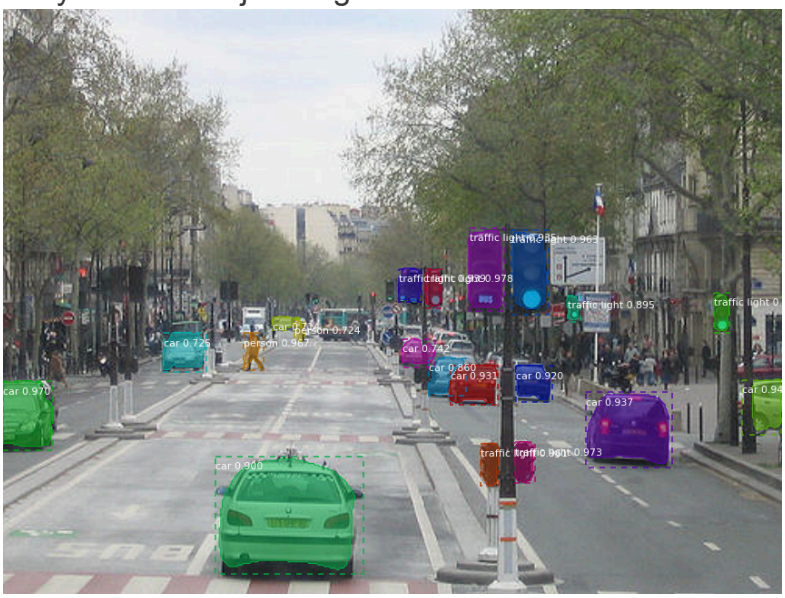
Pre-trained Deep Learning Models

List of Pre trained Models:

- **Object Detection**
  - **Mask R-CNN:**
    - It generates bounding boxes and segmentation masks for each instance of an object in a image
    - It mainly used for object segmentation
    - Ex: 
    - Link: [Mask R-CNN]
  - **YOLOv2**
    - YOLO is an ultra popular object detection framework for deep learning applications. This repository contains implementations of YOLOv2 in Keras
    - Ex: 

    - Link: [YOLOv2]

- **MobileNet**
  - MobileNet is an architecture designed for mobile devices
  - As you can see above, the applications of MobileNet are not just limited to object detection but span a variety of computer vision tasks – like facial attributes, landmark recognition, finegrain classification, etc
  - It has more than 20,000 classes of millions of images
  - Ex:

  - Link: [MobileNet]
- **Ripe/unripe tomato classification**
  - It helps to identify whether the tomato in the given image is grown or unripe using a pretrained Keras VGG16 model.
  - Ex:

  - Link: [Ripe/Unripe Tomato Classification]

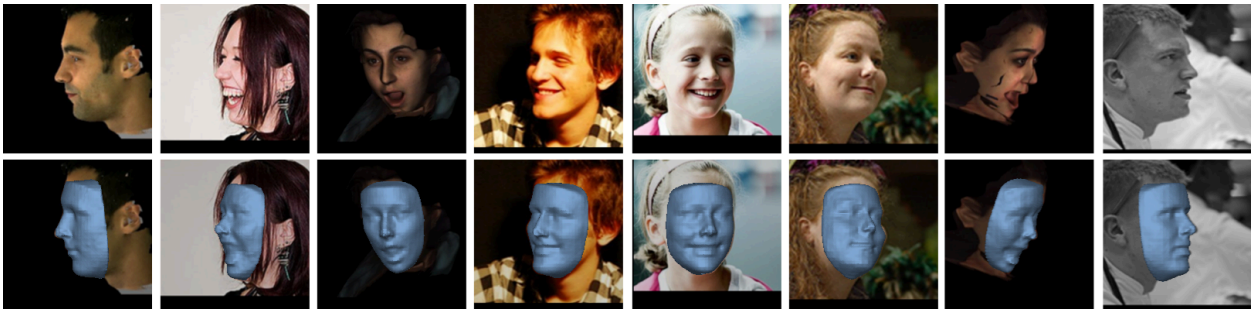
- **Car classification**
  - There are numerous ways of classifying a vehicle – by it's body style, number of doors, open or closed roof, number of seats, etc. In this particular problem, we have to classify the images of cars into various classes. These classes include make, model, year, e.g. 2012 Tesla Model S
  - The model was trained using pretrained VGG16, VGG19 and InceptionV3 models
  - Ex:
  - 
  - Link: Car Classification

- **Facial Recognition and Regeneration**
  - **Facial recognition**
    - It is all the rage in the deep learning community
    - Its applications span a wide range of tasks – phone unlocking, crowd detection, sentiment analysis by analyzing the face, among other things.
  - Face regeneration:
    - itis the generation of a 3D modeled face from a closeup image of a face
  - Types:
    - VGG-Face is a dataset that contains 2,622 unique identities with more than two million faces. This pretrained model has been designed through the following method:
      - vgg-face-keras: Directly convert the vgg-face model to a keras model
      - vgg-face-keras-fc: First convert the vgg-face Caffe model to a mxnet model, and then convert it to a keras model

- Ex:

- Link: VGG-Face Model

- **3D Face Reconstruction from a Single Image**
  - This is a really cool implementation of deep learning. You can infer from the above image how this model works in order to reconstruct the facial features into a 3 dimensional space.
  - This pretrained model was originally developed using Torch and then transferred to Keras.
  - Ex:



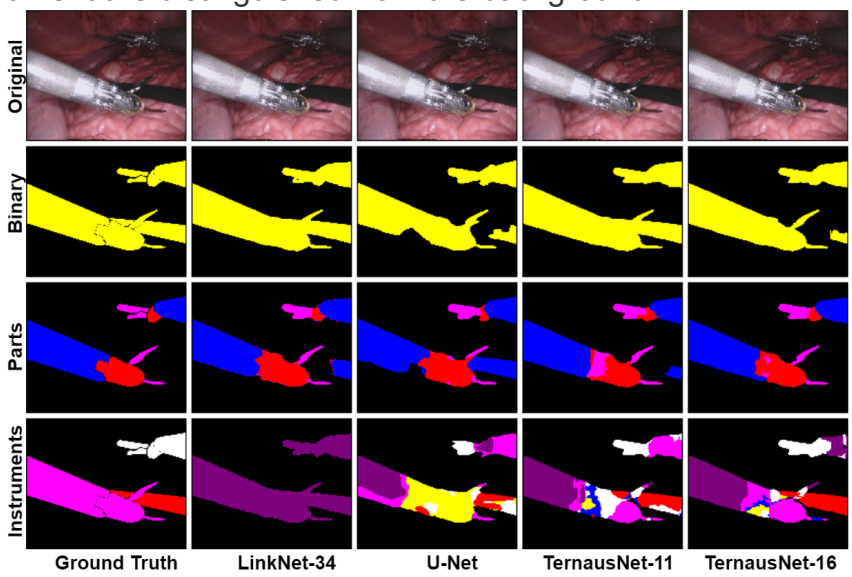
  - Link:3D Face Reconstruction from a Single Image

- **Segmentation**
  - **Semantic image segmentation:**
    - It is the task of assigning a semantic label to every single pixel in an image. These labels can be “sky”, “car”, “road”, “giraffe”, etc. What this technique does is it finds the outlines of objects and thus places restrictions on the accuracy requirements (this is what separates it from image level classification which has a much looser accuracy requirement).

- Types:
- Semantic Image Segmentation – Deeplabv3
  - Deeplabv3 is Google's latest semantic image segmentation model. It was originally created using TensorFlow and has now been implemented using Keras.
- Ex:

- Link: [Semantic Image Segmentation – Deeplabv3+]

- **Robot Surgery Segmentation**
  - This pretrained model is based on the U-Net network architecture
  - This model attempts to address the problem of image segmentation of surgical instruments in a robot-assisted surgery scenario. The problem is further divided into two parts, which are as follows:
  - Binary segmentation: Every pixel in an image is labelled as an instrument or background
  - Multi-class segmentation: Different instruments or different parts of an instrument are distinguished from the background



  - Ex:
  - Link: [Robot Surgery Segmentation]

- Miscellaneous
  - **Image Captioning**
    - It uses a combination of NLP and Computer Vision to produce the captions.

- The solution lies in converting a given input image into a short and meaningful description. The encoder-decoder framework is widely used for this task. The image encoder is a convolutional neural network (CNN).
- This is a VGG 16 pretrained model on the MS COCO dataset where the decoder is a long short-term memory (LSTM) network predicting the captions for the given image.
- Ex:

```
4
true: a Zebra grazing on grass in a green open field.
predict: A zebra standing in a field of grass near a fence .
```

- Link: Image Captioning