

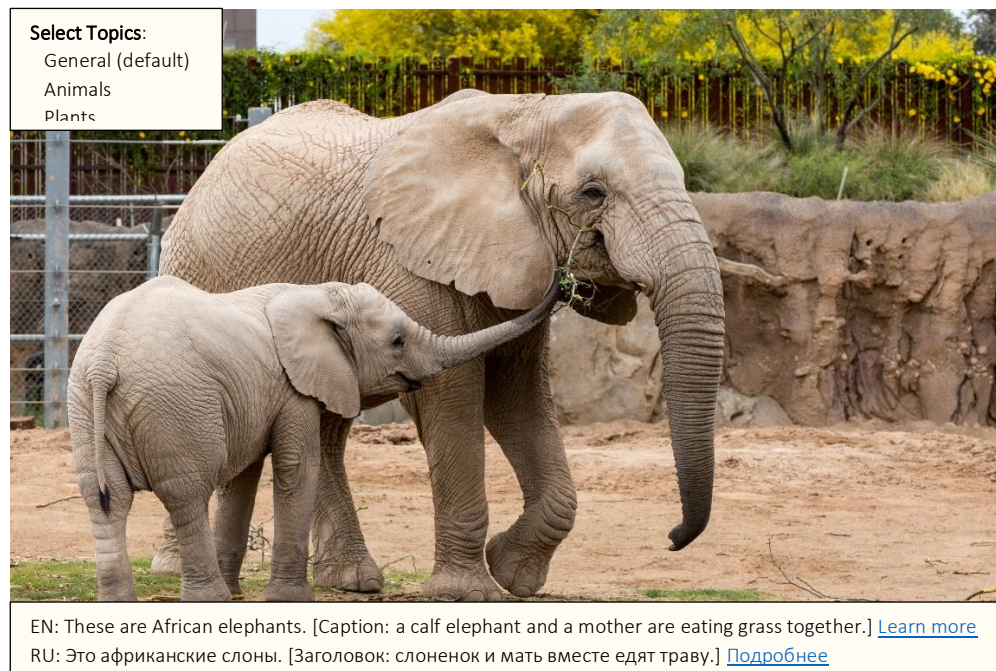
The Project is Based on Image Classification:

Explore Surroundings with Visual Dictionary

Alisher Siddikov

Introduction:

This project is a combination of image classification and translation launched on the iOS edge device. This project is an initial step to build a visual dictionary app by using Create ML and XCode. Create ML helped me train my model, and XCode helped me deploy my model to the iOS edge device. This app should help children and adults to explore their surroundings in their native and any other languages found in Google Translate. My final objective is to use the iPhone camera to detect surrounding objects, populate caption, translate it to the user-selected languages, and provide additional information about the detected object from Wikipedia. For example, if children visit Zoo, they can take their camera and point to the animal, let's say to the elephants. It should detect the elephant and translate it to their selected language. The goal is to combine translation and Wikipedia information with object detection and help people explore their surroundings with a new language. It is an exciting way to learn about surroundings with a live visual dictionary.



Object Detection:

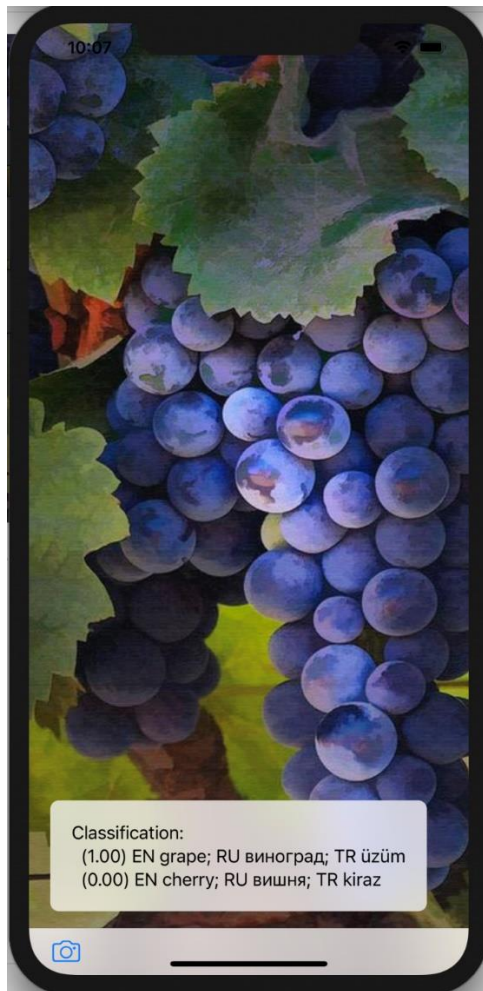
Solution:

I started testing the idea by classifying the fruits. So, I have four classes: cherries, grapes, apples, and strawberries. The training dataset is composed of 100 color images of the fruits with different settings. The validation dataset is composed of 40 images. I trained the fruit data without data augmentation, and I reached 95% accuracy within a few seconds. I also trained the

model with data augmentation such as blurring, cropping, flipping, and rotating the images, and it took about 1.5 hours; I did not see accuracy improvement. Therefore, I select the model which trained without data augmentation - higher accuracy and faster computation. The testing/evaluation with new fruit images was pretty accurate.

The failed solution:

I was not able to deploy the Google AutoML trained model with translation onto the iOS edge device. CoreMLTool model editing and model conversion package did not solve the issue. There were very limited online resources on this issue.



Conclusion and Future Development

My idea was tested successfully. In the future, I am planning to use the pre-trained models such as MobileNet, ResNet, GoogleNet, and apply transfer learning to the selected topics along with translations. Also, Azure computer vision has a high success on vision captioning, which I will integrate into my future project.

GitHub: <https://github.com/Siddikov/MSDS462>