

รายงานการปฏิบัติงานสาขาวิชาศึกษา^{เรื่อง}

ระบบตรวจหาข้อความบนภาพมังงะด้วยเทคนิค Stroke
Width Transform
Detecting Text in Manga using Stroke Width
Transform

ปฏิบัติงาน ณ มหาวิทยาลัยอุบลราชธานี

โดย

นาย บุญฤทธิ์ พริย์โภชินกุล

รหัสประจำตัว 58070077

รายงานนี้เป็นส่วนหนึ่งของการศึกษารายวิชา สาขาวิชาศึกษา^{สาขาวิชาวิศวกรรมซอฟต์แวร์คอมพิวเตอร์} ภาคเรียนที่ 1 ปีการศึกษา 2561
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

รายงานการปฏิบัติงานสาขาวิชากศึกษา

ระบบตรวจหาข้อความบนภาพมังงะด้วยเทคนิค Stroke

Width Transform

Detecting Text in Manga using Stroke Width

Transform

ปฏิบัติงาน ณ มหาวิทยาลัยฮอกไกโด

โดย

นาย บุญฤทธิ์ พริย์โภชินกุล

รหัสประจำตัว 58070077

ปฏิบัติงาน ณ มหาวิทยาลัยฮอกไกโด

Hokkaido University Kita 8, Nishi 5, Kita-ku,

Sapporo, Hokkaido, 060-0808 Japan

Web site : www.global.hokudai.ac.jp

Detecting Text in Manga using Stroke Width Transform

Boonyarith Piriyothinkul

**A REPORT SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENT FOR COOPERATING EDUCATION PROGRAM
THE DEGREE OF BACHELOR OF SCIENCE PROGRAM IN
INFORMATION TECHNOLOGY
FACULTY OF INFORMATION TECNOLOGY
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG**

**COPYRIGHT 2018
FACULTY OF INFORMATION TECHNOLOGY
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG**

วันที่ 10 พฤษภาคม พ.ศ. 2561

เรื่อง ขอส่งรายงานการปฏิบัติงานสหกิจศึกษา
เรียน รองศาสตราจารย์ ดร. กิตติสุชาต พสุภา¹
ที่ปรึกษาสหกิจศึกษาในสาขาวิชาโภชนาศึกษา (วิศวกรรมซอฟต์แวร์)

ตามที่ข้าพเจ้า นาย นุญฤทธิ์ พิริยะธนกุล นักศึกษาสาขาวิชา วิศวกรรมซอฟต์แวร์คณะเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง ได้ปฏิบัติงานสหกิจศึกษาระหว่างวันที่ 23 กรกฎาคม พ.ศ. 2561 ถึงวันที่ 30 พฤษภาคม พ.ศ. 2561 ในตำแหน่ง ผู้ช่วยนักวิจัย ณ สถาบันประกอบการ มหาวิทยาลัยหอการค้าไทย และได้รับมอบหมายจากพนักงานที่ปรึกษาให้ศึกษาและจัดทำรายงาน เรื่อง ระบบตรวจหาข้อความบนภาพมังงะด้วยเทคนิค Stroke Width Transform

บัดนี้ การปฏิบัติงานสหกิจศึกษาได้สิ้นสุดลงแล้ว จึงคร่าวงขอส่งรายงานการปฏิบัติงาน สหกิจศึกษา ดังกล่าวมาพร้อมนี้ จำนวน 1 เล่ม เพื่อขอรับคำปรึกษาต่อไป

จึงเรียนมาเพื่อโปรดพิจารณา

ขอแสดงความนับถือ

.....
(นาย นุญฤทธิ์ พิริยะธนกุล)

กิตติกรรมประกาศ

ตามที่ข้าพเจ้า นาย นุญฤทธิ์ พิริย์ไยชินกุล ได้มาปฏิบัติงานสหกิจศึกษา ณ มหาวิทยาลัยออกไกโอด ตั้งแต่วันที่ 23 กรกฎาคม พ.ศ. 2561 ถึงวันที่ 30 พฤษภาคม พ.ศ. 2561 ทำให้ข้าพเจ้าได้รับความรู้และประสบการณ์ต่าง ๆ ที่มีคุณค่ามากmany สำหรับรายงานสหกิจศึกษานี้บันทึกไว้ดังนี้

1. Professor Dr. Masanori Sugimoto ตำแหน่ง ศาสตราจารย์
2. Jiang Ye ตำแหน่ง นักศึกษาปริญญาเอก ปี 2

นอกจากนี้ยังมีบุคคลท่านอื่น ๆ อีกที่ไม่ได้กล่าวไว้ ณ ที่นี่ ซึ่งให้ความกรุณาแนะนำในจัดทำรายงานสหกิจศึกษานี้ ข้าพเจ้าจึงขอรับขอบพระคุณทุกท่านที่ได้มีส่วนร่วมในการให้ข้อมูลและให้ความเข้าใจเกี่ยวกับชีวิตของการปฏิบัติงาน รวมถึงเป็นที่ปรึกษาในการจัดทำรายงานฉบับนี้จริงเสีย สมบูรณ์

นาย นุญฤทธิ์ พิริย์ไยชินกุล
ผู้จัดทำรายงาน
วันที่ 10 พฤษภาคม พ.ศ. 2561

ชื่อรายงานการปฏิบัติงานสหกิจศึกษา	ระบบตรวจหาข้อความบนภาพมังงะด้วยเทคนิค Stroke Width Transform
ผู้รายงาน	นาย บุญฤทธิ์ พิริย์ไบชินกุล
คณะ	เทคโนโลยีสารสนเทศ
สาขาวิชา	วิศวกรรมซอฟแวร์

.....
**(รองศาสตราจารย์ ดร. กิตติสุชาต พสุภา)
 อาจารย์ที่ปรึกษาสหกิจศึกษา**

.....
**(Professor Dr. Masanori Sugimoto)
 พนักงานที่ปรึกษา**

คณะเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
 อนุมัติให้นับรายงานการปฏิบัติงานสหกิจศึกษาฉบับนี้ เป็นส่วนหนึ่งของการศึกษา
 ตามหลักสูตรวิทยาศาสตรบัณฑิต สาขาวิชาวิศวกรรมซอฟแวร์

ชื่อนักศึกษา	ระบบตรวจหาข้อความบนภาพมังงะด้วยเทคนิค Stroke Width Transform
รหัสนักศึกษา	นาย นุญฤทธิ์ พิริย์ไยธินกุล
สาขาวิชา	วิศวกรรมซอฟแวร์
อาจารย์ที่ปรึกษา	รองศาสตราจารย์ ดร. กิตติสุชาต พสุภา
ปีการศึกษา	2561

บทคัดย่อ

การ์ตูนญี่ปุ่น หรือที่รู้จักกันอย่างแพร่หลายว่า มังงะ (Manga) ถูกเป็นหนึ่งในหัวข้อที่ถูกหยิบมาวิจัย งานวิจัยชิ้นนี้มุ่งเน้นที่ปัญหาการตรวจหาข้อความบนภาพมังงะ เนื่องจากปัญหาการสร้างชุดข้อมูลภาพมังงะและข้อมูลประกอบ (Annotation) อย่างเช่นการระบุขอบเขตข้อความ ซึ่งต้องใช้แรงงานคนและกินเวลาอย่างมาก ดังนั้นระบบอัตโนมัติที่สามารถเข้ามาช่วยในงานส่วนนี้จึงเป็นที่ต้องการอย่างมาก โดยเราได้นำเสนอวิธีการตรวจหาข้อความบนภาพมังงะแบบใหม่ด้วยการใช้ Stroke Width Transform (SWT) ร่วมกับการใช้ Support Vector Machine (SVM) อย่างไรก็ได้ วิธีการเดิมที่ใช้ SWT เพื่อการตรวจหาข้อความบนภาพถ่ายนั้น ไม่สามารถประยุกต์ใช้กับภาพมังงะ ได้ เพราะความแตกต่างระหว่างวัตถุและลักษณะอักษรของข้อความในภาพวาดและภาพถ่าย ไม่ว่าจะในเชิงความคล้ายคลึง, ขนาด, รูปร่าง, และลักษณะของเส้นเหล่านั้นคล้ายคลึงมากกันเกินไป ดังนั้นเพื่อให้สามารถใช้งานกับมังงะ ได้ เราจึงนำวิธีการตรวจหาข้อความด้วย SWT ต้นฉบับมาปรับปรุงและพัฒนาขึ้นเป็นวิธีการใหม่ของเรารา ได้ปรับปรุงกฎเกณฑ์ในการค้นหาวัตถุที่คล้ายคลึงอักษร (Letter Candidates) ซึ่งช่วยเพิ่มประสิทธิภาพในการตรวจจับอักษร ได้ครบถ้วนมากขึ้น และใช้ SVM เพื่อคัดแยกวัตถุอื่น ๆ ออกจากอักษร ช่วยในการลด False Positive ของผลลัพธ์ ในท้ายที่สุดเรานำประสิทธิภาพวิธีการของเรามาเปรียบเทียบร่วมกับวิธีการต้นฉบับและวิธีอื่น ๆ รวมถึงวิธีที่ใช้ Deep Learning เป็นส่วนประกอบ ในท้ายที่สุดประสิทธิภาพของวิธีการใหม่ของเรานั้นสามารถทำคะแนน F-measure ได้สูงสุดเทียบกับวิธีการอื่น ๆ ที่ 0.506

สารบัญ

	หน้า
หนังสือส่งรายงานการปฏิบัติงานสาขาวิชากิจศึกษา	i
กิตติกรรมประกาศ	ii
หน้าอนุมัติรายงาน	iii
บทคัดย่อ	iv
สารบัญ	v
สารบัญตาราง	vi
สารบัญภาพ	vii
บทที่ 1 บทนำ	1
1.1 ที่มาและความสำคัญ	1
1.2 วัตถุประสงค์	2
1.3 ขอบเขตของงานวิจัย	2
1.4 ประโยชน์ที่คาดว่าจะได้รับ	2
บทที่ 2 แนวคิด ทฤษฎีและงานวิจัยที่เกี่ยวข้อง	3
2.1 การตรวจหาข้อความในภาพถ่ายด้วยเทคนิค Stroke Width Transform	3
2.2 Histogram of Oriented Gradients	5
2.3 Support Vector Machine	6
บทที่ 3 วิธีการทดลอง	8
3.1 วิธีการใหม่ที่อุปกรับปรุงและพัฒนาเพิ่มเติม	8
3.2 ชุดข้อมูลสำหรับการ tren โมเดล SVM	12
3.3 การทดลอง	12
บทที่ 4 ผลการทดลอง	14
บทที่ 5 สรุปผล	16
บรรณานุกรม	17
ภาคผนวก ก การใช้ชีวิตในประเทศไทยปัจจุบัน	20
ก.1 ที่อยู่อาศัย	20
ภาคผนวก ข กิจกรรมระหว่างฝึกงาน	22
ข.1 Mirai Symposium	22
ข.2 Lab Meeting	22
ข.3 กิจกรรมอื่น ๆ	23

สารบัญตาราง

หน้า

ตารางที่ 4.1 ตารางแสดงการเปรียบเทียบประสิทธิภาพของวิธีการใหม่ของเราร่วมกับวิธีการ
อื่น ๆ

14

สารบัญภาพ

	หน้า
รูปที่ 2.1 ขั้นตอนการทำงานของ Stroke Width Transform	4
รูปที่ 2.2 ตัวอย่าง Histogram of Oriented Gradients ก่อนแปลงสภาพลายเป็นเวกเตอร์ของตัวอักษรภาษาญี่ปุ่น	5
รูปที่ 2.3 Cell และ Block ในการทำงานของ Histogram of Oriented Gradients	6
รูปที่ 2.4 การแบ่งแยกกลุ่มข้อมูลด้วย Hyper-plane ของ SVM	7
รูปที่ 2.5 ตัวอย่างแสดงคุณสมบัติการเปลี่ยนมิติของข้อมูลด้วย Kernel	7
รูปที่ 3.1 ตัวอย่างผลลัพธ์จากการตรวจหาข้อความบนภาพมังงะคัวบิชิตันฉบับ [12] ซึ่งแสดงให้เห็น False positive จำนวนมาก (ก) นักวาด: Shinoasa (ข) นักวาด: Kousei (Public Planet)	9
รูปที่ 3.2 แผนผังการทำงานของ (ก) วิธีการดึงเดิน [12] และ (ข) วิธีการใหม่ของเรา	9
รูปที่ 3.3 ตัวอย่างแสดงการเปรียบเทียบผลลัพธ์ของขอบเขตข้อความที่ตรวจพบระหว่างวิธีการใหม่ (a) และวิธีการต้นฉบับ (b) ข้อมูลภาพถูกนำมาจากเรื่อง Arisa ©Yagami Ken	11
รูปที่ 3.4 ตัวอย่างของ Patch: (ก) ภาพ positive patches and (ข) ภาพ negative patches	11
รูปที่ 3.5 ตัวอย่างแสดงการจับกลุ่มของตัวอักษร	12
รูปที่ 4.1 ตัวอย่างขอบเขตข้อความในวิธีการของเราราตรวจพบ (ก–ข) Love Hina ©Ken Akamatsu และ (ก–ก) Eva Lady ©Miyone Shi.	15
รูปที่ ก.1 ภาพหอพัก International House Kita 8 East	21
รูปที่ ข.1 ภาพกิจกรรมในงาน Mirai symposium	23
รูปที่ ข.2 ภาพระหว่างงานเลี้ยงต้อนรับตัวผู้	24
รูปที่ ข.3 ภาพงานเลี้ยง野心ตัวผู้และต้อนรับนักศึกษาปีสาม	25
รูปที่ ข.4 ป้ายเชิญชวนชมห้องทดลองโดยการตัวละครจากการ์ตูนประกอบให้น่าสนใจ	26

บทที่ 1

บทนำ

1.1 ที่มาและความสำคัญ

การตูนญี่ปุ่นเป็นที่รู้จักกันอย่างแพร่หลายทั่วโลกในฐานะสื่อบันเทิง หรืออีกชื่อหนึ่งคือ “มังงะ (Manga)” ในปัจจุบันมีงานวิจัยในหัวข้อมังงะอย่างหลากหลาย ในหลาย ๆ งานวิจัย [1–7] มีการใช้ชุดข้อมูลสำหรับการทดสอบอย่างเช่น Manga109 [8] โดยชุดข้อมูลนี้เป็นข้อมูลที่ถูกสร้างขึ้นจากภาพมังงะจำนวน 20,260 หน้า จากมังงะ 109 เรื่อง ถูกวัดโดยนักวัดมืออาชีพชาวญี่ปุ่น นอกจากภาพสแกนของมังงะแล้ว ชุดข้อมูลนี้ยังประกอบไปด้วยข้อมูลอธิบายประกอบ หรือ Annotation ต่าง ๆ เช่น ขอบเขต และตำแหน่งของใบหน้า ร่ายกาย กรอบภาพ เป็นต้น นอกจากนี้ยังมีข้อมูลของเขตและตำแหน่งของข้อความที่ปรากฏในภาพมังงะซึ่งถูกป้อนข้อมูลด้วยแรงงานคน โดยไม่พึงพาระบบอัตโนมัติ ฯ ใน การป้อนข้อมูลดังกล่าวเน้นใช้เวลานานและต้องพึ่งพาแรงงานมนุษย์ ด้วยเหตุนี้ระบบอัตโนมัติที่จะเข้ามาช่วยในการระบุข้อมูล Annotation นั้นจึงมีประโยชน์และลดภาระงานลงได้อย่างมาก

ถึงแม้สำหรับภาพการตูนญี่ปุ่นจะมีทั้งแบบภาพวาดทั่วไปและมังงะ แต่ในงานวิจัยนี้เรามุ่งเน้นไปที่มังงะเป็นหลักเนื่องจากข้อความมักมีบนหนังสือการตูนมากกว่าภาพวาดทั่วไปอย่างที่ทราบกันสำหรับวิธีการตรวจหาข้อความในภาพมังงะนั้นมีการพัฒนามาหากาหളายก่อนหน้านี้ [9, 10] แต่วิธีเหล่านี้ถูกจำกัดให้ทำงานภายในโครงสร้างของภาพมังงะต่าง ๆ เช่น กรอบช่องภาพวาด, ลักษณะของกล่องคำพูด เป็นต้น นอกจากนี้บางวิธียังคงต้องพึ่งพาการป้อนข้อมูลเข้าจากภายนอกทำให้ไม่สามารถทำงานได้อัตโนมัติอย่างสมบูรณ์ อีกทั้งไม่สามารถนำมีการพัฒนาวิธีการใหม่โดยใช้วิธีการ Deep Learning อย่างเช่นเทคนิค Convolutional Neural Network เพื่อช่วยในการสกัดลักษณะเด่น (Feature) ออกจากภาพมังงะเพื่อช่วยในการตรวจหาข้อความในภาพมังงะ [11] ซึ่งวิธีนี้การลดข้อผิดพลาด False Positive ไปได้โดยปราศจากการใช้โครงสร้างในภาพมังงะต่าง ๆ แต่อย่างไรก็ต้อง Deep Learning ยังเป็นการวิธีการที่มีข้อเสียคือใช้ทรัพยากรในการคำนวนของระบบมากกว่าระบบอื่น ๆ [11]

ในงานวิจัยนี้เราต้องการพัฒนาระบบตรวจหาข้อความที่ทำงานได้กับมังงะอย่างหากาหളายและไม่ถูกจำกัดด้วยโครงสร้างหรือลักษณะบางประการของภาพมังงะ เราจึงเลือกใช้ Stroke Width Transform (SWT) ในการสกัดลักษณะเด่นของเส้นต่าง ๆ ที่อยู่ในภาพอุปกรณ์ โดยวิธีการนี้ถูกใช้เป็นขั้นตอนแรกเริ่มในการตรวจหาข้อความบนภาพถ่ายมาก่อน ทำงานด้วยการพึ่งพาสมมติฐานที่ว่าขอบของเส้นอักษรในข้อความนั้นมีขอบที่ชัดเจนและหนาแน่นบนพื้นหลังที่ราบรื่น [12] อีกทั้งไร์กีติการใช้วิธีการนี้กับการตรวจหาข้อความบนภาพมังงะส่งผลให้เกิด False Positive จำนวนมากในการตรวจหาซึ่งส่งผลทางลบต่อประสิทธิภาพการทำงานของระบบ ปัญหานี้เกิดจากความแตกต่างของภาพถ่ายและภาพวาดมังงะ มังงะนั้นเป็นภาพขาวดำ และลักษณะของวัตถุภายในมังงะ เช่น ขนาด, เส้น, พื้นหลัง นั้นมีความคล้ายคลึงกับตัวอักษรของข้อความมาก ดังนั้นเราจึงตั้งเป้าหมายในการปรับปรุงและพัฒนา SWT ที่ถูกใช้ในภาพถ่าย [12] และนำมันมาปรับปรุงพัฒนาเพิ่มเพื่อให้สามารถทำงานกับภาพมังงะได้

วิธีการใหม่ของเรานี้จะนำเสนอในงานนี้สามารถแบ่งออกเป็น 4 ส่วนดังนี้ (i) The Stroke Width Transform (ii) กันไฟวัตถุที่เข้าข่ายลักษณะของตัวอักษร (iii) คัดแยกอักษร โดยใช้ Support Vector

Machine (SVM) ร่วมกับ Histogram of Oriented Gradients Feature (iv) จัดกลุ่มอักษรที่ผ่านการคัดแยกแล้วเป็นบรรทัดเดียวกันหรือกลุ่มข้อความเดียวกัน

1.2 วัตถุประสงค์

พัฒนาระบบที่คำนึงถึงความแม่นยำของข้อความที่มีลักษณะทางกายภาพ เช่น Stroke Width Transform ที่ถูกใช้เป็นกระบวนการแยกเริ่มในเทคนิคตรวจหาข้อความบนภาพถ่ายมาพัฒนาและปรับปรุงเพื่อให้ใช้กับภาพมังงะได้มากขึ้น

1.3 ขอบเขตของงานวิจัย

1. พัฒนาระบบตรวจหาคำแม่นยำที่คำนึงถึงความซึ้งใช้สำหรับภาพมังงะขาวดำ
2. ภาษาของเนื้อหาในมังงะที่นำมาใช้งาน คือ ภาษาญี่ปุ่น
3. ข้อมูลที่ใช้ในการวิจัยนำมาจากฐานข้อมูล Manga109

1.4 ประโยชน์ที่คาดว่าจะได้รับ

1. ได้ Stroke Width Transform ที่ผ่านการปรับปรุงสำหรับภาพมังงะโดยเฉพาะเพื่อใช้ในการใช้งานด้านอื่น ๆ ต่อไป
2. ทำให้ทราบถึงลักษณะที่เป็นเอกลักษณ์ของมังงะซึ่งแตกต่างจากภาพถ่ายทั่วไป

บทที่ 2

แนวคิด ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1 การตรวจหาข้อความในภาพถ่ายด้วยเทคนิค Stroke Width Transform

วิธีการการตรวจหาข้อความในภาพถ่าย [12] นี้เป็นวิธีการที่เรานำมาใช้ศึกษาและเป็นต้นแบบในการพัฒนาเพื่อทำงานร่วมกับภาพมั่งจะ โดยมีขั้นตอนทั้งหมดแบ่งได้เป็น 3 ขั้นตอน ขั้นแรกคือการใช้ Stroke Width Transform ใน การปรับเปลี่ยนข้อมูลให้แสดงลักษณะของความกว้างในแต่ละเส้นภายในภาพ ขั้นที่สอง ค้นหาวัตถุที่คล้ายคลึงกับตัวอักษรในภาพ โดยใช้กฎเกณฑ์ที่กำหนดไว้ขั้นสุดท้าย ถ้าการจัดกลุ่มตัวอักษรเข้าด้วยกันเป็นบรรทัดของข้อความ

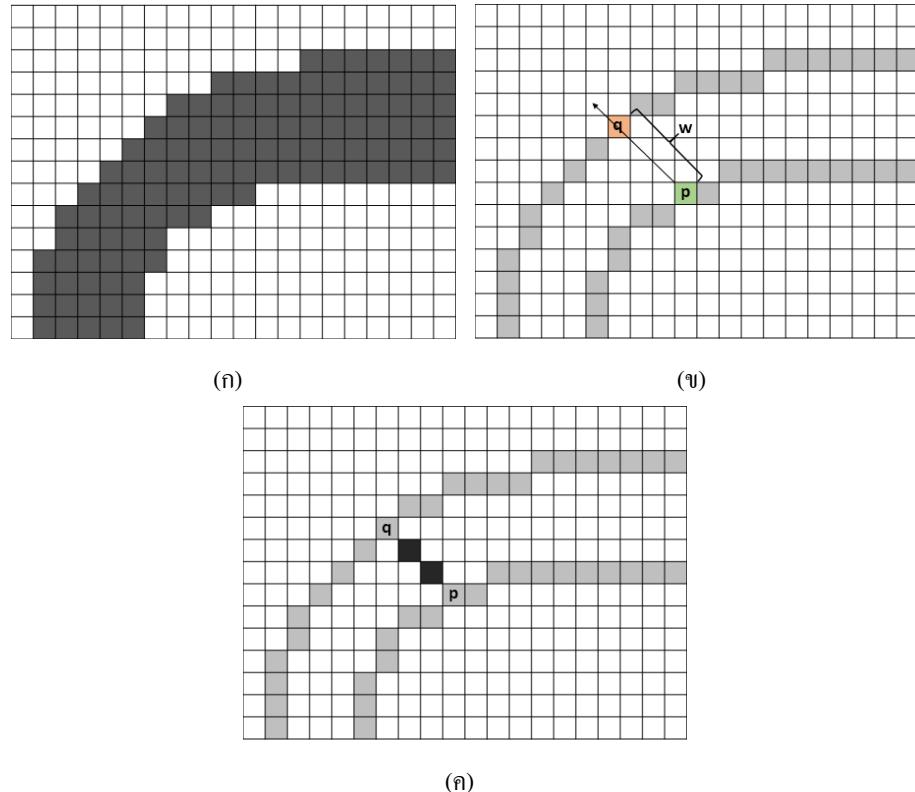
2.1.1 Stroke Width Transform

Stroke Width Transform หรือ SWT เป็นเทคนิคที่ใช้ช่วยในการทำงานของระบบตรวจหาข้อความในภาพถ่าย โดยสกัดลักษณะ (Feature) ของเส้นต่าง ๆ ในภาพ เช่น เส้นของตัวอักษร เป็นต้น [12] ด้วยลักษณะดังกล่าว ทำให้เราสามารถใช้ในการคัดแยกวัตถุที่เป็นตัวอักษรออกจากวัตถุอื่น ๆ โดยเพิ่งพัฒนาลักษณะเด่นเหล่านี้

เริ่มแรกเราสร้างภาพ Output ที่มีขนาดเท่ากับภาพที่ต้องการตรวจหาข้อความ โดยแต่ละ Pixel ในภาพถูกกำหนดให้มีค่าอนันต์ (∞) จากนั้นใช้ Canny Edge Detection [13] ซึ่งเป็นเทคนิคตรวจหาขอบของเส้นหรือวัตถุ ตรวจหาตำแหน่งของขอบของวัตถุในภาพ ต่อมาคือการคำนวณความกว้างของเส้น โดยใช้ขอบที่ได้มา คำนวณความกว้างคำนวนได้จากระยะห่างระหว่างขอบของเส้น โดยพิจารณาทุก Pixel p ของขอบที่ได้จาก Canny Edge Detection เพื่อหา Pixel q ที่เข้าคู่กัน อย่างที่แสดงให้เห็นในภาพ 2.1(บ) การหา q จาก p ทำได้โดยใช้ทิศทางของแนวตั้ง หรือ Gradient Direction ของ p ซึ่งคือ d_p โดย d_p จะซึ่งไปทาง q และหาก d_p และ d_q มีทิศทางตรงกันข้าม โดยประมาณ $d_q = -d_p \pm \pi/6$ ให้กำหนดค่าให้กับแต่ละ Pixel ที่อยู่ภายใต้แนวทางระหว่าง p และ q ให้เท่ากับ $\|\overrightarrow{p-q}\|$ เว้นแต่ Pixel ที่จะระบุค่าให้นั้นมีค่าน้อยกว่าค่าใหม่ที่จะระบุให้ หากค่าใหม่น้อยกว่าค่าเดิมใน Pixel ให้ทำการระบุค่าใหม่แทนที่ค่าเดิมให้กับ Pixel นั้น อย่างที่ปรากฏในภาพ 2.1(ค) ในที่สุดเราจะได้เมทริกซ์ Output ขนาดเท่ากับ Input ที่มีค่าแต่ละ Pixel เป็นขนาดความยาวของขอบในพื้นที่นั้น ๆ

2.1.2 ค้นหาวัตถุที่มีลักษณะใกล้เคียงตัวอักษร

ในขั้นตอนนี้เรานำผลลัพธ์จากขั้นตอนก่อนหน้ามาคำนวณวัตถุที่ไม่เกี่ยวข้องกับอักษรออก เริ่มจากจับกลุ่มแต่ละ Pixel ในผลลัพธ์ของขั้นตอน SWT การจับกลุ่มทำได้โดยเปรียบเทียบแต่ละ Pixel กับ Pixel เพื่อบันทึกขอบข้าง หาก Pixel ส่องตัวที่เปรียบเทียบกันมีค่าของความกว้างเส้นไม่ต่างกันเกิน 3.0 Pixel ทั้งสองจะถูกจัดกลุ่มเข้าด้วยกัน อย่างไรก็ได้หากวัตถุที่เกิดจากการรวมกลุ่มของ Pixel นั้นใหญ่หรือเล็กเกินไปก็จะถูกคัดออก การคัดวัตถุลักษณะดังกล่าวออกไปทำโดยการใช้กฎสองข้อ (i) อัตราส่วนระหว่างเส้นผ่าศูนย์กลางต่อมัธยฐานของความกว้างของเส้นตัวอักษรนั้นต้องน้อยกว่า 10 (ii) ความสูงต้องมากกว่า 10 และน้อยกว่า 300 ตามที่แสดงในสมการ 2.1



รูปที่ 2.1: ขั้นตอนการทำงานของ Stroke Width Transform

$$f(d, h, \tilde{s}) = \begin{cases} 1, & \text{if } \frac{d}{\tilde{s}} < 10 \text{ and } 10 < h < 300 \\ 0, & \text{otherwise} \end{cases}, \quad (2.1)$$

โดย d คือ เส้นผ่าศูนย์กลางของวัตถุ, h คือ ความสูงของวัตถุ, และ \tilde{s} คือ มัธยฐานของความกว้างเส้นตัวอักษร โดยค่าความกว้างได้มามาจากค่าของ Pixel ในพื้นที่ของเส้นที่ถูกระบุไปโดยขั้นตอน SWT

2.1.3 จัดกลุ่มตัวอักษร

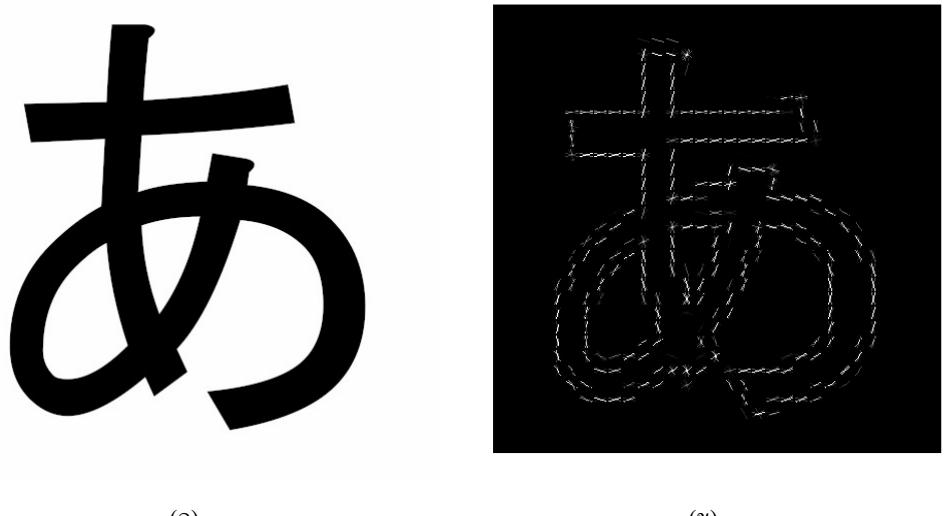
วัตถุแต่ละชิ้นที่ที่ลักษณะคล้ายคลึงกับอักษรซึ่งผ่านการคัดกรองด้วยกฎเกณฑ์จากขั้นตอนก่อนหน้าจะถูกนำมาจับกลุ่มในขั้นตอนนี้ ด้วยการเบริญเทียบความคล้ายคลึงระหว่างลักษณะอักษร ประกอบไปด้วย ระยะห่างระหว่างวัตถุ, อัตราส่วนความกว้างของเส้นอักษร, และความสูงของอักษร โดยสองวัตถุจะถูกจัดกลุ่มกันต่อเมื่อ (i) ห้องสองมีค่ามัธยฐานความกว้างเส้นน้อยกว่า 2 เท่า (ii) ความสูงของอักษรทั้งสองต่างกันไม่เกิน 2 เท่า (iii) ระยะห่างระหว่างสองวัตถุนั้นมีค่าไม่เกิน 3 เท่าของวัตถุที่กว้างที่สุดในสองตัวที่ใช้เบริญเทียบ หลังจากการจัดกลุ่มนี้เราจะได้ใช้ของอักษรที่ถูกจัดกลุ่มเข้าด้วยกันแต่ละ ใช้ประกอบไปด้วยอักษรส่องตัวที่ถูกจัดกลุ่ม ถัดจากนั้นแต่ละ ใช้จะถูกรวมเข้าด้วยกัน ได้ด้วย เช่นกันหากใช้อักษรไม่อักษรในโซ่อื่นๆ และทิศทางของโซ่มีความใกล้เคียงกัน

สุดท้ายขั้นตอนนี้จะจบลงเมื่อไม่มีโซ่อักษรใดๆ ถูกเชื่อมต่อเพิ่มเติม ในที่สุดเราจะได้กลุ่มหรือโซ่อักษรที่เกิดจากการจัดกลุ่มด้วยความคล้ายคลึงของอักษรและทิศทางของข้อความ อิกนัยนึงคือเรา

ได้กลุ่มบรรทัดของแต่ละประโภคของมาจากการถ่ายเรียงร้อยในขั้นตอนนี้

2.2 Histogram of Oriented Gradients

Histogram of Oriented Gradients (HOG) เป็นการสกัด Feature ของภาพโดยอาศัยรูปแบบ Histogram ของทิศทางเฉดสีในภาพ หรือ Gradient direction เพื่อพิจารณาลักษณะของวัตถุต่าง ๆ อาย่างที่แสดงในภาพ 2.2 ด้วยความสามารถนี้ จึงมีการนำ HOG มาใช้สำหรับสกัด Feature เพื่อใช้ในงานจำพวก Object detection อาย่างหลากหลาย ทั้ง การตรวจจับท่าทางของมือ [14] , การตรวจจับถนนท้องถนน [15] , การตรวจจับมุนุษย์ในภาพ [16] และไม่เพียงแค่สามารถใช้กับงานตรวจจับวัตถุ แต่ยังสามารถใช้กับงานด้านตรวจหาข้อความในภาพได้เช่นกัน [17–19]



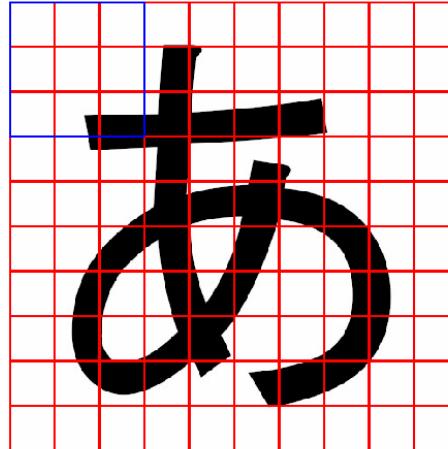
(ก)

(ล)

รูปที่ 2.2: ตัวอย่าง Histogram of Oriented Gradients ก่อนแปลงสภาพลายเป็นເງິນເຕອີ້ນຂອງตัวອັກຍາສູ່ປຸ່ນ

การสกัด Feature ของ HOG นั้นทำได้โดยเริ่มจากแบ่งภาพเป็นส่วนเล็ก เรียกว่า Cell หรือช่องสี แดงตามที่แสดงในภาพ 2.3 จากนั้นสร้าง Histogram สำหรับ Cell นั้น ๆ ด้วยค่า Gradient direction และ Magnitude โดย Histogram นี้จะเป็นตัวแทนของลักษณะของรูปร่างที่อยู่ภายใน Cell นั้น ๆ จากนั้นจะทำการ Normalization กับ Histogram ของแต่ละ cell ด้วยกลุ่มของ Cell หรือเรียกมันว่า Block อาย่างที่เห็นเป็นช่องสีน้ำเงินในภาพ 2.3 สุดท้ายเราจะได้ Histogram ของ Gradient Direction จากทุก ๆ Cell ของภาพซึ่งเป็นตัวแทนของรูปร่างวัตถุแต่ละส่วน ด้วย Histogram ที่ได้มานะก็จะนำไปเข้ากระบวนการ Vectorization เพื่อให้สามารถใช้ในงานอื่น ๆ ได้ต่อไป

อย่างไรก็ได้การที่จะได้ HOG ของวัตถุที่เราต้องการตรวจสอบจำเป็นต้องใช้ภาพของวัตถุนั้น ๆ ใน การคำนวณ แต่ในสถานการณ์จริงภาพของวัตถุอาจอยู่ในภาพถ่ายขนาดใหญ่ที่ประกอบไปด้วยหลายวัตถุ เราจึงต้องตัดภาพของวัตถุเพื่อใช้คำนวณกับ HOG เราเรียกภาพส่วนย่อยที่ถูกตัดออกมานั้นว่า Patch ดังนั้นเราจำเป็นต้องใช้ Patch ที่มีขนาดใกล้เคียงกับวัตถุนั้น ๆ เป็นเหตุให้หากภาพมีขนาดใหญ่ แต่วัตถุที่ต้องการตรวจสอบมีขนาดเล็ก จำนวน Patch ก็จะมากขึ้น



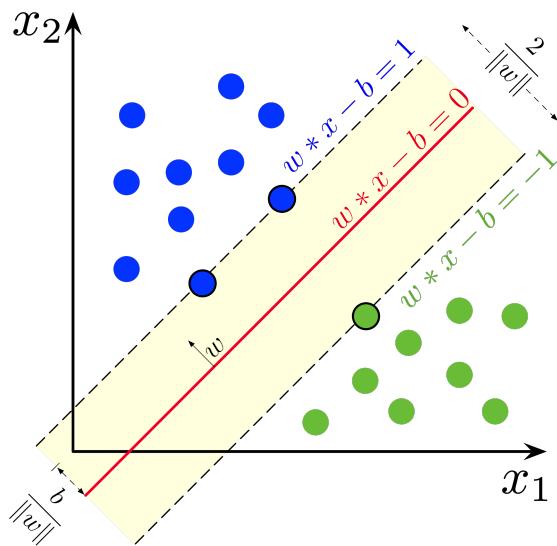
รูปที่ 2.3: Cell และ Block ในการทำงานของ Histogram of Oriented Gradients

ในกรณีของมังงะ ตัวอักษรมักมีขนาดเล็ก ($20\text{px} - 40\text{px}$ โดยส่วนใหญ่ขึ้นอยู่จาก Dataset ของเรา) เมื่อเทียบกับขนาดภาพมังงะ (1170px ขึ้นอยู่จาก dataset ของเรา) ดังนั้นจำนวนของ Patch ที่ต้องสร้างและคำนวนด้วย HOG จึงมีมากตามลำดับ แต่การคำนวน Feature โดย SVM อย่างมาก ซึ่งเป็นสาเหตุที่ทำให้กระบวนการเรียนรู้ช้าลง จึงใช้ SWT สำหรับกำหนดพื้นที่เพื่อใช้สร้าง Patch แทนการสร้าง Patch จากทุกส่วนของภาพ

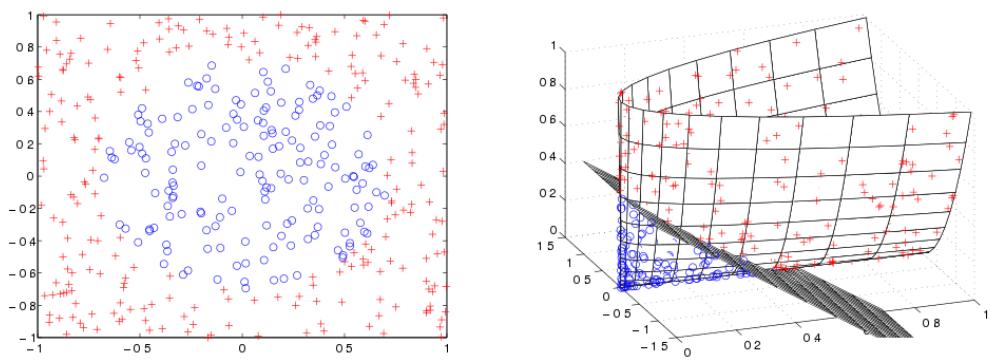
2.3 Support Vector Machine

Support Vector Machine (SVM) [20] นั้นเป็นเทคนิค Pattern Recognition แบบ Supervised Learning ซึ่งถูกใช้ทั้งในงานเพื่อ Classification และ Regression ซึ่งภายในงานนี้ได้ใช้งานเพื่อ Classification โดยทำงานด้วยการสร้าง Hyper-plane ที่เหมาะสมที่สุด (Optimal) เพื่อจำแนกแยกข้อมูลสองกลุ่มอย่างที่แสดงในภาพ 2.4 ซึ่งแตกต่างจาก Logistic regression ที่ไม่สามารถสร้างเส้นแบ่งข้อมูลระหว่างกลุ่มได้เหมาะสมที่สุดเหมือน SVM

เพื่อที่จะแยกข้อมูลทั้งสองกลุ่มด้วย Optimal Hyper-plane นั้น $w \times x - b = 0$ จะทำหน้าที่แบ่งข้อมูลสองกลุ่มออกจากกันโดยมี Support Vector ทำหน้าที่เป็นกันชนระหว่างข้อมูลที่ใกล้กันที่สุด ระหว่างกลุ่มข้อมูลทั้งสอง ซึ่ง SVM นั้นจะสร้างพื้นที่การตัดสินใจขึ้นมา หรือก็คือพื้นระหว่าง $w \times x - b = 1$ และ $w \times x - b = -1$ โดยจะปรับให้ระยะห่างระหว่างทั้งสองนั้นกว้างที่สุดเท่าที่ทำได้ โดยระยะห่างนั้นมีค่าเท่ากับ $\frac{2}{\|w\|}$ โดยใช้การ Minimize $\|w\|$ อย่างไรก็ได้ ในการแบ่งข้อมูลแบบ Non-linear นั้นสามารถใช้ Kernel เข้ามาช่วยในการเปลี่ยนมิติของข้อมูลเพื่อให้สามารถแบ่งแยกข้อมูลทั้งสองกลุ่มได้ด้วย Linear Hyper-plan ตามที่แสดงในภาพ 2.5



รูปที่ 2.4: การแบ่งแยกกลุ่มข้อมูลด้วย Hyper-plane ของ SVM



รูปที่ 2.5: ตัวอย่างแสดงคุณสมบัติการเปลี่ยนมิติของข้อมูลด้วย Kernel

บทที่ 3

วิธีการทดลอง

ในบทนี้เรากล่าวถึงวิธีการใหม่ของเราว่าที่ได้ปรับปรุงและพัฒนาขึ้นมาโดยใช้ SWT เป็นต้น เพื่อทำให้มั่นสามารถใช้งานร่วมกับภาพมังงะได้อย่างมีประสิทธิภาพ และดำเนินการทดลองเพื่อวัดประสิทธิภาพของวิธีการใหม่ของเราว่าสามารถทำงานได้ดีขึ้นหรือไม่อย่างไรเมื่อเปรียบเทียบกับวิธีการต้นฉบับ [12] นอกจากนี้เรายังนำวิธีการอื่น ๆ มาร่วมเปรียบเทียบเพิ่มเติม

3.1 วิธีการใหม่ที่ถูกปรับปรุงและพัฒนาเพิ่มเติม

สำหรับวิธีการอย่างที่ได้กล่าวไว้ในบทที่ 1 วิธีการใหม่ของเราได้ใช้ประโยชน์จาก SWT ร่วมกับความสามารถของ SVM โดยใช้ HOG เป็น Feature อย่างไรก็ตามที่ได้กล่าวไว้ในบทที่ 1 จุดประสงค์หลักของ SWT นั้นที่ถูกนำเสนอในงานวิจัยก่อนหน้าถูกออกแบบเพื่อการตรวจหาข้อความบนภาพถ่าย ด้วยเหตุผลนี้ทำให้การทำงานร่วมกับภาพมังงะไม่สามารถทำงานได้อย่างที่ควร ก่อให้เกิด False Positive จำนวนมาก ตามที่แสดงให้เห็นในภาพ 3.1 สาเหตุหลักคือความแตกต่างเชิงเอกลักษณ์ของวัตถุในภาพจริงและภาพวาดมังงะ นอกจากนี้องค์ประกอบในภาพวาดของมังงะนั้นยังมีความคล้ายคลึงกับลักษณะของตัวอักษรในภาพมากเกินไป เช่น เส้นของตัวหน้า, เส้นผ่านศูนย์กลาง, และรายละเอียดบนพื้นหลัง อย่างที่แสดงในภาพ 3.1(ก) และภาพ 3.1(ข) ด้วยเหตุนี้วิธีการของเรางานได้ปรับปรุงขึ้นตอนบางส่วนของการค้นหาตัวอักษรที่คล้ายคลึงกัน เช่น การจับกุญแจอักษร, และเพิ่มขั้นตอนใหม่บางส่วนเพื่อให้สามารถใช้งานกับมังงะได้อย่างมีประสิทธิภาพมากขึ้น

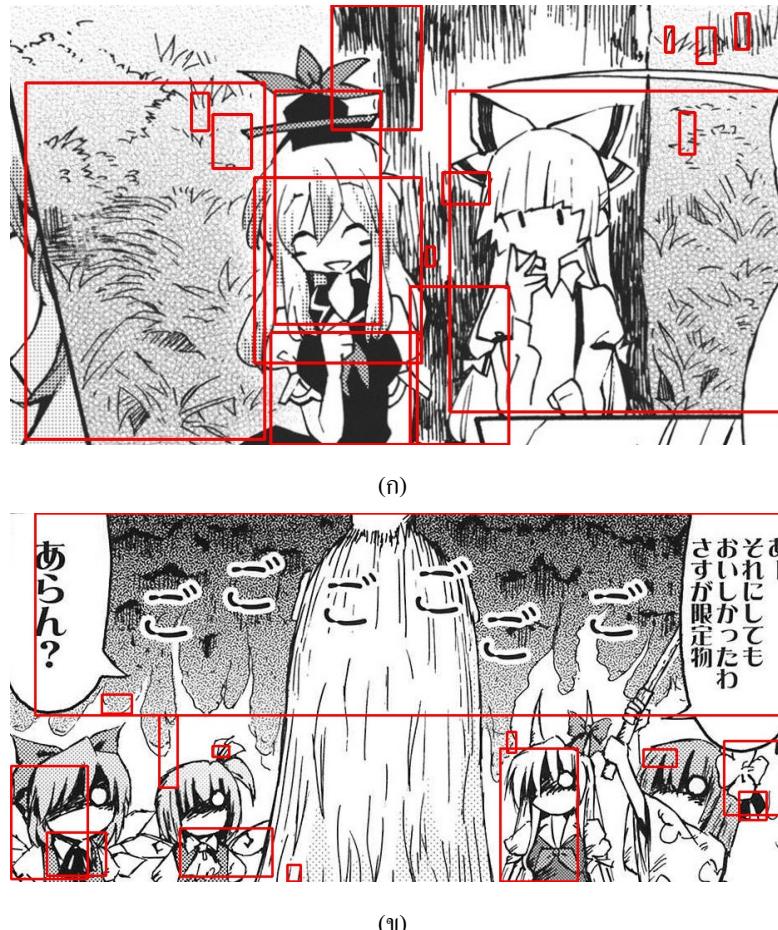
ความแตกต่างของวิธีต้นฉบับและวิธีการใหม่ของเรานั้นถูกแสดงให้เห็นในภาพ 3.2 อย่างที่เห็นในภาพ 3.2(ข) เราได้เพิ่มขั้นตอนการคัดแยกตัวอักษรใหม่ขึ้นมาก่อนการจับกุญแจตัวอักษรเข้าเป็นประโยชน์ ขั้นตอนการคัดแยกนี้ใช้ความสามารถของ SVM classification เพื่อช่วยลด False positive ของผลลัพธ์ การค้นหาตัวอักษรที่คล้ายตัวอักษรจากขั้นตอนก่อนหน้า

3.1.1 The Stroke Width Transform

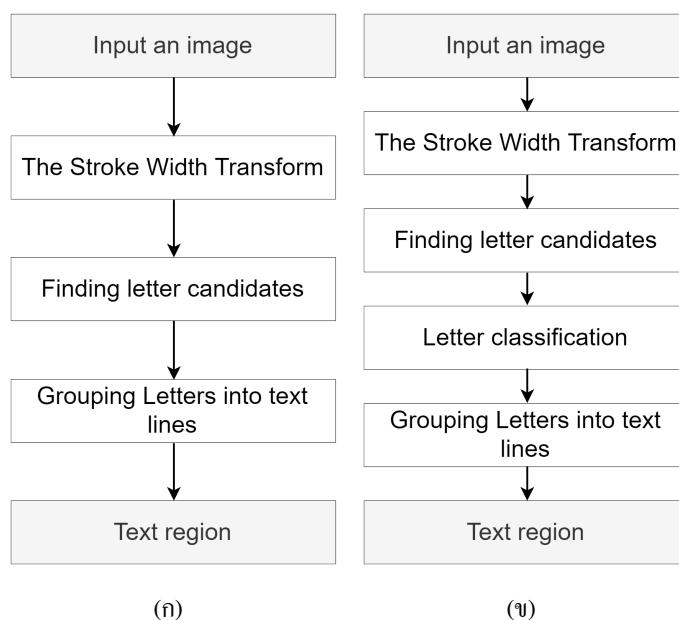
ในขั้นตอนนี้ใช้วิธีการเดียวกับงานวิจัย [12] ที่กล่าวไว้ในบทที่ 2 โดยเรานำ SWT มาดำเนินการบนภาพมังงะเพื่อให้อ่ายรูปแบบตัวดำเนินการ SWT ซึ่งข้อมูล Output ของ SWT นั้นเป็นแมทริกซ์มีขนาดเท่ากับภาพ Input ซึ่งจะใช้ในขั้นตอนต่อไป

3.1.2 ค้นหาตัวอักษรที่ใกล้เคียงกัน

ในมังงะนั้น ข้อความหรืออักษรทั้งหลายมีขนาดที่หลากหลายและแตกต่างไปจากภาพถ่าย เราจึงต้องนำกฎเกณฑ์ที่ใช้ในการคัดกรองตัวอักษรบนภาพถ่ายมาดัดแปลงให้เหมาะสมกับสภาพลักษณะของอักษรในมังงะ โดยกฎดังกล่าวถูกดังแปลงให้อ่ายรูปแบบสมการ 3.1



รูปที่ 3.1: ตัวอย่างผลลัพธ์จากการตรวจหาข้อความบนภาพมังงะด้วยวิธีด้านบน [12] ซึ่งแสดงให้เห็น False positive จำนวนมาก (ก) นักวาด: Shinoasa (ข) นักวาด: Kousei (Public Planet)



รูปที่ 3.2: แผนผังการทำงานของ (ก) วิธีการดั้งเดิม [12] และ (ข) วิธีการใหม่ของเรา

$$f(d, h, w, \tilde{s}) = \begin{cases} 1, & \text{if } 1 < \frac{d}{\tilde{s}} < 15 \text{ and } \tilde{s} \leq 80 \text{ and} \\ & 5 < h, w < 50 \\ 0, & \text{otherwise,} \end{cases} \quad (3.1)$$

โดยที่ตัวแปรที่เพิ่มเข้ามาคือ w คือ ความกว้างของวัตถุนั้น ๆ

เมื่อเราจำจัดวัตถุที่ไม่มีลักษณะแตกต่างจากอักษรอย่างสิ้นเชิงออกไปแล้ว เราจะได้กลุ่มของวัตถุที่มีลักษณะใกล้เคียงอักษร อย่างที่แสดงในภาพด้านล่าง ขึ้นตอนนี้ของวิธีการตัดตอนบันทึกไม่สามารถรวมรวมวัตถุที่คล้ายคลึงอักษร ได้ครบถ้วนเพียงพอตามที่แสดงให้เห็นในภาพ 3.3 แต่กลุ่มที่ถูกปรับปรุงนั้นสามารถรวมรวมวัตถุที่คล้ายคลึงอักษรได้ครอบคลุมมากขึ้น อย่างไรก็ได้ถูกใหม่นั้นสร้าง False Positive ที่มากขึ้นตาม ซึ่งปัญหาดังกล่าวจะถูกแก้ไขด้วยขั้นตอนคัดแยกอักษรด้วย SVM ต่อไป

3.1.3 คัดแยกอักษรด้วย SVM

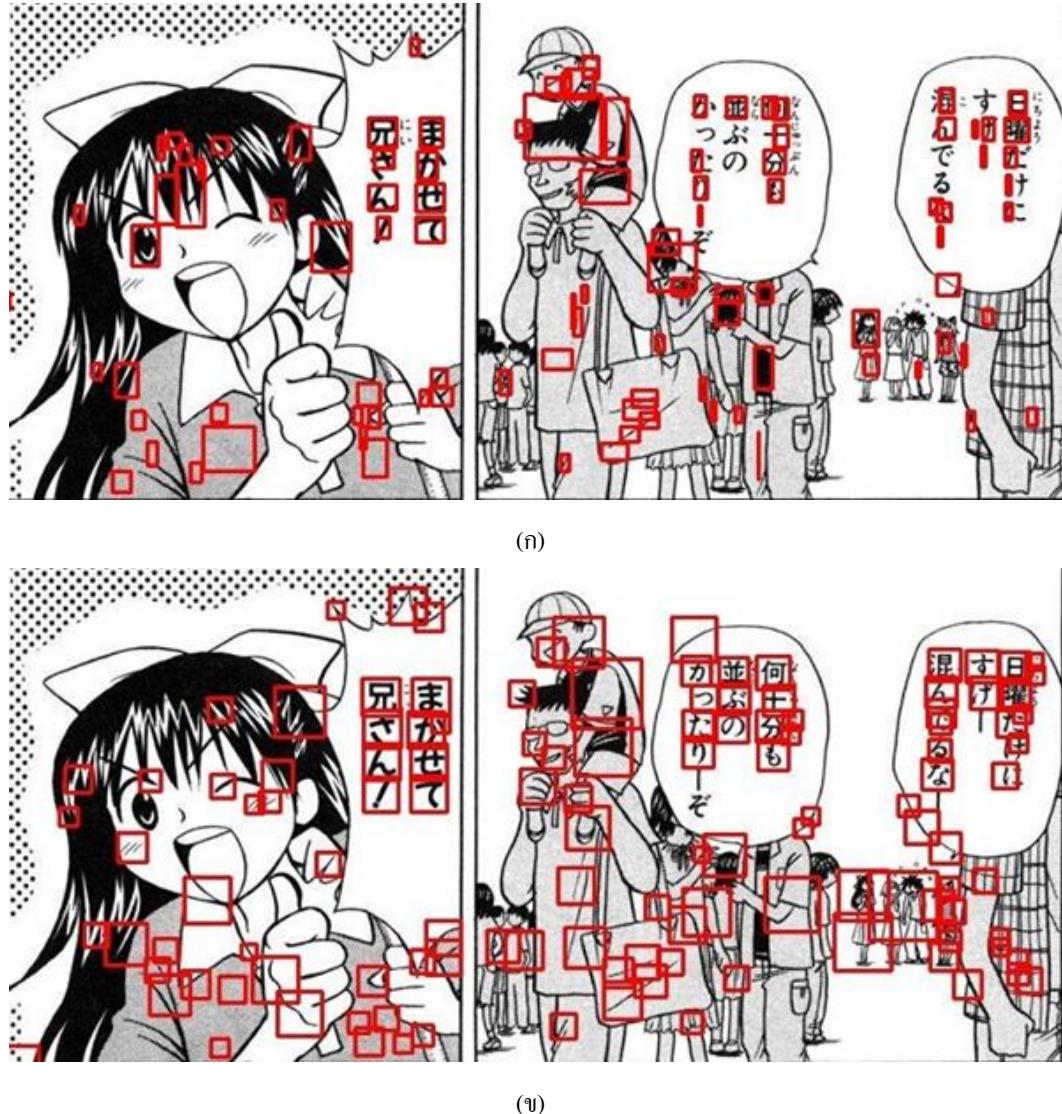
ในขั้นตอนนี้เราสร้างภาพขนาดเล็ก (Patch) ซึ่งใช้ตำแหน่งและขนาดของวัตถุที่คล้ายคลึงอักษร จากขั้นตอนก่อนหน้า โดยภาพขนาดเล็กเหล่านี้จะถูกคัดแยกเป็นกลุ่มที่เป็นอักษรและกลุ่มที่ไม่ใช่ซึ่งจะช่วยลด False Positive ให้ต่ำลง

เราได้นำ SVM มาดำเนินการ ในขั้นตอนนี้ SVM คือ เทคนิค Supervised Learning แบบหนึ่งซึ่งมักถูกใช้ในงานคียนคัดแยก (Classification) และ สมการลดด้อยต่อเนื่อง (Regression) [20] สำหรับชุดข้อมูลสำหรับเทรนโมเดลของ SVM ในขั้นตอนนี้นั้นสร้างจากภาพขนาดเล็ก หรือ Patch โดยแบ่งออกเป็นภาพที่เป็นอักษร (Positive) และ ภาพที่ไม่ใช้อักษร (Negative) อย่างที่แสดงในภาพ 3.4 ภาพขนาดเล็กเหล่านี้สร้างจาก Manga109 ซึ่งจะถูกตัดลักษณะเด่นหรือ Feature ด้วย HOG [14] ซึ่งเป็นเทคนิคการสกัดข้อมูลเชิงรูปของวัตถุในภาพด้วยการพิ่งพากรกระจายตัวของทิศทาง โโนนสี โดย Feature ที่ถูกสกัดมานั้นมีเป็นข้อมูลรูปแบบเวกเตอร์ 2,916-dimension

เช่นเดียวกับข้อมูลสำหรับเทรน ภาพขนาดเล็กของวัตถุที่คล้ายคลึงอักษรจากขั้นตอนก่อนหน้าจะถูก HOG สกัด Feature ออกมานแล้วจึงนำไปให้ SVM ดำเนินการคัดแยกภาพที่เป็นอักษรและไม่ใช้อักษรจากกัน หลังจากเสร็จสิ้นกระบวนการ ส่วนภาพที่เป็นอักษรจะถูกนำมาใช้ในการจัดกลุ่มอักษรในขั้นตอนต่อไป

3.1.4 จัดกลุ่มอักษรเป็นข้อความ

ในวิธีการตัดตอนบันทึก [12] ขั้นตอนจัดกลุ่มวัตถุที่คล้ายคลึงอักษรเข้าด้วยกัน ได้ใช้หลักการเปรียบเทียบความคล้ายคลึงของตัวอักษร ประกอบไปด้วย ของความสูง ขนาดเส้น ทิศทาง และ ระยะห่าง เพื่อกำจัดข้อมูลรบกวนอื่น ๆ เช่น วัตถุที่ไม่ใช้อักษรที่กระจายอยู่ในภาพ โดยใช้สมมติฐานว่าประโยค หรือข้อความมักเกิดจากการรวมตัวกันของอักษรมากกว่าหนึ่งตัวและจัดเรียงอยู่ในทิศทางเดียวกันกับตัวอักษรอื่น ๆ ที่ขนาดใกล้เคียงกันตามที่ได้กล่าวไปใน 2 อย่างไรก็ดี วิธีการของเรา ได้กำจัดข้อมูลรบกวนเหล่านี้ออกไปแล้วในขั้นตอนการคัดแยกอักษรด้วย SVM ดังนั้นเราจึงใช้เพียงระยะห่างระหว่างอักษรเป็นปัจจัยในการจับกลุ่มอักษร

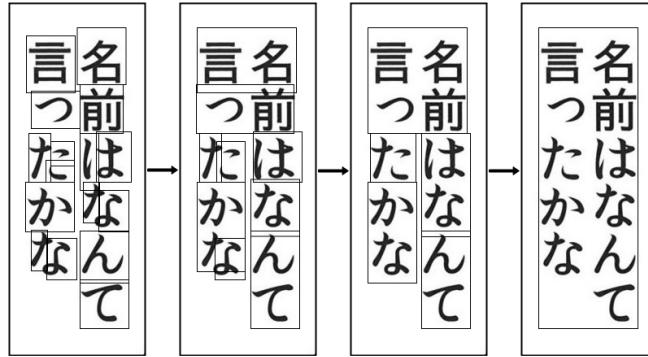


รูปที่ 3.3: ตัวอย่างแสดงการเปรียบเทียบผลลัพธ์ของขอบเขตข้อความที่ตรวจพบระหว่างวิธีการใหม่ (a) และวิธีการต้นฉบับ (b) ข้อมูลภาพถูกนำมาจากเรื่อง Arisa ©Yagami Ken



รูปที่ 3.4: ตัวอย่างของ Patch: (ก) ภาพ positive patches and (ψ) ภาพ negative patches

วิธีการจัดกลุ่มอักษรของเรานั้นจะใช้อักษรที่ถูกคัดแยกแล้วมาจัดกลุ่มเป็นประโภคโดยการจัดกลุ่มแต่ละอักษรที่อยู่ห่างกันไม่เกิน 1.5 เท่าของตัวอักษรที่แคบที่สุดของคู่อักษรที่ใช้เปรียบเทียบ หากตัวอักษรใดที่ห่างจากกันเกินกว่าค่าที่กำหนดจะถือว่าเป็นอักษรของคนละประโภคซึ่งจะไม่ถูกจับกลุ่มเข้ามา โดยตัวอย่างในภาพ 3.5 แสดงถึงตัวอย่างขั้นตอนการจับกลุ่มด้วยระยะห่าง นอกเหนือจากนี้แต่ละ



รูปที่ 3.5: ตัวอย่างแสดงการจับกลุ่มของตัวอักษร

ประโยชน์ที่ถูกจับกลุ่มด้วยระยะห่างแล้วจะถูกพิจารณาขนาดคำข่ายเข่นกัน โดยแต่ละข้อความต้องมีพื้นที่ (ความกว้าง \times ความสูง) เกินกว่า 2,550px อ้างอิงจากการทดลองกับชุดข้อมูลของเรา สุดท้ายเราจะได้กลุ่มของอักษรหรือประโยชน์ข้อความจากภาพในมังงะออกมา

3.2 ชุดข้อมูลสำหรับการเทรนโมเดล SVM

เราได้นำภาพจากชุดข้อมูลภาพมังงะขนาดใหญ่ Manga109 [8] ประกอบไปด้วยภาพมังงะพร้อมข้อมูลประกอบ (Annotation) ของมังงะ 109 เรื่อง จัดทำโดยห้องทดลอง Aizawa Yamasaki แห่งมหาวิทยาลัยโตเกียว มังงะทั้งหมดในชุดข้อมูลนี้ถูกวัดโดยนักวัดมังงะมืออาชีพชาวญี่ปุ่นและถูกจัดจำแนยในช่วงปี 1970 ถึงปี 2010 แต่ละหน้ามังงะถูกระบุตำแหน่งของข้อความในภาพซึ่งหมายความว่าสำหรับการใช้เทรนโมเดลและทดสอบวิธีการของเรานะ

3.3 การทดลอง

เราดำเนินการทดลองในรูปแบบที่คล้ายคลึงกับงานวิจัยของAramaki et al. [11] ซึ่งจะช่วยให้เราสามารถเปรียบเทียบผลการทดลองประสิทธิภาพวิธีการของเรากับงานวิจัยอื่น ๆ ก่อนหน้าได้ เราได้เลือกภาพมังงะด้วยวิธีการสุ่มเลือก 100 หน้าสำหรับเทรน และอีก 100 หน้าสำหรับทดสอบประสิทธิภาพโดยภาพมังงะทั้งหมดนี้ถูกสุ่มเลือกจากมังงะ 6 เรื่อง ได้แก่ *Aosugiru Haru, Arisa 2, Bakuretsu Kung Fu Girl, Dollgun, Love Hina, and Uchiha Akatsuki EvaLady*

เนื่องจากวิธีการของเรานี้ SVM ซึ่งต้องสร้างโมเดลสำหรับใช้งานคัดแยกภาพระหว่างภาพขนาดเล็ก (Patch) ระหว่างกลุ่มที่เป็นอักษรและไม่ใช้อักษรตามที่แสดงไปในภาพ 3.4 เราจึงได้สร้างชุดข้อมูลประกอบด้วยภาพอักษร 5,201 ภาพ และภาพขนาดเล็กอื่น ๆ ที่ไม่ใช้อักษรอีก 5,201 ภาพ กล่าวคือแบ่งเป็นข้อมูล Positive และ Negative ส่วนละ 50% เท่า ๆ กัน โดยภาพเหล่านี้ได้รับจากขั้นตอนการหาวัตถุที่คล้ายคลึงกับอักษรโดยใช้ภาพ 100 ภาพสำหรับเทรนเป็นข้อมูลนำเข้า

สำหรับ SVM เราใช้ Radial Basis Function Kernel โดย Hyperparameter ที่ร่วมใช้งานประกอบไปด้วย C และ γ เราใช้ Grid Search บนเครื่องคอมพิวเตอร์ *Google Cloud Compute Engine n1-highcpu-8* ในการทั่นหาค่า Hyperparameter ที่ดีที่สุด ในช่วง 2^{-10} ถึง 2^{10} โดยให้ค่า C และ γ ที่ดีที่สุดที่ 2^5 และ

$2^{-6.75}$ ตามลำดับ โมเดลที่ถูกปรับปรุงให้เหมาะสม (Optimized) และนี้จะถูกนำไปใช้ในขั้นตอนคัดแยกอักษร

การประเมินวิธีการตรวจหาข้อความที่เราปรับปรุงขึ้นใหม่นั้น เราได้ใช้รูปแบบการประเมินเดียวกับที่ใช้ใน ICDAR 2013 Robust Reading Competition [21] โดยถ้าอัตราส่วนระหว่างพื้นที่ Overlapped ต่อ พื้นที่ Ground-truth นั้นมากกว่าค่า t_p และอัตราส่วนระหว่างพื้นที่ Overlapped ต่อ พื้นที่ Detected region มากกว่า t_r ให้ถือว่า พื้นที่ระบุขอบเขตอักษรที่ถูกทำนายไว้นั้นถูกต้อง โดย t_p และ t_r นั้นมีค่าเท่ากับ 0.5 โดยอ้างอิงค่าตามงานวิจัยของ Aramaki et al. [11] สำหรับ Precision และ Recall เราได้คำนวณตามสมการดังต่อไปนี้ 3.2, 3.3 ตามลำดับ

$$P = \frac{\text{#Correctly Detected Rectangles}}{\text{#Detected Rectangles}} \quad (3.2)$$

$$R = \frac{\text{#Correctly Detected Rectangles}}{\text{#Rectangles of the Ground-truth}} \quad (3.3)$$

สำหรับ F-Measure เราคำนวนด้วยสมการดังนี้ 3.4

$$F = 2 \cdot \frac{P \cdot R}{P + R} \quad (3.4)$$

เราได้เปรียบเทียบผลการทดลองของวิธีการใหม่ของเราร่วมกับวิธีการต้นฉบับ [20] ซึ่งถูกใช้กับภาพถ่าย นอกจากรูปที่เปรียบเทียบกับงานวิจัยก่อนหน้านี้แล้ว ด้วย ดังนี้ Basic Grouping+ImageNet Classification model (BG+ImN) [11], Basic Grouping+Illustration2Vec model (BG+I2V) [11], Scene Text Detection (STD) [22], Speech Balloon Detection (SBD) [9], and Text Line Detection (TLD) [23]

วิธีข้างต้นที่เรากล่าวถึงมีเทคนิคในการตรวจหาข้อความที่ต่างกัน เช่น การยึดหลักสมมติฐานพื้นฐาน (ทิศทางของข้อความ, รูปแบบการจัดวาง, ลักษณะของกล่องคำพูด) และ Convolutional neural network สำหรับ BG+ImN, BG+I2V, STD, SBD, และ TLD นั้นเราได้นำผลลัพธ์การทดลองจากงานวิจัยของ Aramaki et al. [11] มาใช้ในการเปรียบของเรารอยalty ซึ่งสามารถทำได้เนื่องจากเราได้ดำเนินการทดลองในสภาพแวดล้อมเดียวกับงานวิจัยดังกล่าว โดยผลลัพธ์การเปรียบเทียบและตัวอย่างขอบเขตข้อความที่วิธีการของเรานั้นสามารถตรวจพบถูกแสดงให้เห็นที่ 4

บทที่ 4

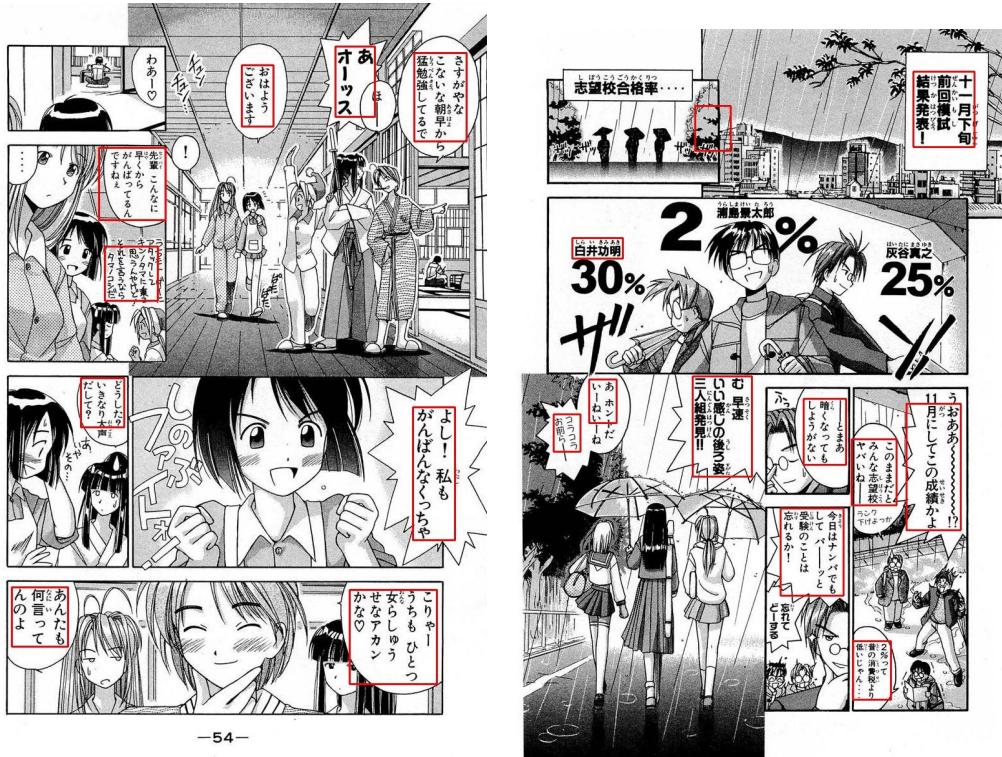
ผลการทดลอง

ผลลัพธ์การทดสอบประสิทธิภาพในวิธีการใหม่ของเราและการเปรียบเทียบกับงานวิจัยอื่น ๆ แสดงในตารางที่ 4.1 จากตารางดังกล่าว วิธีการของเราได้รับ F-measure สูงที่สุด ที่ 0.506 ซึ่งแสดงให้เห็นชัดเจนว่าวิธีการของเรามีประสิทธิภาพดีกว่าวิธีการต้นฉบับ [12] ยิ่งไปกว่านั้นวิธีการของเรา yang ได้รับ F-measure ที่กว่าวิธีการ BG+ImN [11] และ BG+I2V [11] ซึ่งทั้งสองวิธีนี้ล้วนใช้เทคนิค Deep Learning เป็นส่วนหนึ่งในการตรวจหาข้อความในภาพอย่างไรก็ได้ ค่า Precision และ Recall สูงสุดของการทดลองนี้อยู่ที่ 0.715 และ 0.481 เป็นของ BG+I2V และ BG+ImN ตามลำดับ สำหรับตัวอย่างขอบเขตของข้อความที่วิธีการใหม่ของเราระบุถูกแสดงในภาพ 4.1

Method	Precision	Recall	F-measure
STD [22]	0.165	0.051	0.078
SBD [9]	0.180	0.102	0.130
TLD [23]	0.095	0.095	0.095
BG + ImN [11]	0.451	0.481	0.466
BG + I2V [11]	0.715	0.191	0.301
Baseline [12]	0.068	0.336	0.113
วิธีการของเรา	0.564	0.458	0.506

ตารางที่ 4.1: ตารางแสดงการเปรียบเทียบประสิทธิภาพของวิธีการใหม่ของเราร่วมกับวิธีการอื่น ๆ

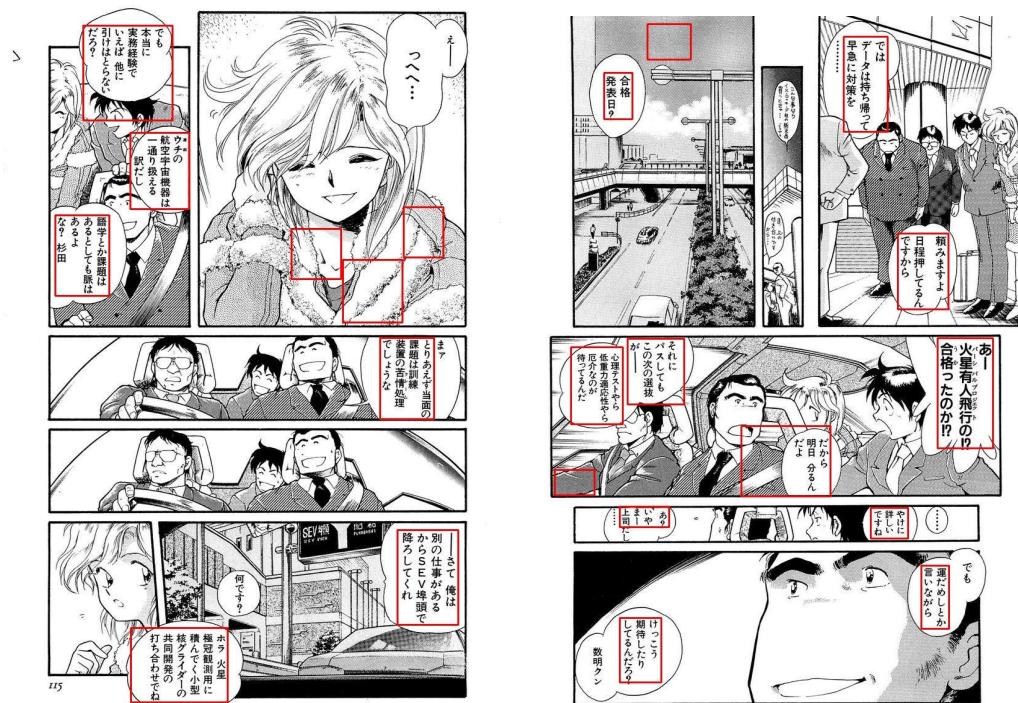
เป็นที่น่าสนใจอย่างมากที่วิธีการของเราสามารถทำงานได้ดีกว่าเทคนิค Deep Learning ทั้งสองวิธี สมมติฐานแรกคือ BG+ImN นั้นใช้ ImageNet Classification Model [24] ซึ่งถูกเทรนบนภาพถ่ายของวัตถุจริง อย่างไรก็ได้ภาพความมั่งคงของวัตถุต่าง ๆ นั้นมีความแตกต่างจากภาพวัตถุจริงอย่างชัดเจนซึ่ง ณ จุดนี้ทำให้วิธีการนี้ไม่สามารถทำงานได้เต็มประสิทธิภาพ อีกวิธีหนึ่งที่ใช้ Deep Learning คือ BG+I2V ถึงแม้ว่าวิธีการนี้จะได้รับ Precision สูงที่สุดในการทดลองของเรา แต่คะแนน Recall นั้นต่ำกว่าทั้ง BG+ImN และวิธีการของเรา วิธีการนี้ใช้โมเดล Illustration2Vec [25] เป็นโมเดลสำหรับคัดแยกข้อความจากวัตถุอื่น ๆ ส่วนของโมเดลนี้ถูกเทรนบนภาพวาด Anime (ภาพการ์ตูนแบบญี่ปุ่น) และภาพมั่งจะจากหลากหลายแหล่ง ประกอบไปด้วย Danbooru และ Safebooru ซึ่งมีลักษณะงานคล้ายกับวิธีการของเรานำเชิงข้อมูล แต่โมเดลนี้ถูกออกแบบมาเพื่อนการทำนายป้ายกำกับ (Tag Prediction) และค้นหาภาพที่คล้ายคลึงกัน ดังนั้นนี่จึงอาจเป็นเหตุผลว่าทำไมโมเดลนี้จึงไม่มีประสิทธิภาพเท่าที่ควรในการทดลองนี้



-54-

(n)

(o)



(p)

(q)

รูปที่ 4.1: ตัวอย่างขอบเขตข้อความในวิธีการของเราราจพน (ก-ภ) Love Hina ©Ken Akamatsu !! และ (ก-ภ) Eva Lady ©Miyone Shi.

บทที่ 5 สรุปผล

ในการทดลองนี้ เราได้เสนอวิธีการตรวจหาข้อความบนภาพมังงะด้วยเทคนิค SWT ร่วมกับการใช้ SVM และ HOG ในการลด False Positive ที่เกิดขึ้น การทดลองของเราดำเนินการบนข้อมูลจาก Manga109 ซึ่งเป็นชุดข้อมูลภาพมังงะที่ถูกระบุ Annotation มาเรียบร้อยแล้ว วิธีการของเรานั้นสามารถทำงานได้ผล F-measure ที่ดีที่สุดในการเปรียบเทียบกับวิธี Baseline และวิธีอื่น ๆ รวมถึงวิธีการที่ใช้ Deep Learning ถึงแม้ว่างานของเรานั้นซึ่งได้ทดสอบบนการตัดสินใจป้อนสามารถทำงานได้เป็นผลดีเยี่ยมอย่างไรก็ตาม การของเรานั้นยังต้องมีการพัฒนาและค้นคว้าเพิ่มเติมเพื่อปรับปรุงประสิทธิภาพและสามารถใช้งานร่วมกับภาษาอื่น ๆ ได้

បរចាំនូវក្រម

- [1] H. Yanagisawa, T. Yamashita, and H. Watanabe, “A study on object detection method from manga images using CNN,” in *Proceedings of the International Workshop on Advanced Image Technology (IWAIT 2018)*, Chiang Mai, Thailand., Jan 2018, pp. 1–4.
- [2] X. Liu, C. Li, H. Zhu, T.-T. Wong, and X. Xu, “Text-aware balloon extraction from manga,” *The Visual Computer*, vol. 32, no. 4, pp. 501–511, Apr 2016. [Online]. Available: <https://doi.org/10.1007/s00371-015-1084-0>
- [3] X. Pang, Y. Cao, R. W. Lau, and A. B. Chan, “A robust panel extraction method for manga,” in *Proceedings of the 22nd ACM International Conference on Multimedia (MM 2014)*. New York, NY, USA: ACM, 2014, pp. 1125–1128. [Online]. Available: <http://doi.acm.org/10.1145/2647868.2654990>
- [4] Y. Aramaki, Y. Matsui, T. Yamasaki, and K. Aizawa, “Interactive segmentation for manga using lossless thinning and coarse labeling,” in *Proceedings of the Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA 2015)*, Hung Hom, Kowloon, Hong Kong, Dec 2015, pp. 293–296.
- [5] T. Ogawa, A. Otsubo, R. Narita, Y. Matsui, T. Yamasaki, and K. Aizawa, “Object detection for comics using manga109 annotations,” *CoRR*, vol. abs/1803.08670, 2018. [Online]. Available: <http://arxiv.org/abs/1803.08670>
- [6] S. Kovánen and K. Aizawa, “A layered method for determining manga text bubble reading order,” in *Proceedings of the IEEE International Conference on Image Processing (ICIP 2015)*, Quebec City, QC, Canada, Sept.
- [7] Y. Matsui, T. Shiratori, and K. Aizawa, “Drawfromdrawings: 2D drawing assistance via stroke interpolation with a sketch database,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, no. 7, pp. 1852–1862, 2017.
- [8] Y. Matsui, K. Ito, Y. Aramaki, A. Fujimoto, T. Ogawa, T. Yamasaki, and K. Aizawa, “Sketch-based manga retrieval using manga109 dataset,” *Multimedia Tools and Applications*, vol. 76, no. 20, pp. 21 811–21 838, Oct 2017. [Online]. Available: <https://doi.org/10.1007/s11042-016-4020-z>
- [9] H. Tolle and K. Arai, “Manga content extraction method for automatic mobile comic content creation,” in *Proceedings of the International Conference on Advanced Computer Science and Information Systems (ICACSIS 2013)*, Bali, Indonesia, Sept 2013, pp. 321–328.

- [10] C. Rigaud, T. Le, J. . Burie, J. Ogier, S. Ishimaru, M. Iwata, and K. Kise, “Semi-automatic text and graphics extraction of manga using eye tracking information,” in *2016 12th IAPR Workshop on Document Analysis Systems (DAS)*, Santorini, Greece, April 2016, pp. 120–125.
- [11] Y. Aramaki, Y. Matsui, T. Yamasaki, and K. Aizawa, “Text detection in manga by combining connected-component-based and region-based classifications,” in *Proceedings of the IEEE International Conference on Image Processing (ICIP 2016)*, Phoenix, AZ, USA, Sept 2016.
- [12] B. Epshtain, E. Ofek, and Y. Wexler, “Detecting text in natural scenes with stroke width transform,” in *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR 2010)*, San Francisco, CA, USA, Jun 2010, pp. 2963–2970.
- [13] J. Canny, “A computational approach to edge detection,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, no. 6, pp. 679–698, 1986.
- [14] W. T. Freeman, W. T. Freeman, M. Roth, and M. Roth, “Orientation Histograms for Hand Gesture Recognition,” in *Proceedings of the International Workshop on Automatic Face and Gesture Recognition*, 1994, pp. 296–301. [Online]. Available: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.6.618>
- [15] S. Bougarriou, F. Hamdaoui, and A. Mtibaa, “Linear SVM classifier based HOG car detection,” in *Proceedings of the 18th International Conference on Sciences and Techniques of Automatic Control and Computer Engineering (STA 2017)*, Monastir, Tunisia, Dec 2017, pp. 241–245.
- [16] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2005)*, San Diego, CA, USA, Jun 2005, pp. 886–893.
- [17] D. Wang, H. Wang, D. Zhang, J. Li, and D. Zhang, “Robust scene text recognition using sparse coding based features,” *CoRR*, vol. abs/1512.08669, 2015. [Online]. Available: <http://arxiv.org/abs/1512.08669>
- [18] S. Tian, S. Lu, B. Su, and C. L. Tan, “Scene text recognition using co-occurrence of histogram of oriented gradients,” in *2013 12th International Conference on Document Analysis and Recognition*, Washington, DC, USA, Aug 2013, pp. 912–916.
- [19] A. K. Sah, S. Bhowmik, S. Malakar, R. Sarkar, E. Kavallieratou, and N. Vasilopoulos, “Text and non-text recognition using modified hog descriptor,” in *2017 IEEE Calcutta Conference (CALCON)*, Dec 2017, pp. 64–68.
- [20] J. Suykens and J. Vandewalle, “Least squares support vector machine classifiers,” *Neural Processing Letters*, vol. 9, no. 3, pp. 293–300, Jun 1999. [Online]. Available: <https://doi.org/10.1023/A:1018628609742>

- [21] D. Karatzas, F. Shafait, S. Uchida, M. Iwamura, L. G. i. Bigorda, S. R. Mestre, J. Mas, D. F. Mota, J. A. Almazàn, and L. P. de las Heras, “Icdar 2013 robust reading competition,” in *2013 12th International Conference on Document Analysis and Recognition*, Kolkata, India, Aug 2013, pp. 1484–1493.
- [22] L. Gómez and D. Karatzas, “Multi-script text extraction from natural scenes,” in *Proceedings of the 12th International Conference on Document Analysis and Recognition (ICDAR 2013)*, Washington, DC, USA, Aug 2013, pp. 467–471.
- [23] C. Rigaud, D. Karatzas, J. Van De Weijer, J.-C. Burie, and J.-M. Ogier, “Automatic text localisation in scanned comic books,” in *Proceedings of the 9th International Conference on Computer Vision Theory and Applications*, Barcelona, Spain, Feb 2013. [Online]. Available: <https://hal.archives-ouvertes.fr/hal-00841492>
- [24] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in *Proceedings of the 25th International Conference on Neural Information Processing Systems (NIPS 2012)*, vol. 1. Lake Tahoe, Nevada, USA: Curran Associates Inc., Dec 2012, pp. 1097–1105. [Online]. Available: <http://dl.acm.org/citation.cfm?id=2999134.2999257>
- [25] M. Saito and Y. Matsui, “Illustration2Vec: A semantic vector representation of illustrations,” in *SIGGRAPH Asia 2015 Technical Briefs*. Kobe, Japan: ACM, Nov 2015, pp. 5:1–5:4. [Online]. Available: <http://doi.acm.org/10.1145/2820903.2820907>

ภาคผนวก ก

การใช้ชีวิตในประเทศญี่ปุ่น

สำหรับการฝึกงานร่วมกับมหาลัยจากไทยในญี่ปุ่น ได้ฝึกอบรมกันเป็นระยะสี่เดือนกับอีกสิบวัน โดยได้เข้ามาทำงานในห้องทดลองภาษาได้การคุ้มครองอาจารย์ประจำห้องทดลอง Masanori Sugimoto และในช่วงเวลาฝึกงานนี้ก็ยังได้รับความช่วยเหลือจากสมาชิกภายในห้องทดลองอีกท่าน Jiang Ye ซึ่งเป็นนักศึกษาปริญญาเอก ปี 2 จากประเทศจีน ได้ช่วยเหลือในการเรียนการสอนภาษาที่จำเป็นต่างๆ เช่น เอกสารการเขียนภาษาญี่ปุ่นและการเขียนภาษาจีน นอกจากนี้ยังได้รับความช่วยเหลือในเรื่องความเป็นอยู่และสิ่งของทั่วไปในการใช้ชีวิตจากกลุ่มนักศึกษาชาวไทยทั้งบริษัทและปริญญาโทซึ่งล้วนศึกษาอยู่ที่มหาลัยจากไทย

การใช้ชีวิตในต่างแดน เช่น ประเทศญี่ปุ่นนั้นมีอุปสรรคหลายอย่าง แต่สำคัญที่สุดคือเรื่องภาษาอย่างที่ทราบกันดีว่าชาวญี่ปุ่นส่วนใหญ่ไม่สามารถสนทนาภาษาอังกฤษได้ ดังนั้นผมจึงต้องเรียนรู้ภาษาญี่ปุ่นพื้นฐานเพื่อช่วยให้การดำเนินชีวิตประจำวันด้วยเช่นกัน รวมถึงสมาชิกภายในแล้วที่ส่วนใหญ่คือนักศึกษาบริษัทและเอกชนส่วนใหญ่มีปัญหาในการสื่อสารด้วยภาษาอังกฤษเช่นกัน โดยส่วนมากจะสามารถสื่อสารได้เป็นประกายง่าย ๆ สัก ๆ ไม่สามารถสนทนาเป็นบทสนทนาที่ยาวหรือเป็นการเล่าเรื่องได้อย่างไรก็ดีภายในห้องทดลองก็มีสมาชิกชาวญี่ปุ่นบางคนที่มีความสามารถภาษาอังกฤษดีมาก ซึ่งดีกว่าตัวผมที่มาฝึกงานที่นี่ อย่างเช่นนักศึกษาบริษัทชาวญี่ปุ่นท่านหนึ่งที่เคยได้ฝึกงานในนิวเจลแลนด์และฟิลิปปินส์มาก่อนหน้านี้ทำให้สามารถสื่อสารภาษาอังกฤษได้เป็นอย่างดี แต่ถึงจะมีปัญหาในการสื่อสารระหว่างกัน เมื่อเกิดปัญหาใด ๆ และต้องการความช่วยเหลือ หากได้ขอความช่วยเหลือจากเพื่อนชาวญี่ปุ่น พากษาที่พร้อมเข้ามาช่วยเหลืออย่างเต็มที่โดยไม่รีรอ

นอกจากตัวผมที่เป็นนักศึกษาฝึกงานจากไทย ภายในห้องทดลองที่ผมเข้ามาอยู่ก็มีนักศึกษาต่างชาติอีกสองคน คนแรกอย่างที่กล่าวไปในตอนต้น คือ นักศึกษาบริษัทจากบริษัท สำหรับชาวจีนนั้นใช้ภาษาญี่ปุ่นเป็นภาษาหลักและสามารถสื่อสารกับคนไทยได้อย่างราบรื่น และบังพูดภาษาอังกฤษได้ดีอีกด้วย แต่สำหรับชาวบริษัทนี้ โดยส่วนตัวผมมีปัญหาในการฟังสำเนียงภาษาอังกฤษที่มาจากการผู้ใช้ภาษาเดนิสุโรปที่ไม่ใช่ประเทศอังกฤษ ซึ่งประเทศบริษัทนี้ใช้ภาษาโปรตุเกสเป็นหลัก ทำให้สำหรับผมการสื่อสารภาษาอังกฤษกับชาวบริษัทในແລ້ວนั้นมีปัญหาเสมอเนื่องจากศัพท์บางคำผมฟังแล้วไม่สามารถเข้าใจได้ เพราะสำเนียง แต่ก็สามารถสื่อสารกันได้ลำเรื่จแม้ว่าจะใช้เวลานานมากกว่าเมื่อเทียบกับสื่อสารกับชาวจีน นอกจากสองคนนี้ยังมีชาวมองโกเลียที่เข้ามาฝึกงานระยะสั้นหนึ่งสัปดาห์อีกด้วย

ก.1 ที่อยู่อาศัย

หอพักที่อาศัยอยู่จัดทำให้โดยทางมหาลัยจากไทย หอพักซึ่งอยู่ที่ International House Kita 8 East เป็นหอช雅ล์วันจัดสำหรับนักศึกษาชาวต่างชาติเท่านั้น โดยภายในห้องจะมีเพียงห้องเดียว โต๊ะ โคมไฟ ไฟเพคาน ตู้เย็น ชีฟเตอร์ และ ถังขยะ ให้อย่างที่แสดงในภาพ ก.1(ก) และ ก.1(ข) สำหรับห้องน้ำ และห้องซักอบตาก็ต้องใช้ของส่วนกลางเท่านั้น โดยจะจัดแยกในแต่ละชั้น ให้ผู้อยู่อาศัยใช้ร่วมกัน

สำหรับห้องครัวและห้องรับประทานอาหารมีจัดเตรียมให้ที่ชั้นแรกของหอพักโดยต้องใช้รวมทั้งหอ เดอะเก๊สและอ่างล้างจานถูกติดตั้งไว้เรียบร้อยสามารถใช้งานได้ทันทีแต่ต้องซื้อของใช้ส่วนตัวรวมถึงเครื่องครัวมาเอง ในกรณีของผมถือว่าโชคดีที่ซื้อแค่ชุดจานชามช้อนช้อมและแก้ว สำหรับอุปกรณ์ทำอาหารได้รับมาจากคนเงินที่กำลังจะกลับประเทศไทยอีก ก.1(ค) และ ก.1(จ)



(ก) ห้องนอน



(ข) ห้องนอน



(ค) ห้องอาหาร



(ง) ห้องครัว

รูปที่ ก.1: ภาพหอพัก International House Kita 8 East

ภาคผนวก ข

กิจกรรมระหว่างฝึกงาน

นอกจากการมาทำโครงการแล้วนั้น ระหว่างสี่เดือนนี้ผมยังได้ร่วมกิจกรรมด้านวิชาการต่าง ๆ ดัง เช่น การประชุมแล่เปรยสปดาห์ และ การร่วมนำเสนอผลงานระหว่างห้องทดลอง เป็นต้น โดยรายละเอียดจะกล่าวต่อจากนี้

ข.1 Mirai Symposium

กิจกรรมนี้จัดขึ้นที่เรียนกังหรือเรื่อสอร์ทแบบญี่ปุ่น ซึ่งมีอ่อนเช็นหรือบ่อน้ำแร่ร้อนให้ได้แช่อิ่อกด้วยค่าใช้จ่ายทั้งหมดทางห้องทดลองออกให้ทั้งหมด ภายในงานจะมีกิจกรรมหลักคือการให้นักศึกษาทุกคนของห้องทดลองทุกห้องที่มาร่วมงานนำผลงานตนเองมาเสนอตัวยไปสู่เตอร์ในห้องประชุม เวียนไปเป็นช่วงเวลาคันละ 30 นาที หากสนใจงานของครูก็เดินไปสอบถามได้ อย่างที่เห็นในภาพ ข.1(ก) การนำเสนอไม่ได้จำกัดว่างานต้องเป็นผลงานที่เสร็จล้วนแล้ว อาจเป็นความก้าวหน้าของงานที่ทำอยู่ได้ เช่นเดียวกัน สำหรับครั้งนี้มีห้องทดลองร่วมงานสามห้องทดลอง สำหรับนักศึกษาต่างชาติอย่างผมและชาวบริษัท ไม่ค่อยมีคนเข้ามาสอบถามมากนัก แต่จากการดูแลของอาจารย์และนักศึกษาญี่ปุ่น ส่วนใหญ่พอดังกฤษไม่ได้สำหรับผมก็นำผลงานที่ทำในโครงการนี้ไปนำเสนอตัวเองเช่นเดียวกัน

เมื่อช่วงนำเสนอเสร็จสิ้น ทุกคนจะได้พักผ่อนตามอัธยาศัย โดยส่วนใหญ่จะไปเปลปอน้ำพุร้อนกัน สำหรับบ่อน้ำพุร้อนก็มีทั้งภายนอกอาคารและภายใน จากนั้นเป็นมือเย็นตามฉบับอาหารชุดญี่ปุ่น ซึ่งเน้นความหลากหลายเป็นหลักตามภาพ ข.1(ข)

สำหรับวันที่สองก็จะมีกิจกรรมสร้างสรรค์ เป็นกิจกรรมที่ให้แยกกลุ่มกันและมีโจทย์ปัญหาให้แก่ภายในเวลาที่กำหนด ตัวอย่างคำาถาม เช่น มีเครื่องอยู่ร้อยเครื่องมีเที่ยวที่หงายหน้าก้อยและหัวอย่างละครึ่ง ต้องการแบ่งกลุ่มเครื่องอยู่สองกลุ่ม โดยที่แต่ละกลุ่มนี้จำนวนหน้าเครื่องและก้อยเท่า ๆ กัน โดยที่ผู้แบ่งมองไม่เห็นเครื่องอยุจะทำได้อย่างไร เป็นต้น เมื่อเสร็จกิจกรรมสร้างสรรค์ก็ถือเป็นการจบกิจกรรม Mirai Symposium แต่เพียงเท่านี้ และเดินทางกลับมายังไทย

ข.2 Lab Meeting

สำหรับห้องทดลองที่ผมมาทำงานนั้นจะมีกิจกรรม Lab Meeting ทุกวันพุธ โดยจะกำหนดตารางให้แต่ละคนที่ถูกกำหนดในวันนั้นนำ Conference Paper หรือ Journal ที่ถูกตีพิมพ์ในงานประชุมอันดับต้น ๆ ของโลกมาร่วมกันนำเสนอแบบสรุปพร้อมสไลด์ กิจกรรมนี้ไม่มีคะแนนหรือรางวัลใด ๆ แต่เป็นการแลกเปลี่ยนความรู้ที่เกี่ยวข้องกับงานของตนเองที่ทำอยู่

สำหรับผมได้มีโอกาสนำเสนอผลงานวิจัยที่ถูกตีพิมพ์ใน Journal เรื่อง *Text-aware balloon extraction from manga* โดยหลังจากนำเสนอเสร็จก็ได้รับคำถามและความคิดเห็นจากสมาชิกและอาจารย์ของห้องทดลองอย่างหลากหลาย ซึ่งถือเป็นโอกาสที่ดีในการเรียนรู้เรื่องที่แตกต่างจากการของเราและความคิดเห็นที่เปลี่ยนใหม่ต่องานของเรา



(ก) ภาพระหว่างการนำเสนอผลงานด้วยโปสเตอร์



(ข) ชุดอาหารจัดเลี้ยงมีอิ่มหลังจากแข่งขันน้ำพุร้อน

(ค) ห้องอาหารจัดเลี้ยง

รูปที่ ข.1: ภาพกิจกรรมในงาน Mirai symposium

ข.3 กิจกรรมอื่น ๆ

นอกจากกิจกรรมเชิงวิชาการแล้ว ผู้เข้าร่วมในกิจกรรมสังสรรค์อื่น ๆ เช่น งานเลี้ยงตามโอกาสต่าง ๆ โดยผู้เข้าร่วมงานเลี้ยงต้อนรับผู้เข้าร่วงเดือนตุลาคมที่ผ่านมา สาเหตุที่จัดขึ้นไปหลายเดือนเนื่องจากกำหนดการในตอนแรกนั้นคือหนึ่งเดือนหลังจากผู้เข้าร่วมเดือนกันยายนที่ญี่ปุ่น แต่ในช่วงกำหนดการเกิดเหตุการณ์แผ่นดินไหวใหญ่บุนเบิกทางใต้ทำให้ต้องเลื่อนกำหนดการออกไป โดยงานเลี้ยงนี้จัดที่ร้านเนื้อย่างเจกิสบ้าน เป็นเนื้อแกะย่างบนเตาเหล็กลักษณะคล้ายหมูกระทะในประเทศไทยแต่ต่างกันเพียงเน้นไฟที่เนื้อแกรเป็นหลักสำหรับเจกิสบ้าน โดยค่าใช้จ่ายทั้งหมดทางอาจารย์ชาวญี่ปุ่นเป็นผู้ออกให้สำหรับผู้ที่เข้าร่วมกิจกรรมแสดงในภาพ ข.2

นอกจากงานเลี้ยงต้อนรับผู้เข้าร่วมแล้วยังมีงานเลี้ยงต้อนรับนักศึกษาปีสามที่เข้าเป็นสมาชิกใหม่ในห้องทดลองนี้อีกด้วย ซึ่งงานนี้จัดรวมเป็นงานเลี้ยงสำราญพร้อม ๆ กันเนื่องจากจัดในวันที่ต้องกลับ



(ก) สมาชิกในห้องทดลองระหว่างกินเลี้ยง



(ข) โต๊ะอาหารในร้านอาหารเจงกิสخ่าน



(ค) กระทะร้อนเจงกิสخ่าน

รูปที่ ข.2: ภาพระหว่างงานเลี้ยงต้อนรับตัวผู้

ประเทศไทย งานเลี้ยงจัดในห้องทดลองและทางอาจารย์ญี่ปุ่นก็เป็นคนจัดการค่าใช้จ่ายในส่วนของพม อีกเช่นเคย งานไม่มีกิจกรรมอะไรมากมายเป็นเพียงการกินเลี้ยงเพื่อทำความรู้จักกับสมาชิกใหม่ของ ห้องทดลองและนั่งเล่นกันคุยกันตามที่แสดงในภาพ ข.3

อธิบายเพิ่มเติมสำหรับงานเลี้ยงต้อนรับนักศึกษาปีสาม สำหรับห้องทดลองของคณะวิศวกรรมศาสตร์ ที่ผมมาอยู่นี้จะมีการเปิดห้องแล็บให้นักศึกษาในคณะชั้นปีสามได้เข้าเยี่ยมชมห้องทดลองที่ตอนอง สนใจก่อนจะเลือกเข้ามาเป็นสมาชิกในห้องทดลองที่เกี่ยวข้องกับงานที่สนใจจะทำ โดยงานที่สนใจจะ ทำจะถูกนำเสนอเป็นชิ้นงานจากการศึกษาเหมือนกับ โปรเจกต์จบ ในมหาลัยของไทย โดยในช่วงนี้แต่ละห้อง แล็บก็จะมีการโฆษณาห้องทดลองของตัวเองและเปิดโอกาสให้เข้าเยี่ยมชมงานภายในห้องทดลองว่า ทำวิจัยเรื่องอะไร อย่างเช่นภาพ ข.4 ที่เป็นป้ายโฆษณาห้องแล็บด้านเลี้ยงและมีการใช้ตัวละครจากเรื่อง Kemono Friend ประกอบให้น่าสนใจและดึงดูดมากขึ้น



(ก) อาหารต่าง ๆ ในงานเลี้ยง



(ข) สมาชิกในห้องทดลองระหว่างเล่นเกมในงานเลี้ยง

รูปที่ ข.3: ภาพงานเลี้ยงอาหารตามสั่งและต้อนรับนักศึกษาปีสาม



รูปที่ ข.4: ป้ายเชิญชวนชมห้องทดลองโอดิการตัวละครจาก การ์ตูนประกอบให้น่าสนใจ