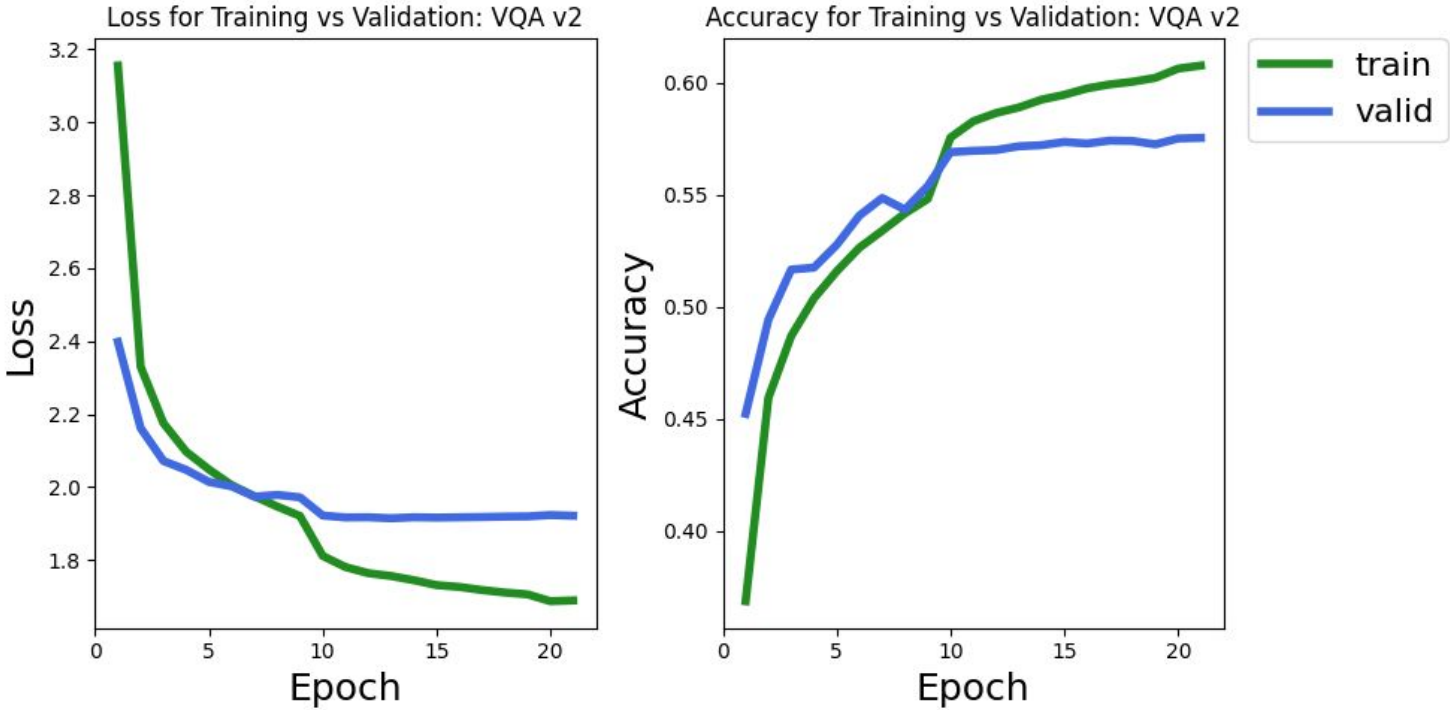
Sprint 3

VQA for Visually Impaired

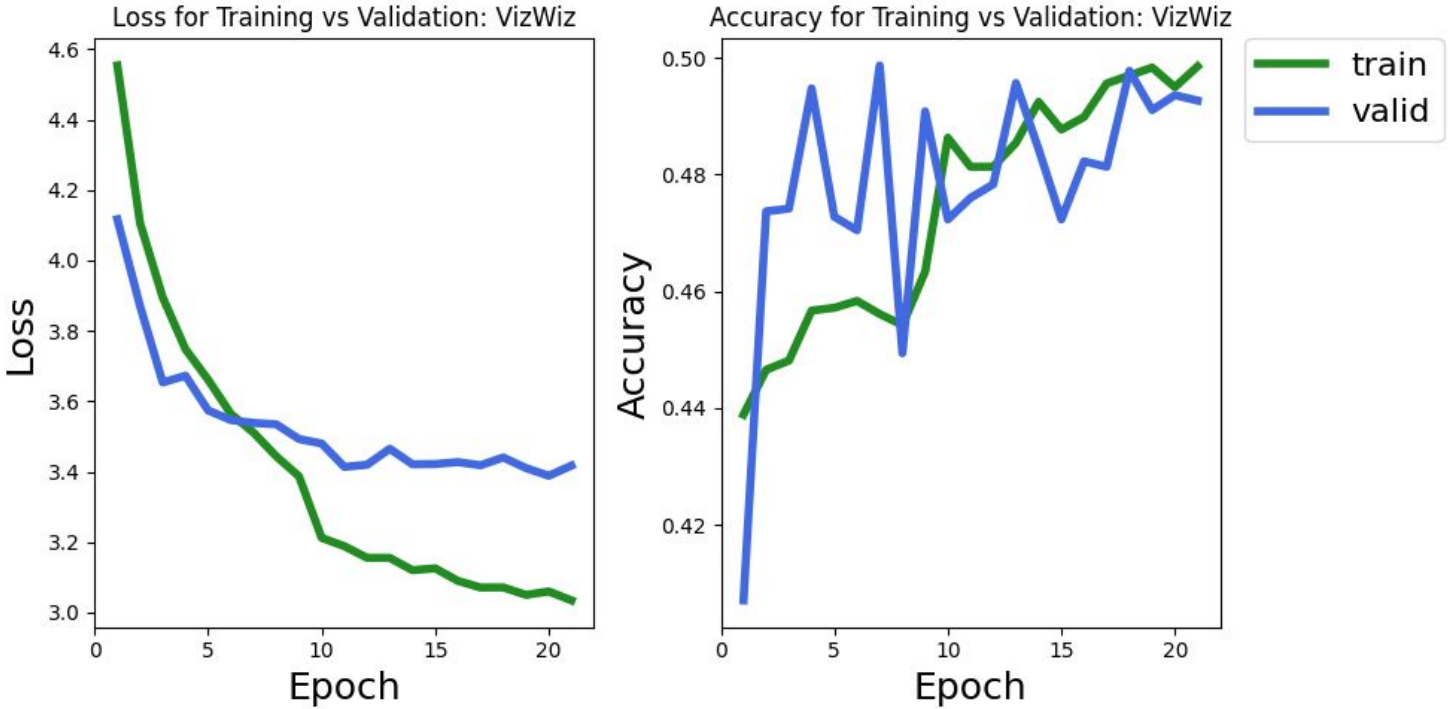
What have we done:

- Reformatted VizWiz data set to fit the structure of VQA v2
- Reformatted VQA v2 data set to fit the structure of VizWiz
- Trained and validated the VizWiz data on one of our models
- Continued working on the state of the art model VinVL [https://arxiv.org/abs/2101.00529]

Original VQA: Results - VQA v2



Original VQA: Results - VizWiz



Conclusions: Results - VizWiz

- Size: The first aspect that we can notice when comparing the two datasets is the large difference in size. VQA v2.0 contains almost 35x more visual questions than VizWiz.

- Questions: in VQA v2.0 have a small set of common initial words (e.g. “What”, “How”, “is” . . .). While in VizWiz often they start with a rare word: 28% of VizWiz questions starts with a term that occurs less than 5% of the times

- Answers: Average answers in VizWiz are longer.

Conclusions: Quality of Images

Original VQA: Test - VizWiz


```
Question:
what is the temperature?
Answer:
 unanswerable <unk> thermostat unsuitable 79
```


```
Question:
Is this my cough and cold medicine?
Answer:
 unanswerable unsuitable vitamins <unk> soup
```

Testing Procedure:

- Train and validate on VQA - v2

- Train and validate on VizWiz

- Train on VQA - v2 and validate on VizWiz

- Train on VQA -v2, fine tuning on VizWiz and validating on VizWiz

Sprint 4 Goals:

- Continue with the testing procedure
- Compare results with different models
- Try to fine tune model