Assignment 5: Ear Detection and Classification

Stefanela Stevanović
IBB 22/23, FRI, UL
ss51676@student.uni-lj.si

I. INTRODUCTION

The goal of this assignment is to combine methods used in assignment 2 and assignment 1 to perform object detection, feature extraction and image classification. Since YOLO detector gave good result in assignment 2, this detector was also used in this assignment. In the assignment 1 local binary pattern (LBP) texture descriptor was used and rank-1 accuracy was computed using Euclidean distance measure. In this assignment we try to get improvement in image classification using Support Vector machine on LBP features. Support Vector Classification (SVC) is a supervised learning algorithm that can be used for image classification task. It works by finding the best hyperplane to separate the data into different classes.

II. METHODOLOGY

Firstly, ground-truth boxes with the ear coordinates were computed for all the images in the dataset, by drawing a rectangle around the white areas in masks. Then, YOLO detection method was set-up and evaluated by calculating by calculating the Intersection over Union between the detected object box and the obtained ground-truths. IoU (Intersection over Union) is a term used to describe the extent of overlap of two boxes and is calculated as:

$$IoU(\%) = \frac{area\ of\ overlap\ of\ two\ boxes}{area\ of\ union\ of\ two\ boxes} \cdot 100 \quad (1)$$

IoU has value between 0% and 100%, where 0% means no overlap and 100% means complete overlap of two detection boxes. The accuracy of predictions for one method was calculated as:

$$ACC\ (\%) = \frac{TP}{TP+FP+FN} \cdot 100 \quad (2)$$

Where TP represents number of true positives (number of correct predictions), FP number of false positives (detected object is not a wanted object) and FN number of false negatives (wanted object is not detected on an image). Number of true positives is determined based on selected IoU threshold, such that predictions with IoU ≥ threshold are considered true positives and predictions with IoU ≤ threshold are false.

Rectangle coordinates from YOLO detector were used to crop the images, so that they would only contain detected object – ear. On the cropped images Skiimage's local binary pattern was performed and the rank-1 recognition rate was computed. Since the images are two-dimensional matrices of pixels, they were transformed into one-dimensional feature vectors that were compared with each other. For each vector the closest vector to the currently observed one was found using Euclidean distance measure and it was checked whether they belong to the same class. This was repeated for all the vectors and the percentage of correct predictions gave a rank-1 recognition rate. With the intention to improve class prediction, supervised classifier Support Vector Machine from sklearn library was used on LBP features and on the original input images, for the comparison purpose. Prior to implementing SVC, LBP features and original image pixels were normalized, to help reduce the impact of the scale of the features and to find a more optimal decision boundary, which leads to better performance of the support vector classifier. Since some of the folders in the dataset are missing, LabelEncoder was used to encode output (class) values to ordinal variables.

III. EXPERIMENTS

The dataset contains around 14300 images belonging to 569 classes. 7 folders that contained less that 6 images were removed, leaving us with 562 classes. The code was written in the Python programming language (version 3.9.7) in the Spyder environment. All images were loaded in grayscale format in order to overcome memory issues.

YOLO v5 detector was set up by running *detect.py* script from git repository https://github.com/ultralytics/yolov5 and the detection was evalued by calculating IoU (1) between returned labels and ground truth boxes for each detection. To make sure YOLO detects an object on each image confthres parameter was set to 0.00 and max-det parameter was set to 1, to make sure we only have one detection per image. Afterwards detection accuracy was calculated with different thresholds between 0.1 and 1.0, using formula (2).

Original images were cropped based on labels (boxes) returned by YOLO detector and resized to 64x64 image size. SVC with the linear kernel was implemented on normalized cropped images for the comparison purpose. Each SVC experiment was conducted by separating the data into train and test set, with the training set containing 80% of the data. When separating training and testing data, stratification was used to ensure that the distribution of target variable is similar in the training and testing sets. Following experiments were conducted:

1. LBP (image size = 64x64, R = 1, L = 8) + SVC
2. LBP (image size = 32x32, R = 1, L = 8) + SVC
3. LBP (image size = 32x32, R = 2, L = 8) + SVC

IV. RESULTS AND DISCUSSION

In this part of assignment the results obtained in the experiments described in chapter III are presented and discussed. YOLO detection method performed with 99.7% accuracy with the IoU threshold of 0.5. Figure 1 shows YOLO accuracies with different thresholds between 0.1 and 1.0 with the step of 0.1. We see that with thresholds lower than 0.4 accuracy is 100%, while with thresholds greater than 0.5 accuracy starts dropping rapidly.

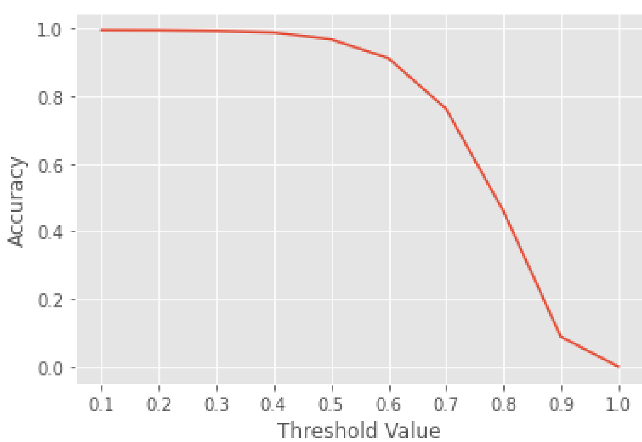

Fig.1. YOLO detection accuracy for different IoU thresholds.

Figure 2 is a histogram showing the number of detections with each IoU value, where we see that most of the detections have IoU value between 0.8 and 0.9, indicating good YOLO performance.

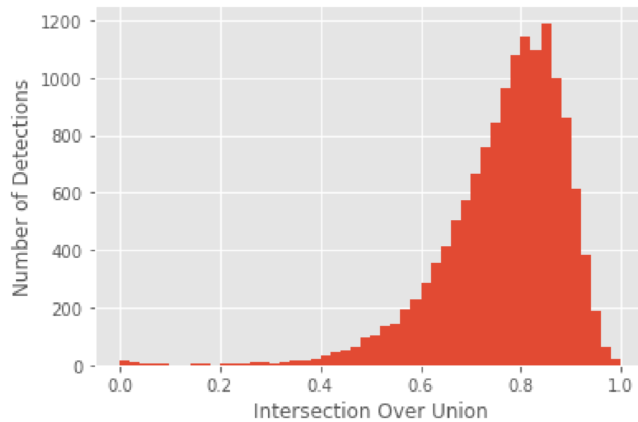

Fig.2. Intersection Over Union distibution.

Table 1 shows the results of the experiments with the parameters specified in section III. This table shows the rank 1 accuracy when using Euclidean distance for finding the most similar image and the classification accuracy when using SVC. From the table we can see that for the LBP ranking by Euclidean distance measure, we get much lower rank-1 accuracy than in the first assignment, which was expected considering that now more classes are present in the dataset used in experiments.

TABLE 1
Rank-1 accuracy with Euclidean distance measure and SVC classification accuracy with different input features

Features	**Image size**	**Rank-1 accuracy (%)**	**Classification accuracy with SVC (%)**
Original image pixel vector	64x64	-	0.3
$LBP_{1,8}$ feature vector	64x64	9.5	13.9
$LBP_{1,8}$ feature vector	32x32	11.6	13.4
$LBP_{2,8}$ feature vector	32x32	15.6	16.0

We also see that the SVC classifier has a very poor performance when the original image is used as input, which changes drastically by taking the LBP feature vector as the classifier input.Using a combination of LBP and SVC improved the accuracy significantly, especially in the first experiment. It’s interesting to note that the ranking by euclidean distance measure had a higher accuracy on the 32x32 images with the same LBP parameters, while the SVC classifier had a better performance on the 64x64 images. For LBP with the radius of 2, we see that both rank-1 accuracy and SVC classification accuracy have improved significantly.

V. CONCLUSION

In this assignment, it was shown that with the combination of LBP and SVC, it is possible to achieve greater classification accuracy than by using plain distance measure between LBP features. Although the highest achieved accuracy of 16% is not good enough for direct use, it shows that improvement are certainly possible. Given that LBP + SVC model has been shown to work better on larger images, somewhat better performance could probably be achieved using larger image size with LBP radius od 2 or more. However, to achieve more significant improvements, we would probably need a CNN.