

**INSTANCIFY**

(Team Number - 7)

**Team Members :-**

*Aditya Varshney E18CSE010*

*Amit Patel E18CSE013*

*Anushka Upadhyaya E18CSE026*

*Arnav Bansal E18CSE028*

*Bhavya Batra E18CSE035*

**Mentor :-**

*Tanveer Ahmed*

*Tapas Badal*

**INTRODUCTION**

The basic ideas of our project is to achieve object detection. The Object Detection allows the recognition, detection and the localization of multiple objects within an image. As the name suggests the feature we are trying to explore is instantiation of an image, means generating a mask for all objects detected. This can also be done on a video or in real time but for now we are limiting ourselves to images. Hence, object detection provides us a much better understanding of an image or a frame being processed as compared to human visual recognition.

Object Detection is the process of finding instances of real world objects such as faces, bicycles and buildings in images or videos. Currently, object detection algorithms typically use extracted features and learning algorithms to recognize instances of an object category (object class) and it is commonly used in applications such as image retrieval, security surveillance and advance deriver assistance systems which is known as DES (Detection with Enriched Semantics). For now, our vision is just to instantiate all object categories in an image accurately, and then we can plan to extend this for a video or real time.

We will be doing instance segmentation using a Mask-RCNN model which we will retrain on our own dataset.

Semantic segmentation makes dense predictions inferring labels for each pixel so that every pixel in the image is labeled with the class of its enclosing object. Object detection provides not only the classes but also indicate the spatial location of those classes. It takes into account the overlapping of objects. Instance segmentation includes identification of boundaries of the objects at the detailed pixel level.

We are studying and analyzing a Python Model which we aim to retrain and make an api out of it which would run at the backend of our web.

**PROBLEM STATEMENT AND OBJECTIVES**

Concretely, our task was to develop Instancify which could work as an Online Image Instance Segmentation Application using a highly accurate Mask-RCNN Model to perform the process of segmentation. In the end our model will be able to instantiate over 81 different classes of objects.

OBJECTIVES:

* Properly define objects which needs to be detected or instantiated.
* Make the website easy to use and fast.
* Reduce the process time as much as we can so as to provide user a better experience.
* Increase the accuracy of existing algorithm according to our needs.
* Learn something new while working on this project.
* Integrate the frontend with backend efficiently.

Areas in which our model can be improved in future are (these can be considered as the future objectives of our project and we will try our level best to complete what we have started):

1. Improve Open Street Map by identifying areas of interest from drone or aerial footage which improves map quality and helps to identify the features of maps.

2. Helps in implementing the program for autonomous driving car.

3. Helps in applying colors and filters on images.

4. Threat detection using computer vision (image segmentation) can be much more responsive then their human counterparts.

5. 3D reconstruction of objects in image.

6. Semantic analysis of human emotions using image.

Some of these application might be way ahead of our knowledge but they are obviously possible to implement.

**WHY AND HOW THE PROJECT WAS CHOSEN?**

The holy grail of modern computer vision techniques (Image Segmentation to be precise) would be to mimic the human visual perception. Or in other words, understand the visual data by Image Segmentation and make complex conclusions about this data is the basic idea behind the Innovation of this project.

Also, coming to the third semester, the subjects we are being taught are all core to our stream which means we get to learn a lot of new things which will become the base for our future. Taking this as an opportunity we wanted to make something that would help us brainstorm and open up a broad spectrum of ideas. We discussed a lot of ideas and finally decided upon Instance Segmentation. We knew it wouldn't be an easy task but a little challenge is always a motivator. Also making a project using the knowledge of each of the subject that we have in this semester was a difficult task and instance segmentation was what we found really interesting and suitable. Our project is a part of machine learning which we all wanted to work on hence our interests were also satisfied. And this is mostly why we went for the Instancify.

**LITERATURE SURVEY / BACKGROUND STUDY / STUDY OF SIMILAR PROJECTS**

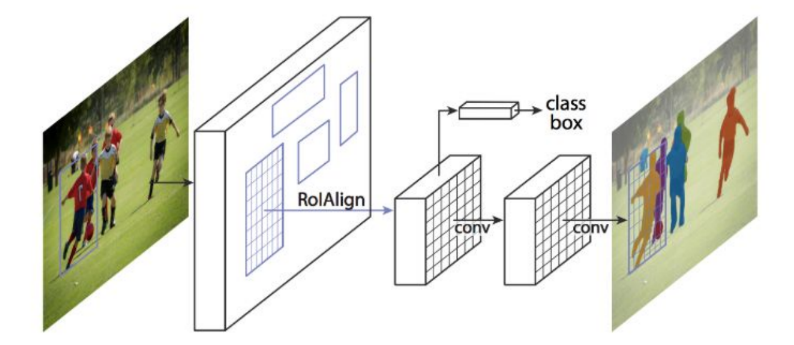
There are various techniques that are used in computer vision tasks. Some of them include classification, semantic segmentation, object detection, and instance segmentation. Classification tells us that the image belongs to a particular class. It doesn’t consider the detailed pixel level structure of the image. It consists of making a prediction for a whole input. Semantic segmentation makes dense predictions inferring labels for each pixel so that every pixel in the image is labeled with the class of its enclosing object. Object detection provides not only the classes but also indicate the spatial location of those classes. It takes into account the overlapping of objects. Instance segmentation includes identification of boundaries of the objects at the detailed pixel level. Mask R-CNN is an instance segmentation technique which locates each pixel of every object in the image instead of the bounding boxes. It has two stages: region proposals and then classifying the proposals and generating bounding boxes and masks. It does so by using an additional fully convolutional network on top of a CNN based feature map with input as feature map and gives matrix with 1 on all locations where the pixel belongs to the object and 0 elsewhere as the output. To help with debugging and understanding the model we can break our model into 3 parts namely,(inspect\_data.ipynb, inspect\_model.ipynb, inspect\_weights.ipynb) that provide a lot of visualizations and allow running the model step by step to inspect the output at each point. 1.Anchor sorting and filtering Visualizes every step of the first stage Region Proposal Network and displays positive and negative anchors along with anchor box refinement. 2.Bounding Box Refinement This is an example of final detection boxes (dotted lines) and the refinement applied to them (solid lines) in the second stage. 3.Mask Generation Examples of generated masks. These then get scaled and placed on the image in the right location. We studied how to run models on a web browser using FLASK and how to save images in a database.

There are many Instance Segmentation models available on the internet but all of them require heavy setups and installations. Because of this only people with the required knowledge are able to use such technologies. We at Instancify aim on bringing such technologies to the general public. With our easy to use UI, users will be able to experience Instance Image Segmentation just by opening up webpages. We chose Instance Image Segmentation because of its unique method of identifying its target pixels. Data scientists and researchers at Facebook AI Research (FAIR) pioneered a deep learning architecture, called Mask R-CNN that can create a pixel-wise mask for each object in an image. This is a really cool concept so follow along closely! Mask R-CNN is an extension of the popular Faster R-CNN object detection architecture. Mask R-CNN adds a branch to the already existing Faster R-CNN outputs. The Faster R-CNN method generates two things for each object in the image:

WHY ARE WE USING A MASK-RCNN?

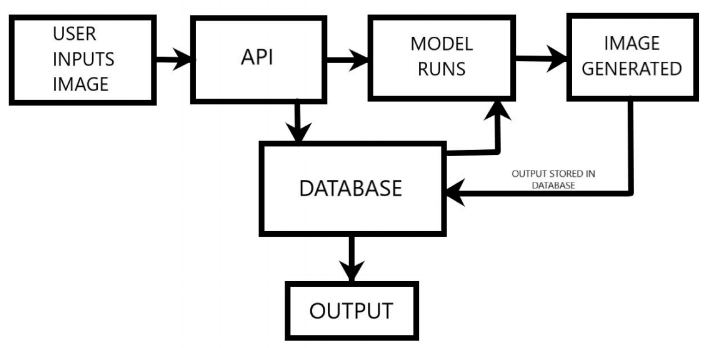
1. Gives three outputs for each object in the image: its class, bounding box coordinates, and object mask
2. Simple, flexible and general approach
3. It is also the current state-of-the-art for image segmentation

**MODEL WORKING DIAGRAM**



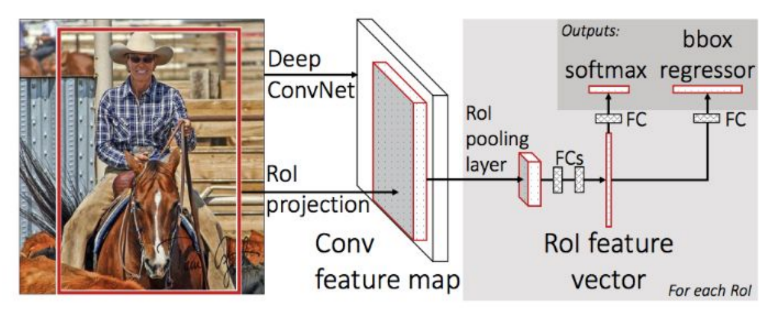
Basic working of how the model will target the pixels and construct a mask over them

**FLOW DIAGRAM**



Overall working of the project

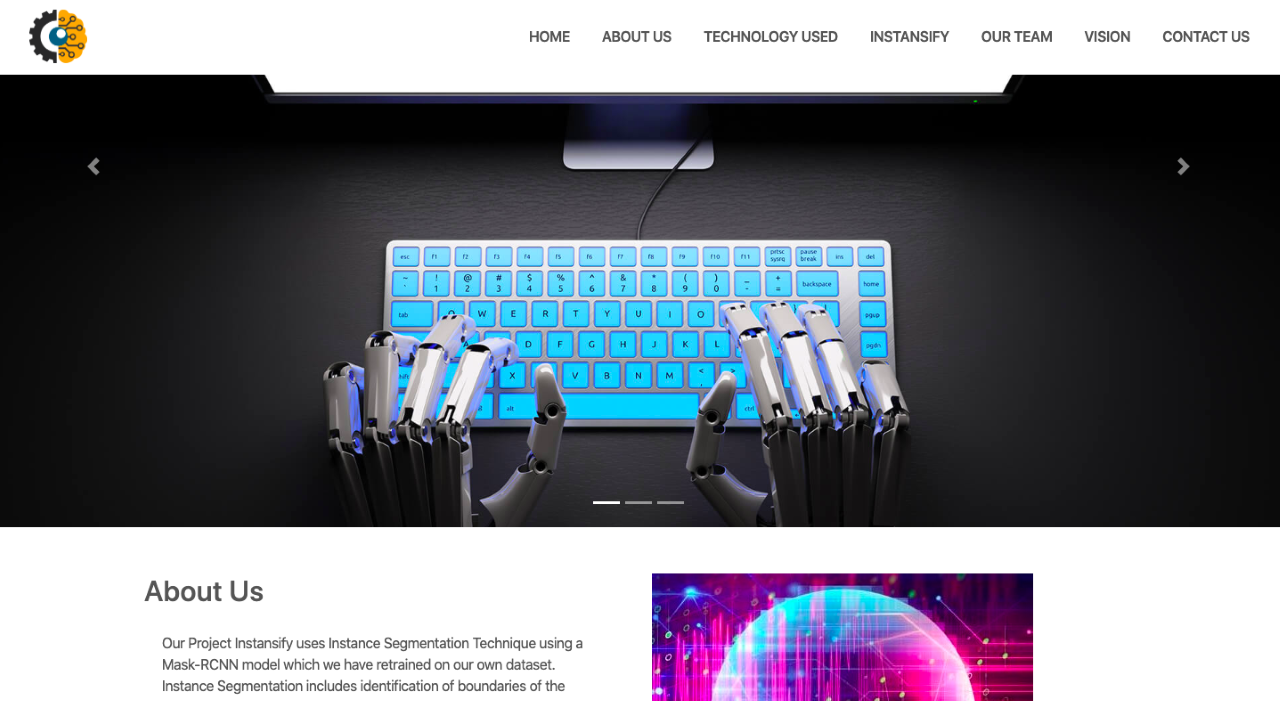
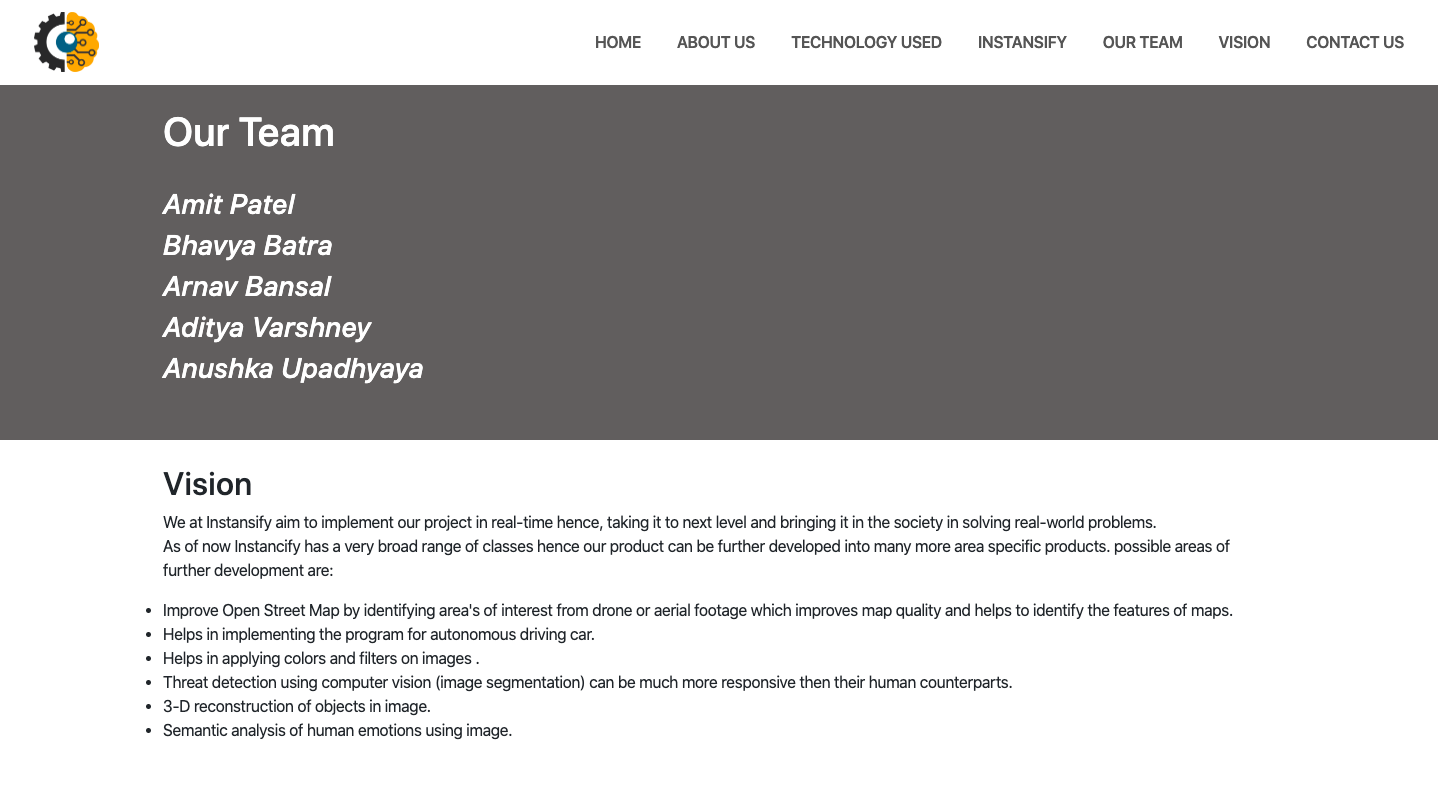
**CONVOLUTIONAL DIAGRAM**

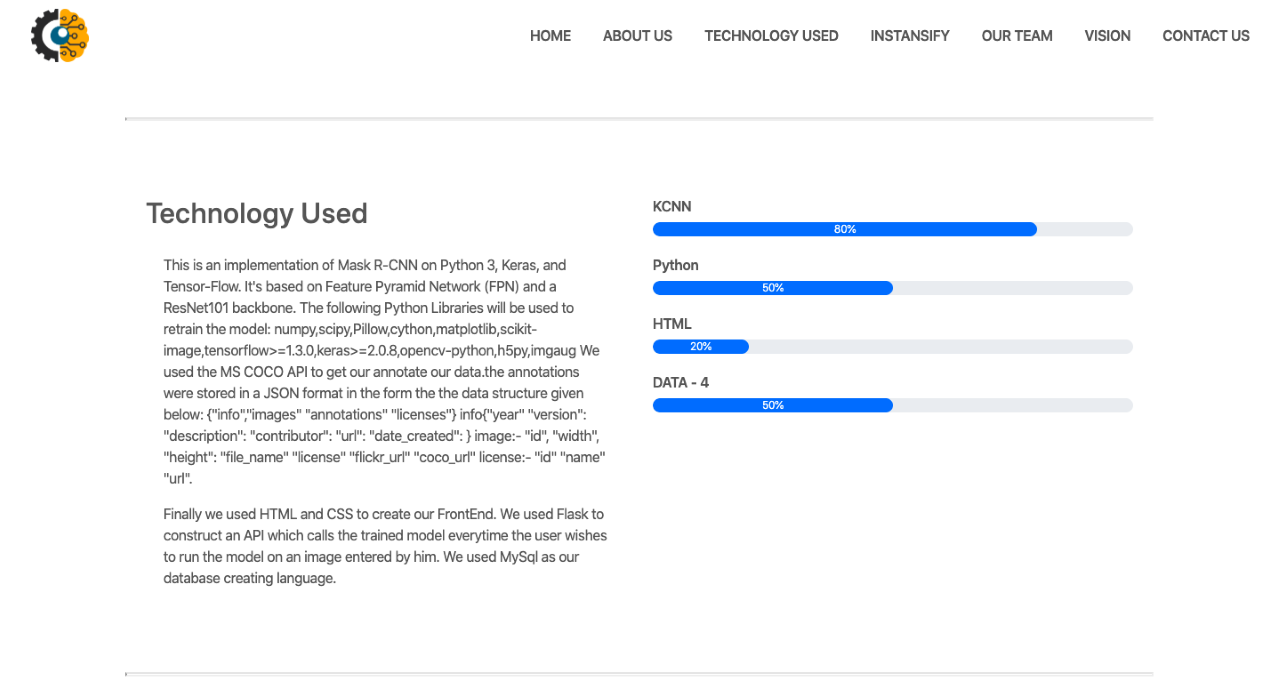


Working of the Convolutional Network in creating the required instances

**FUNCTIONALITY (SNAP SHOTS)**

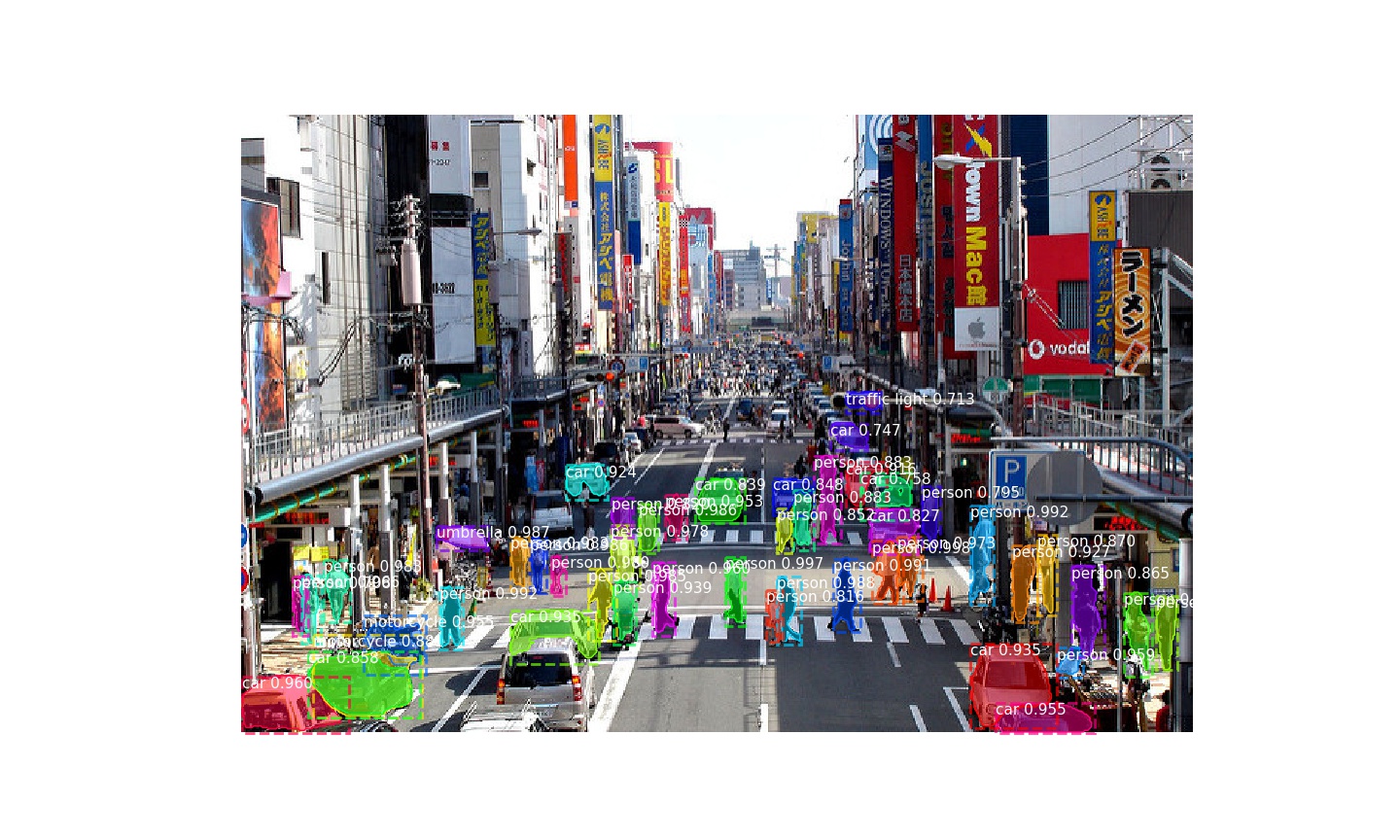
*WEBSITE*

** **

****

****

*An example of processed image*

****

**PROJECT DEVELOPMENT TIME SCHEDULE**

**Week 1-2: Initial Consult,**

Brainstorming, think for the idea. Multiple ideas must be discussed with the concerned people to get a proper guidance which could help us organize our work in a simpler manner.

**Week 3-4: Model Study,**

Explore different methods targeted to solve the problem. In this case we are needing a model to instance an image. Compare all the models and find a suitable model which could support our system and give us the required output.

**Week 5-6: Retrain model,**

The model chosen must have been trained on a specific model. So, retraining that model on our own dataset to modify the outputs in our accordance.

**Week 7: Frontend Design,**

Develop an interface which could take an image input and give the user an output image having classified objects according to their categories. Systematically presenting the outline functioning of our idea.

**Week 8: Backend,**

Linking the Frontend with Backend to make the site interactive and work properly without any bugs and errors.

**Week 9-10: Compilation and Review,**

Attempt to meet all the objectives which we had thought to achieve at the starting of the project. Compile the project as whole and review it with different types of systematic inputs to avoid any kind of irregularity. Consider taking suggestions of concerned people and improve accordingly.

**LEARNINGS, REFLECTIONS AND CONCLUSION FROM THE PROJECT**

This group object not only developed our mentality to work as a group. This group taught us how to collectively go for the solution for a problem. The project had a great impact on learning and creative implementations of our ideas. We explored various libraries in python related to our project. We gained the knowledge of database management which will come in handy at a point of time when we will need to manage the database of users.

We implemented various models as to get the best model possible which really helped in understanding the minute differences between different models. Now, we know how some of the part of a model works and how it is able to do what it does. Doing instance segmentation accurately of an image is no doubt a pretty difficult task, but just trying and trying continuously, applying different models, different algorithms, and help from everybody else including our mentors and fellow batch mates just made us capable enough to do it.

This project is a very interesting but has many things to do. Our strategy was to work together by dividing the work between the team members, so as to get good and effective solutions to our problems in time. Medical complications, if any one of our team members is not well or unable to complete his/her work on time, it was decided that further divide his work amongst the remaining in order to maintain the team spirit. We first studied a pre-trained mode, which in itself is a tedious task given our current knowledge about the field of work. Which is another risk factor that we experienced during our work on the project. Later on we worked on retraining the same model and build a user friendly interface for it.

**LIMITATIONS AND FUTURE ENHANCEMENTS**

As a team, we have come to the conclusion to implement this idea of instantiation only on images until it could really give us some sensible output. If we are able to achieve our objectives, then we can plan to extend this project on video and real time segmentation.

Also, the model can be less accurate than we expect it to be, the main reason for this would be less complex data. More the data, more is the accuracy. More complex images will be needed to have a near perfect model, though we will be trying our best in order to showcase what we have planned.

Instancify can be further implemented to start-up ideas like in Self-driving cars, Object Detection, Recognition Tasks, Machine Vision, Traffic Control Systems and much more. By its applications it can further be used in good form of monetization. In the long run, we see various upcoming opportunities for us in this field like Content-based image retrieval, is the application of computer vision techniques to the image retrieval problem, that is, the problem of searching for digital images in large databases, Machine vision (MV) is the technology and methods used to provide imaging-based automatic inspection and analysis for such applications as automatic inspection, process control, and robot guidance, usually in industry. Machine vision refers to many technologies, software and hardware products, integrated systems, actions, methods, and expertise. Machine vision as a systems engineering discipline can be considered distinct from computer vision, a form of computer science. Object detection is a computer technology related to computer vision and image processing that deals with detecting instances of semantic objects of a certain class (such as humans, buildings, or cars) in digital images and videos. Well-researched domains of object detection include face detection and pedestrian detection. Object detection has applications in many areas of computer vision, including image retrieval and video surveillance.

We can do further research in the field of computer vision and write a few research papers on techniques used in our project. With further studies we believe we can get better results from our model by experimenting and modifying the Convolutional Layers. "We can use Image Segmentation in the Form Video Segmentation. Video (temporal) segmentation is the process of partitioning a video sequence into disjoint sets of consecutive frames that are homogeneous according to some defined criteria. In the most common types of segmentation, video is partitioned into shots, camera-takes, or scenes. In the community, we would present a decent presentation and spread about our project as far as we can. We will conduct conferences related to the technique of Instance Segmentation and tell the learners more about it in detail and make it as interesting as possible by telling its results and long term goals. Also, we would make courses on its online series and upload them on YouTube also our results. We will train learners to make models and conduct a hands-on workshops."

**REFERENCES**

1. Matterport
   * <https://github.com/matterport/Mask_RCNN>
2. Cornell University Research Paper
   * <https://arxiv.org/abs/1703.06870>
3. YouTube Link For Mask RCNN Image Segmentation
   * <https://www.youtube.com/watch?v=5ZStcy7NWqs>

* Matterport

**VIDEO PRESENTATION**

https://youtu.be/-S8gpn8ACvw

**THANKYOU**