HACETTEPE UNIVERSITY
ENGINEERING FACULTY
DEPARTMENT OF COMPUTER ENGINEERING

BBM 325
INTERNSHIP REPORT

Ataberk ASAR
[redacted]

Performed at
BOTAŞ

18.07.2022 – 12.08.2022
20 Work Days

Report template version: v1. Feb 27, 2019.

TABLE OF CONTENTS

1 Introduction

I have done my BBM325 Summer Internship at BOTAŞ (Boru Hatları ile Petrol Taşıma Anonim Şirketi) Information Technologies Department. Founded in 1974, BOTAŞ mainly focuses on the construction and operation of oil and natural gas pipelines. Storing, administering, and transporting the resources are in the control of the company. I chose BOTAŞ for my BBM325 Summer Internship because of its wide variety of research areas and deep-rooted company structure.
During my summer internship, I developed a mobile application to make inventory management easier and foolproof, while learning about cyber security, networking, and software solutions that are used in BOTAŞ. The current inventory management system in BOTAŞ is error-prone due to human errors, such as inexperienced workers that do not know the product code of the incoming items, or even do not know the item's proper name. Thus to make the system uniform and easy to use, the task of developing a machine learning mobile application that performs image classification assigned to me.

2 Company Information

2.1 About the company

In 1974, BOTAŞ was founded for the construction and operation of the Kirkuk-Ceyhan Oil Pipeline. In 1987, BOTAŞ extends its operations to natural gas transportation and trade activities. Aside from Kirkuk-Ceyhan Oil Pipeline, BOTAŞ operates Ceyhan-Kırıkkale, Batman-Dörtyol, and Irak-Türkiye, Bakü-Tiflis-Ceyhan oil pipelines, and the national gas grid of Turkey with a total length of over 18000 kilometers. This wide transportation network also allows BOTAŞ to construct and utilize its own fiber optic cable network along with its pipelines. The headquarters of BOTAŞ is located at Bilkent/ ANKARA. BOTAŞ joined to the Turkey Wealth Fund on January 24, 2017. [1]

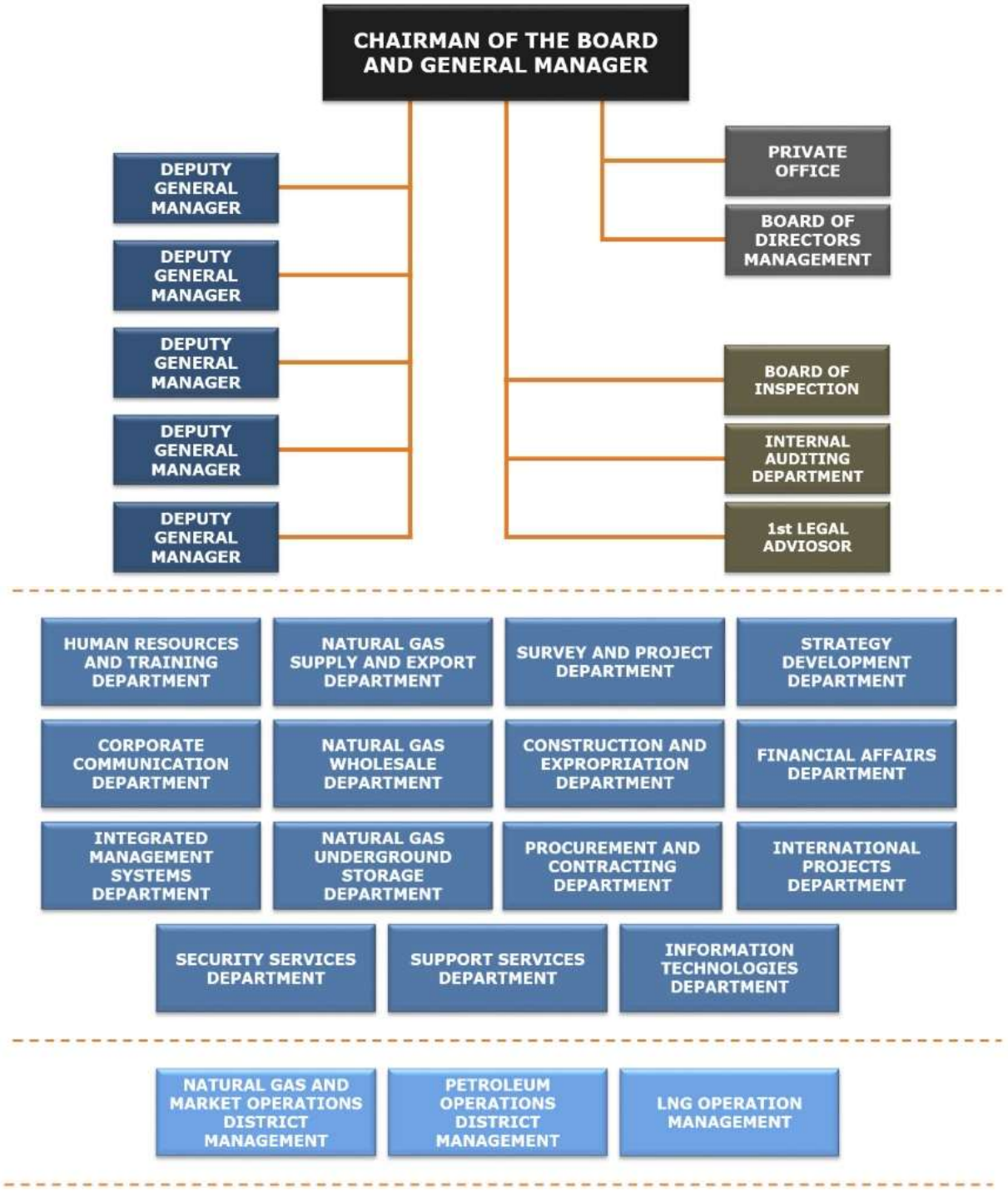

Fig1. Organizational Chart of BOTAŞ

2.2 About your department

I did my BBM 325 summer internship in BOTAŞ Information Technologies Department. During my internship, I gain experience in Software Department, Cyber Security Department, and Network Department. Although my training in cyber security and network departments can be considered as passive learning, I actively worked on a software project with a team of two in the software department. My work mainly focused on building a machine learning model and using the model in a mobile application to make image classification. The mobile application will be used to identify new incoming products and put them into the inventory management database system correctly.
Aside from the project that is given to me, senior developers in my department mainly focused on Enterprise Resource Planning (ERP) systems.

2.3 About the hardware and software systems

.NET: Free, cross-platform, open source developer platform for building many different types of applications. [2]

MySQL and MSSQL: Relational database management systems, which store the data in tables, based on structured query language (SQL).

Java: Used mainly for BOTAŞ's own internal communication, information, and ERP systems.

Flutter: Used mainly for developing BOTAŞ's own mobile ERP applications.

Oracle Exadata: Computing platform optimized for running Oracle Databases, which are optimized in enterprise-class databases. [3]

2.4 About your supervisor

- Salih KOCABAŞ,
- BOTAŞ Genel Müdürlüğü A1 Plaza, Bilkent/ Ankara,
- [redacted]
- [redacted]
- Karadeniz Technical University, Computer Engineering, 2000

3 Work Done

3.1 Introduction to Cybersecurity

Cybersecurity is defending data, devices, and networks against damage, loss, or unauthorized access. Although the threats facing BOTAŞ contain the typical threats, such as data theft, fraud, or ransomware, several attributes of the company heighten the risks. Thus, an introduction to cybersecurity training was provided to me, during the first week of my internship.

BOTAŞ has thousands of industrial control systems, that are used to manage pipelines (oil and gas) around all over the country. Systems designed without security concerns, physical security weaknesses, limited security built into Supervisory Control and Data Acquisition (SCADA) systems, and large attack surfaces of IoT devices are among the main threat vectors. Sabotages against control and safety processes, damaged devices, power outages, threats to personnel and public safety, theft of customer information, and even environmental disasters are among possible troubles if an attacker gains access to those control systems.

3.2 Introduction to Networking

Multiple computers that share information and resources via a connection is called computer network. The computer networking department is responsible to maintain, monitor, and optimize the computer network structure of BOTAŞ. Application backups, recovery tools, and communication systems between other teams and departments of BOTAŞ are some of the main focus areas of this department. An introduction to networking training was provided to me, during my internship.

3.3 Main Project

During my summer internship, I developed a mobile application to make inventory management easier and foolproof. The current inventory management system in BOTAŞ is error-prone due to human errors, such as inexperienced workers that do not know the product code of the incoming items or even do not know the proper name of the coming product. This problem leads to data inconsistencies in the system. Thus to make the inventory system uniform and easy to use, my supervisor assigned me with the task of developing a machine learning mobile application that performs image classification.
The given task is important to me since it introduced me to artificial intelligence and showed me the process of developing a software product in a large company like BOTAŞ.

I used Python, TensorFlow for the machine learning part, and used Flutter to develop a mobile application with a modern graphical user interface.

Fig2. Mobile Application Home Screen

Collecting and Preprocessing the Data

Because of BOTAŞ's privacy policies, the real dataset of product images does not shared with me. First of all, I had to find a sufficient dataset for the given task. Thus, to train the machine learning model a customized version of the DomainNet dataset [4] is used. I altered the dataset in order to make it suitable for the given task. Customizations include subtracting unnecessary labels (irrelevant labels like zebra, bird, whale, etc.) or merging similar labels (labels like mug and coffee cup merged into cup, etc.). After alterations, my custom dataset has 50 different classes with a total of 23,707 images. Images are preprocessed with TensorFlow data generation pipeline. Also, data augmentation was applied via the TensorFlow pipeline. Data augmentation

techniques used in the project include rotation, cropping, horizontal flip, brightness augmentation, and saturation augmentation.

Model Creation

During my internship, I learned general machine learning techniques used for image classification task.

Convolution Layer

Basic principle of convolutional layers arise from upgrading the conventional neural networks, so that it can work with two dimensional image data better. Convolution operation is a linear multiplication of two dimensional weight arrays, called kernel, with the input pixels of the given image. In convolutional neural networks, multiple weigth kernels (usually used kernel number is in the form of powers of 2, such as 32, 64, 128, etc.) used at the same time. Since the model can learn from more than one kernel simultaneously, it can learn and generalize easily. Hence, enhances the training process.

Batch Normalization Layer

When gradient-based learning methods used in machine learning models, the models face with problems like vanishing gradient problem, exploding gradient problem, etc. To solve this problems, models need input data to be in a common scale. Normalization pre-process the input data into this state, without distorting its characteristic features. Batch Normalization layer applies rescaled and normalized filters over input data. Therefore model gains more stability, in addition to solving problems like vanishing gradient problem and exploding gradient problem.

Pooling Layer

Outputs of convolution layers are dependent to positional features of the input. Thus, small changes (rotations, translations, etc.) in the given image result in big inconsistencies in neural network's predictions. A possible solution is that using lower resolution versions of input images. Pooling layers used to achive this effect. Average pooling uses a filter to calculate the average of each small part of the feature map and using it to create a new output from the input matrix. Max pooling uses another filter to take maximum value of each small part of the feature map. Pooling layers also help to reduce computational cost of the machine learning model.

Zero Padding

After applying layers, dimensions of input matrices decays due to convolution operation. Hence valuable information can be lost as the image data passes through the network. In order to solve this issue, zero padding layers applied to the inputs.

Depthwise Convolution Layer

Computers store the image data as three dimensional arrays (width x height x channel), thus every image has three diferent channels (red, green, and blue). In a depthwise convolution layer, image data seperated into its channels, different kernels applied to different layers, and the outputs combined again to form a three channel image output.

Activation Functions

Activation function introduces non-linearity into outputs of neurons, by restricting activation of the neurons based on their weighted sum. Basically, activation function decides if information should be passed into next neuron or not.

Dropout Layer

During the trainig of model, some of the neurons (20% in my model) deactivated randomly. This help model to learn from different neurons at every iteration. In return, dropout layer helps model to generalize easily.

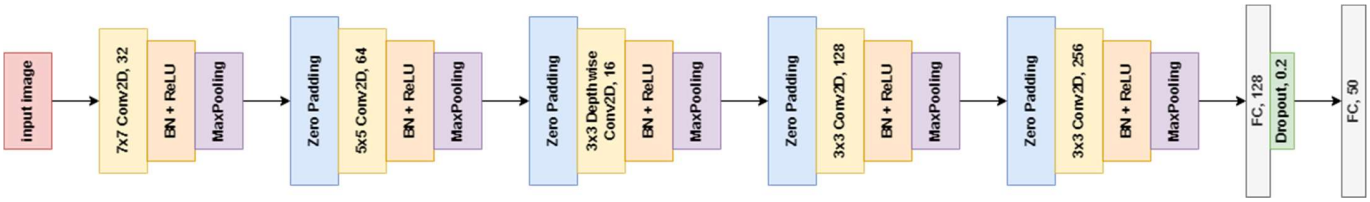

Fig3. My CNN Model Structure

Training the Model

Machine learning model created and trained on the custom dataset, by using TensorFlow library. In the end model has 80% training accuracy, and 73% validation accuracy. The trained model was then transformed into a TensorFlow Lite model format, in order to work with the Flutter framework.

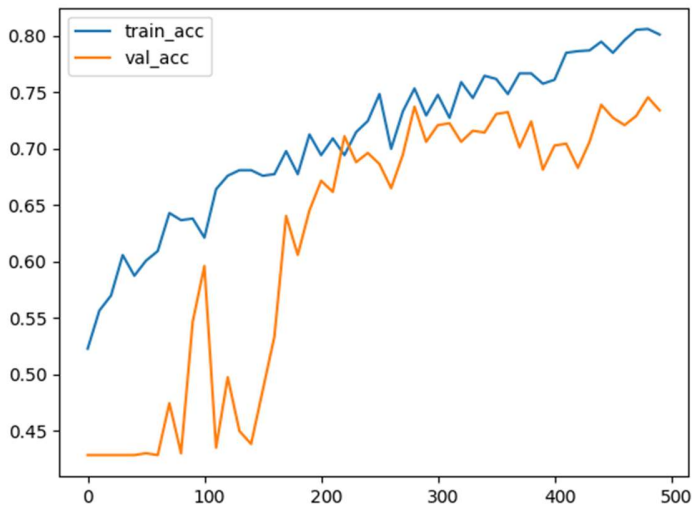

Fig4. Trainig Accuracy and Validation Accuracy Over Training Epochs

Mobile Application Development

Flutter, open-source mobile UI framework, is chosen for developing a stylish mobile application. I developed the mobile application in a way that it can make predictions on a given image, based on saved TensorFlow Lite model in the previous step, by using tflite and tflite_helper packages. Both gallery or device camera can be used as an image source. Mobile application returns five guesses with highest probability, in order to minimize false predictions.

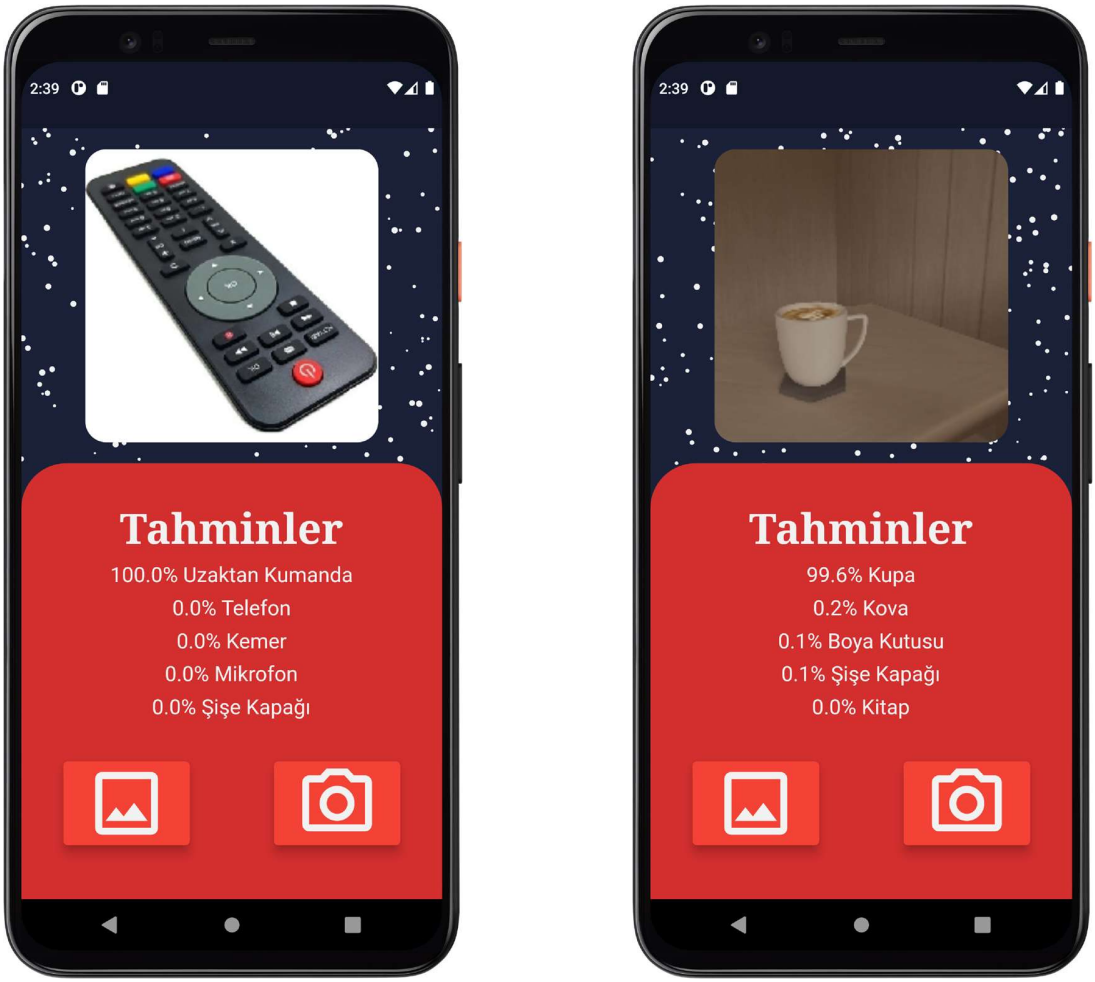

Fig5. Image Classification over Gallery Image (Left) and Camera Image (Right)

4 Performance and Outcomes

4.1 Applying Knowledge and Skills Learned at Hacettepe

Introduction to Programming lectures (BBM 101 – BBM 102) helped me to develop a machine learning application in Python when following object oriented programming principles. Also Introduction to Computer Engineering (BBM105) helped me to develop a modern looking mobile application.

4.2 Solving Engineering Problems

In order to solve overfitting problem in the model, data augmentation techniques applied to the existing dataset. Data augmentation techniques used in the project includes rotation, cropping, horizontal flip, brightness augmentation, saturation augmentation.
Another problem was about model's accuracy. My own image classification model performed with 80% accuracy on the training data. This was not a big problem since my mobile application shows the most probable five prediction instead of returning only the most probable one. However, in order to improve accuracy, I used transfer learning. Using a pre-trained model on a new machine learning task is called transfer learning. Since the model has some ideas about the task before (classification tasks, ie. used datasets, should be similar in order for transfer learning to work), it performs faster on the new task. This method also improves the accuracy. I choosed EfficientNet B0 [5] for transfer learning. After model selection, model training and application development parts were intuitive, since I already implemented them for my custom model.

4.3 Teamwork

During my main project at the software department, I worked with another computer engineering intern. Together, we developed a mobile application for an inventory management system. For the project, I created a dataset, a model for image classification, and developed a mobile application. A problem about the task is that the dataset is not constant, in other words new items come to the inventory. Since those items never seen by the trained model, application cannot make plausible predictions about them. Our supervisor advised us to retrain the model when a new item comes. Since one or two new items added to the inventory every month, retraining the model every month is reasonable. Thus my teammate worked to solve this problem when needed. Thanks to this, a new item can be added to the system with its label and image data, with the approval of the authorized personnel. Model then retrained and mobile application updated automatically.

4.4 Multi-Disciplinary Work

I was not a part of a multi-disciplinary work or team during my internship.

4.5 Professional and Ethical Issues

I did not encounter any professional or ethical issues during my summer internship.

4.6 Impact of Engineering Solutions

During my internship at BOTAŞ, I once again recognized the importance of engineering solutions. BOTAŞ is a really large company with so many different element working together, and without proper engineering approaches, it would be really hard to maintain all fragments of it. Moreover, thanks to the cyber security and networking training that is given to me, I can understand the responsibilities of a computer scientist better. A developer's actions have huge impacts on a company's economic, social, and global elements.

4.7 Locating Sources and Self-Learning

For machine learning training I followed official TensorFlow tutorials [6]. For developing a mobile application, I followed Flutter tutorials [7], and Flutter cookbook [8].

4.8 Using New Tools and Technologies

At the end of my internship, I was familiarized with TensorFlow and able construct simple image classification models. Also, I could train them on custom datasets. Moreover, I could develop basic mobile applications with Flutter.

5 Conclusions

To sum up, I developed a mobile application to automate inventory management system of BOTAŞ. I was not familiar with machine learning or mobile application development before, but Hacettepe University taught me how to improvise, revise, and solve a software problems. Thus I could learn new things fast and adapt to the challenging situations as a computer scientist.

References

[1] “Lisans Öğrencileri İçin | BOTAŞ - Boru Hatları İle Petrol Taşıma Anonim Şirketi”. https://www.botas.gov.tr/Sayfa/lisans-ogrencileri-icin/196. [Accessed; August 11, 2022].

[2] “What is dotnet?”. https://dotnet.microsoft.com/en-us/learn/dotnet/what-is-dotnet. [Accessed; August 12, 2022].

[3] “Oracle Exadata”. https://www.oracle.com/engineered-systems/exadata/. [Accessed; August 12, 2022].

[4] “DomainNet Dataset”. http://ai.bu.edu/M3SDA/#dataset. [Accessed; August 8, 2022].

[5] “EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks”. https://arxiv.org/pdf/1905.11946v5.pdf. [Accessed; August 10, 2022].

[6] “TensorFlow Tutorials”. https://www.tensorflow.org/tutorials. [Accessed; August 1, 2022].

[7] “Flutter Tutorials”. https://docs.flutter.dev/reference/tutorials. [Accessed; August 9, 2022].

[8] “Flutter Cookbook”. https://docs.flutter.dev/cookbook. [Accessed; August 9, 2022].