4.Discussion & Results

4.1 Object Classification

4.1.1 Metrics

After conducting all the experiments described in the methodology, the next step is to extract the results from each of the. In most studies, researchers opted to create a confusion matrix, where all the True Positives, True Negatives, False Positives and False Negative classifications would be noted. This deemed to be problematic because to identify such results, these had to be done manually by going through each image one by one. With a dataset of around four hundred and fifteen test images, this option would not be possible since it would take a very long time to do so, therefore this wasn’t possible.

Instead of using a confusion matrix, three tables each representing a configuration were created. Each table would show the number of “Actual Classes” which are the number of labelled food objects in all of the images and the number of “Predicted Classes” for each of the six food items. This would give a general idea of how the model performed because it would be possible to see if the model classified more objects, less or maybe even amount. These can be seen in the tables labelled “Configuration 1”, “Configuration 2” & “Configuration 3”.

Configuration 1

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Coin | Ghaq Tal-Ghasel | Imqaret | Pastizz | Zalzett Malti | Gbejniet | Qassatat | Totals |
| **Actual** | 413 | 138 | 184 | 137 | 92 | 0 | 160 | 1124 |
| **Predicted** | 416 | 137 | 170 | 146 | 46 | 3 | 166 | 1084 |
| **Difference** | +3 | -1 | -14 | +9 | -46 | +3 | +6 | -40 |

Configuration 2

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Coin | Ghaq Tal-Ghasel | Imqaret | Pastizz | Zalzett Malti | Gbejniet | Qassatat | Totals |
| **Actual** | 413 | 69 | 138 | 114 | 138 | 92 | 183 | 1147 |
| **Predicted** | 433 | 78 | 192 | 140 | 145 | 95 | 183 | 1266 |
| **Difference** | +20 | +9 | +54 | +26 | +7 | +3 | 0 | +119 |

Configuration 3

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Coin | Ghaq Tal-Ghasel | Imqaret | Pastizz | Zalzett Malti | Gbejniet | Qassatat | Totals |
| **Actual** | 413 | 92 | 92 | 137 | 161 | 92 | 137 | 1124 |
| **Predicted** | 413 | 93 | 93 | 206 | 160 | 92 | 137 | 1194 |
| **Difference** | 0 | +1 | +1 | +69 | -1 | 0 | 0 | +70 |

4.1.2 Result Extraction

These tables where generated by extracting the Actual classes from the mask found in the annotation files, whilst also getting the Predicted classes by looping through each image and trying to detect the items in each loop. These results were saved to a data frame so that it would be a bit easier to identify the classifications found in each image. Refer to *Table 1* for a visualization of these extracted results.

Table 1

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Image\_Id | Actual Class | Predicted Class | Total Actual Class | Total Predicted Class |
| 0 | Coin | Coin | 1 | 1 |
| 0 | Ghaq Tal-Ghasel | Ghaq Tal-Ghasel | 1 | 1 |
| 1 | Coin | Coin | 1 | 1 |
| 1 | Ghaq Tal-Ghasel | Ghaq Tal-Ghasel | 1 | 1 |
| 2 | Coin | Coin | 1 | 1 |
| 2 | Ghaq Tal-Ghasel | Ghaq Tal-Ghasel | 2 | 2 |
| 3 | Coin | Coin | 1 | 1 |
| 3 | Ghaq Tal-Ghasel | Ghaq Tal-Ghasel | 2 | 2 |
| 4 | Coin | Coin | 1 | 1 |
| 4 | Imqaret | Imqaret | 3 | 3 |

4.1.3 Discussion of Results

It can be noticed that from each of the configurations, the foods detected varied a lot between one another. There were two main reasons why this occurred. The first being that some objects where not even identified. In some cases, such as *figure 2*, food objects which were part of the six being used for this study, were present but not classified as one of the objects. As previously mentioned in the methodology section of this research, “fake objects” where also included as part of the dataset. This meant that there would be a chance that some of these fake objects would be falsely identified as one of the six focused items. One of these fake objects was a Green Apple, and this could be seen confusing the model, because when one of the images which had predicted an access amount of “Pastizzi” was visualised, it could be noticed that in most images where there is a green apple present, it was identified as a “Pastizz”. *Figure 3* shows an example of these images. This is the other main reason why the discrepancies between the Actual and Predicted values where present. The totals for these configurations where then added together and divide by three to get the one single value as an average of how well the models performed. The average of Actual Total objects was that of 1132 (1132.667, rounded upwards) and the average Predicted Total was 1181 (1181.333, rounded downwards). However, with this data it was still very difficult to understand how well the classification performed, but at least this gave an idea of which classes proved to be the most problematic to classify due to the difference in values. The values for “Qassatat”, “Gbejniet” and “Ghaq Tal-Ghasel” showed that these where most probably identified with the least number of problems encountered because they had the least difference in all three scenarios. This is most probably because, in configuration 2 and 3, in most cases whenever a green apple was present, this was miss classified as a “Pastizz”, hence why the +69 and + 26 difference can be seen in both scenarios respectively. For the large difference of “Imqaret” in configuration 2, this was because there where cases where the kiwi was also being detected as the class “Imqaret” with a lesser confidence level of around 80%. The “Zalzett Malti” on the opposite had less predictions in the first configuration, because in some images, multiple of class were present, but not all instances where identified. For example, in certain images where two of the “Zalzett Malti” class where visible, only one was identified.

4.2 Calorie Estimation

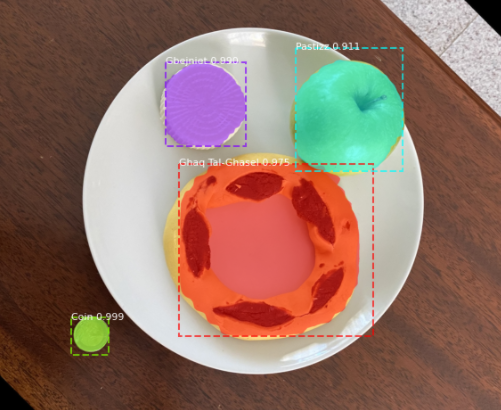
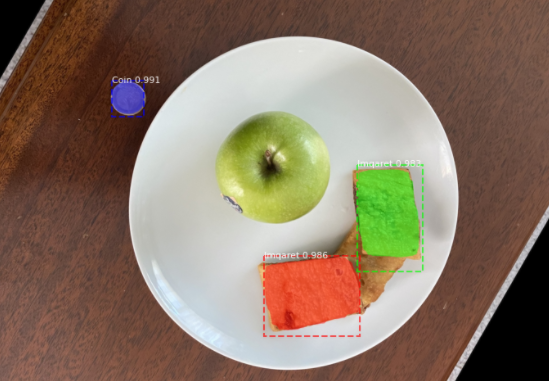


Figure 2

Figure 3

Since the predicted classes in each image where identified, the next step was to extract information from the results, this helped in estimating the calories for each class detected. Previously in the Methodology section of this paper, it was stated that the neural network, “Mask R-CNN” was used to generate the models because it was capable of segmenting instances whilst providing the exact number of pixels detected. This value was mainly used to generate the nutritional value, because in previous sections, the pixels per cm2 and the calorie per cm2 were found, thus by using those values and the number of detected pixels it was possible to generate these results.

4.2.1 Metrics & Results

To get the best idea possible of how the model performed, multiple approaches had to be taken to make sure that a better understanding of what effects the final metric. The first approach which was taken included a simple comparison of the total pixels from each item’s mask with the predicted pixels value. Since each of the testing images had a mask included in the annotation file, it was simple to extract the mask’s total pixels value from the polygon and the predicted mask was also part of the results when detecting the food items from every image. The issue in this was that, if pixels were detected outside of the original mask, those pixels shouldn’t be counted, because of this it wasn’t possible to just compare the values together. Here it was opted to use the Intersect over Union (IoU) evaluation metric to get the accuracy of pixels detected whilst only considering the pixels correctly detected on the mask. So, by using this metric the accuracy of the model’s instance segmentation was determined. Configuration 1 yielded an IoU percentage of 87.13%, configuration 2 resulted to 73.66% and configuration 3 showed 80.57% accuracy, therefore the average of all 3 configurations was that of 80.45%. This value shows that model performed quite well since it managed to identify a very high percentage of pixels correctly, this was the first indication that there would be a greater chance for the estimation of calories to be accurate.

The next approach was to calculate the calories predicted based off the detected mask, then the same would be done for the actual calories but by using the original mask as the reference point. To calculate the calories, the reference object had to be used to scale up or scale down the predicted pixel values depending on the IoU value for the coin (the reference object in this case). Once this was calculated, this new value was divided by 3349.092, which is the previously found pixels per cm2, to find an estimation of the actual size of the detected object. The same procedure was applied to the actual value so that two values would be generated, the “Actual Kcal” and “Predicted Kcal”. To compare these values together, it was opted to use the evaluation metric of “Root Mean Square Error” to get a single value which would be able to represent the difference between the “Actual” and “Predicted” Kcal’s. The formula for this metric used was, . By applying this formula for each configuration, three RSME values where generated, 80.36, 79.30 and 99.5. As previously done, these values where then averaged, which resulted in a final RMSE value of 86.39. This meant that on average the models created would have a calorie estimate of plus or minus 86.39. It is important to note that whenever multiple items from one food class are identified, the total calories for that food item are considered as a single aggregated value. For example, if two “Imqaret” where to be detected in a single image, only one pixel value for the total area of both would be extracted. The above information could be better understood in *table 2,* which includes the name and description of each column in the results sheet to better understand the process of calculating these results.

The metrics which have been previously presented show give a very good idea how the model performed overall, but these calculations still had something missing, this being a comparison with the real calorie value. This was achieved by weighing the objects used in the dataset and manually calculating how much calories each item had, to see how the real value would compare with both the actual and the predicted calorie value. When looking at the results, it could be noticed that the “Actual Kcal” and “Real Kcal” do not match even though in theory these should match due to the calculations which were done. This happened because in certain cases objects have different weights, but their size might be very similar. Since throughout this research the calories where always calculated based on the cm2 of the object, if weight varies but the size doesn’t, the calculated “Actual Kcal” will not change, whilst the “Real Kcal” is based off manually weighing them, making it a more accurate value. The same steps where repeated to calculate the RMSE between “Real Kcal” and “Predicted Kcal”, which resulted in the three values of 94.2, 138.24, 75.5 and an average of 102.65. *Table 3* shows a table recap of all the three mentioned metrics used to show how the chosen approach performed in calorie estimation.

Table 2

|  |  |
| --- | --- |
| **Column Name** | **Description** |
| Image\_Id | The id for each image when being prepared to be used by the classifier |
| Actual Class | The class of the food item which was annotated |
| Predicted Class | The class of the food item which was predicted by the classifier |
| Total Actual Class | The total number of each Actual Class found in every image |
| Total Predicted Class | The total number of each Predicted Class predicted in every image |
| Total Actual Pixels | The total number of pixels from the mask previously annotated. |
| Total Predicted Pixels | The total number of predicted pixels for the detected class. |
| Average Pixels Predicted | The IoU percentage of correctly detected pixels from the classifier |
| Pixels Per cm2 | The total number of pixels per cm2 previously gathered in the methodology section |
| Kcal Per cm2 | The number of calories per cm2 previously gathered in the methodology section |
| Coin Ratio | The Total Actual Pixels divided by the Total Predicted pixels for the coin in each image |
| Actual \* Ratio | The Total Actual Pixels multiplied by the Coin Ratio to scale up or down the Actual mask |
| Predicted \* Ratio | The Total Predicted Pixels multiplied by the Coin Ratio to scale up or down the Predicted mask |
| Actual Kcal | The (Actual \* Ratio) divided by the (Pixels Per cm2), multiplied by the (Kcal Per cm2) |
| Predicted Kcal | The (Predicted \* Ratio) divided by the (Pixels Per cm2), multiplied by the (Kcal Per cm2) |
| Real Kcal | The calorie value of each item manually weighed |
| Difference (Actual-Predicted) | The difference between the Actual Kcal and Predicted Kcal (Used as part of the RMSE equation) |
| Difference Squared | The difference squared (Used as part of the RMSE equation) |
| Total Average | The average value of the difference squared (Used as part of the RMSE equation) |
| RMSE | The square root of the total average |

Table 3

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Metric** | **Config1** | **Config2** | **Config3** | **Average** |
| IoU (%) | 87.13 | 73.66 | 80.57 | 80.45 |
| RMSE (Actual - Predicted) | 80.36 | 79.30 | 99.50 | 86.39 |
| RMSE (Real – Predicted) | 94.20 | 138.24 | 75.50 | 102.65 |

4.2.2 Discussion of Results

The results for each of the configurations showed some interesting results because there was not a single configuration which had the best results in all three metrics presented. Due to this, it was difficult to decide which of the three configurations would be chosen. As can be seen from *Table 3*, Config1 had the best IoU score and the second best RMSE (Actual – Predicted) value with just a difference of 1.06, but it was not as accurate as Config3 for the RMSE (Real – Predicted) portion of the results. This proved to be an issue because if someone wanted to compare which of the three configs detected the most correct pixels, it would be the first config. This is because 87.13% were correctly detected when compared to its respective mask, which was the highest value. However, for the purpose of this study, the most ideal configuration would be the 3rd configuration because it had the most accurate RMSE when comparing the real calories (which have been manually weighed) with the predicted ones, whilst also providing a decent IoU score. This meant that most of the pixels where correctly detected, and the predicted calories were very close to the real weighed values. The second configuration was not taken into consideration, even though it had the best RMSE (Actual – Predicted) value, it had the worst values in both IoU and RMSE (Real – Predicted). Something to also note is that even though only a limited number of four epochs where used, Mask R-CNN performed quite well when it came down to both classification and segmentation as could be identified from the average results. The main concern which was thought of upon starting the study was that, since some items look very similar in colour, the algorithm might have a hard time detecting the different classes, but this was in fact not the case as could be seen in these results. Having said this, as previously mentioned, there where cases where two objects which are not of the same shape and not of the same colour where miss identified for each other (the “Pastizzi” and the Green Apple). This could have been because of other factors such as undertraining or maybe that particular random split had images which were not varied enough.

Comment