02466 Project work in Artificial Intelligence and Data LOGBOOK

Khalil Ahmadi, s204161@student.dtu.dk
Michael Damholt, s204151@student.dtu.dk
Mathias Hadi Dyhr, s204156@student.dtu.dk
~~(Christoffer Wejendorp, s204090@student.dtu.dk)~~ - left the course

The main purpose of the logbook is that it serves as a tool for you to keep track of the project and document project meetings.

Project Meetings

Week 2: 09.02.22
We made a draft for our "Samarbejdsaftale".

We had questions about general info regarding the project, e.g. 'how does the data set look like?', and decided to write a mail to our supervisor, Maxim.

We agreed that we would meet next Wednesday, having the necessary material read such that the project plan can be completed. (Chapter 2 to 5 in Styrk Projektarbejdet and the Canvas Manual)

We agreed to research about Gantt charts.

Week 3: 16.02.22
We completed as much of the ProjectCanvas as possible. Still needs milestones and actions. We assume this will be easier to complete after the data has been explored more.
We initiated the Gantt Chart but we need more clarification on some of our activities.

Doing the meeting we received a part of the complete data set from the supervisor. Based on this we agreed to look through data and get a better understanding of the structure of the data and think of ways to solve the problem.

Week 4: 22.02.22

We met today for the planned Q&A session (for all the project groups) with our course supervisor, Morten Mørup, and he answered some doubts about what our project plan should include. We spent much time after the Q&A session looking into what our project description should include and went through some of the papers our supervisor sent us to form research questions. Assessing the main problem of our project, we found out the solution could be split into two parts - Detecting the barcode (a Scene Text Recognition (STR) model is needed) and recognizing the numbers (an Optical Character Recognition (OCR) model is needed). We discussed a lot about what specific STR- and OCR models we would like to implement and compare for this problem and decided it would be best to make the decision tomorrow, after reading through all the research papers that our supervisor had sent us yesterday (21.02.22). Barring this important decision, we worked on everything else, which we could work on, on the project description and found that we got a much better (more concrete) overview of how our plan would look like.

Week 4: 23.02.22

We finished the project plan including: the Project canvas, the Gantt Chart, Learning Outcome and the Project description. Completing this plan included long discussions with different opinions on how to conduct the project and the end-goal.

We will use the project plan as the basis for the discussion with our supervisor in 2 days.

Week 5: 02.03.22

The meetings this week will be skipped in favor of working on assignments in other courses. (This week was filled with deadlines of assignments) This was planned and is also a part of the Gantt chart.

Week 6: 09.03.22

We delegated responsibilities.
- Khalil/Christoffer - Will attempt to create models to crop down the pictures. (Cut out unimportant content in the pictures)
- Mathias/Michael - Attempt to use pretrained OCR-models to extract the necessary information. (So far this seems to be slow and includes too much information)

Week 6: 11.03.22

Unfortunately, one of our group members, Christoffer has withdrawn from the course due to personal reasons.

Week 7: 16.03.22

We have decided to be more structural in our planning of work assignments.
That means that we will in the future will be better at delegating the assignments with better deadlines – We have now updated our Gantt-chart and will make sure to keep it up to date as it is a good tool for keeping structure. We have written our job list down as:
- Michael will research influence of resolution on accuracy
    - Deadline at 18/3
- Khalil, Mathias will retrain a pretrained model (EasyOCR) on our data set to see the performance difference. We want something to present for our next supervisor meeting.
    - Deadline 18/3

After 18/3, we plan on focusing on our mid-way report.

Week 8: 23.03.22

We finished the midterms assignment and pre-planned questions for our supervisor.

Week 9: 30.03.22

Meeting was skipped. (Because of the many assignment deadlines in other courses)

Week 9: 06.04.22

The feedback meeting was held today. (With group 2 of Generating Speech from Transcripts)
The feedback received was: (The negative part)
- Construct more precise and shorter sentences.
    - They gave a few examples of long and convoluted sentences in our report.
- Better descriptions on why we chose the state-of-the-art tools that we did.
    - In the rapport we only mention that they are state-of-the-art and do not go into details
- The introduction was short
- Missing description on how to handle the issues we have in our dataset
    - In the rapport we mention that some of the pictures are useless. How do we handle this? Do they get removed?
- Some references are missing
- The order of the rapport could be changes up a bit
    - Some parts of the rapport makes use of stuff that is only introduced on later pages
- We should mention any potential bias in the dataset and any ethical issues.
    - Even if there isn't any
- What data was used to train the pretrained models that we use.

The positive:
- The part about the project structure was a good idea

- The figures were well made
- The method section was well written
  - The feedback group was able to tell what we were trying to do

Week 11: 19.04.22

We had a status meeting regarding what had been researched in the Easter break, as there are difficulties with the EasyOCR framework and how to further train using our data set. The framework finds bounding boxes using CRAFT(STR-part) and recognizes the text in the bounding boxes with a CRNN, as seen below.
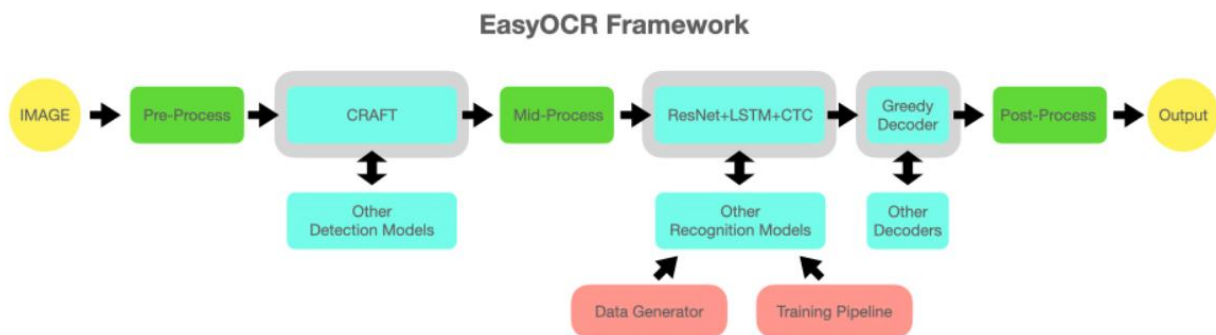
However, our data is quite messy and filled with a lot of noise, meaning CRAFT can return hundreds of bounding boxes, where only a single one of them is of relevance for training our recognition model. This is a challenge, as the EasyOCR framework assumes that the recognition model is trained on clean text, corresponding to only inputting cropped barcodes.

Furthermore, there are also examples where CRAFT is not able to pick up barcodes (especially thin fonts) and assign bounding boxes for them, and this might mean that training CRAFT could also be of importance.

We are thinking of the possibility of labeling the barcodes/cropping where they are on the images and thus training the recognition model using the cropped barcodes. If this does not improve accuracy, it can mean that the detection of the barcodes (CRAFT) is lacking/bottlenecks the process.

Week (1 day before 3-week): 6/1-2022
We had a status meeting in regard to how far we were in our assignments and what problems we should address with our supervisor. We have all been busy with our exams, so the number of hours used on the project has been a little reduced.

- Michael has looked into weights&biases, but it is not finished, and will also look into old-school computer-vision techniques for barcode-scanning. Michael will also look into confidence scores.
- Mathias is working on re-training CRAFT and has formulated questions for our supervisor in regard to the SynthText-dataloader which the CRAFT has been trained on in the reimplementation by backtime92 (https://github.com/backtime92/CRAFT-Reimplementation)
- Khalil still needs to label the last images in the 2. Data dump (and going through the 1. Data dump with stricter criterias in regard to the readability of barcodes (no blurred images)), and then comparing different recognition models on cropped data.

3 Weeks Course 9/6-2022 - 19/6-2022
In the 3 weeks course, we decided to have daily meetings while writing the report and deciding the last parts to include and exclude. We met everyday at 1pm to have a status meeting where each of us got assigned task to write on the report.

Supervisor Meetings

Week 2: 11.02.22

We had a kick-off meeting with our supervisor, where we were introduced to the setting in a more in-depth presentation and discussed some of the details of the project. To inspect the dataset, which contains sensitive data, the rules require us to sign an NDA and our supervisor will send it sometime next week. Albeit we were able to discuss the form of the problem and what tools we could use. The scope of the project can be scaled such that it fits our skills and time.

The structure of the problem can be set up as:
- Performing OCR to match barcodes to products
- Image recognition (Score how well the installation was performed?)
- Performing NLP on conversation between customer and technician to correlate with score.

We agreed to meet every week, Friday 14:00, building 321-221.

To next Friday, we can look at packages and look at how to set up a virtual environment.

Week 3: 18.02.22

Unfortunately, the planned meeting was not held as our supervisor could not meet. At this moment, we are not sure of the cause, it might be some miscommunication.
However, we spent some time looking and discussing the given dataset to tackle the problem, one of the challenges we talked about was noise/pictures of wrong objects and blurry barcodes.
We have questions about the Gantt diagram and Canvas, but we plan to wait for a Q&A-session with our course supervisor, Morten Mørup, presumably next Tuesday (22.02.22).

Week 4: 25.02.22

We had a discussion which clarified how this project would be used in a real-world setting if this project was to be used by Otovo. In the real world Otovo wouldn't have the correct label for the electricity meter. We got suggestion on how to calculate the confidence score using the concept of "dropout layers" in our neural network.

Week 5: 04.03.22

This meeting was skipped. See log entry week 5 in "project meetings".

Week 6: 11.03.22

Unfortunately, one of our group members, Christoffer has withdrawn from the course due to personal reasons.
We decided to continue with the plan in hand and have the regular meeting with our supervisor. We received practical information on transfer learning and how we can continue training existing

models(finetuning). We got suggested to visualize the distribution of the different confidence levels when reading numbers and visualize the confidence in terms of different resolutions. We also got an introduction to computing PCA on higher dimensional data if we wanted to head towards that direction (Resnet18).

Week 7: 18.03.22
This meeting was cancelled. Supervisor was busy.

Week 8: 25.03.22
We started out the meeting talking about the next feedback phase that we must go through with our peer group. We informed our supervisor that it could be a good idea for him to be included, if the scheduling allows it, but that it is not a requirement. Then we discussed how much we can blackbox for the models. He continued to explain us about how CRAFT works with a UNet architecture - including its encoding and decoding sections. We needed to find out how the pre-processing is done since a VVG architecture needs a specific size input(224x224).

After the meeting, we decided to each look further into pytorch.

Week 9: 01.04.22
Meeting was skipped due to this project being put on hold this week. (Because of the many assignment deadlines in other courses)

Week 10: 08.04.22
This meeting happened online since our supervisor was in Norway. We started out this meeting talking about some problems we had with our coding towards improving the EasyOCR Model, we learned to look more into the code with debugging. After that we also talked about pre-processing – and what was wrong with the results of our transfer learning model. We suspect that it may have something to do with imgH and imgW or with how we handle the outputs in prediction. We also talked about importing a barcode scanner model and using that to find the barcode labels.

We delegated responsibilities.
- Khalil - Will debugged the code for the outcome of the prediction in the EasyOCR model(CTC part)
- Michael – Will research into barcode detectors so it's easier for us to locate the barcode labels.
- Mathias – Will research into imgH and imgW(Image Dimensions) on how we can get a better prediction result.

We presented some of the results from further training the recognition model on our cleaned dataset with 2k iterations and a high learning rate of 0.05. The fact that validation loss decrease with input is a good indicator and experimenting we expect to experiment with other hyper-parameters after we get the further training of the CRAFT up and running.

We shared some of the concerns about data cleanliness – around half of the 2063 images are more or less noise – we have sorted through them all and labeled all of the images that are not noise and end up with around 1000 labeled images. As of this moment, images that were kind of unclear but where the space between the letters are distinct have been included like this,
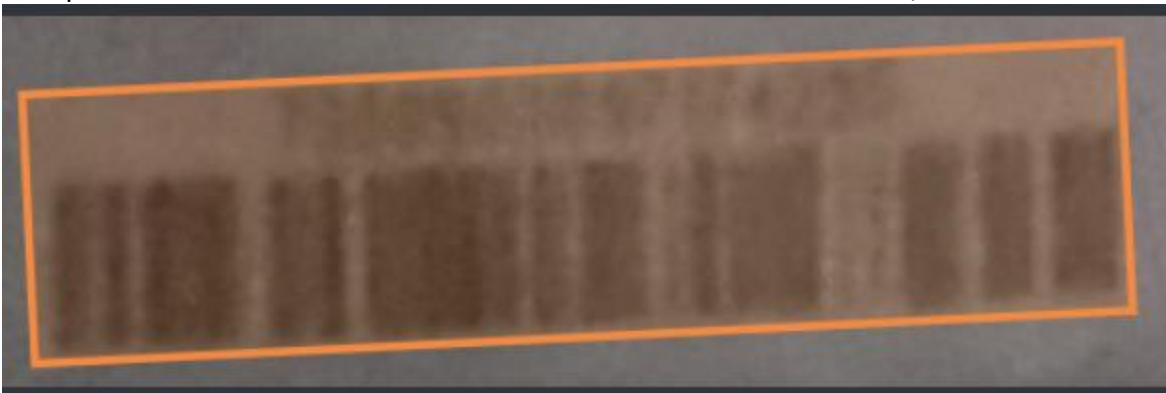
Where the threshold for not including the image is something like this:

However, when we discussed with our supervisor, he made a good point by explaining that it is best fi we train on (wholly) clean data only so we might want to exclude both images – and in this regards our supervisor will send more data since we do not have that much. There are also images without

the barcode(only serial numbers) and this is also something that our supervisor suggest that we exclude.

We talked about VisualStudio for Remote Desktop and how it might save us the hassle of using ThinLinc by directly connecting to a DTU PC from our own Desktops IDE.

For handling model results (Loss/Accuracy, etc.) our supervisor recommended us a plug-in API in python that do the visualization (Insert here, I forgot what it was called something with widgets&labels or smfn.).

We also presented an issue we had with properly loading observations using the labels.csv file into label-studio and here our supervisor looked over it and suggested that we try using another delimiter than tab ('\t') and that it might solve the problem. As of this moment, the images were manually inserted with no meter-number annotated in label-studio, and only after the Export of BB's in a .csv (including ID and bounding box info) did we collect the results in python using labels.csv (including ID and meter_number).

Week 12: 29.04.22
Meeting was skipped due to this project being put on hold this week. (Because of the many assignment deadlines in other courses)

Week 13: 06.05.22
Today's meeting consisted of 2 different topics: Performance evaluation and barcode detection.

Performance evaluation:
Our supervisor gave us suggestions on how to evaluate the performance of our model. He suggested techniques like: Average recall/precision, confusion matrices with numbers as "classes" and IOU scores. We need to determine the error on the detection- and the recognition part. The detection part included determining the loss between the predicted values (a point and a height/width) and the ground truth and the recognition part included pairwise digit comparison (The first predicted digit is compared to the first digit in the ground truth). This should be displayed in a confusion matrix.

Barcode detection:
Our supervisor suggested an alternative way of detecting barcodes. Instead of using neural networks (CRAFT) we could classic computer vision techniques. Barcodes are a unique greyscale object which can be deterministically found with classic computer vision. These classic techniques are simpler than neural networks and may be capable to solve the problem in our project. Classic computer vision will be the topic at the next group meeting.

Day 2 of 3 Weeks (June): 03.06.22

Today's meeting consisted of 3 different topics: Different recognition models, Re-implementing CRAFT and classical barcode scanning.

Different recognition models:

We came up with the idea of creating a table where we compare all these different recognition models:

#	Trans.	Feat.	Seq.	Pred.	Acc. %	Time ms
T1	None	VGG	None	CTC	69.5	1.3
T2	None	**ResNet**	None	CTC	80.0	4.7
T3	None	ResNet	**BiLSTM**	CTC	81.9	7.8
T4	**TPS**	ResNet	BiLSTM	CTC	82.9	10.9
T5	TPS	ResNet	BiLSTM	**Attn**	84.0	27.6

Our supervisor advised that it would be a good idea to just start out simple and if there was time, we could look at more advanced recognition models. We took the decision to start with comparing T1, T2 and T3.

(The picture is taken from: https://towardsdatascience.com/pytorch-scene-text-detection-and-recognition-by-craft-and-a-four-stage-network-ec814d39db05)

Re-implementing CRAFT

For re-implementing CRAFT, we asked our supervisor for advice on what approach to take when looking at someone else's code. He informed us that the most important parts to understand the code was to look at the input and use the method of debugging the code. We took the decision to follow our supervisor's advice on the approach.

Old-school Barcode scanning

Our supervisor advised us to test on the whole dataset before we decide to include the Old-school Barcode scanning in our final model. We took the decision to follow our supervisor's advice.

Day 7 of 3 Weeks (June): 10.06.22

Today's meeting consisted of doing Bayesian Optimization in Weight & Biases, Training CRAFT and the "results" section in our rapport.

Bayesian Optimization in Weight & Biases

We had trouble doing BO and our supervisor helped us locate the error. the error - the reason we weren't getting any logs of validation loss was because the run crashed and our supervisor advised us that it might have something to do with the implementation of the sweep function

Training CRAFT

Our supervisor gave advise on how to proceed in training CRAFT and how to discover the input and output dimensions of the models in the stack. This included instantiating random tensors and debugging the code.

Results

We got advice on what metrics to include in the results section in our rapport. This included Accuracy, validation and training loss, confusion matrix over the numbers in order to determine which numbers the model struggles with and which it does not, Jaccard similarity to determine the overlap between the predicted bounding boxes and the ground truth

Day 14 of 3 Weeks (June): 17.06.22

Today's meeting consisted of how to handle "bad" results in our project. Our supervisor gave us advice on how to proceed given we have" non-optimal" results. Some results become worse after we have further trained the pre-trained weights. We got advised to come up with hypothesis on why this happens and to give suggestion on what would be done different given we were to do this project again.