*Rephrase  
Noted to myself*

**Title**

Traditional Maltese Food and Snacks Classifier

**Abstract**

With Malta’s tourism sector on continuous growth annually and being a very small island with rich culture, a tourist can engage and interest their selves into study our culture. Using a classification tool such as this study proposes makes it easier for a foreigner to learn basic concepts in Malta’s traditional food. This has been achievable using YOLOv3 neural network to detect and classify different articles of food, from cheeselet to deep fried date roll’s, labelled in Maltese ‘*ġbejniet*’ and ‘*imqaret*’ consecutively. This study has achieved (%) classification accuracy on a dataset of (?) images.

**Keywords**

Convolutional Neural Network (CNN), YOLOv3, Algorithm, Image Annotation.

**Introduction**

Using computer vision as one of deep learning methods, it is possible to classify various traditional Maltese food. The idea behind this study is to have a more user-friendly introduction and easier access for foreigners to traditional Maltese food, how to identify different articles of food and their name. Hence this study discusses and explore the field of computer vision and the reasoning behind a regression based YOLOv3 algorithm high speed’s performance. This study also highlights the collection and construction of a dataset to train a YOLOv3 model and what such dataset consists. Finally this study will question and answer whether or not computer vision be able to reach a high level of accuracy, which algorithms I s best suited to classify food and how a reliable dataset is constructed. The significance of this research is to provide a tool that assists people who are not familiar with traditional Maltese food such as tourists to distinguish between various food easily.

**Literature Review**

Computer vision has become a very demanding technology that contributed to various sectors to identify various aspects; for instance, computer vision contributes a lot in the medical field and helps practitioners such as nurses and doctors in the process of identifying, study or treat diseases. (Human Bone Localization in Ultrasound Image Using YOLOv3 CNN Architecture, 2019) This technology is often found in agriculture, the tourism sector, sport, manufacturing amongst others. This technology works by taking an image, understand what the image is and interpret its content with a view to solve tasks and resolve queries.

(Tumas) discuss how computer vision includes two operations; to filter the image by reducing noise that could hinder the object segmentation algorithms performance, and the second operation is subtracting the background from the region of interest (ROI). The main concern about this method is that the captured image conditions may have different lighting, shadows and level of reflections that might influence the overall performance of subtracting the background. This step could be quite challenging and plays a very important role in object detection systems, a solution is to use a sliding window that moves across an image and collect window view to the histogram of oriented gradients (HOG) however, this is an expensive operation. Therefore a less expensive solution are algorithms based on deep neural networks that are proven to have a better success rate.

Other relevant studies take into variable shape, size, the colour of the object to be detected. During a study regard vehicle intelligence and understanding of its surroundings (Wang), the researched stars to question various aspects in relation to a typical driving scenario. Other variables discussed where the dynamic environment a vehicle is found in, thus the researcher takes the lighting of the environment and the background into consideration.

Throughout, various researchers have done intense study in the field of computer vision, some of which have solved major tasks. One vital study in the medical sector has researched the use of computer vision in medical imaging to localize and examine human long bones. Images also come in the form of sound waves which are emitted by a B-scan sensor and are translated into a digital image more understandable to the human eye. However, a B-scan sensor is not able to distinguish between long bones and other specimens such as muscles, soft tissue, veins and other internal organs under the skin tissue. Hence, this study implements a deep learning system to be able to recognize bone specimens. This study discusses various other algorithms but makes use of YOLOv3 convolutional neural network algorithm to differentiate the said specimens in ultrasound images. (Human Bone Localization in Ultrasound Image Using YOLOv3 CNN Architecture, 2019)

Similarly, my research makes use of the same concepts this study does which classifies long bones, or in my case traditional Maltese food, in digital images.

***YOLOv3 Algorithm***

YOLOv3 is a regression-based one-stage target detection algorithm just like SSD (Single shot multibox detector) which can detect and classify objects in an image at the same time (Djebbar). YOLOv3 divides a digital image into S x S non-overlapped grids and determines whether each grid has a target to predict bounding boxes around that target (Zheng). Considering that YOLOv3 is a single end-to-end network that performs feature extraction, location and classification in a network, makes the algorithm extremely fast i.e. at 320x320 YOLOv3 runs at 22ms at 28.2mAP. (Miao)

A convolution neural network (CNN) is a deep learning network that assigns importance to an image such as weights and can differentiate one aspect of an image from another. YOLOv3 is one of the fastest object detection algorithms that make use of a CNN in real-time detection without loss of accuracy.

***Relevant Work***

A study by (Wang) highlights and compares other object detection algorithms with a view of identifying which is the better algorithm.

Computer vision is categorized into two types; region-based such as R-CNN and Faster RCNN and regression-based such as SSD and YOLOv3. Region-based target detection algorithms extract region proposals from top to bottom of a given image as a candidate region for the model to analyse, extract features in each proposal, classifies them and performs border regression on the region proposals thus this process takes a lot of time since it is divided into several stages. R-CNN is a region-based algorithm that uses a selective search algorithm that makes the process run too slow therefor making the overall object detection run slow.

Faster RCNN was introduced to improve the quality of R-CNN. This algorithm makes use of region proposal network (RPN) to select the candidate regions instead of the selective search algorithm which results in faster detection and enables end to end detection by a neural network. RPN reduces the number and improves the quality of region proposals (Miao). This method has a limited ability to extract other features and generalization hence a neural network-based algorithm using the deep learning model solves several problems.

A one-stage algorithm, SSD, results that the algorithm achieves better detection on smaller objects since the SSD generates more anchor points to make the object position more accurate. (Wang)

***Dataset***

Another study in the medical field using computer vision by (Djebbar) who researched cancer tumour in mammogram images using a convolutional neural network. This study explains the use of a digital database for screening mammographs to train and tests the CNN system. The dataset used contains 2,620 cases of breast imaging, including four mammograms for each case made up from non-malignant to critical cases in order to have accurate results. Indeed, deep learning requires a complex dataset to train a model to be more accurate. During this study, researchers have discussed techniques of augmentation to the training dataset where such techniques require for the dataset to generate new instanced using different transformation methods such as rotation of images, translation and scale (Heath). Considering that very limited images of traditional Maltese food are available online to scrape which results in a small dataset, this technique could improve the training process of my model by expanding and generating more images. These researchers discuss how the original dataset was augmented by rotating the mammograms three times with angles 90⁰, 180⁰ and 270⁰.

**Methodology**

To start the process of classifying different articles of Maltese food I had to establish which articles I would like my model to classify hence I choose the traditional Maltese cheeselet commonly known as ‘*ġbejniet*’ and another Maltese snack deep fried date roll’s commonly known as ‘*imqaret*’. Knowing which article of food my model can classify I could start the training process which requires a large set of annotated images to train. However, since classifying these types of food is not a very popular concept it was difficult to find a ready annotated dataset that consists of hundreds of images of cheeselet and date roll’s, thus I had to create my own.

**Dataset**

To acquiring a number of images I made use of an online tool called ‘*Fatkun Bach Download*’, this enabled me to search and scrape a number of images found on google then filter trough each image and decide their relevance under some circumstances i.e. angle of article of food, lighting, resolution, other food articles overlapping each other etc. Filtering through the scraped images I was able to delete irrelevant images downloaded unintentionally or duplicates of the same image. This process cause the dataset to have a smaller number than desired to train the model which results into a less accurate prediction. *solution: techniques of augmentation (rotate, scale)* *, taking pictures myself*

Annotation process enables the images to be labelled for the algorithm of choice to train the model. Before feeding an annotation tool with images to label, YOLOv3 requires images to all be the same size, this step was easily achieved by using ‘*Batch Resize*’ online tool to resize all images at 416x416. Making use of ‘*BBox-Label-Tool*’ is a great tool that annotated images ready for YOLOv3 by providing a JPG, PNG or JPEG image and highlighting a border around the article of food. In this process I made sure to highlight each induvial article of food alone as grouping a number of foods all together will influence the training. Image annotation is another way of filtering through each image and deciding their relevance. After highlighting the regions of food, the tool will provide a text file populated with the image label which includes the label of class (ġbejniet, imqaret), x-axis, y-axis, width and height.

**Configure and Train**

Training a YOLOv3 model requires a great amount of processing power hence I trained my model using Google Colab as this environment provides Googles own GPU’s for use. I opted for this method as my personal computer doesn’t have a graphics card thus training a model would have taken a long time. Before training the model, colab required a setup of installing darknet and uploading the dataset to darknet where two text files where generated for training and testing. Configuring darknet enabled me to provide the environment with the number classes and the destination to their names, destination of the backup folder to save weights and paths to train and test files. Finally it was time to train in darknet using yolov3.wieghts according to the yolov3\_custom.cfg file which consists of configurations and arithmetic solutions related to our model.

**Test (and Results?)**

# References,

Djebbar, K. (n.d.). *Deep convolutional neural networks for detection and classiﬁcation of tumors in mammograms.*

Heath, M. (n.d.). *Current Status of the Digital Database for Screening Mammography.*

(2019). *Human Bone Localization in Ultrasound Image Using YOLOv3 CNN Architecture.*

Miao, F. (n.d.). *Vehicle direction detection based on yolov3.*

(n.d.). *Single shot multibox detector.*

Wang, H. (n.d.). *A Comparative Study of Stateof-the-Art Deep Learning Algorithms for Vehicle Detection.*

Zheng, Y. (n.d.). *A Method of Detect Traffic Police in Complex Scenes.*