Practical ML & DL. Deliverable 1.1

Lada Morozova, Rizvan Iskaliev, Daniil Igudesman

October, 2022. GitHub

The proposed project is consisting from two main parts: object detection and image classification. Mainly our team was focused on researching and developing the most suitable classification model for our task. Moreover. through the first iteration of the project the data preparation and additional research was performed.

Data research

During the first iteration our team researched the available and suitable data for the project. Finding data for image classification was quite easy, since a sufficient number of good datasets for ASL are available. As the main dataset for image classification the ASL alphabet dataset was selected. It contains 87000 images with resolution 200x200. The following dataset contains 29 classes (A-Z letters, space, delete, nothing).

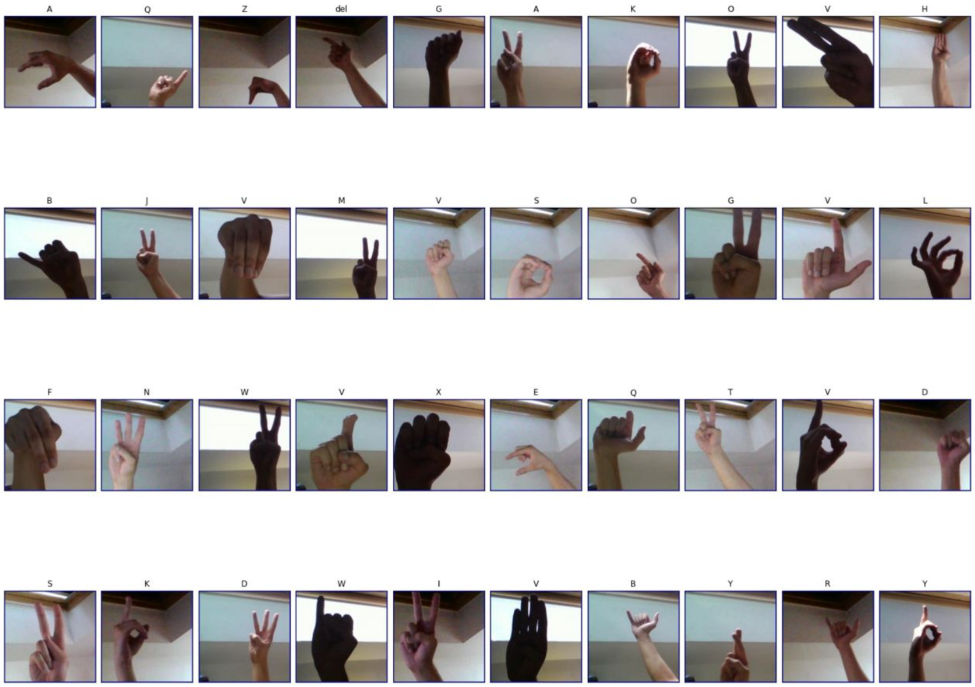


Figure 1: Sample from the classification dataset

During the research of the data, we discard some datasets for detection, e.g. EgoHands and COCO-WholeBody, as the content is not suitable for our task. However, we found the dataset which possible will be more effective for our task. ASL detection dataset contains bounding box annotations of ASL classes. However, for our purposes, we may convert it to a one-class dataset, in order to only detect a hand in the wild. In order to understand suitability of the following dataset to our task we perform analysis of its distribution.

The sample images from the dataset with annotations:

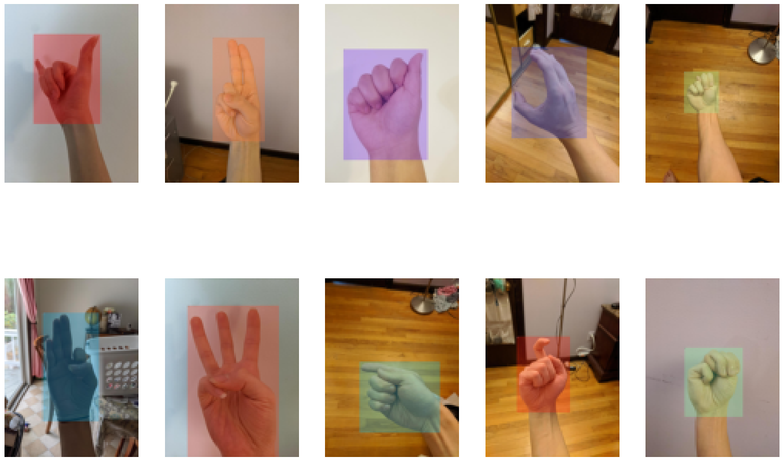


Figure 2: Sample from the detection dataset

The following graphs present a distribution of the images' sizes in the dataset, which gives us an information about the best possible resolution for test data. The heatmap of the bounding boxes' centers allow to understand the quality of the dataset in terms of the relative spatial position of the bounding box on the image. The bottom plot allows to estimate the relation between class of the bounding box and its size, which give an insight of the suitable anchors for object detection model.

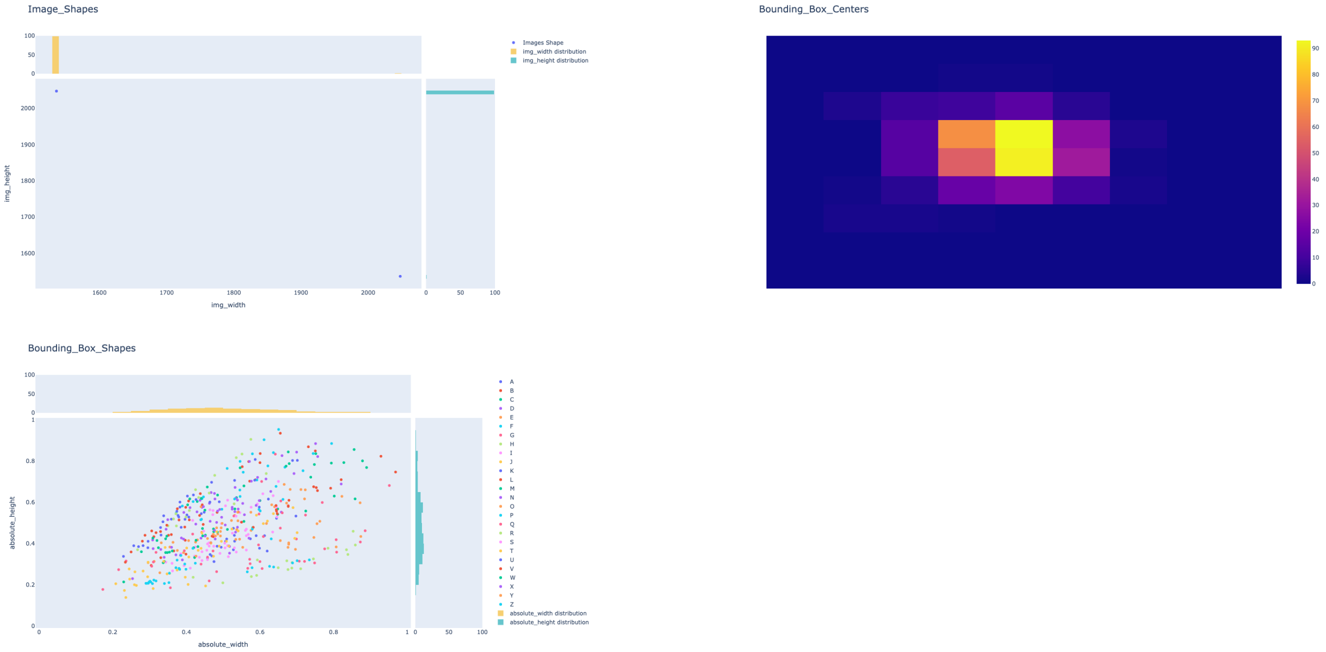


Figure 3: Distributions in the data

Image classification

As implementation of the image classification model two models were tested MobileNetV2 and VGG16. The following models were trained on the ASL alphabet dataset with pretrained weights. MobileNetV2 and VGG16 are convolutional neural networks with classification heads.

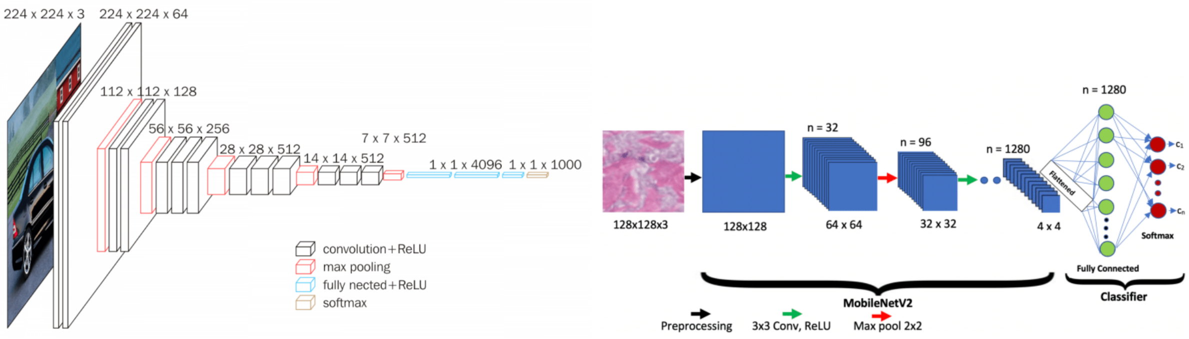


Figure 4: VGG16 and MobileNetV2 architectures

MobileNetV2 training

We used pretrained ImageNet weights, Adam optimizer with learning rate=0.001 and categorical cross entropy loss. We trained MobileNetV2 model for 5 epochs.
The following model obtained 96.22% accuracy on the test set.

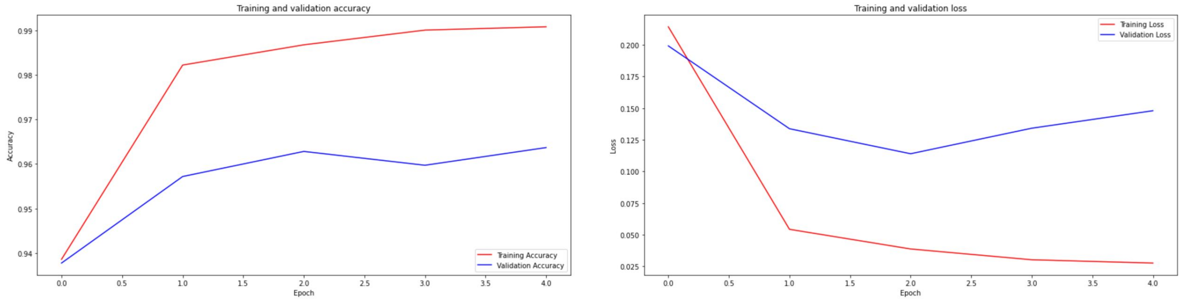


Figure 5: Accuracy and loss of the MobileNetV2 during training and validation steps

VGG16 training

We used pretrained ImageNet weights, Adam optimizer with learning rate=0.0001 and categorical cross entropy loss. We trained VGG16 model for 10 epochs.
The following model obtained 99.1% accuracy on the test set.

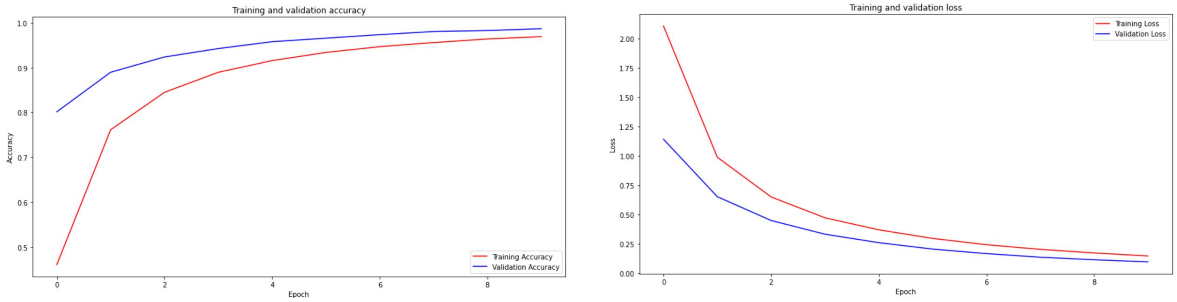
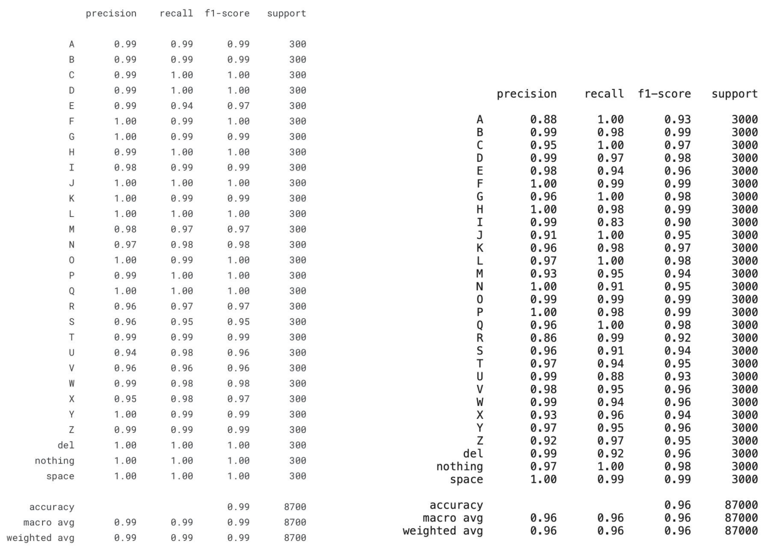


Figure 6: Accuracy and loss of the VGG16 during training and validation steps

Results



```
              precision    recall  f1-score   support

           A       0.99      0.99      0.99       300
           B       0.99      0.99      0.99       300
           C       0.99      1.00      1.00       300
           D       0.99      1.00      1.00       300
           E       0.99      0.94      0.97       300
           F       1.00      0.99      1.00       300
           G       1.00      0.99      0.99       300
           H       0.99      1.00      1.00       300
           I       0.98      0.99      0.99       300
           J       1.00      1.00      1.00       300
           K       1.00      0.99      0.99       300
           L       1.00      1.00      1.00       300
           M       0.98      0.97      0.97       300
           N       0.97      0.98      0.98       300
           O       1.00      0.99      1.00       300
           P       0.99      1.00      1.00       300
           Q       1.00      1.00      1.00       300
           R       0.96      0.97      0.97       300
           S       0.96      0.95      0.95       300
           T       0.99      0.99      0.99       300
           U       0.94      0.98      0.96       300
           V       0.96      0.96      0.96       300
           W       0.99      0.98      0.98       300
           X       0.95      0.98      0.97       300
           Y       1.00      0.99      0.99       300
           Z       0.99      0.99      0.99       300
         del       1.00      1.00      1.00       300
     nothing       1.00      1.00      1.00       300
       space       1.00      1.00      1.00       300

    accuracy                           0.99      8700
   macro avg       0.99      0.99      0.99      8700
weighted avg       0.99      0.99      0.99      8700
```

```
              precision    recall  f1-score   support

           A       0.88      1.00      0.93      3000
           B       0.99      0.98      0.99      3000
           C       0.95      1.00      0.97      3000
           D       0.99      0.97      0.98      3000
           E       0.98      0.94      0.96      3000
           F       1.00      0.99      0.99      3000
           G       0.96      1.00      0.98      3000
           H       1.00      0.98      0.99      3000
           I       0.99      0.83      0.90      3000
           J       0.91      1.00      0.95      3000
           K       0.96      0.98      0.97      3000
           L       0.97      1.00      0.98      3000
           M       0.93      0.95      0.94      3000
           N       1.00      0.91      0.95      3000
           O       0.99      0.99      0.99      3000
           P       1.00      0.98      0.99      3000
           Q       0.96      1.00      0.98      3000
           R       0.86      0.99      0.92      3000
           S       0.96      0.91      0.94      3000
           T       0.97      0.94      0.95      3000
           U       0.99      0.88      0.93      3000
           V       0.98      0.95      0.96      3000
           W       0.99      0.94      0.96      3000
           X       0.93      0.96      0.94      3000
           Y       0.97      0.95      0.96      3000
           Z       0.92      0.97      0.95      3000
         del       0.99      0.92      0.96      3000
     nothing       0.97      1.00      0.98      3000
       space       1.00      0.99      0.99      3000

    accuracy                           0.96     87000
   macro avg       0.96      0.96      0.96     87000
weighted avg       0.96      0.96      0.96     87000
```

Figure 7: Resulted metrics of the VGG16 and MobileNetV2 trainings

Example predictions

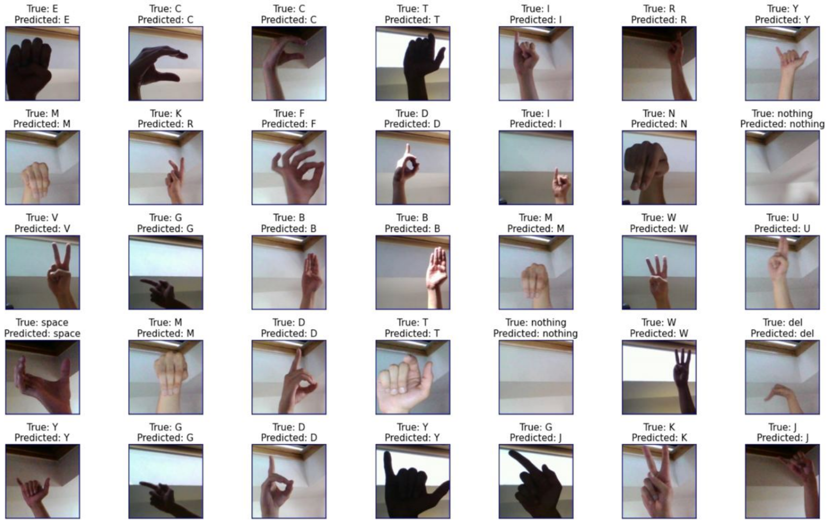


Figure 8: Example predictions

Contribution:

- Daniil Igudesman - classification data exploration, VGG16 training
- Lada Morozova - MobileNetV2 training, report
- Rizavan Iskaliev - data research, object detection data exploration