上海交通大学
学生实习报告

实习题目：Text Recognition via OCR
实习单位：上海交通大学

实习时间：2023-06-19 至 2023-07-14

学院(系)：电院自动化系

专　　业：自动化

学生姓名：JOONGWON CHAE 学号：518030990043

2023年 7 月 17日

实习报告主要内容包括实习目的与任务、实习单位、实习内容、实习收获等

OCR and image processing project report

1. Introduction

Over the past four weeks, I conducted a practical exercise on deep learning and image processing. The greatest achievements during this process included the enhancement of debugging skills, a mathematical understanding of existing image processing algorithms, and the mastery and learning of new

deep learning structures and techniques. The main content of the practice was to extract text using OCR (Optical Character Recognition) and create CSV files. The final goal was to complete image processing within 30 seconds.

2. Practice Process

The practice was carried out in the following three stages.

2.1. Initial Attempt using YOLO

Initially, we explored the potential for text extraction based on YOLO. We used YOLO to create pt files and used them to create bounding boxes. However, there were problems such as YOLO taking too long to detect more than 100 boxes at once, and needing nearly 1000 bounding box detections to detect combinations of characters and numbers consisting of 8 digits. There were unsolvable issues like overlapping bounding box regions, long detection time due to a large number of box detections, and errors occurring when text detection failed. Despite attempts to improve this by debugging and adjusting various parameters like anchor boxes and batch size, the problem was not solved, so we decided to use OCR open source for processing.

2.2. Study on OCR and Deep Learning Structure

We studied various deep learning-based OCR tools, including Naver's EasyOCR and Baidu's PaddleOCR, to apply OCR. During this process, we studied image preprocessing techniques such as Affine and Perspective, applied them, and increased the text recognition rate from 70% to 80%. Additionally, we also learned about deep learning structures like VGG16 and U-Net and implemented them.

2.3. Application and Optimization of PaddleOCR

Due to the unsatisfactory performance of Naver's EasyOCR, we decided to apply Baidu's PaddleOCR. We used PaddleOCR to extract text from images, and the text recognition rate reached 94%. Various image preprocessing techniques were applied, including Threshold, adaptive threshold, canny edge, shadow removing, and perspective. However, there were problems where the recognition rate decreased when applying other techniques. Ultimately, we confirmed that images with Perspective applied were the most effective.

Thereafter, while the recognition rate of the image was good, the instability of the algorithm that automatically finds Perspective was an issue. It was because the algorithm finding the contour through the edge of the image was affected by the amount of light and changes in angles, which influenced the performance of Perspective. To solve this, we tried to find the edges of the image using YOLO, but failed to find the feature points.

Afterwards, we assigned black stickers and cross-shaped features to the edges of the boxes to be detected and applied Perspective using Template Matching, and then conducted OCR based on this. However, the problem of template matching being sensitive to light and angle kept occurring, and edge detection through perspective transform failed when slight condition changes occurred. Finally, we conducted feature matching and successfully performed edge detection under various conditions.

During this process, we created a GUI and added functions to take pictures with a camera, perform OCR extraction, and create CSV files. However, there was a problem where multithreading did not occur simultaneously from the camera. To solve this, we modified the code so that when the button is pressed, pictures can be taken sequentially from two cameras and displayed on the UI. This was an instance where we could understand why C# and C++ are used more often than python in UI creation.

3. Conclusion

Through practice, we improved debugging skills and conducted mathematical learning on image processing algorithms. This allowed us to understand how mathematical knowledge can assist when coding. We also mastered and learned new deep learning structures and techniques. In particular, we successfully extracted text using OCR technology and converted it into CSV files. However, the processing speed was slower than expected. This was due to the performance limit of the CPU and the decrease in recognition rate when reducing image size. We expect this problem can be solved by using a GPU or high-performance CPU.

During this project, the advancement of AI was truly astonishing. I used CHAT GPT4 and Github Copilot, which allowed me to design algorithms easily based on ideas.

The main takeaway from this project was that AI is not a panacea but should be used as a tool. To use AI proficiently, the user's basic knowledge is critically important. AI can write code proficiently based on user commands and find problems, but if the user does not understand the functions of the library and does not give detailed commands, it cannot produce satisfactory results.

The development of China's AI technology was also very surprising. After conducting research using all three of Tesseract, EasyOCR, and Paddle, Paddle's basic performance was truly astonishing. It showed better performance than custom-trained Tesseract and EasyOCR. It was also well-customized under various conditions, there were many materials in Chinese, and the quality of the materials was good.

Considering that the most important skill when coding is finding ideas and materials, I believe English and Chinese have become essential elements for programmers.

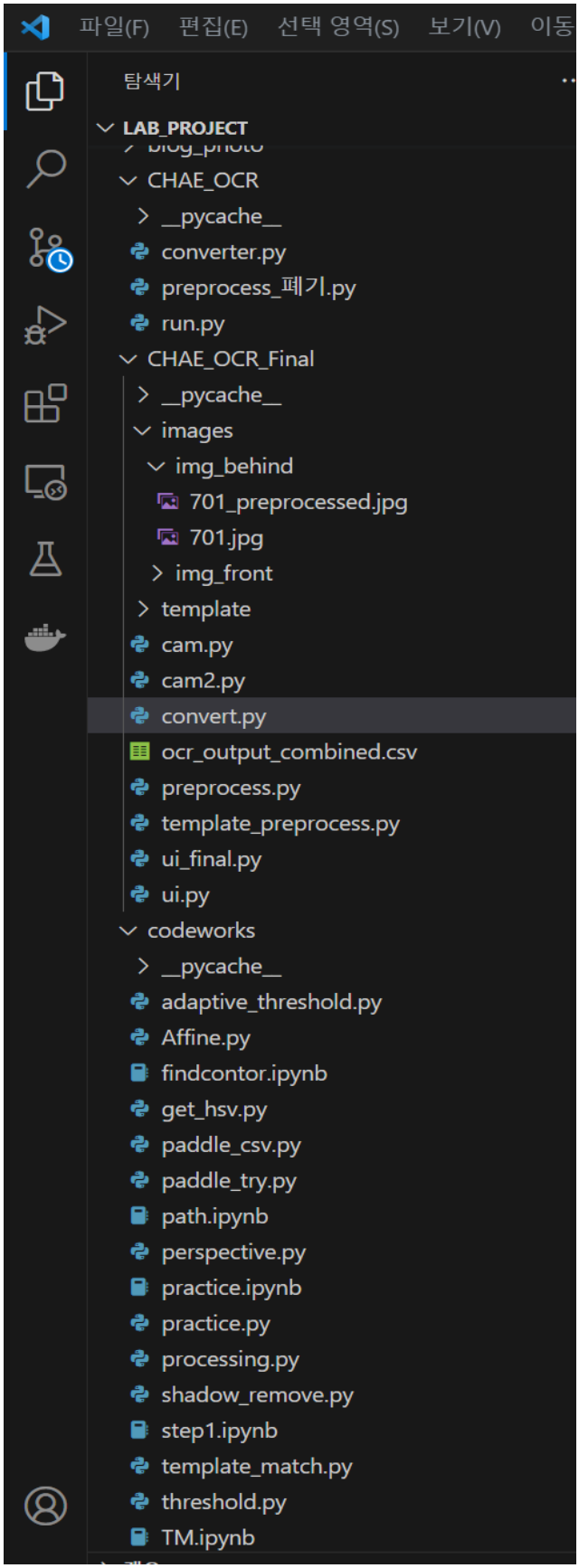


Figure 1. studied Algorithms

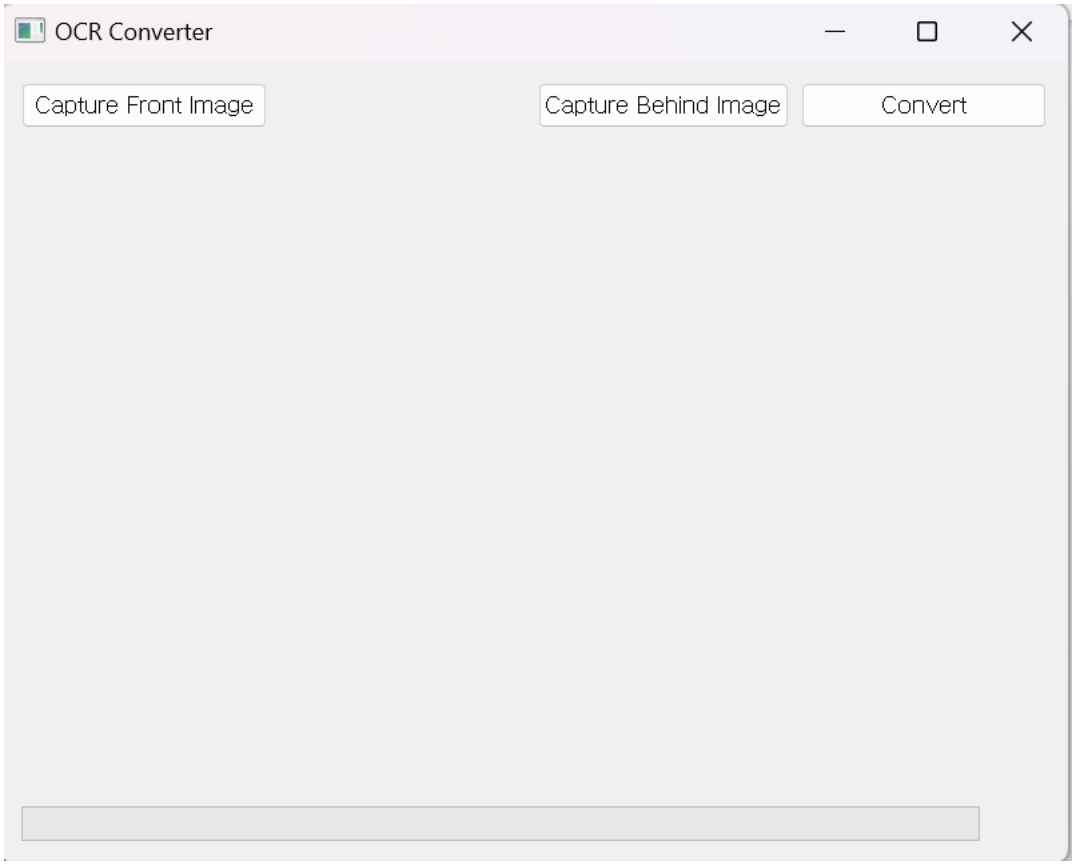


Figure 2. UI Design

Figure 3. Image before preprocessing

Figure 4. Image after preprocessing

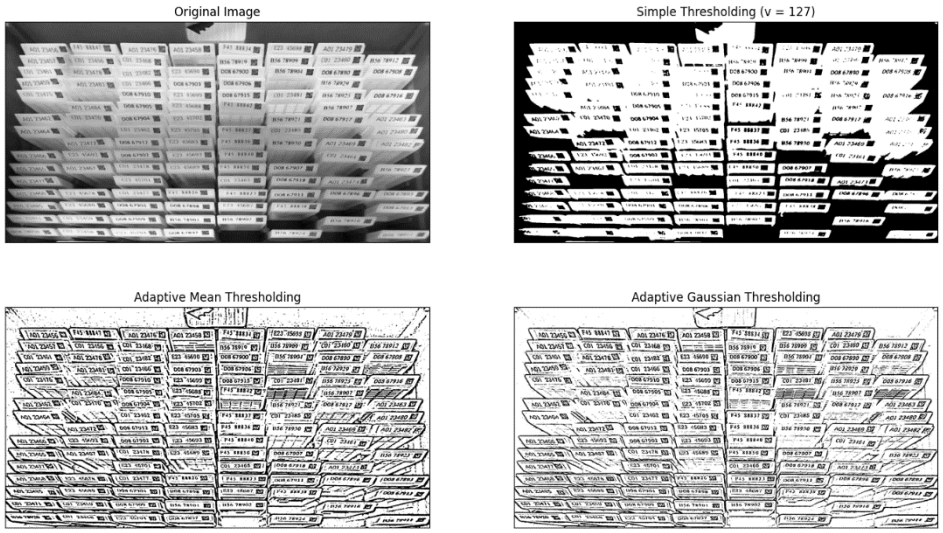
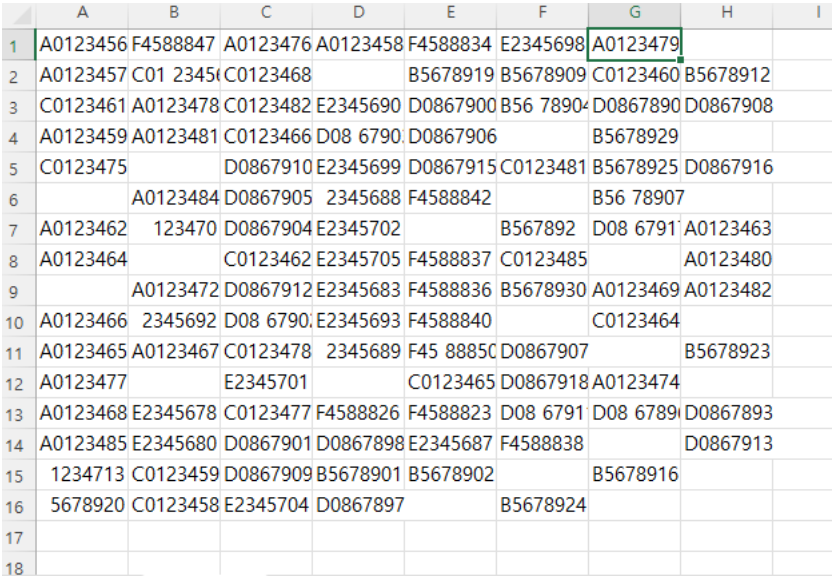


Figure5. Algorithms for bad results



	A	B	C	D	E	F	G	H	I
1	A0123456	F4588847	A0123476	A0123458	F4588834	E2345698	A0123479		
2	A0123457	C01 2345	C0123468		B5678919	B5678909	C0123460	B5678912	
3	C0123461	A0123478	C0123482	E2345690	D0867900	B56 7890	D0867890	D0867908	
4	A0123459	A0123481	C0123466	D08 6790	D0867906		B5678929		
5	C0123475		D0867910	E2345699	D0867915	C0123481	B5678925	D0867916	
6		A0123484	D0867905	2345688	F4588842		B56 78907		
7	A0123462	123470	D0867904	E2345702		B567892	D08 6791	A0123463	
8	A0123464		C0123462	E2345705	F4588837	C0123485		A0123480	
9		A0123472	D0867912	E2345683	F4588836	B5678930	A0123469	A0123482	
10	A0123466	2345692	D08 6790	E2345693	F4588840		C0123464		
11	A0123465	A0123467	C0123478	2345689	F45 8885	D0867907		B5678923	
12	A0123477		E2345701		C0123465	D0867918	A0123474		
13	A0123468	E2345678	C0123477	F4588826	F4588823	D08 6791	D08 6789	D0867893	
14	A0123485	E2345680	D0867901	D0867898	E2345687	F4588838		D0867913	
15	1234713	C0123459	D0867909	B5678901	B5678902		B5678916		
16	5678920	C0123458	E2345704	D0867897		B5678924			
17									
18									

Figure6. results

专业实习企业导师考核评分表

实习生姓名	JOONGWON CHAE	指导教师	于文彬
实习单位	上海交通大学	团队负责人	于文彬

考核项目及说明	对应的毕业要求指标点	项目分值	考核得分	备注
1、严格遵守公司各项规章制度，无违纪情况，理解并遵守工程职业道德和规范，履行责任	8-3	10	10	
2、出勤		10	10	
3、了解相关的方针、政策与法律法规。了解相关工程项目的行业规范与标准	7-1	10	10	
4、能在规定时间内完成任务，遇到问题迅速反馈，工作质量达到预期效果		15	14	
5、正确理解工作目标，善于总结、学习，具有一定的自我学习能力	12-2	15	14	
6、了解本专业的科学技术发展前沿和趋势，了解专业领域的工程最新应用背景	6-1	10	9	
7、能够综合运用所学知识，对工程问题进行识别、分析、表达，获得有效结论	2-4	10	10	
8、能够针对具体问题设计解决方案,并考虑社会、健康、安全、法律、文化以及环境等因素	3-3	10	9	
9、了解自动化工程实践与环境保护的关系，能评价工程实践对社会可持续发展的影响	7-2	10	9	
总　分		100	95	

实习单位综合意见：

项目进展顺利，效果良好。实生在实习中努力刻苦，较好完成了项目

实习单位人事签章

考核人员签名		考核日期	2023.7.17

11

专业实习企业导师考核评分表

实习报告评阅人意见	评阅人（签名）： 年　　月　　日

系主任意见	评　定　成　绩：_____ 系　主　任(签名)：_____ 年　　月　　日