HEAT NUMBER DIGITIZATION SYSTEM

PHASE 1 PRESENTATION

Trainee Name : Ling Zhao Jia, Danny

Company Name : Keppel FELS Ltd

Company Supervisor Name & Email : Mr. Liew Wing Leong (wingleong.liew@keppelfels.com)

NUS-ISS Supervisor Name & Email : Dr. Tian Jing (tianjing@nus.edu.sg)

AGENDA (PHASE 1)

- Problem Background & Business Objectives
- Project Scope & Success Criteria
- Knowledge/Data collection and preparation
- Technical Progress
- System Overview
- Project plan

PROBLEM BACKGROUND & BUSINESS OBJECTIVES

Problem/Domain background

- Logistics Steel Stock supply chain process is highly manual and time consuming

Project's business objectives/goals;

- To automate Steel Stock supply chain process
- To identify and track Steel Stock
- Overall reduction in No. of manhours for Steel Stock Receiving & Issuing

PROJECT SCOPE

Technical Development

- Use Object Recognition and OCR to automatically classify Heat Numbers labelled on Steel Stock and digitize it

Vendor Evaluation

- Evaluate 3rd Party Image Capturing Devices and/or OCR Vendor solution and benchmark against in-house technical solution for management's review

PROJECT SUCCESS CRITERIA

Project success criteria (technical)

- Demonstrate Proof of Concept using Python script to automatically run images through Object Recognition model and output classification and predicted text onto .csv

Project success criteria (business)

- Eliminate manual identification of Heat Numbers
- Eliminate the need to use handwriting and manual data-entry for Heat Numbers
- Vendor evaluation for management's review

PROJECT BUSINESS USERS & STAKEHOLDERS

Business user profile

- KOM Digi Team
- Logistics Operations/Administration Team

Stakeholders

- Logistics Department

KNOWLEDGE/DATA COLLECTION AND PREPARATION

- **Image Acquisition**
  - Estimated 10-20 images from 10 different Suppliers (Total 100 – 200 images)
    - Different format for labelling the different classes: Heat Number, Dimensions, Country of Origin etc.

Supplier 1 – 10 images

Supplier 2 – 10 images

Supplier 3 – 10 images

KNOWLEDGE/DATA COLLECTION AND PREPARATION

- **Image Labelling**
  - Label images with bounding boxes

Image_0001.jpg

Image_0002.jpg

Class 1 - Heat Number

Image_0003.jpg

Filename	x1	y1	x2	y2	Label
Image_0001.jpg	251	51	444	300	heat_number
Image_0002.jpg	100	25	400	120	heat_number
Image_0003.jpg	111	41	414	250	heat_number

Image_0001.jpg

Image_0002.jpg

Class 2 - Dimension

Image_0003.jpg

Filename	x1	y1	x2	y2	Label
Image_0001.jpg	111	21	300	212	dimension
Image_0002.jpg	333	22	231	111	dimension
Image_0003.jpg	243	35	234	11	dimension

KNOWLEDGE/DATA COLLECTION AND PREPARATION

- **Image Augmentation**
  - Colour Space Manipulation
    - Random Brightness, Random Contrast, Noise/Blur
      - To simulate different lighting conditions and camera exposure
      - To simulate different camera image quality
      - To simulate camera user hand movements / out-of-focus image

Original

High Brightness & Contrast

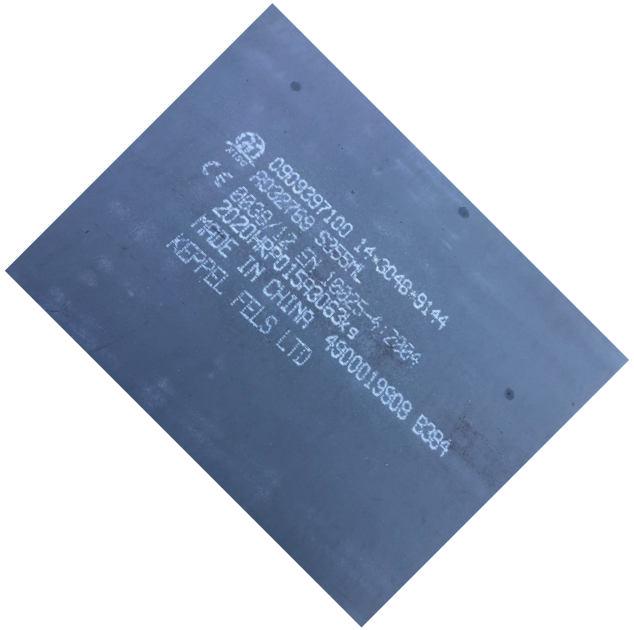
Low Brightness & Contrast

Gaussian Blur

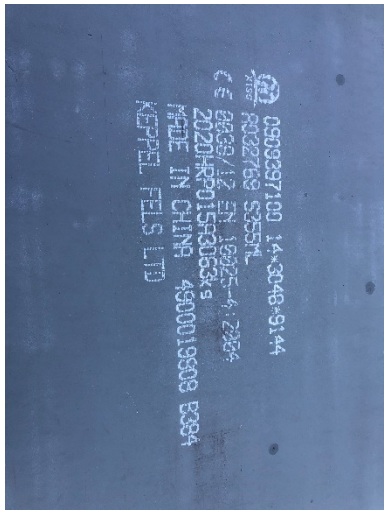
KNOWLEDGE/DATA COLLECTION AND PREPARATION

- **Image Augmentation**
  - Rotation
    - 45, 90, 135, 180, 225, 270, 315 degrees rotation
      - To simulate different image orientation

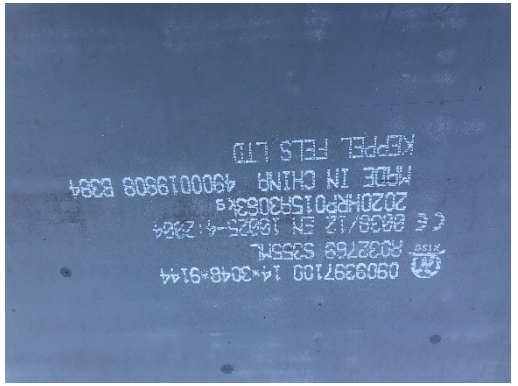
Original



45 degrees



90 degrees



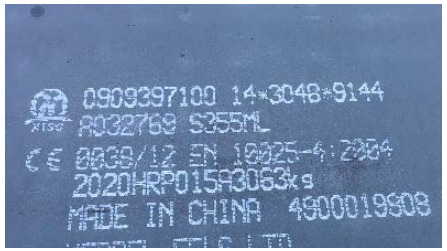
180 degrees

KNOWLEDGE/DATA COLLECTION AND PREPARATION

- **Image Augmentation**
  - Scaling / Cropping / Translation
    - To simulate camera zoom (near image vs far image)
    - To simulate incomplete images taken
    - To simulate different position of texts

Original

Scaling



Cropping

Translation

TECHNICAL PROGRESS

Image Acquisition

- Acquire100 -200 images using camera
- 10 – 20 images each from 10 different suppliers (Different labelling format)

Image Labelling & Augmentation

- Label Images & Object bounding boxes
- Image Augmentation
- Visually verify actual text on images and tabulate into Excel for OCR accuracy comparison

Classification Model

- Develop CNN model for Image Classification
- Save Keras model (HDF5 format) with trained weights for script implementation

OCR Model

- Paddle OCR vs. pytesseract
- Compare accuracy

Script Development

- Python script in .py format
- To run in Windows environment
- User Guide & Documentation

EXPECTED PROOF OF CONCEPT

Image_0001.jpg

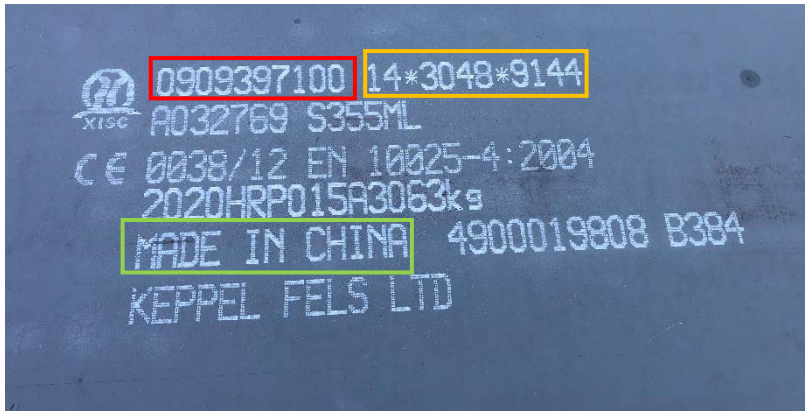


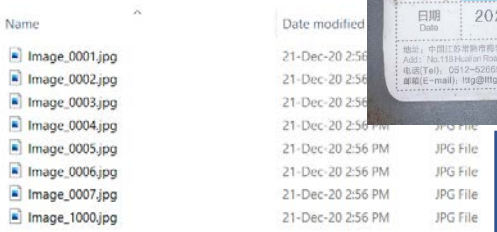
Image_0002.jpg

Classification Labels

- Heat Number
- Dimension
- Country of Origin

Filename	Heat Number	HN Confidence	Dimension	Dim Confidence	Country of Origin	COO Confidence
Image_0001.jpg	909397100	92%	14*3048*9144	95%	China	98%
Image_0002.jpg	2003A2933	94%	150*90*9	95%		
Image_0003.jpg	A12345B	93%	14*2048*9144	90%	Germany	97%
Image_0004.jpg	C123456D	80%	150*90*9	77%		
..	..	..	..	..	..	..
Image_1000.jpg	E1234567G	94%	14*3048*9144	94%	Germany	97%

System Overview



Tesseract OCR

PaddlePaddle
PaddleOCR

Open-source Optical Character Recognition (OCR) Engine

OCR Engine
- For Text Extraction

Legend:
- Development Tasks
- Heat Number Digitization Process

Text Extraction Engine

1 Images captured and stored in folder

Input

2 Run Python Script

Output

3 Output to .csv format

Inference Engine

Labelled Images for Model Training

Trained CNN Model

Classification Labels
- Heat Number
- Dimension
- Country of Origin

Deep Learning Model
- Convolutional Neural Network (CNN)
- For Object Recognition

Class 1 - Heat Number

Class 2 - Dimension

Filename	Heat Number	HN Confidence	Dimension	Dim Confidence	Country of Origin	COO Confidence
Image_0001.jpg	909397100	92%	14*3048*9144	95%	China	98%
Image_0002.jpg	2003A2933	94%	150*90*9	95%		
Image_0003.jpg	A12345B	93%	14*2048*9144	90%	Germany	97%
Image_0004.jpg	C123456D	80%	150*90*9	77%		
..	..	..	..	..	..	..
Image_1000.jpg	E1234567G	94%	14*3048*9144	94%	Germany	97%

UiPath + SAP

Use Case 1: RPA for data entry into SAP

Infinity Scan

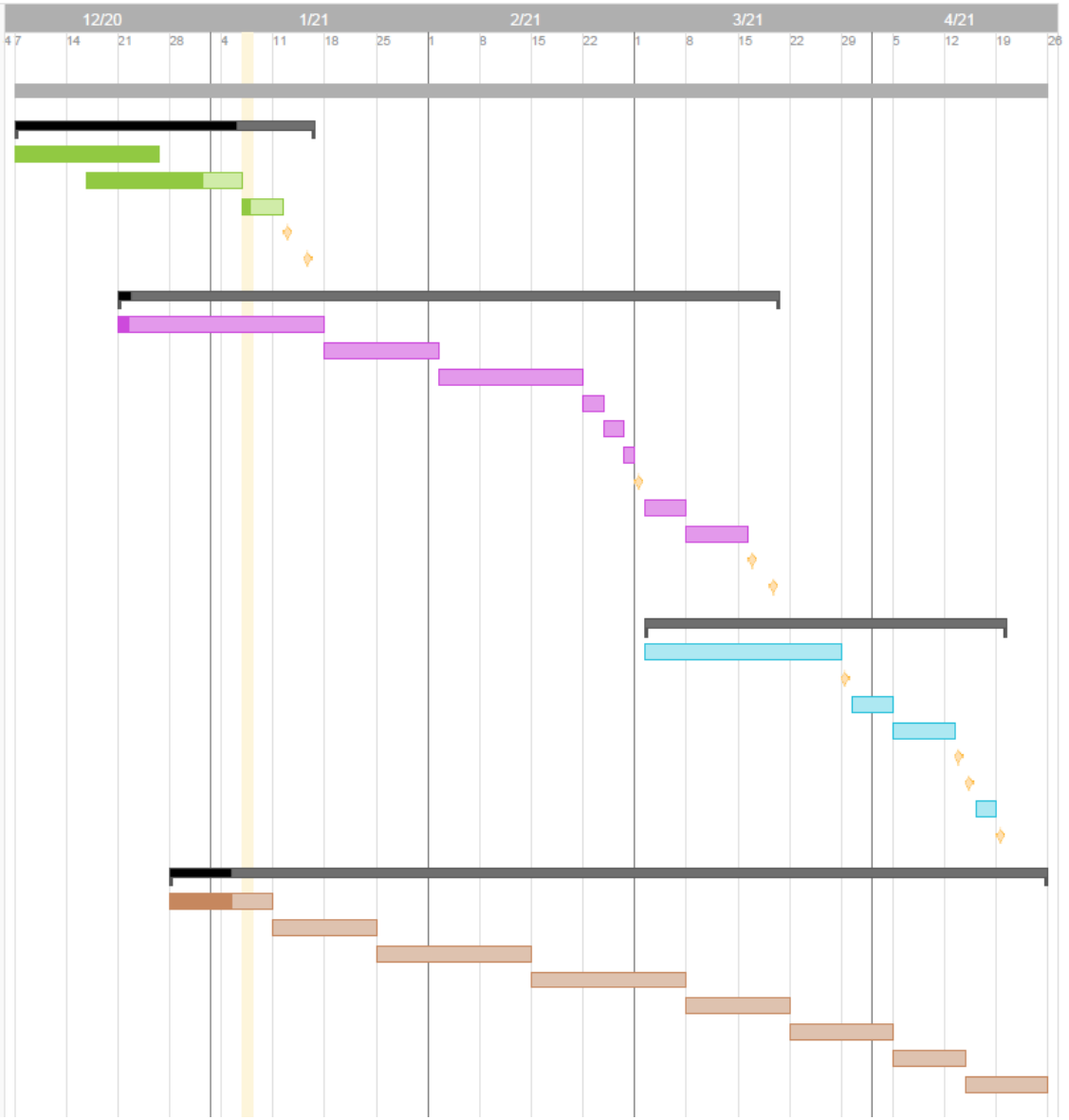
Use Case 2: Work with Infinity Scan for Automatic Mill Cert Validation

SAP

Use Case 3: For Heat Number Tracking using Steel Stock Management System

PROJECT PLAN

Keppel Steel Stock Automation OCR Project	start	end	15%
Business Understanding, Project Planning and Goal Setting (4 Weeks)	07/12/20	14/01/21	74%
Domain Familiarization	07/12	24/12	100%
Formulation of Project Plan	16/12	05/01	75%
Preparation & Draft Submission for Phase 1 Presentation	06/01	11/01	20%
Phase 1 Presentation Submission (Bef 11:59pm)	12/01	12/01	0%
Phase 1 Presentation (10:00am)	14/01	14/01	0%
Knowledge Acquisition and Initial Design & Development (8 Weeks)	21/12/20	18/03/21	2%
Data and Image Acquisition (200 - 400 Images)	21/12	15/01	5%
Data Processing - Image Labelling and Augmentation	18/01	01/02	0%
Development of Classification model for Object Detection	02/02	19/02	0%
Training and Testing of Classification model	22/02	23/02	0%
Optimization of Classification model	24/02	25/02	0%
Test OCR engine on Predicted Data	26/02	26/02	0%
Initial Proof of Concept & Evaluation of Object Detection Model	01/03	01/03	0%
Further Optimization of Classification + OCR model	02/03	05/03	0%
Preparation & Draft Submission for Phase 2 Presentation	08/03	15/03	0%
Phase 2 Presentation Submission (Bef 11:59pm)	16/03	16/03	0%
Phase 2 Presentation (10:00am)	18/03	18/03	0%
System Development & Deployment (4 Weeks)	02/03/21	19/04/21	0%
Development of Python Script	02/03	26/03	0%
Demonstrate Final Proof of Concept	29/03	29/03	0%
User Guide & Documentation	30/03	02/04	0%
Preparation of Final Project Report & Phase 3 Presentation	05/04	12/04	0%
Phase 3 Presentation Submission (Bef 11:59pm)	13/04	13/04	0%
Phase 3 Presentation (10:00am)	14/04	14/04	0%
Final Refinement of Project Report	15/04	16/04	0%
Final Report Submission (Bef 11:59pm)	19/04	19/04	0%
Third Party Vendor Evaluation	28/12/20	23/04/21	7%
Research on suitable vendors for engagement	28/12	08/01	60%
Propose 2-3 OCR and/or Storage Solution vendors	11/01	22/01	0%
Invite Vendors to present Proof of Concept / Site demo	25/01	12/02	0%
Vendors to submit quotation	15/02	05/03	0%
Vendor Cost-Benefit Evaluation	08/03	19/03	0%
Vendor Engagement Presentation/Proposal to Keppel	22/03	02/04	0%
Refinement of Presentation/Proposal	05/04	13/04	0%
Final Presentation/Proposal to Keppel	14/04	23/04	0%



Step-by-Step Deliverables (Technical Development)

No.	Item Description	Platform Used	Date	Deliverables / Expected Results
1	Data and Image Acquisition	Handheld/Phone Camera	15-Jan-20	1) 10 – 20 images of labels from 10 different suppliers
2	Data Processing - Image Labelling and Augmentation	Jupyter Notebook - LabelImg - OpenCV MS Excel	02-Feb-21	1) .txt or .xml file indicating coordinates for object detection bounding boxes 2) Labelled images (800 or more images after augmentation) 3) Actual heat numbers, dimensions and country of origin (to be visually verified) in excel
3	Development of Classification Model for Object Detection with training and testing	Jupyter Notebook - CNN model	19-Feb-21	1) Classification Accuracy Score > 80%
4	Optimization of Classification Model	Jupyter Notebook - CNN model vs Other models	25-Feb-21	1) Select optimal Classification model for use 2) Saved Keras model (HDF5 format) with trained weights
5	Test OCR engine on Predicted Data	Jupyter Notebook - pytesseract vs. PaddleOCR	26-Feb-21	1) Predicted Text Accuracy > 80% and select optimal OCR engine for use
6	Initial Proof of Concept & Evaluation of Object Detection Model	Jupyter Notebook	**01-Mar-21**	1) Demonstrate object detection and text extraction using test data
7	Further Optimization of Classification + OCR model	Handheld/Phone Camera Jupyter Notebook	05-Mar-21	1) Take additional images for training and check if model accuracy can be further improved 2) Additional image augmentation for training and check if model accuracy can be further improved
8	Development of Python Script for Object Detection	Spyder IDE	26-Mar-21	1) Python code to run script 2) Transfer of Jupyter Notebook code to Spyder IDE 3) Final Python script in .py
9	Demonstrate final Proof of Concept	Windows - Python script	**29-Mar-21**	1) Run through a folder of unseen images 2) Output predicted text and classification in .csv
10	User guide & documentation	MS Word / PDF	02-Apr-21	1) User guide and documentation in .pdf

Step-by-Step Deliverables (Vendor Evaluation)

No.	Item Decription	Date	Deliverables / Expected Results
1	Research on suitable vendors for engagement	15-Jan-21	Propose 2-3 OCR and/or storage solution providers
2	Invite Vendors to present Proof of Concept / Site Demo	12-Feb-21	2 - 3 Vendors to come in for presentation and site demo
3	Vendors to submit quotation	05-Mar-21	2 - 3 Vendor quotations to be received
4	Vendor Cost-Benefit Analysis	12-Feb-21	Tabulate vendor cost vs benefit
5	Vendor Engagement Presentation/Proposal to Keppel	19-Mar-21	Present Cost-Benefit Analysis
6	Refinement of Presentation/Proposal	02-Apr-21	Update presentation/proposal after feedback from Keppel
7	Final Presentation/Proposal to Keppel	23-Apr-21	Final Vendor Engagement Presentation

Final Deliverables

Technical Development Deliverables

- 1x Trained CNN Model for Object Recognition
- 1x Python Script
- 1x User Guide & Documentation

Vendor Evaluation Deliverables

- 2 – 3x 3rd Party Vendor Quotations

Q&A

Any Questions?

END