Project plan for degree project

Version 1.6 – June 12, 2020

COURSE CODE (DV2572): MASTER THESIS COURSE IN COMPUTER SCIENCE

February 3, 2023

Thesis	Tentative title	Real time OCR recognition in Steel bars using Deep Learning methods.
	Classification	Data Collection, Deep Learning, YOLO, OCR libraries
Student	Name	Monica Gattupalli
	e-Mail	moga20@student.bth.se
	Social security nr	19991130-5002
Supervisor	Name and title	Hüseyin Kusetogullari
	e-Mail	Huseyin.kusetogullari@bth.se
	Department	Department of Computer Science
Co-advisor	Name and title	Krister Ekström
	e-Mail	Krister.ekstrom@swerim.se
	Company/HEI**	SWERIM
Co-advisor	Name and title	Privschek Edwin
	e-mail	Edwin.Privschek@ovako.com
	Company	OVAKO

1 Introduction

It is impossible to picture a future without industries because they have been expanding enormously over the last few decades. Most industries are concentrating on product quality and developing by utilizing new technologies. Many conventional methods are being replaced by cutting-edge procedures, which boosts industry quality and profits. Product number, Digital QR code, barcode, and printed characters on the items that aid to distinguish between excellent and poor-quality products can all be used to separate the products. These printed characters can either be photographed or remembered.

A well-researched topic of computer vision, sometimes known as optical character recognition (OCR), is the detection and recognition of text on images or camera-captured photos. In general, image data is unstructured, making it difficult to convert an image into digital text and recognize it. Using a typical machine learning technique also makes it challenging to achieve high object detection accuracy[1] [2]. When compared to machine learning algorithms, many Deep learning text detection methods have shown promising outcomes with higher accuracy and efficiency.

OCR is a complicated process that uses a variety of learning and modelling techniques to extract the text from an image and convert it into digital form. Traditional OCR is used to recognize the text on fixed spots in photos (like license plates), but the images may differ in shape, brightness, orientation, and other ways. The robust algorithm must be adapted to respond to these types of photos. However, there are distinct issues and difficulties with object detection and text detection[1]. To deal with the realistic images, various techniques including R-CNN and SSD (single stage detectors) are employed[3].

The most recent method, YOLO (you only look once), is highly advised for object detection. Prior to 2015, numerous deep learning algorithms, including R-CNN, fast R-CNN, and SSD, were used to detect objects, but they are slow and ineffective at doing so. There have been eight versions of YOLO released to date, and each one has undergone significant upgrades and adjustments. Version 4 of YOLO was created in C, whereas version 5 was created in Python[4]. YOLO receives a picture as input, passes it through all convolutional layers in the algorithms, then outputs the result with a bounding box. YOLO employs the Intersection over union (IOU) technique to produce a single bounding box instead of the usual numerous bounding boxes based on the number of convolution layers.[5].

Along with object detection, text detection is a classic challenge in this era in some of the application both the object detection and the text detection need to be worked together. In text detection many OCR services are developed and they are available in free version some of them are Tesseract, ABBYY FineReader, Google Docs OCR and Transym all the OCR services can be applied on scanned documents, camera captured images, written document and so on[6].

The thesis' main objective is to investigate the best OCR method for resolving the practical issue. In OVAKO, it is necessary to identify the characters on the Steel bars. Using the advancements in deep learning in object and text recognition, this thesis investigates the best OCR techniques that accurately predict the characters on the Steel bars. It also seeks to pinpoint the uncertainty conditions that still require improvement to achieve accurate text prediction on the Steel bars. The company provides real-world data for the investigation as its data source.

1.1 Problem Definition

Having traceability of unique production items is essential for the manufacturing quality. Data analysis of collected industrial data often becomes unnecessarily difficult due to the challenges of uniquely tracing manufacturing details over production steps, and thereby associating collected detailed production measurements with an item. Good item traceability allows finding quality issues back the root cause, such that production improvements devised can improve the entire production. A typical approach to traceability is to set a unique identification onto each item, such as marking a unique number in the material itself, and the read the marking at each production station.

Such a number is often hard to read automatically (by machines), since the translation from an image of the stamp to the identification number needs to be robust enough for a wide distribution in the image quality. Recent Deep Learning-based image analysis algorithms can be made very robust, given that there is a large enough data set to train on. Further, some approaches for better robustness uses "transfer learning" to first train on a high-volume image set for a basic understanding, then refining the training by case-specific image training examples.

Image data

All the data provided by the company is in image format and needle printed on them. Typically, background noise will be present in the photos along the region of interest. The Steel bars photographs from the company feature several problems and many places that need improvement when compared to normal images.

Figure 1: Normal image.

Figure 2: Steel bar

In the above images figure 1 is the combination of different types of images like scanned document, handwritten which has text on them captured with camera, figure 2 is the steel bar image which is needle printed and capture with a camera in production unit.

When we compare the figure 1 with figure 2, we observe that figure 2 have much background noise like rough edges on the Steel bar, extra background other than Steel bar, due to the dark green on the steel rods the character is little hard to identify with the necked eye, we can also see there is patterns on the Steel bars which make the text identification much harder. The light focus is constant in all images. Sometimes the shadow of the rough edges will affect the text on the Steel bar. the characters are printed with a needle so, there will be variation of sharpness to the needle which will affect the marking on the steel bars. The Steel bars will be on the rolling mill, thus there will be a rotation where the rods will occasionally go close to the needle and occasionally away from the needle which causes the causing a difference in the depth of the text on the steel bars. Due to the rotation of the steel bar in 360 degrees, not every image of a Steel bar will be in the same orientation.

By considering all the aspects of the figure 2 it is important that algorithms selected should be good enough to localize the text position without losing any kind of data or information from the image.

1.2 Ethical, societal and sustainability aspects

In this experiment ethical aspects are important, and it is followed throughout the thesis. The image dataset is provided by the OVAKO and the information about the steel production in OVAKO, all this information is maintained confidentially.

2 Related work

This section gives the overview of the different research articles and previous works in object detection and text detection. also discussed about the different OCR libraries.

Haifeng Wang, Changzai Pan, Xiao Guo, Chunlin Ji, Ke Deng discussed various deep learning algorithms associated with object detection, as well as key points related to OCR. The statistical model in the field of text detection and recognition, as well as how the deep learning era has changed object detection since 2010. Many algorithms, such as CNN, R-CNN, Fast RCNN, Faster-RCNN, YOLO, and SSD, have been developed to detect objects in images/videos. Many software packages, such as Detectron2, Tesseract, and CRNN, are used in text detection, and they eventually concluded with different applications that require both object and text detection [1].

Peiyuan Jiang, Daji Ergu, Fangyao Liu, Ying Cai, and Bo Ma reviewed the YOLO object detection technique and versions. They also compared CNN with the YOLO algorithm. Major advancements between YOLO versions are also discussed. When comparing YOLO versions, they concluded that YOLO V5 is more flexible, as it employs the Hardswish activation function [5].

Myung-Cheol, Ju-young Lee proposed RF-RCNN, a faster RCNN architecture with a refining block. The refining block is made up of the regressor and the classifier. Humans and license plates are detected using the proposed architecture. The new architecture outperformed faster RCNN in detecting humans and license plates [7].

Xingyu Zeng, Wanli Ouyang, Bin Yang, Junjie Yan, and Xiagang Wang proposed gated bi-directional CNN (GBD-Net) for object detection. This GBD-Net can be merged after the convolution layer in any architecture to help retain the features of the region of interest and images in different resolutions. The proposed methodology has been validated using ImageNet, Pascal VOC2007, and Microsoft COCO [8].

Rayson Laroca, Evair Severo, Luiz Zanlorensi , Luis S. Oliveria, Gabriel Resende Goncalves, William Robson Schwartz ,David Menotti has used the YOLO object detector to identify license plates, and trained the network using 4500 fully annotated photos from 150 automobiles using two different CNNs. Up to 6 characters on the license plate may be recognized by the suggested model with an accuracy of 93.53% [9].

Jian Han , Yaping Liao, Junyou Zhang, Shufeng Wang and Sixian Li has employed YOLO to identify color and LiDAR depth pictures. The suggested model is broken down into two parts. In the first step, the YOLO method is applied to the images, in second step convolution and pooling layers are applied. To obtain the precise object detection, the bounding boxes from two phases are combined [10].

Jia-Ping Lin, Min-Te Sun has designed a technique for counting vehicles. The vehicle counter is in charge of counting after the YOLO algorithm identifies the vehicle and creates bounding boxes around the objects, these bounding box co-ordinates are then saved in the buffer. The video is provided as the YOLO algorithm's input. When trained with the COCO dataset, YOLO typically recognizes 80 moving things. Therefore, they modified YOLO to only detect the car, bus, and truck while masking out all other moving objects. The studies use footage taken in the morning, the afternoon, and the evening, and all three videos are made with 100% accuracy [11].

Rachel Huang, Jonathan Pedoeem, Cuixian Chen has created the real-time YOLO-LITE object detection system, which was trained using the PASCAL VOC and COCO datasets. Personal computers can be used to run the model. YOLO-LITE performed well on a small system, averaging 21 frames per second. They also talked about how to make the existing system better in the future [12].

Wei Liu, Dragomir Anguelov, Dumitru Exhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, Alexander C.Berg has proposed a new model for detecting objects which is named as Single Shot Multibox Detector SSD and this model is experimented upon PASCAL VOC, ILSVRC and COCO datasets which showed the promising results and justified that SSD is good at object detection.it also achieved 74.3% mAP on 300X300 input size, when 512 X 512 input is given it outperformed than Faster RCNN with 76.9% mAP [3].

Michal Buta, Luka Neumann, Jiri Matas has proposed a scene text detector which is able to detect the text in a video stream. It can also detect the different language like Latin, Hebrew and Chinese along with English. Designed detector is good at detecting the text with have stroke ending, but it failed to detect the text which have no corner. They concluded that proposed model detects the 25% more characters than the MSER and 3 times faster than the MSER detector [2].

Yunong Tian, Guodong Yang, ZheWang , Hao Wang, En Li, Zize Liang has developed an real time detection system which detects the growth of the apple developed using YOLO-V3. Images of immature apples, expanding apples, and ripples are initially gathered, then data is increased using augmentation techniques. The YOLO -V3 feature layers are employed with the DenseNet technique, which improved network performance. Comparing the performance of the YOLOv3-dense net model to that of the YOLOV3 and Faster R-CNN with VGG-16 model [13].

Chirag Patel, Atul Patel, Dharmendra Patel conducted a comparison between the Transym OCR and Tesseract OCR. Text detection is performed on images in both color and grayscale. Tesseract produced text detection accuracy of 61% and 70% on color and grayscale photos, respectively. It also performed better than Transym, taking 1 second and 0.82 seconds to process color and grayscale images, respectively [14].

Malathi T, Selvamuthukumaran D, Diwaan Chandar C S, Niranjan V, Swashthika A k has performed an comparative study between Microsoft OCR and tesseract OCR on different type of images which have white background, dark background, different fonts, colored font, grey scaled in most of the cases Microsoft OCR provided better results than tesseract OCR, but in some cases vice versa, but when coming to time of recognition tesseract OCR is better [15].

Rifiana Arief, Achmad benny Mutiara, Tubagus Maulana Kusuma, Hustinawaty applied Google vision OCR on the document image to recognize the text in Hadoop environment. The OCR library is 2 times faster when runned in Hadoop environment also extract the exact text by reaching the 100 % accuracy [16].

3 Aim and objectives

This project aims to investigate and implement a robust text-classification algorithm that that is preferably as accurate as a human reader in reading the text on the ends of the steel bars.

Objective 1: Detection and localization of text on Steel bars, resulting in images with bounded boxes.

Objective 2: Extracting text from the images produced by Objective 1.

Objective3: Report success rates and uncertainty estimates broken down by marking/surface/image quality.

4 Research questions

RQ1: How good is the proposed model at recognizing the characters on the Steel bars?

Motivation: The RQ1 is quite important to determine the how well our model is detecting the text on the steel bars, the accuracy of the model helps to justify whether the proposed model is good fit or not.

RQ2: What are the uncertainty conditions in the images that must be overcome to make better predictions?

Motivation: This RQ helps the company to understand the exact conditions which need to be improved in the aspects of marking, camera quality to get the better recognition of the text.

5 Method

In general, methodologies can be carried out in a variety of ways, including literature reviews, experiments[17], survey analysis[18], and case study research[19]. The research questions in this thesis are addressed using experimental methodology.

The thesis experimental methodology is divided into three steps 1. Dataset preparation 2. YOLOv5 algorithm 3. Making use of OCR libraries.

Dataset Preparation:

1. Data Augmentation is performed on the dataset to increase the number of images, YOLO trains well when we train with high number of images.
2. Converting the data according to YOLOv5, that means the annotation are converted to ".txt" files for each image in the dataset.
3. The annotations are tested to see if they are working properly.
4. The data set is divided into three sections: training, validation, and testing, where they are divided into 70:15:15 ratio respectively.

YOLOV5 algorithm:

1. The dataset file's details will be saved in a file called the Data Configuration file.
2. If necessary, the neural network's hyperparameters can be configured.
3. If necessary, a custom architecture is designed to achieve the desired result; otherwise, the default architecture can be used.
4. Training the YOLOv5 algorithm on how to extract data from images. The text in the images is localized by the bounding box.

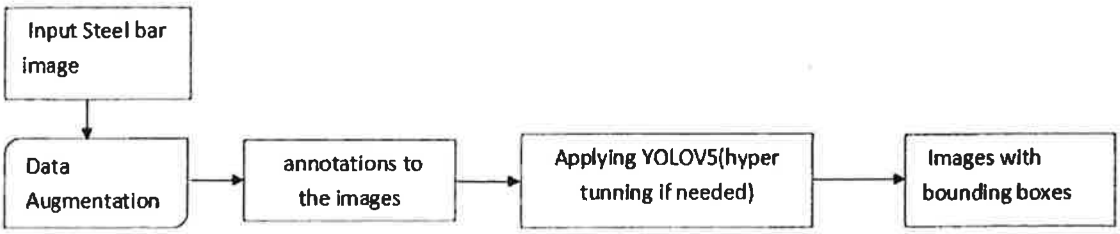

Figure 3: Applying YOLO algorithm.

Applying OCR libraries:

Steps involved in applying the OCR libraries.

1. The resulting images from the yolov5 algorithm which are fed into the OCR libraries.
2. Pre-processing techniques are applied on the images to improve the accuracy rate.
3. Running the OCR library and extracting the text.

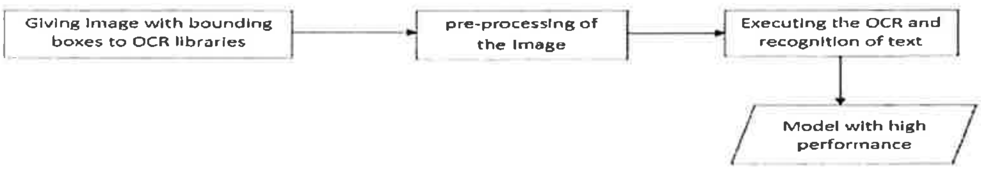

Figure 4: Applying OCR libraries.

The results of using OCR libraries are examined and analyzed to determine the best approach. RQ1 is answered by the experimental approach, and uncertainty conditions can be identified by the corns of the proposed methodology, implying RQ2 is answered.

Data collection:

1. The images of the Steel bars are taken in the Rolling area and the text on the Steel bars are printed there with a needle in production unit of the OVAKO in Hofors and those images are used for the experiment and to answer the RQ1 and RQ2.

2. The cameras are fixed in the production unit and the cameras will capture one end of the Steel bars after the needle printing, but the bars always have the marked side towards the camera.

3. OVAKO has already provided 1000 images of steel bars of size 1920 X1080 from previous years and 30% of these have already been labeled by OVAKO, which are subjected to a yolo algorithm before being subjected to OCR libraries to determine how well the text is detected.

4. Around 3000 new images with new font layout and with additional checksum characters will be produced every week once the image system has been repaired. Then those 3000 images are labeled manually to conduct the experiment.

6 Expected outcomes

Outcome1: The RQ1 results a function which takes the image as the input and results the text on the steel bar, also able to determine how effective the proposed model in detecting the text on the steel bars.

Outcome2: From the RQ2, we will be able to identify the uncertainty conditions in the image quality, camera angle, and marking on the Steel bar, which will improve prediction.

7 Time and activity plan

This section provides the time and activity plan of the thesis.

Activity	Start date	End date
Thesis topic research	2023-01-16	2023-02-05
Research phase: 1. Literature review of the study. 2. Research about the suitable deep learning algorithm. 3. Referring to the existing OCR libraries. 4. Identifying the best approach. Project plan update	2023-01-16	2023-02-05
Project plan submission	2023-02-05	

Experiment 1. Dataset preparation. 2. YOLO v5 algorithm. 3. Applying OCR libraries 4. Finding the uncertainty conditions 5. Analyzing the results	2023-02-06	2023-04-15
Thesis writeup	2023-04-16	2023-05-07
Thesis Draft submission	2023-05-07	
Thesis presentation and defense	2023-05-22	2023-05-26
Final thesis submission	2023-06-04	

7.1 Supervision plan

In the entire thesis period, we are going to have the meeting frequently according to the updates in the project, and the suggestions from the supervisor are updated according.

8 Limitations and risk management

Limitation	Level of Severity	Reason	Mitigation
Labeling the images	medium	From the camera we will only get the images but not labels, they need to be labeled manually.	I personally label the images which are collected from the company.

Risk	Level of Severity	Reason	Mitigation
Production of new images is delayed	Low	The image capturing system stopped working before Christmas and will need to be repaired. Replacements parts have been ordered but additional actions mighty be needed to get the system operational	Construct the solution on existing images wit the old layout, font and no checksum in a way makes it easy to adapt it to new images when they arrive.
No Production of new images	Medium	Ovako are unable to repair the image capturing system in time	Complete the project with images of the old layout and font. There are some more of these available that can be used to increase the size of the

			dataset.

References

[1] H. Wang, C. Pan, X. Guo, C. Ji, and K. Deng, "From object detection to text detection and recognition: A brief evolution history of optical character recognition," Wiley Interdiscip. Rev. Comput. Stat., vol. 13, Jan. 2021, doi: 10.1002/wics.1547.

[2] M. Buta, L. Neumann, and J. Matas, "FASText: Efficient Unconstrained Scene Text Detector," in 2015 IEEE International Conference on Computer Vision (ICCV), Dec. 2015, pp. 1206–1214. doi: 10.1109/ICCV.2015.143.

[3] W. Liu et al., "SSD: Single Shot MultiBox Detector," vol. 9905, 2016, pp. 21–37. doi: 10.1007/978-3-319-46448-0_2.

[4] "How to Train YOLO v5 on a Custom Dataset | Paperspace Blog." https://blog.paperspace.com/train-yolov5-custom-data/ (accessed Jan. 28, 2023).

[5] "Elsevier Enhanced Reader." https://reader.elsevier.com/reader/sd/pii/S1877050922001363?token=BBD7CB30837088 40D488D7C3342934B66720D2A7366DEFD441220DCF01D9413DEEBB06DB9208029 CF0FC5A79471D1C0B&originRegion=eu-west-1&originCreation=20230128131327 (accessed Jan. 28, 2023).

[6] M. Assefi, "OCR as a Service: An Experimental Evaluation of Google Docs OCR, Tesseract, ABBYY FineReader, and Transym," ISCV, Dec. 2016.

[7] M.-C. Roh and J. Lee, "Refining faster-RCNN for accurate object detection," in 2017 Fifteenth IAPR International Conference on Machine Vision Applications (MVA), May 2017, pp. 514–517. doi: 10.23919/MVA.2017.7986913.

[8] X. Zeng, W. Ouyang, B. Yang, J. Yan, and X. Wang, "Gated Bi-directional CNN for Object Detection," in Computer Vision – ECCV 2016, Cham, 2016, pp. 354–369. doi: 10.1007/978-3-319-46478-7_22.

[9] R. Laroca et al., "A Robust Real-Time Automatic License Plate Recognition Based on the YOLO Detector," 2018 Int. Jt. Conf. Neural Netw. IJCNN, pp. 1–10, Jul. 2018, doi: 10.1109/IJCNN.2018.8489629.

[10] J. Han, Y. Liao, J. Zhang, S. Wang, and S. Li, "Target Fusion Detection of LiDAR and Camera Based on the Improved YOLO Algorithm," Mathematics, vol. 6, no. 10, p. 213, Oct. 2018, doi: 10.3390/math6100213.

[11] J.-P. Lin and M.-T. Sun, "A YOLO-Based Traffic Counting System," in 2018 Conference on Technologies and Applications of Artificial Intelligence (TAAI), Nov. 2018, pp. 82–85. doi: 10.1109/TAAI.2018.00027.

[12] J. Pedoeem and R. Huang, "YOLO-LITE: A Real-Time Object Detection Algorithm Optimized for Non-GPU Computers," 2018 IEEE Int. Conf. Big Data Big Data, 2018, doi: 10.1109/BigData.2018.8621865.

[13] Y. Tian, G. Yang, Z. Wang, H. Wang, E. Li, and Z. Liang, "Apple detection during different growth stages in orchards using the improved YOLO-V3 model," Comput. Electron. Agric., vol. 157, pp. 417–426, Feb. 2019, doi: 10.1016/j.compag.2019.01.012.

[14] C. Patel, A. Patel, and D. Patel, "Optical Character Recognition by Open source OCR Tool Tesseract: A Case Study," Int. J. Comput. Appl., vol. 55, pp. 50–56, Oct. 2012, doi: 10.5120/8794-2784.

[15] https://iopscience.iop.org/article/10.1088/1757-899X/1059/1/012004/pdf (Accessed Feb. 02, 2023).

[16] https://media.proquest.com/media/hms/PFT/1/5odfM?_s=q1XR6lS2WdiD94QzMpuOUspuy6s%3D (Accessed Feb. 02, 2023).
[17] A. Jedlitschka and D. Pfahl, "Reporting guidelines for controlled experiments in software engineering," in 2005 International Symposium on Empirical Software Engineering, 2005., Nov. 2005, p. 10 pp.-. doi: 10.1109/ISESE.2005.1541818.
[18] B. A. Kitchenham and S. L. Pfleeger, "Personal Opinion Surveys," in Guide to Advanced Empirical Software Engineering, F. Shull, J. Singer, and D. I. K. Sjøberg, Eds. London: Springer, 2008, pp. 63–92. doi: 10.1007/978-1-84800-044-5_3.
[19] P. Runeson and M. Höst, "Guidelines for conducting and reporting case study research in software engineering," Empir. Softw. Eng., vol. 14, no. 2, p. 131, Dec. 2008, doi: 10.1007/s10664-008-9102-8.

A. Signature enclosure

The company/ organization **SWERIM** accepts the description of the "work" in this document ("Project plan"). SWERIM also accepts to offer the student/ students a reasonable amount of supervision in connection with the future degree project in relation to the parts of the degree project that are related to the" work" at **SWERIM**. **SWERIM** is aware that BTH is not part of any agreement with the company and does not guarantee that the "work" is carried out in a satisfactory manner.
The part of the degree project that is carried out by the student/ students "on behalf of" **SWERIM** is only a commitment between the student/ students and the company. BTH will evaluate the work in the form of documents that are submitted for evaluation.
All submitted works can be made public on request. According to regulations, the passed final report will be published in the document management system "DIVA" and as a part of this publication will then become a generally accessible public document.
This signature enclosure is not a legal agreement and is therefore not binding for any part (BTH nor **SWERIM**) but should be considered as a "declaration of intention" from **SWERIM**.

Karlskrona, Month Day, Year

Kista 2023-02-03

Signature by supervisor at **SWERIM**

Krister Ekström

Printed name

KRISTER EKSTRÖM

The company/ organization **OVAKO** accepts the description of the "work" in this document ("Project plan"). OVAKO also accepts to offer the student/ students a reasonable amount of supervision in connection with the future degree project in relation to the parts of the degree project that are related to the" work" at **OVAKO**. **OVAKO** is aware that BTH is not part of any agreement with the company and does not guarantee that the "work" is carried out in a satisfactory manner.

The part of the degree project that is carried out by the student/ students "on behalf of" **OVAKO** is only a commitment between the student/ students and the company. BTH will evaluate the work in the form of documents that are submitted for evaluation.

All submitted works can be made public on request. According to regulations, the passed final report will be published in the document management system "DIVA" and as a part of this publication will then become a generally accessible public document.

This signature enclosure is not a legal agreement and is therefore not binding for any part (BTH nor **OVAKO**) but should be considered as a "declaration of intention" from **OVAKO**.

Karlskrona, Month Day, Year

Hofors 2023-02-03

Signature by supervisor at **OVAKO**

Printed name

EDWIN PRINSCHEK

I got the thesis in the company OVAKO in association with swerim, the main aim of the company is to identify the characters on the steel bars. Under the supervision of Huseyin from BTH, Krister from swerim and Edwin from Ovako I completed my project proposal, by conducting the meetings I completed the project proposal on Friday I sent the final version to all the supervisors and got the acceptance from them . Under Signature enclosure section both the supervisors from the swerim and Ovako signed the document. They signed the document and scanned it and send it back to me. I am submitting the scanned version of the project proposal.