Deep learning based segmentation of tomatoes images and recognition of tomatoes using classical methods of Computer Vision

Iva Eftimska

March 21, 2023

Abstract

UNET segmentation model was used for segmenting the tomatoes from the Laboro Tomato dataset with different maturity levels. The original dataset is relatively small consisting of 643 train images and 161 test images. To avoid overfitting of the model, data augmentation techniques were applied to increase the number of images in the training set. Recognition of tomatoes was done using HOG features and 1-NN classifier. For this problem from the Laboro Tomato dataset croppped images were generated for tomatoes and backgrounds using the bounding boxes included in the dataset, so the localization problem, here is excluded. UNET's segmentation model showed performance of 0.89 IoU score. For the recognition problem the obtained results are, TPR: 0.9 and FPR 0.1845.

1 Introduction

Tomato segmentation is a crucial step in automatic harvesting systems, disease detection, different maturity levels classification. Due to its rich healthy ingredients tomato is one of the most used fruits. With the help of deep learning methods different tasks within agriculture, as harvesting systems, will significantly faster the process and avoid time-consuming tasks of a farmer. Researchers worked hard to develop accurate methods for the fruit segmentation task. The segmentation process becomes more challenging when the fruit is on the natural background mingled with leaves and stems and green tomatoes has the same color as leaves and stems, which is exactly the case in our dataset. Since that fact, recognition of tomatoes plays an important role in automatic harvesting systems, so the system will be able to correctly pick up the tomato and not the leaves that the tomato was mingled with. Localization and detection of tomatoes in an image is a crucial part of the segmentation process, since it allows us to evaluate the performance of the segmentation method, with the bounding boxes of the tomatoes from the ground truths and outputs of the model. The bounding boxes of the tomatoes in ground truth could be obtained automatically using localization and detection methods or manually.

2 Related work

Several authors worked on fruit segmentation task. In paper [1] authors used multi operation of edge detection and reconstruction for tomato segmentation on natural background. Performance of the mentioned method with the various maturity level were also evaluated in that paper. Authors in paper [2] used U-Net and Modified U-net for detection and segmentation of disease-affected regions in leaf images for tomato plant disease. Blueberry fruit was recognized using HOG descriptors and color features in outdoor scenes in paper [3]. Newly developed TMWE (Template Matching with Weighted Euclidean Distance) classifier was proposed in that paper, which gave a higher accuracy at lower computational cost. This paper regarding to recognition task of tomatoes is most similar to [4], while regarding to the segmentation task of tomatoes is most similar to [2].

3 Methods

U-Net was firstly used for biomedical image segmentation [5]. The U-Net architecture in this paper is shown below.

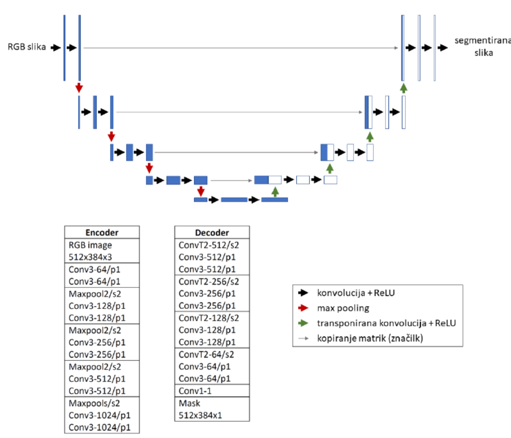

Figure 1: U-Net architecture

For training the U-Net network read matrices for RGB tomato images and masks were converted to tensors and normalization was done on RGB images. Adam optimization method was chosen with a learning rate of 0.0001. For loss function was chosen Binary Cross Entropy and network was learned 100 epochs with batch size of 2. For the tomato recognition problem HOG features were obtained for every image in training and test set and using the 1-NN classifier the unknown sample was classified based on the Euclidean distance of the known training samples. HOG features for backgrounds and tomatoes were concatenated for both training and test set. Before obtaining the HOG features the contrast of images was enhanced using histogram equalization, to improve the performance, so the algorithm can distinguish better green tomatoes from background images that are also green as stems and leaves. Also all images in train and test set were resized to 64x128 before obtaining the HOG features. The number of orientations was chosen to be 12, to better describe the shape of the tomatoes, in general, all tomatoes from the same type have the same shape, but some of them are also occluded by other tomatoes and leaves, increasing the number of orientations helps to get an overall good shape for the occluded tomatoes as well.

4 Experiments and results

Laboro Tomato dataset consists of 643 train images and 161 test images. Dataset was gathered at a local farm with two separate cameras with its different resolution and image quality. The images were captured at bright conditions by day, with sufficient lighting. The dataset used in this paper differs from the original dataset in that way that all images are resized to 512x384 pixels. Images contain tomatoes with different size and maturity level. In the mask folder, we have binary images, where tomatoes represent white pixels, while the background is represented by dark pixels. The ground truth folder contains for every image segmentation masks and bounding boxes of the tomatoes. More information about the annotation details could be found here [6]. The size of the training set was increased using data augmentation techniques. Transforms that were applied in order to generate new train images were chosen in that way that will represent better real world examples. So for instance when applying RandomBrightnessContrast transform, the limits for changing the brightness and contrast of the image were chosen carefully so the newly generated images are not supposed to be excessively dark or bright, the model should be able to distinguish the tomatoes from background. The same goes for applying the ShiftScaleRotate transform, where the limits for shifting in x and y direction were chosen carefully so the tomatoes must be visible in the image after applying the transform. For every image HueSaturationValue transform was applied, hoping that the model will be able to better distinguish green tomatoes from green leaves when changing randomly hue, saturation and value of every image, so the green colors of the leaves and tomatoes will be more distinguable. The total number of images in the train set are 2572, while in the test set remain the same 161. The model achieved the best accuracy in the 86 epoch with an IoU (Intersection of union) of 0.891. The best accuracy of the model shown through epochs is shown below.

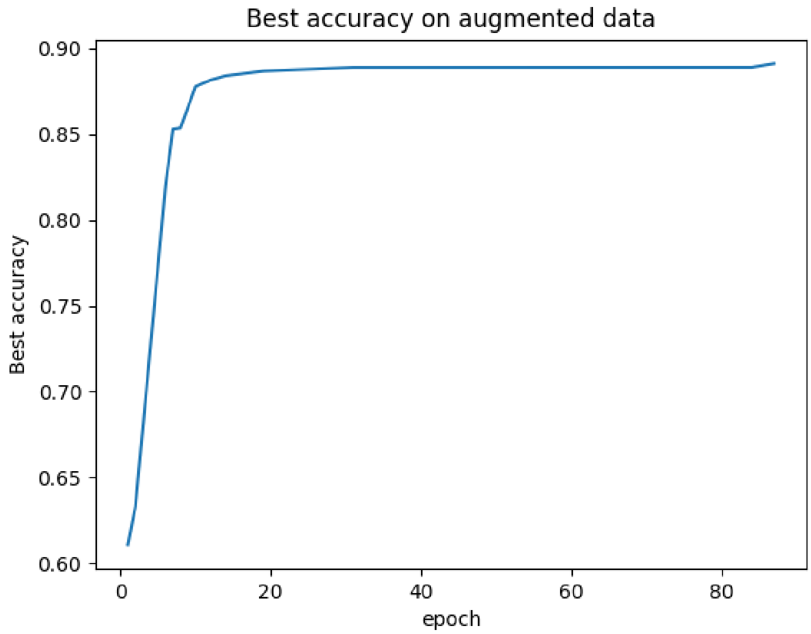

Figure 2: Best accuracy of the model through epochs

The performance of the model should be taken with a grain, since the model segmented tomatoes that were not present in the ground truths or has segmented part of the leaves. Some examples are shown in figure 3.

For the recognition of tomatoes problem the obtained True Positive Rate is 0.9 and False Positive Rate is 0.1845. Here the tomato images were generated based on the box coordinates of tomatoes given in the dataset and the background images have the same size as tomato images with no overlapping regions with the bounding boxes of the tomatoes. The total images in train set for both backgrounds and tomatoes are 7781, while in test set are 1996. Some green tomatoes the classifier recognized as backgrounds due to the similar color with stems or leaves, therefore the HOG features were similar, some examples of false negatives and false positives are shown in figure 4 and 5.

5 Conclusion

The U-Net model gave good results since it was meant for semantic segmentation tasks. It should be considered that maybe the ground truths of the images in the dataset need to be more accurately labeled, so all tomatoes in an image should be labeled, to compare with the outputs of the model gave, because

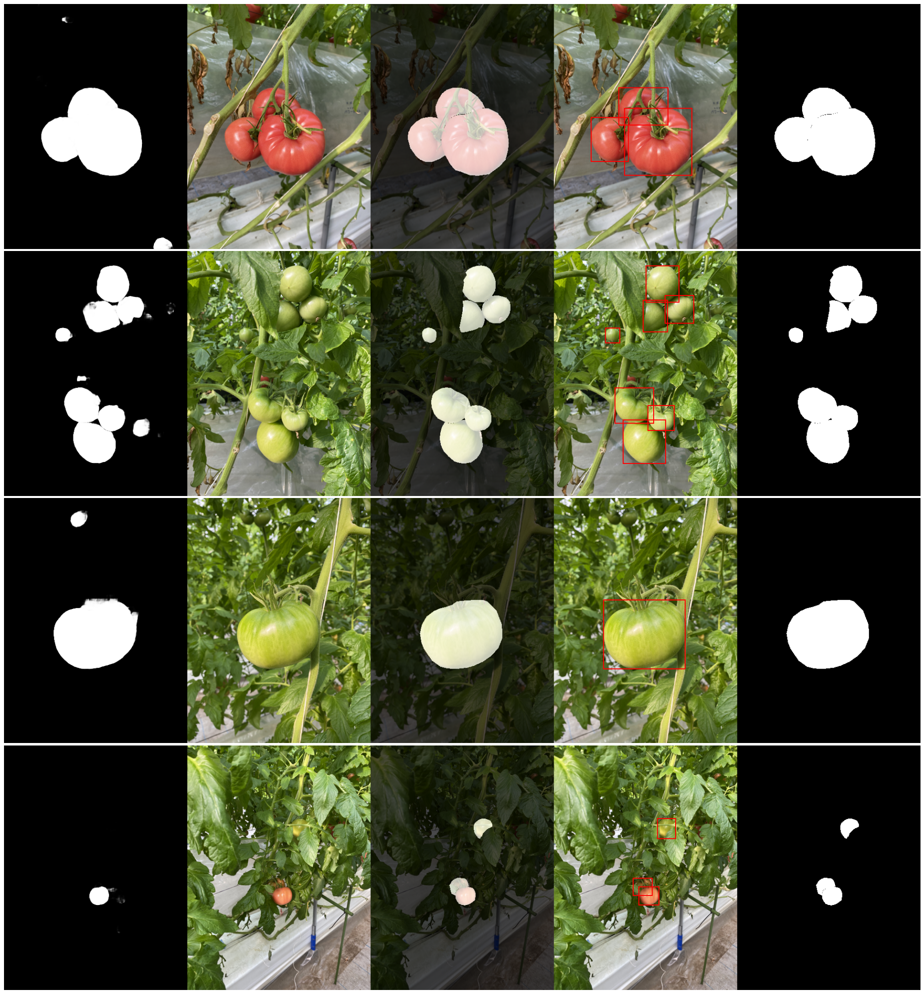

Figure 3: The most left image is the output of the model, the following four are the ground truths

Figure 4: False negatives

the model also segmented tomatoes that were not labeled in the ground truths. The classification of tomatoes with HOG features and 1-NN classifier gave good results. It gave wrong classifications of green tomatoes due to the same color of some background images as stems and leaves. Some combined color features and HOG features will maybe solve this problem, so the classifier should be able to distinguish between the green colors of tomatoes and stem/leaves.

Figure 5: False positives

6 References

[1] Anindita Septiarini and others. Tomato Segmentation on Natural Background Using Multi Operation of Edge Detection And Reconstruction.

[2] Muhammad Shoaib and others. Deep learning-based segmentation and classification of leaf images for detection of tomato plant disease.

[3] Kezhu Tan and others. Recognising blueberry fruit of different maturity using histogram oriented gradients and colour features in outdoor scenes.

[4] Guoxu Liu and others. A Robust Mature Tomato Detection in Greenhouse Scenes Using Machine Learning and Color Analysis.

[5] Olaf Ronneberger and others. U-Net: Convolutional Networks for Biomedical Image Segmentation

[6] https://github.com/laboroai/LaboroTomato