-Learn package contains the splitter function train test-split().

9. **Training the model :** The most crucial phase of machine learning is training. To detect patterns and generate predictions, you feed the prepared data to your machine learning model during training. As a result, the model gains knowledge from the data and completes the task

10. Applying the algorithms : The K-Nearest-Neighbors and Feature Weighting algorithms can now be used on our model. The K-NN algorithm places the new instance in the category that resembles the current categories the most, presuming that the new case and the previous cases are comparable. After storing all the previous data, a new data point is classified using the K-NN algorithm based on similarity.

When compared to other features, Feature Weighting will make sure that these criteria are given a larger voice in selecting the final list of institutions. This is made possible via normalisation, which aids in calculating the similarity score. When considering how to use a set of parameters in the best possible way to achieve the best possible result, feature weighting is the most straightforward and logical idea that comes to mind (here features). In real life, a select few chores are always prioritised over others. The same is true for this undertaking.

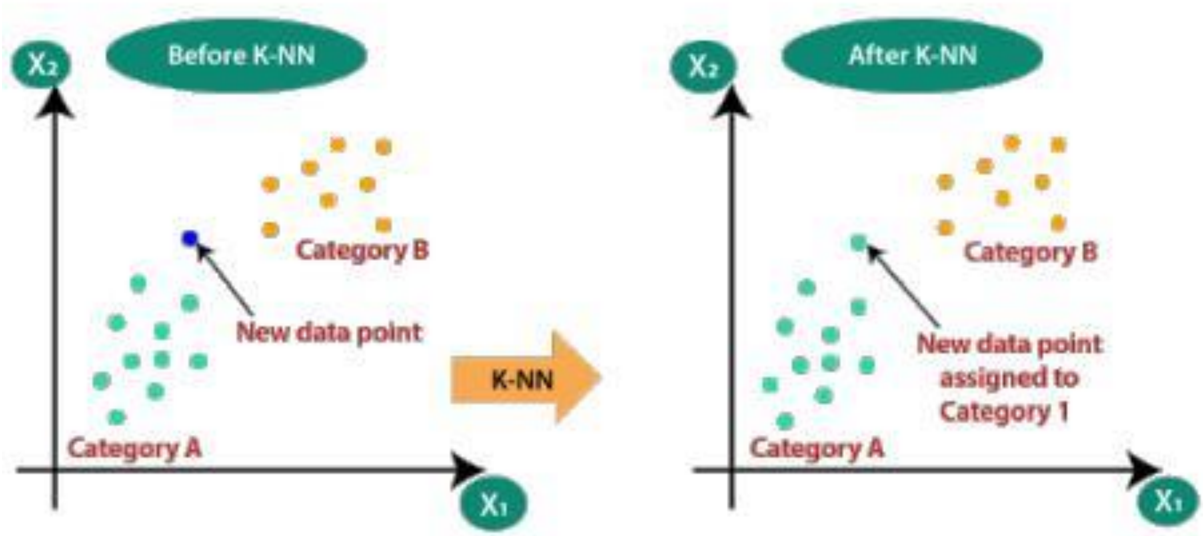


Fig 1.11 Application of algorithm

6. Making predictions and testing accuracy :

Model accuracy is defined as the ratio of the total number of predictions made to the total number of classifications a model correctly predicts. Businesses use machine learning models to make pragmatic decisions, and more accurate model results result in better decisions. Model correctness is essential for firms because a mistake could significantly harm their operational system. Even though increasing a model's accuracy doesn't always lead to greater revenues, doing so is typically beneficial. For instance, a patient who obtains an incorrect sickness diagnosis may cause problems for the hospital as well as the patient. Businesses and techies spend money on precise modelling to reduce stress, time, and costs.

Overall Methodology followed :

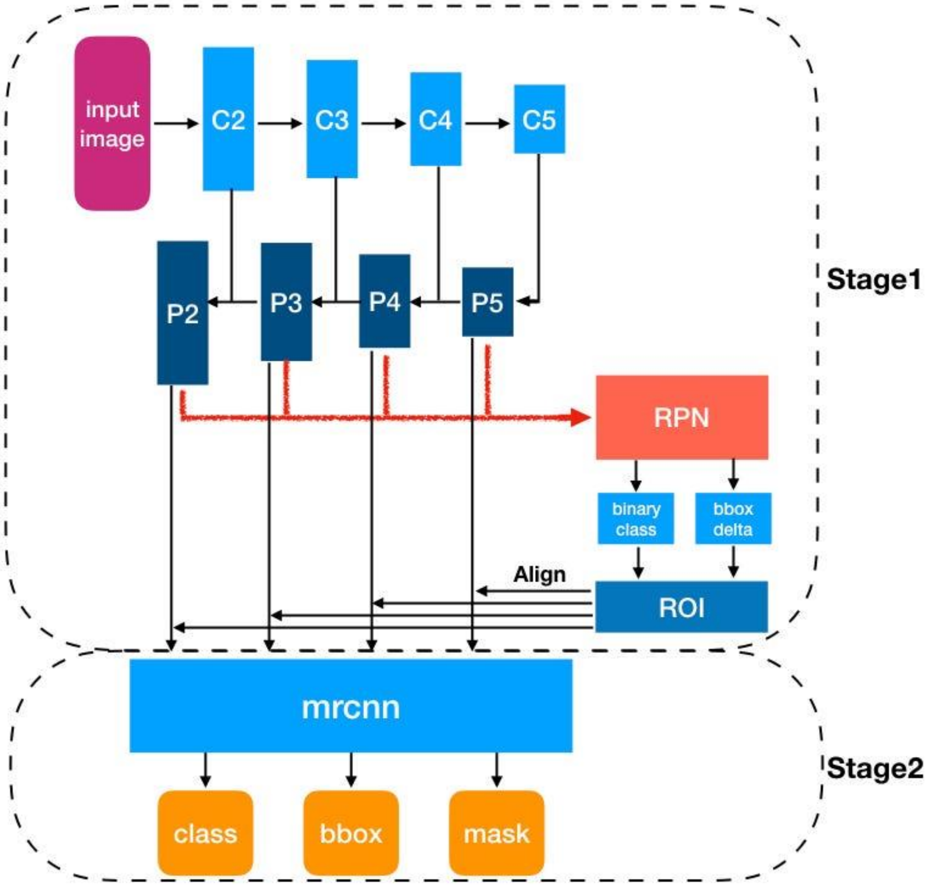


Fig 1.12 Architecture of Mask R-CNN

Chapter 2: Image segmentation and video segmentation

IMAGE SEGMENTATION:

In a picture, each pixel is individually identified and labelled during the process off image segmentation, and pixels with the same label have certain properties. It has granular object detection capabilities and can also determine the shape of an object. It is a more sophisticated and precise method of detecting an object's edge and contour.

A picture is divided into various parts, or segments, through image segmentation. A mask or a labelled image serves as the representation for this group of segments. In this manner, we may process only the crucial portions rather than the complete image. Several distinct fields employ image segmentation. The medical profession is one of the best examples. A medical image's segmentation can be utilized to identify damaged cells. Sometimes, in order to process an object further, we not only need to know its position but also its shape. If we can locate cancer cells at an early stage, cancer may be curable. Consequently, we can swiftly identify cancer cells using picture segmentation, which can help save countless lives.

USING IMAGE SEGMENTATION IN PRACTICE:

- Imaging In Medicine
- Computer Vision
- Detecting Circuit Board Defects.
- Traffic Management Program.
- Autonomous Vehicle.
- Video Monitoring.

TYPES OF IMAGE SEGMENTATION: Mainly there are three types of image segmentation: -

- Semantic Segmentation:
- Instance Segmentation:
- Panoptic Segmentation

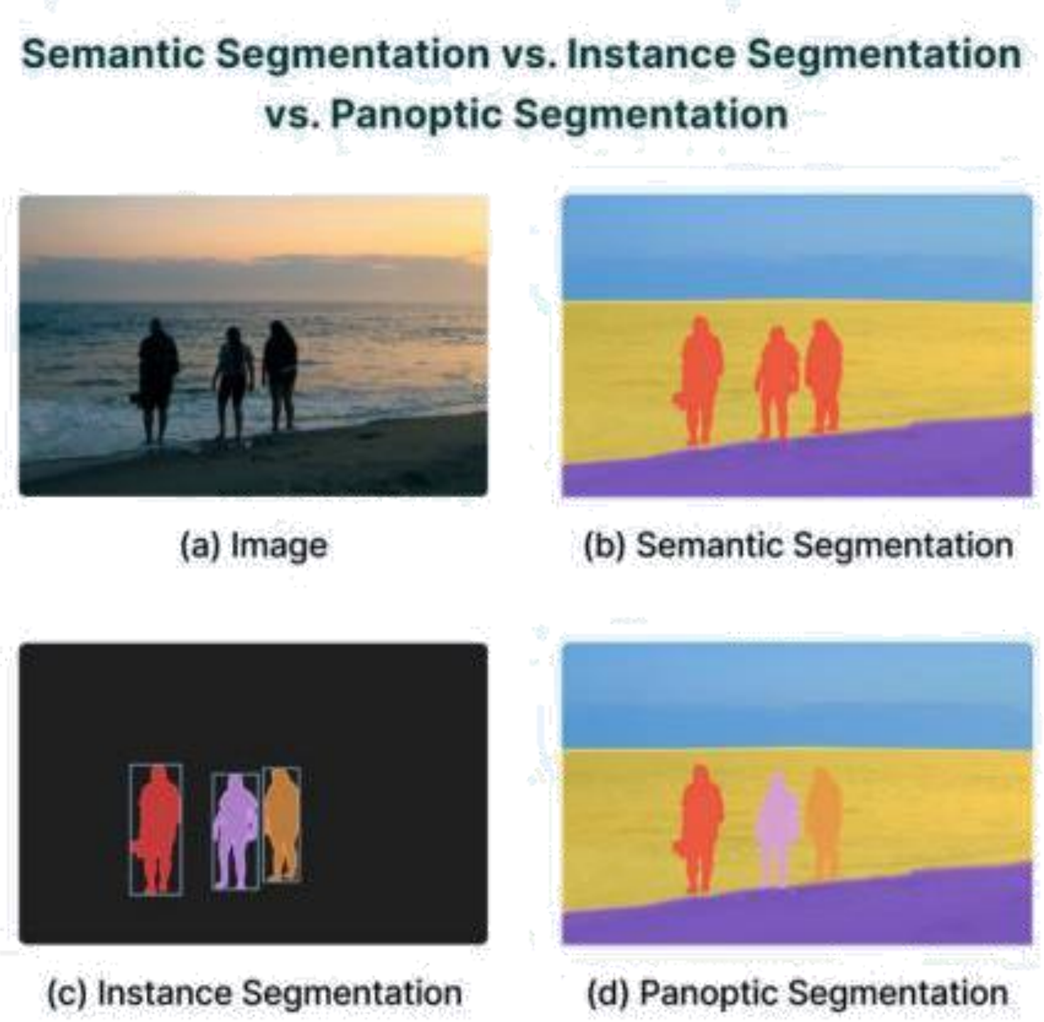

Semantic Segmentation vs. Instance Segmentation
vs. Panoptic Segmentation
(a) Image
(b) Semantic Segmentation
(c) Instance Segmentation
(d) Panoptic Segmentation

Fig 2.1 Different types of Segmentation

SEMANTIC SEGMENTATION:

The grouping procedure together segments of a picture that belong to the same object class are known as semantic image segmentation, often referred to as pixel-level categorization. Picture level categorization and detection are two additional important image tasks. Treating every image as belonging to the same category is classification. Object localization and identification are examples of detection. Because image segmentation assigns each pixel to a category, it can be thought of as pixel-level prediction. Additionally, there is a task called instance segmentation that combines segmentation and detection. Road sign detection, colon crypt segmentation, and the classification of land use and land cover are just a few examples of the many uses for semantic picture segmentation.

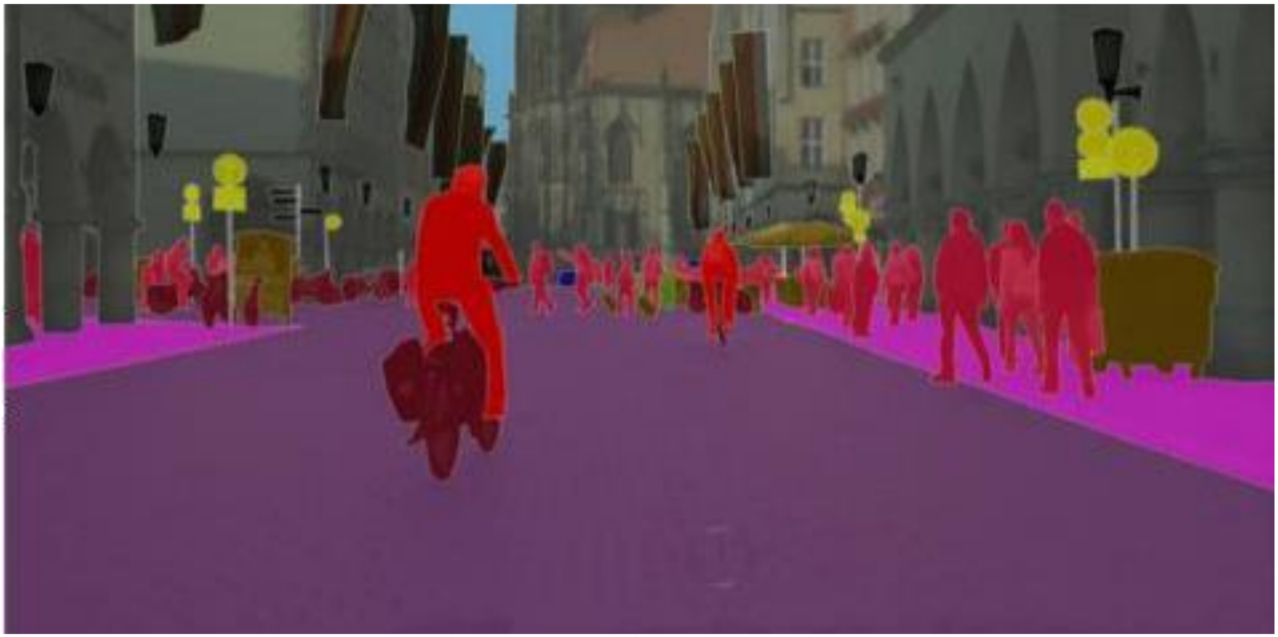


Fig 2.2 Semantic Segmentation

EVALUATION PARAMETRICS:

Regular performance evaluation measures for scene parsing and picture segmentation include frequency weighted IU (FWIU), area intersection following union (IU) MIU, and pixel accuracy (PACC), among others. Let ti=jnij depict the class's amount of pixels I where ncl are present distinct classes, and let nij represent the quantity of pixels in class I that wore properly expected to be of class j .The four metrics are each described as follows.

RECENT DNN IN SEGMENTATION:

The biological neurons that are the basis of artificial neural networks (ANN). Artificial neurons are the foundational component of ANN. Each artificial neuron receives inputs that are weighted and added together. A scale quantity is output by the neuron., which is A transfer function or activation function comes next. Figure below depicts a neural model as an illustration.

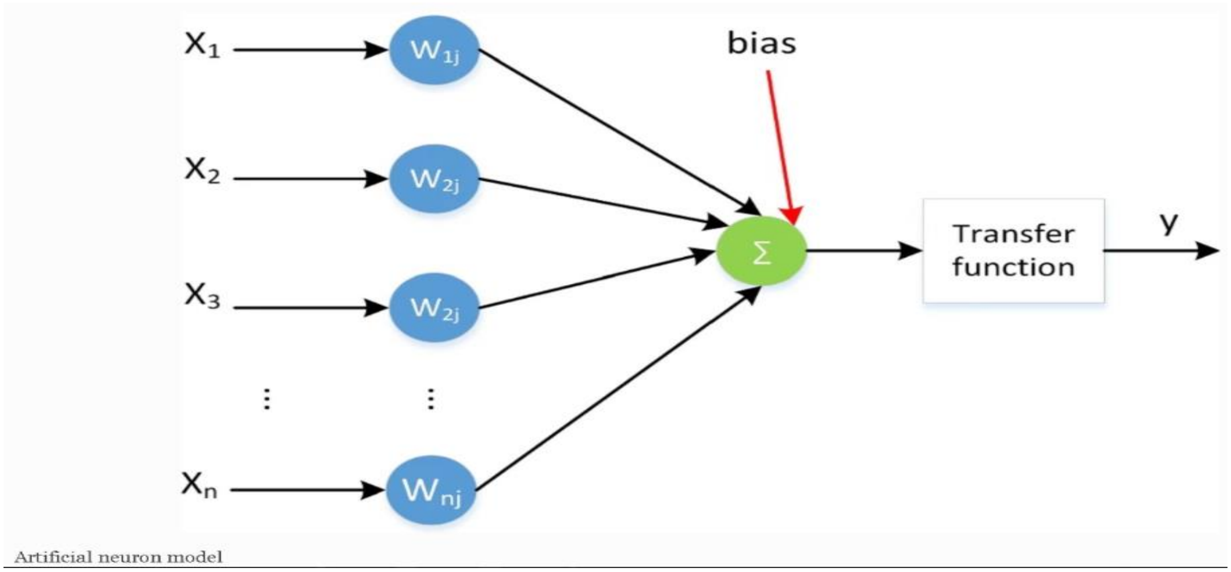


Fig 2.3 DNN in Segmentation

Different ways of stacking artificial neurons result in a variety of artificial neural networks (ANNs), including the Auto-encoder , Restricted Boltz-man Machine (RBM), LongShort Term Memory (LSTM), Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), and others. The fundamental architecture is shown in Fig. below.

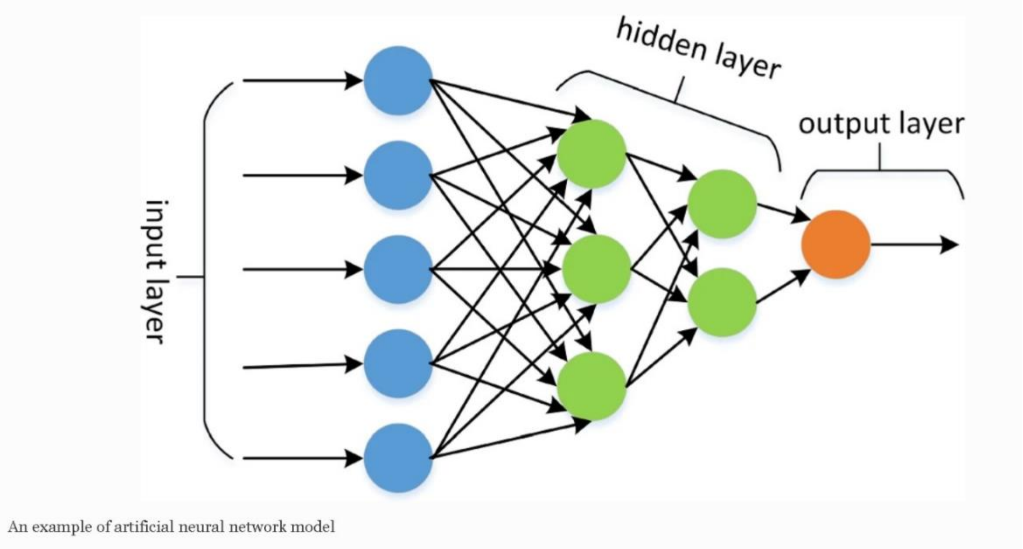


Fig 2.4 Example of ANN

INSTANCE SEGMENTATION:

Locating instances of items and establishing their limits are the focus of instance segmentation, an advanced technique for image segmentation. Large-scale applications can be made for self-driving cars, medical imaging, aerial crop monitoring, and other real-world scenarios.

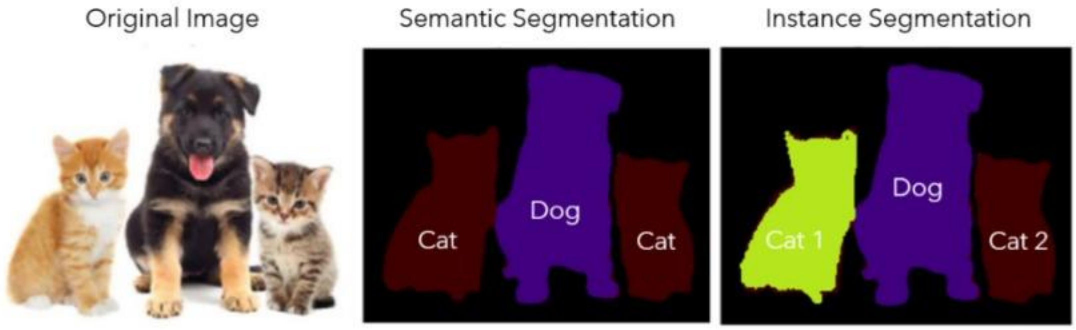


Fig 2.5 Instance Segmentation

When there are various items of a similar type present and they need to be monitored independently, we find instance segmentation to be especially helpful.

By combining Object identification and semantic segmentation (finding all occurrences of a category in a picture), instance segmentation adds the capability of distinguishing distinct instancees of any given segment classes to the basic segmentation task.

Compared to both semantic segmentation networks and object detection networks, instance segmentation generates output formats that are richer. For semantic segmentation, FCNs (Fully Convolutional Networks) have been employed extensively.

The same image pixel obtains the same responses (therefore classification scores) regardless of its relative position in the environment, despite convolutional networks' inability to do instance segmentation, which involves the detection and segmentation of unique object instances.

The same pixel can have various semantics in different regions, which means that Semantic Segmentation must function at the region level in order for Instance Segmentation to be effective. This cannot be modeled by a single FCN on a single image.

The likelihood score of "the pixel belongs to some category" is predicted for each pixel in conventional FCNs by a classifier that has been taught to do so.

To introduce the translation variant property, we employ k2 position-sensitive score maps that react to kxk equally partitioned cells of the object. The possibility of "the pixel belonging to some object instance at a relative position" is now represented by each score.

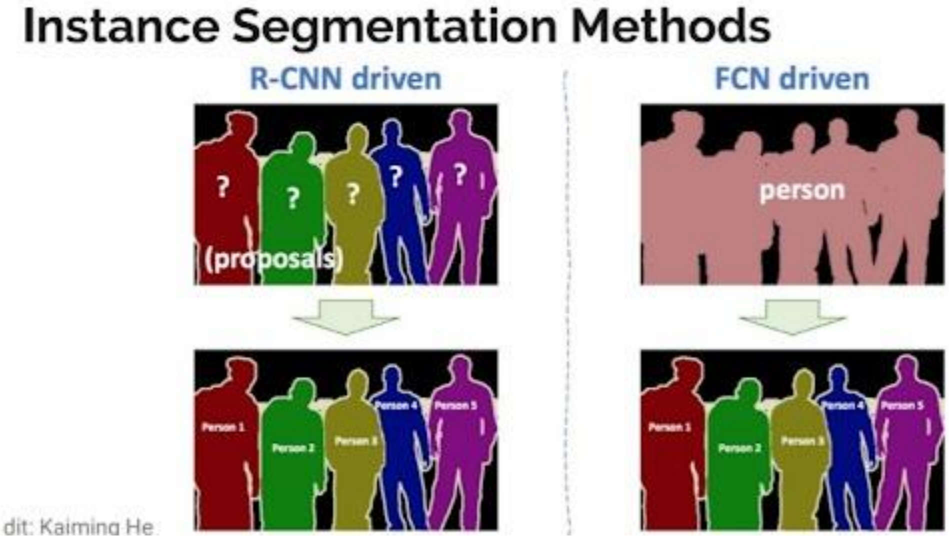


Fig 2.6 Methods of Instance Segmentation

For instance, a person-related pixel can serve as both the background and foreground for another person in different semantic positions (instance). Currently, object detection and segmentation are done in concert and at the same time.

The most recent model for instance segmentation, with three outputs, is Mask R-CNN. Similar to Faster R-CNN, it has a class label and bounding box offset. A third branch produces the object mask, which requires the extraction of a much finer spatial layout of an object .

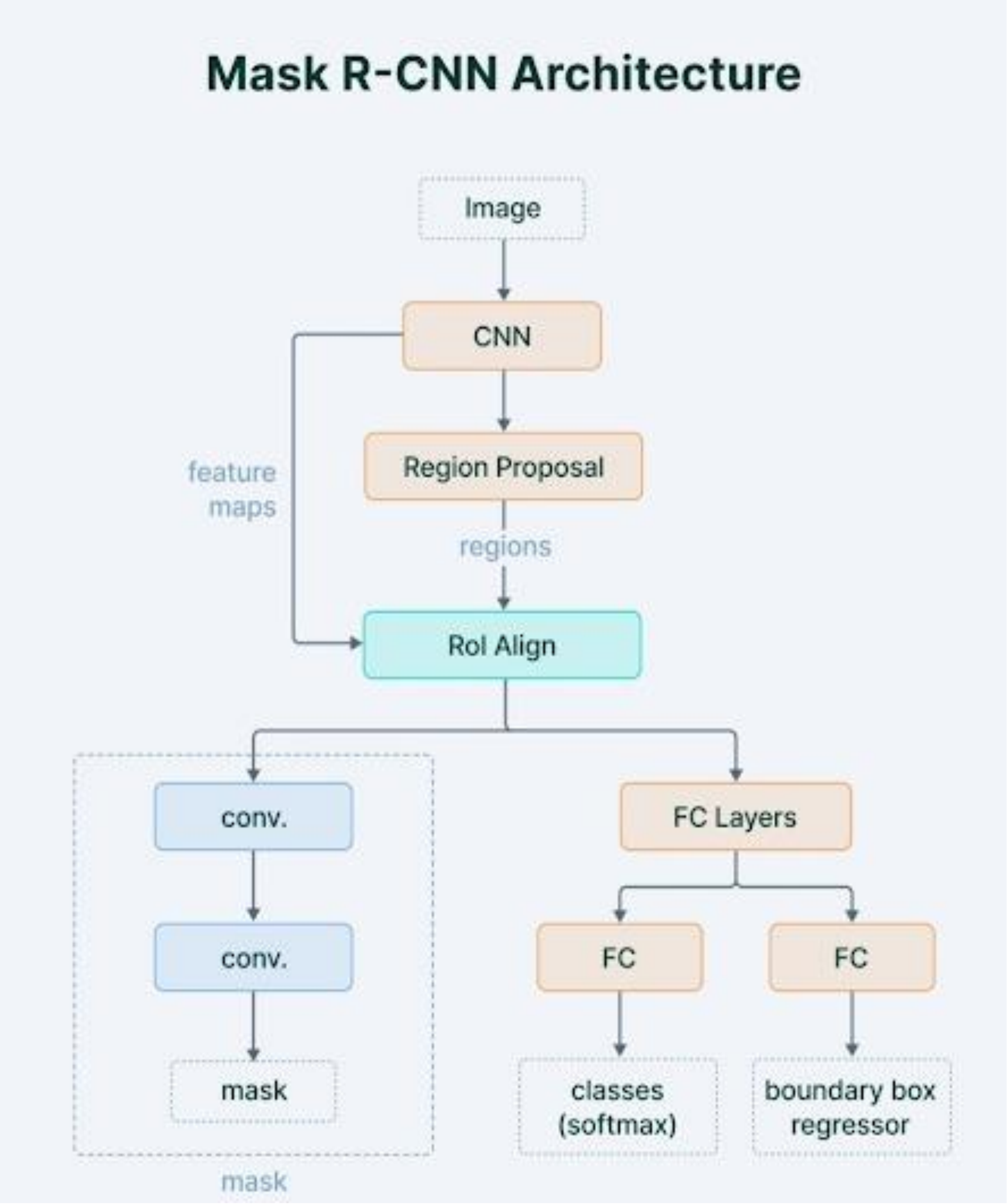


Fig 2.7 Architecture of Mask R-CNN

PANOPTIC SEGMENTATION:

Pan and optic are the roots of the word panoptic.

Both pan and optic refer to vision. Thus, the term "panoptic segmentation" essentially translates to "everything observable in a specific visual field."

Panoptic segmentation in computer vision can be accomplished in three easy steps

- Dividing each picture item into distinct portions that stand alone from one another.
- Labeling involves painting a different color on each distinct component.
- categorizing the Objects.

As you can see, panoptic segmentation is a hybrid technique that combines instance and semantic segmentation. In 2018, Alexander Kirillov and his group released it.

Panoptic segmentation, or as the authors termed it, "the unified or global view of segmentation," aims to comprehensively generalize the work of image segmentation as opposed to employing two independent methodologies separately.

Using panoptic segmentation, it is possible to divide objects into two groups: Things and Stuff.

Things: Things in computer vision are typically described as objects with countable geometry, such as people, vehicles, animals, etc.

Stuff: The term "stuff" is used to describe objects that lack precise geometry but are strongly recognisable by their texture and material, such as the sky, a road, a body of water, etc.

Output

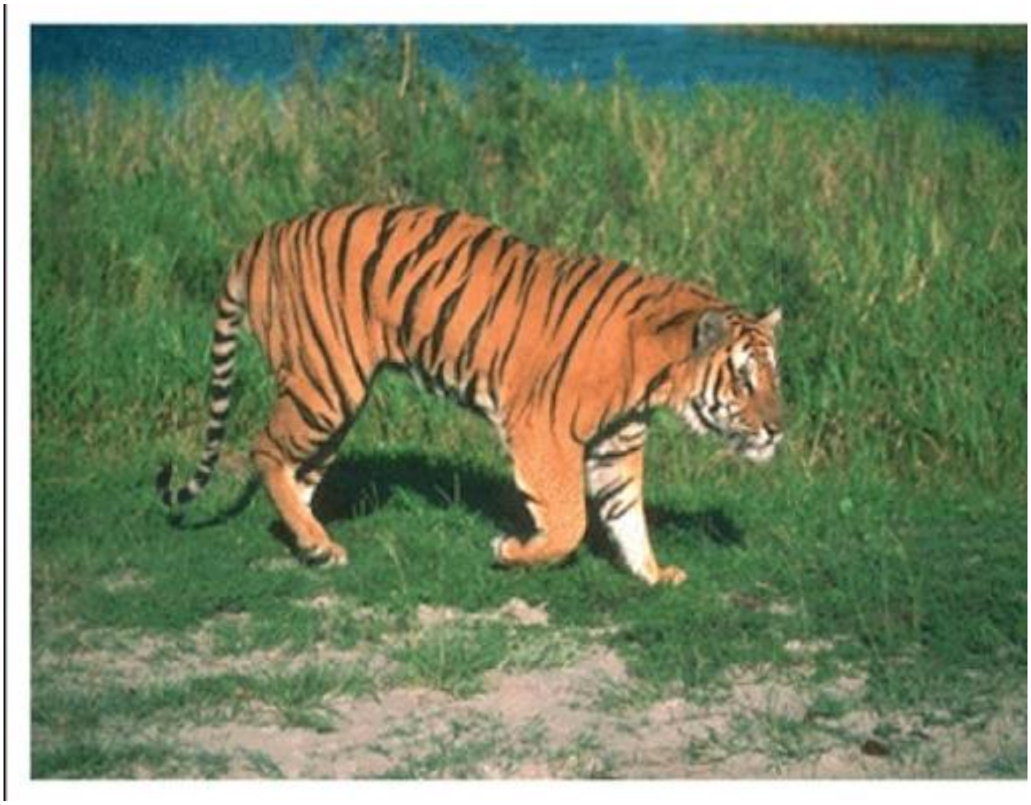


Fig 2.8 Input Image

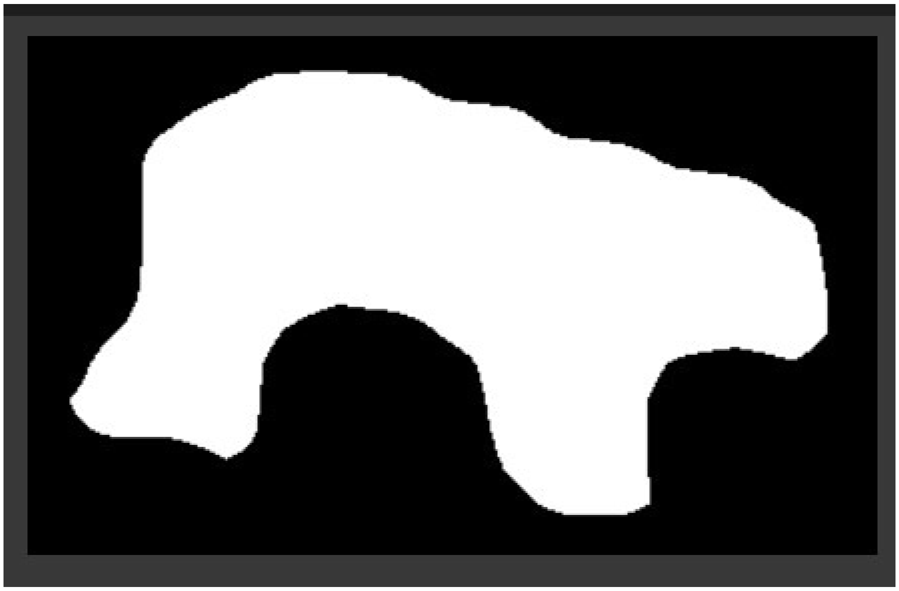


Fig 2.9 Output Image

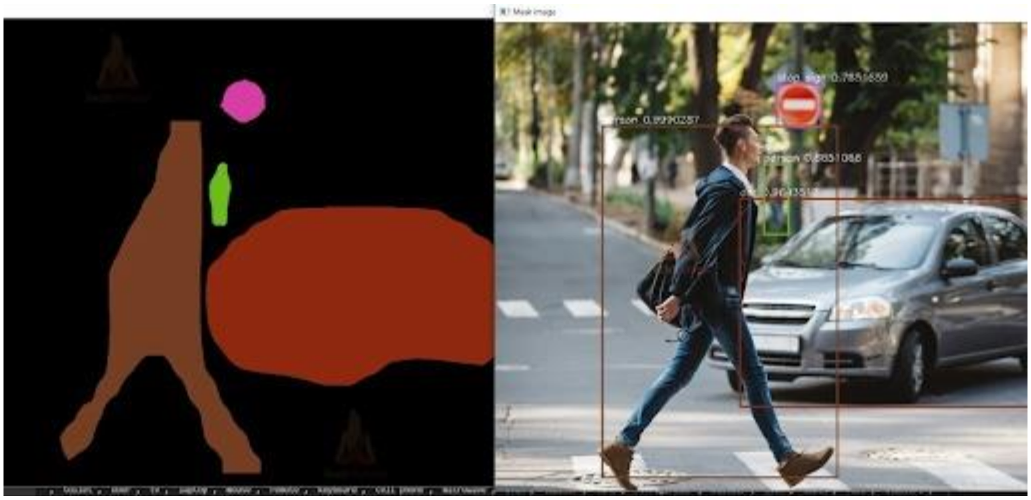


Fig 2.10 Input Image

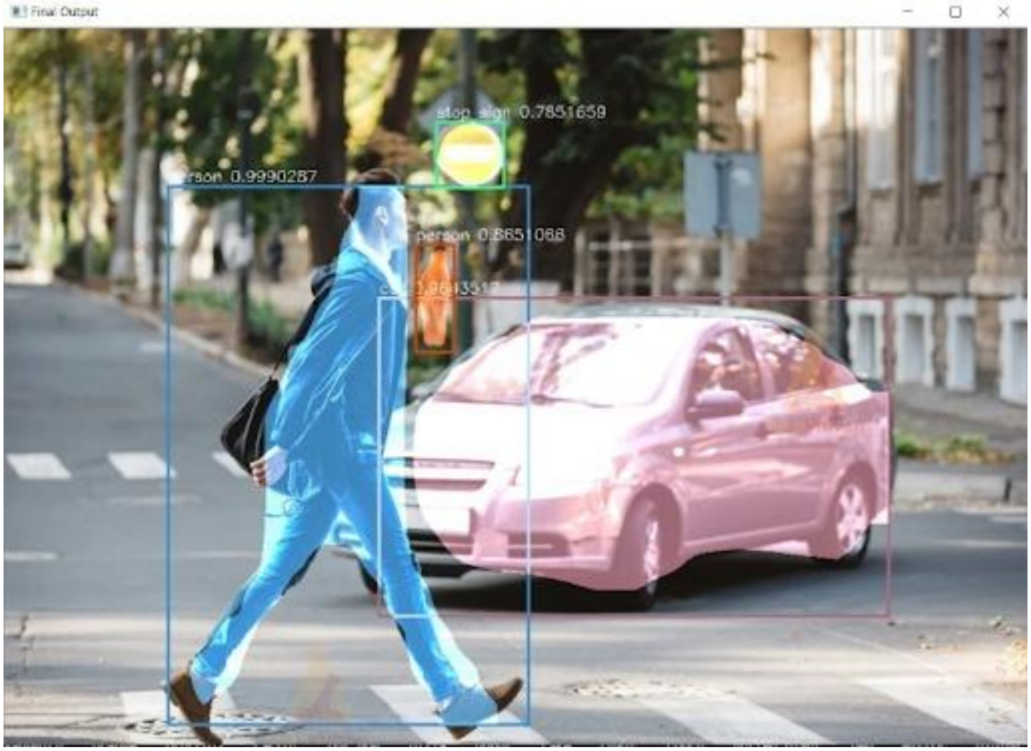


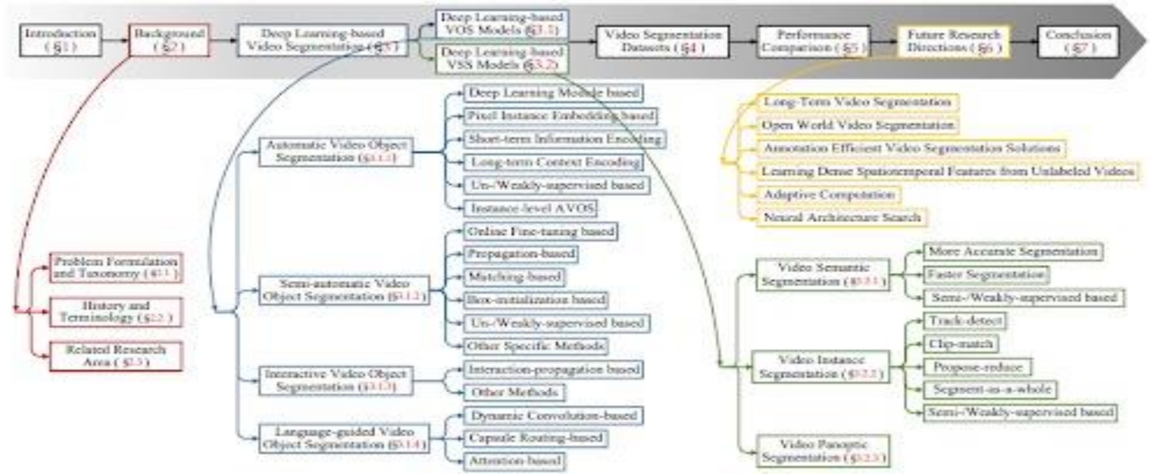
Fig 2.11 Output Image

Chapter 3: Literature Survey

3.1 Analysis

A fundamental and difficult topic in computer vision is video segmentation, which involves finding the group of important objects (with certain specified feature attributes or semantic value) in a video scene. It has long attracted considerable attention and have been extensively looked into the the importance of computer vision and graphics in a variety of range for example, autonomous driving, robotics, automated surveillance, social media, augmented reality, filmmaking, video conferencing, etc.) of application domains..

Various conventional machine learning and computer vision methods, such as hand-craft characteristics, have been used in the past to address the problem of video segmentation (e.g., color, histogram statistics, optical flow, etc.)



Introduction (§1)
Background (§2)
Deep Learning-based Video Segmentation (§3)
Deep Learning-based VOS Models (§3.1)
Deep Learning-based VSS Models (§3.2)
Video Segmentation Datasets (§4)
Performance Comparison (§5)
Future Research Directions (§6)
Conclusion (§7)
Problem Formulation and Taxonomy (§2.1)
History and Terminology (§2.2)
Related Research Area (§2.3)
Automatic Video Object Segmentation (§3.1.1)
Deep Learning Module based
Pixel Instance Embedding based
Short-term Information Encoding
Long-term Context Encoding
Un-/Weakly-supervised based
Instance-level AVOS
Semi-automatic Video Object Segmentation (§3.1.2)
Online Fine-tuning based
Propagation-based
Matching-based
Box-initialization based
Un-/Weakly-supervised based
Other Specific Methods
Interactive Video Object Segmentation (§3.1.3)
Interaction-propagation based
Other Methods
Language-guided Video Object Segmentation (§3.1.4)
Dynamic Convolution-based
Capsule Routing-based
Attention-based
Long-Term Video Segmentation
Open World Video Segmentation
Annotation Efficient Video Segmentation Solutions
Learning Dense Spatiotemporal Features from Unlabeled Videos
Adaptive Computation
Neural Architecture Search
Video Semantic Segmentation (§3.2.1)
More Accurate Segmentation
Faster Segmentation
Semi-/Weakly-supervised based
Video Instance Segmentation (§3.2.2)
Track-detect
Clip-match
Propose-reduce
Segment-as-a-whole
Semi-/Weakly-supervised based
Video Panoptic Segmentation (§3.2.3)

Fig 3.1 Process of Video Segmentation

Video foreground/ (Video Object Segmentation, VOS) Background segmentation Since most traditional scenario for video segmentation, VOS refers to the division of video sequences into the dominating, general objects (of undefined categories).

As a result, Y is a binary space for foreground and background segmentation. The split items' precise semantic categories are unimportant to VOS, which is commonly utilised in some applications for video analysis and editing situations, such as virtual backgrounds, content-based video coding, and object removal in movie editing construction at a video conference.

Video Semantic Segmentation (VSS)

VSS tries to extract objects from videos that fall into preset semantic categories (such as automobile, building, pedestrian, and road) as a direct extension spatiotemporal domain of picture semantic segmentation. Y is As a consequence, we obtain a multi-class, semantic parsing space. Many application fields, including autonomous driving, human-robot interaction, and robot sensing which demand a high level of having knowledge of the physical surroundings, use VSS as a perception foundation.

Remark

Fast motion and object occlusion are two issues that both VOS and VSS face. However, many of the issues are different because of their unique application circumstances. For instance, VOS may focus more on sequences from human-produced media, which frequently have significant camera motion, distortion, and appearance changes.

Instead, VSS frequently focuses on applications like autonomous driving, which call for accurate small object identification, model parallelization, and cross-domain generalisation capability.

Inference modes for Video Segmentation

The three categories of VOS methods—automatic, semi-automatic, and interactive—can be further divided based on how much human interaction is required during inference. • **Automatic Video Object Segmentation(AVOS)**. Without any manual initialization, AVOS, also known as unsupervised video segmentation or zero-shot video segmentation, automatically completes VOS (Fig. 1(a-b)). Only the video domain V is referred to by the input space X. Video editing that needs segmenting arbitrary objects or their components flexibly is not appropriate for AVOS; the most common application is the construction of virtual backgrounds for video conferences.

Semi-automatic Video Object Segmentation (SVOS)

A small amount of human inspection (usually provided in the first frame) is used in SVOS, sometimes referred to as one-shot or semi-supervised video segmentation, to identify the desired items. As a result, for SVOS, we have X = V M, where V denotes the video space and M denotes human involvement. First-frame object mask is the most prevalent type of human intervention. SVOS is also known as pixel-wise tracking or mask propagation in such a situation. Bounding boxes and scribbles are a couple of additional human input methods.

From this vantage point, **language-guided video object segmentation (LVOS)** is a subbranch of SVOS that enables effective human-computer interaction by providing language descriptions of the desired items in place of human intervention. SVOS compensates for AVOS approaches' rigidity in target object definition, albeit at the expense of more manual labour. SVOS is generally used in settings that don't require specific technology, such making videos on mobile devices. How to completely utilise target information from limited human intervention is one of the main issues in SVOS.

- **Interactive Video Object Segmentation (IVOS)**

While systems for IVOS include user instruction throughout the analysis process, SVOS models are built to run immediately following the hit on the target determined. Thus, IVOS can produce ideal for computer-generated images and post-production video, when rigorous human inspection is possible, and has high-quality portions.

In the graphics industry, IVOS is also researched as video cut-out. IVOS's input space X is VS, where S commonly stands for human handwriting. The following provide significant challenges: 1) facilitating the easy specification of segmentation restrictions by users; 2) incorporating human specified constraints into the segmentation algorithm; and 3) responding quickly to the constraints.

Only a few early systems addressed the semi-automatic setting, known as labelling, while VSS methods normally operate in an automated mode, i.e., X V.

Learning Paradigms for Video Segmentation

Current deep learning-based video segmentation models can be divided into three types based on the learning technique to approximate f* supervised, unsupervised, and poorly supervised learning-based.

Supervised Learning-based Methods

Modern video segmentation models typically require training samples with N inputs and the desired results, and they are learned in a fully supervised way.

The following is the recommended approach for assessing learning results a formula for empirical risk/loss minimization :

Chapter 4: Conclusions

5.1 Conclusions

The process of dividing a video "digital video browsing and retrieval's first step toward automatic annotation. Several research communities, including those in image processing, computer vision, pattern recognition, and art, are paying attention to this busy area of study "intellectual capacity.

A wide range of strategies have been developed as a consequence of more than eight years off video segmentation research. Early research centered on cut detection, but more modern methods address gradual transition detection. Almost all algorithms handle uncompressed video

Numerous algorithms that directly work on the compressed video stream were described since it is anticipated that the video will be stored in a compressed format. They can be categorized into six kinds based on the type of information used.

Our tests show that our techniques can successfully meet the goal of image segmentation. The outcome is fairly good for straightforward photographs with only a little internal roughness, and this technique may also work well for natural and landscape images.

The tests mentioned above use RGB or luminance-based similarity measurements. The results got worse when we tested our technique on texture images using the same similarity metric. The primary cause of this is that the texture region's pixel value variance is quite high, which implies that multiple groups of pixels with slight similarities may be segregated into the same texture region's pixels.

5.2 Future Scope

AFTER GETTING THE SEGMENTED IMAGES OF PARTICULAR FRAME OF A VIDEO:

We can use it for the analysis and better understanding of any particular scenario of the videos which is the primary objective of the segmentation of image or video. Further understanding of the video or image leads us to perform many important tasks which help the Human and surroundings in a very profound manner.

Some of the Application of image or video segmentation is listed below:

1. Self-Driving Cars: To distinguish between numerous objects clearly, self-driving cars can use picture segmentation. be it billboards, traffic lights, pedestrians, or vehicles. It might make it possible for the algorithm that creates driving instructions to assess the environment more precisely before producing the next command.

2. Face Recognition: As already said, the majority of modern smartphones have portrait mode. Technically speaking, image segmentation produces portrait mode. Aside from this, security surveillance will be significantly more effective when faces can be separated from distracting objects.

3. Circuit Board Defect Detection: The manufacturer must be held accountable for defective goods. If a camera backed by an Image Segmentation model continuously scans for issues caused in the finished product, a lot of money and time can be saved on fixing a problematic device.

5.3 Applications of the project:

In artificial vision, image segmentation is a crucial stage. Segment-specific processing requires the segmentation of visual data by machines. In well-known such as driverless cars, robots, medical imaging, and intelligent video analytics, image segmentation is thus used. Along with these uses, satellites also use image segmentation to separate out roads, buildings, and trees in aerial imagery.

Here are a few of thee most well-liked image segmentation use cases from the real world

This particular approach has various uses, particularly for students looking to enrol in:

1. ,Robotics (Machine Learning): By highlighting objects in their line of motion, image segmentation supports machine perception and locomotion. This enables machines to change routes successfully and comprehend the context of their environment. In addition to assisting robots with movement, segmentation of images enables them to distinguish between the objects they are working with and allows them to interact with actual objects using only vision as a guide. Because of this, the machine can be used practically anyplace without many limitations.
    - Instance segmentation for robotic grasping
    - Autonomous navigation and SLAM

2. Medical Imaging: One significant area of computer vision is medical imaging, which focuses on making illness diagnoses using visual data, including both straightforward visual data and biological scans. In medical imaging, segmentation plays a crucial role in the quick and precise identification of potentially cancerous areas in pictures. With the use of image segmentation, disease detection can be expedited and made more affordable, helping countless people all over the world.

    - X-Ray segmentation
    - C-T scan organ segmentation
    - Digital pathology cell segmentation

References

[1] V. Badrinarayanan, F. Galasso, and R. Cipolla, “Label propagation in video sequences,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2010, pp. 3265–3272.

[2] L. G. Roberts, “Machine perception of three-dimensional solids,” Optical and Electro-Optical Information Processing, 1965.

[3] J. Chang, D. Wei, and J. W. Fisher, “A video representation using temporal superpixels,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2013, pp. 2051–2058.

[4] R. Jain and H.-H. Nagel, “On the analysis of accumulative difference pictures from image sequences of real world scenes,” IEEE Trans. Pattern Anal. Mach. Intell., no. 2, pp. 206–214, 1979.

[5] C. R. Wren, A. Azarbayejani, T. Darrell, and A. P. Pentland, “Pfinder: Real-time tracking of the human body,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 19, no. 7, pp. 780–785, 1997.

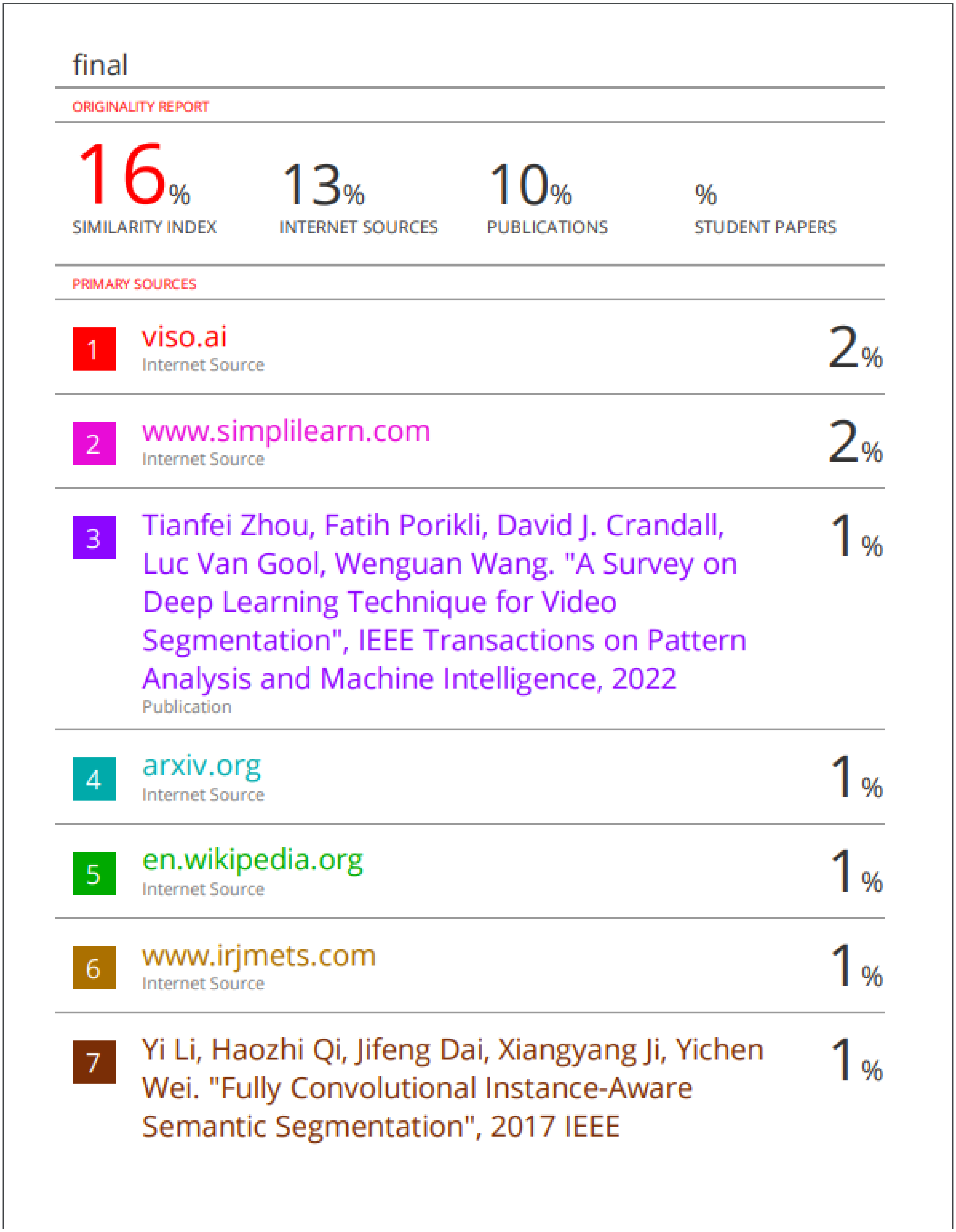
[6] A. Criminisi, G. Cross, A. Blake, and V. Kolmogorov, “Bilayer segmentation of live video,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., vol. 1, 2006, pp. 53–60.

[7] S. Brutzer, B. Hoferlin, and G. Heidemann, “Evaluation of back- ¨ ground subtraction techniques for video surveillance,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2011, pp. 1937–1944.

[8] E. Hayman and J.-O. Eklundh, “Statistical background subtraction for a mobile observer,” in Proc. IEEE Int. Conf. Comput. Vis., vol. 2, 2003, pp. 67–67.

[9] M. Irani and P. Anandan, “A unified approach to moving object detection in 2d and 3d scenes,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 20, no. 6, pp. 577–589, 1998.

[10] C. Yuan, G. Medioni, J. Kang, and I. Cohen, “Detecting motion regions in the presence of a strong parallax from a moving camera by multiview geometric constraints,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 29, no. 9, pp. 1627–1641, 2007



final

ORIGINALITY REPORT

16%	13%	10%	%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	viso.ai Internet Source	2%
2	www.simplilearn.com Internet Source	2%
3	Tianfei Zhou, Fatih Porikli, David J. Crandall, Luc Van Gool, Wenguan Wang. "A Survey on Deep Learning Technique for Video Segmentation", IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022 Publication	1%
4	arxiv.org Internet Source	1%
5	en.wikipedia.org Internet Source	1%
6	www.irjmets.com Internet Source	1%
7	Yi Li, Haozhi Qi, Jifeng Dai, Xiangyang Ji, Yichen Wei. "Fully Convolutional Instance-Aware Semantic Segmentation", 2017 IEEE	1%

	Conference on Computer Vision and Pattern Recognition (CVPR), 2017 Publication	
8	onlineresource.ucsy.edu.mm Internet Source	1 %
9	Pantea Keikhosrokiani, Moussa Pourya Asl. "chapter 1 Introduction to Artificial Intelligence for the Analytics of Literary Works and Social Media", IGI Global, 2022 Publication	1 %
10	www.geeksforgeeks.org Internet Source	1 %
11	Mohamed Mostefai, Mhamed Sekour, Mohamed Amara. "Optimization of Photovoltaic Installation Based on Machine Learning for Water Pumping System Using a BLDC Motor", Journal of Electrical Engineering & Technology, 2023 Publication	1 %
12	V. Rajathi, Syed Abdul Azeem A, Selvakannan S P. "Earlier Detection of Diabetes Using Random Forest Algorithm", 2022 1st International Conference on Computational Science and Technology (ICCST), 2022 Publication	<1 %
13	Luis Augusto Silva, André Sales Mendes, Héctor Sánchez San Blas, Lia Caetano Bastos	<1 %