American International University Bangladesh (AIUB)



Reports Topic

**“A report on a web page testing using an automation testing tool -Katalon Studio”**

Course Tittle: Computer Vision and Pattern Recognition

Department of Computer Science

Name of Group Members

Joyti Saha (21-92083-2)

Ishtiak Hussain (21-92063-2)

**Submitted To**

**DR. DEBAJYOTI KARMAKER**

**Associate Professor**

**1 Introduction**

Computer vision is the field of computer science that focuses on creating digital systems that can process, analyze, and make sense of visual data (images or videos) in the same way that humans do. The concept of computer vision is based on teaching computers to process an image at a pixel level and understand it. Technically, machines. A few common tasks that computer vision systems can be used for: Object classification. The system parses visual content and classifies the object on a photo/video to the defined category. For example, the system can find a dog among all objects in the image. Object identification. The system parses visual content and identifies a particular object on a photo/video. For example, the system can find a specific dog among the dogs in the image. Object tracking. The system processes video finds the object (or objects) that match search criteria and track its movement.

Computer vision algorithms that we use today are based on pattern recognition. We train computers on a massive amount of visual data—computers process images, label objects on them, and find patterns in those objects. For example, if we send a million images of flowers, the computer will analyze them, identify patterns that are similar to all flowers and, at the end of this process, will create a model “flower.” As a result, the computer will be able to accurately detect whether a particular image is a flower every time we send them pictures. Golan Levin, in his article Image Processing and Computer Vision, provides technical details about the process that machines follow in interpreting images. In short, machines interpret images as a series of pixels, each with their own set of color values. For example, below is a picture of Abraham Lincoln. Each pixel’s brightness in this image is represented by a single 8-bit number, ranging from 0 (black) to 255 (white). These numbers are what software sees when you input an image. This data is provided as an input to the computer vision algorithm that will be responsible for further analysis and decision making.

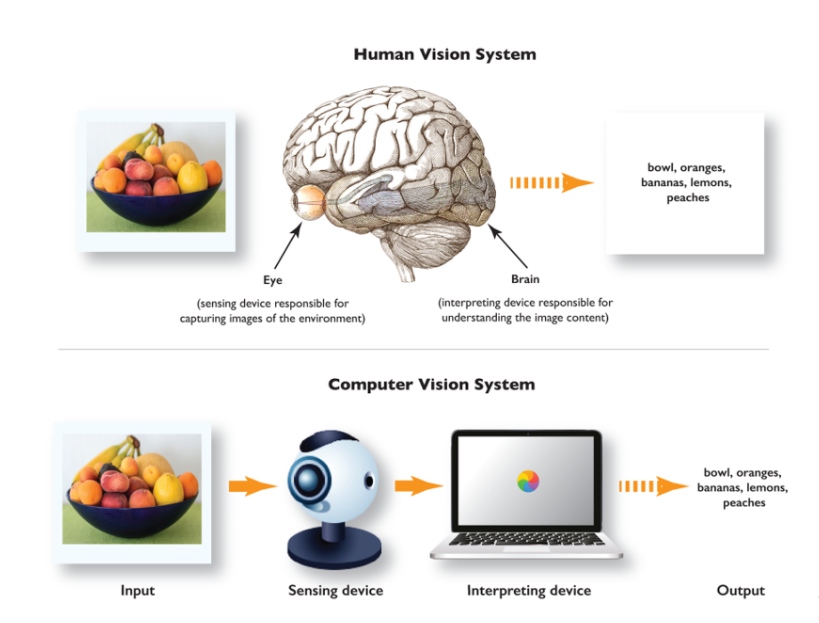
****

Image processing is a quite board research area, not just filtering, compression, and enhancement. Besides, we are even interested in the question, “what is in images?”,i.e., content analysis of visual inputs, which is part of the main task of computer vision. The study of computer vision could make possible such tasks as 3D reconstruction of scenes, motion capturing, and object recognition, which are crucial for even higher-level intelligence such as image and video understanding, and motion understanding.

Image processing is often regarded as improperly exploiting the image in order to achieve a level of beauty or to support a popular reality. However, image processing is most accurately described as a means of translation between a human viewing system and digital imaging devices. The human viewing system does not see the world in the same way as digital cameras, which have additional sound effects and bandwidth. Significant differences between human and digital detectors will be demonstrated, as well as specific processing steps to achieve translation. Image editing should be approached in a scientific way so that others can reproduce, and validate human results. This includes recording and reporting processing actions and applying the same treatment to adequate control images.

**2 System Description**

Semantic Segmentation is classifying each pixel of the image to its class label. A general semantic segmentation architecture can be broadly thought of as an encoder network followed by a decoder network:

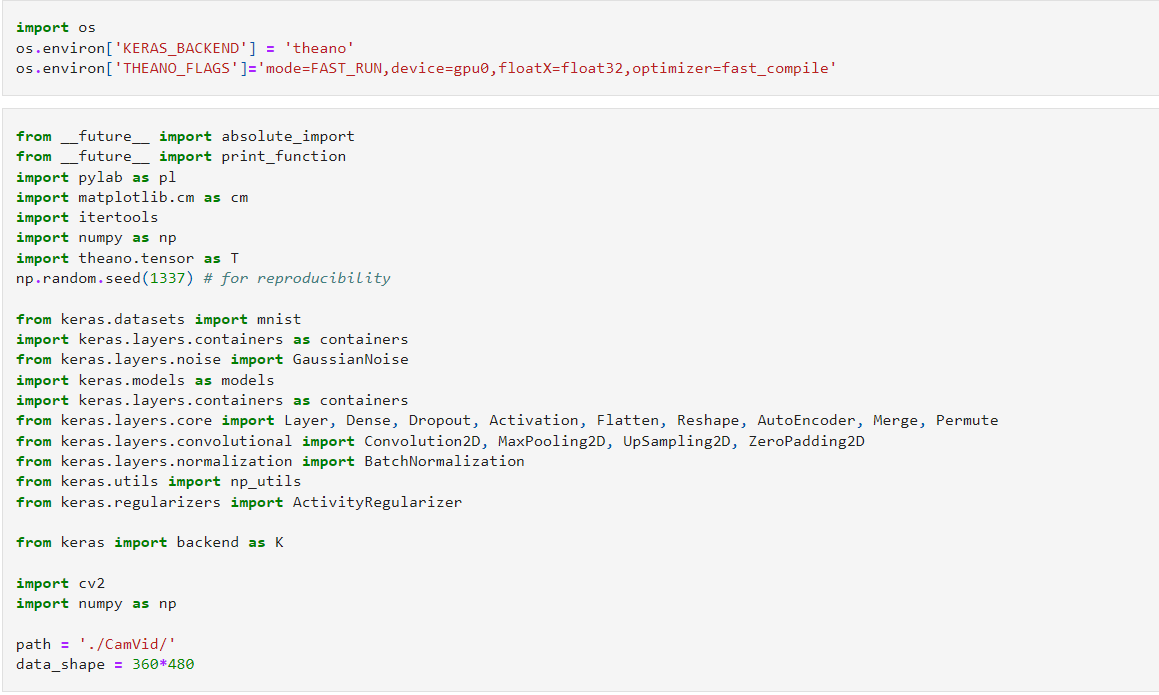
* The encoder is usually is a pre trained classification network like VGG/ ResNet followed by a decoder network.
* The task of the decoder is to semantically project the discriminative features (lower resolution) learnt by the encoder into the pixel space (higher resolution) to get dense classification.

Semantic Segmentation not only requires discrimination at pixel level but also a mechanism to project the discriminative feature as learnt at different stages of the encoder onto the pixel space. Different approaches employ different mechanisms as a part of the decoding mechanism.

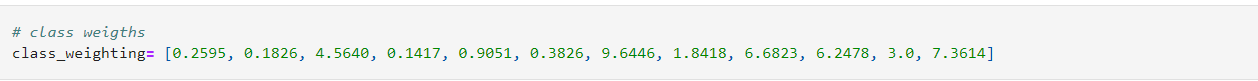
SegNet is designed to be an efficient architecture for pixel-wise semantic segmentation. It is primarily motivated by road scene understanding applications which require the ability to model appearance (road, building). Shape (cars, pedestrians) and understand the spatial-relationship between different classes such as road and side-walk. In typical road scenes, the majority of the pixels belong to large classes such as road, building and hence the network must produce smooth segmentation. SegNet has an encoder network and a corresponding decoder network, followed by a final pixel wise classification layer. This architecture is illustrated in above figure. The encoder network consists of 13 convolutional layers which correspond to the first 13 convolutional layers in the VGG16 network designed for object classification. They discard the fully connected layers in favor of retaining higher resolution features maps at the encoder output. This also reduces the number of parameters in the SegNet encoder network significantly (from 134 m to 14.M) as compared to other recent architectures. Each encoder layer has a corresponding decoder layer and hence the decoder network has 13 layers. The final decoder output is fed to a multi-class soft-max classifier to produce class probabilities for each pixel independently.

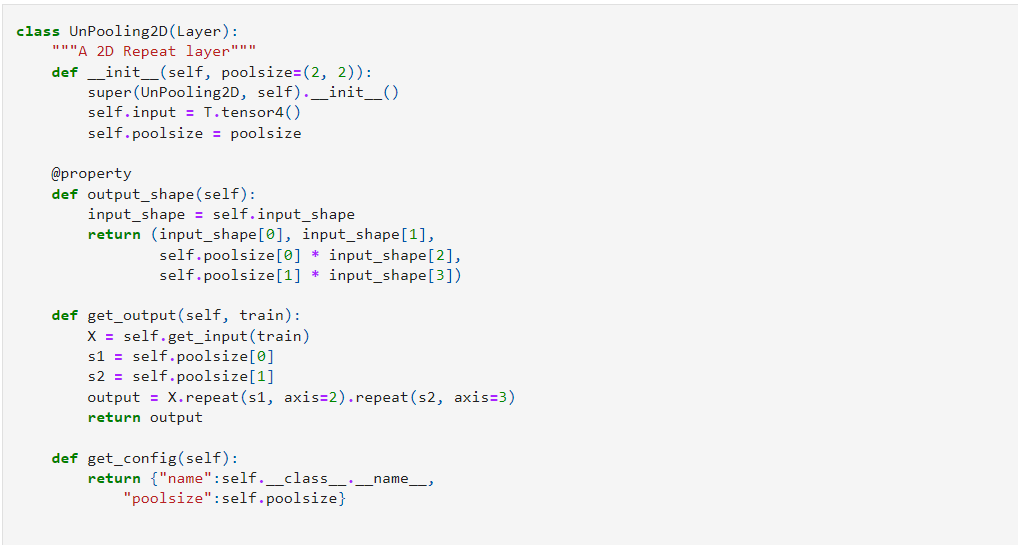
Encoder Network performs convolution with a filter bank to produce a set of features maps. These are then batch normalized. Then an element-wise rectified linear non-linearity max (0, x) is applied. Following that, max-pooling with a 2x2 window and stride 2(non-overlapping window) is performed and the resulting output is sub-sampled by a factor of 2.Max-pooling is used to achieve translation invariance over small spatial shifts in the input image. The increasingly lossy (boundary detail) image representation is not beneficial for segmentation where boundary delineation is vital. Therefore, it is necessary to capture and store boundary information in the encoder feature maps before sub-sampling is performed. If memory during inference is not constrained, then all the encoder features maps (after sub sampling) can be stored. This is usually not the case in practical applications and hence we propose a more efficient way to store this information. It involves storing only the max-pooling indices, i.e. and the locations of the maximum features value in each encoder feature map. In principle, this can be done using 2 bits for each 2x2 pooling window and is thus much more efficient to store as compared to memorizing feature , map(s) in float precision.

Decoder network up-samples its input feature map using the memorized max-pooling indices from the corresponding encoder feature map(s). This step produces sparse feature map (s0. This SegNet decoding techniques is illustrated in figure. These feature maps are then convolved with a trainable decoder filter bank to produce dense feature maps. A batch normalization step is then applied to each of these maps. The decoder corresponding to the first encoder (closest to the input image) produces a multi-channel features maps, although its encoder input has 3 channels (RGB). This is unlike the other decoders in the network which produce features maps with the same number of size and channels as their encoder inputs. The high dimensional feature representation at the output of the final decoder is fed to a trainable soft-max classifier. This soft-max classifies each pixel independently. The output of the soft-max classifier is a K channel image of probabilities where K is the number of classes. The predicted segmentation corresponds to the class with maximum probability at each pixel.

****

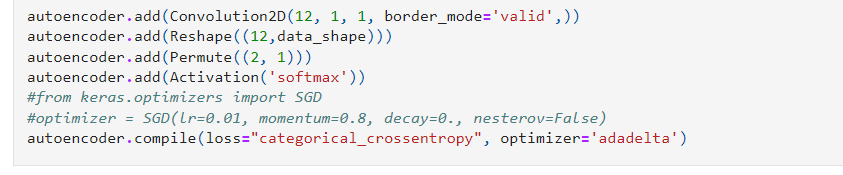
****

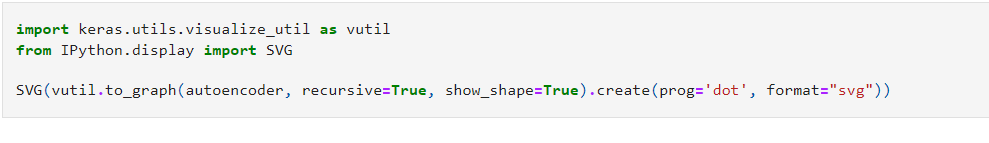
****

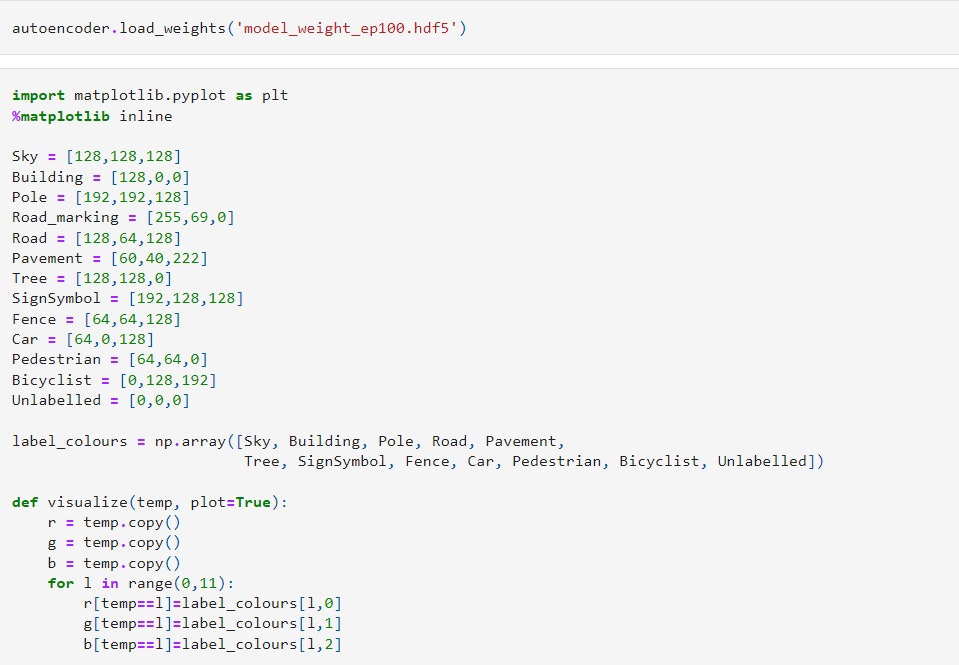
****

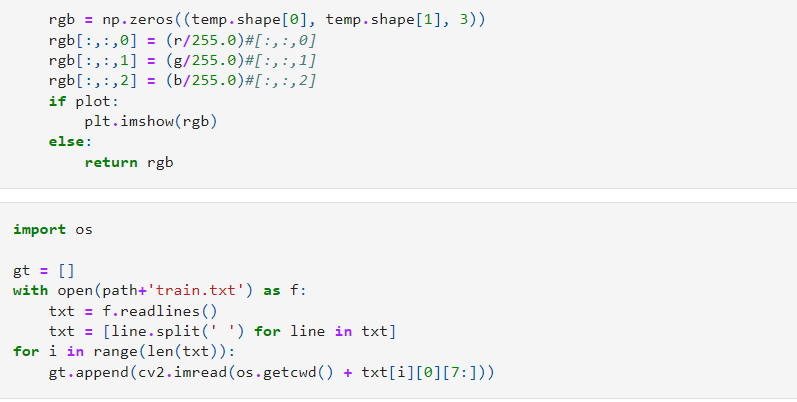
****

****

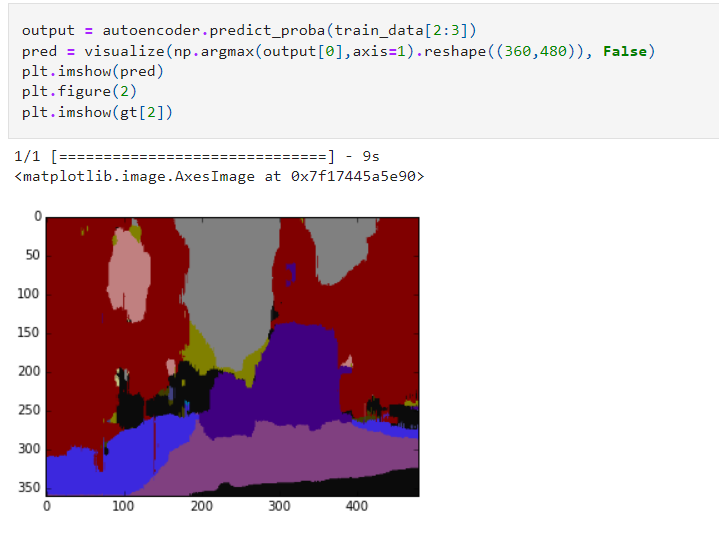
****

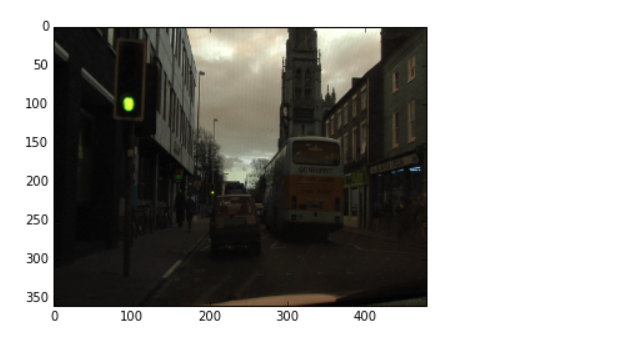
****

****

****

**3 Results**

****

****

**4 Conclusions**

[Computer vision](https://www.sciencedirect.com/topics/computer-science/computer-vision) technology based on color-image processing and analysis is a useful tool for the evaluation of fresh meat, including beef, pork, and lamb. The image features extracted can be used to effectively quantify and characterize quality attributes such as muscle color, marbling, maturity, and muscle texture, and quality and yield grades and cooked-beef tenderness can be predicted with satisfactory accuracy. Therefore, [computer vision](https://www.sciencedirect.com/topics/computer-science/computer-vision) is a promising technology for objective meat quality grading. The extensive research results published have formed a foundation for further investigation of the ability of computer vision systems to better provide quantitative information that is unobtainable subjectively, leading to the eventual replacement of human graders. Despite the above progress and successes, many challenging issues still remain and require continued in-depth research – among these, developing effective methodologies for consistent meat-image segmentation and discovering new quality indicators are the priorities.

**5. Referances**

1. <https://www.mygreatlearning.com/blog/introduction-to-image-processing-what-is-image-processing/>

2. <https://www.tutorialspoint.com/dip/image_processing_introduction.htm>

3. <https://blog.superannotate.com/introduction-to-computer-vision/>

4. <https://www.sciencedirect.com/topics/food-science/computer-vision-technology>

5. <https://www.jeremyjordan.me/semantic-segmentation/>

6. <https://medium.com/@fezancs/understanding-of-semantic-segmentation-how-segnet-model-work-to-perform-semantic-segmentation-5c426112e499>