ข้อเสนอโครงการ

รหัสโครงการ 18p32c0626

การแข่งขันพัฒนาโปรแกรมคอมพิวเตอร์แห่งประเทศไทย

ชื่อโครงการ (ภาษาไทย) ข้อความเอย เจ้าอ่านว่าอย่างไร, ระบบระบุตำแหน่งและรู้จำข้อความบนภาพถ่ายทั่วไป

(ภาษาอังกฤษ) O' Text How to read you?

ประเภทโปรแกรมที่เสนอ BEST 2016 : การแข่งขันสุดยอดการหาตำแหน่งและรู้จำข้อความในภาพถ่าย (Text Localization and Recognition Contest)

หัวหน้าโครงการ

ชื่อ-นามสกุล นายธนานพ กอบชัยสวัสดิ์

วัน/เดือน/ปีเกิด 15/10/2531 **ระดับการศึกษา** ปริญญาเอก **สถานศึกษา** จุฬาลงกรณ์มหาวิทยาลัย

ที่อยู่ตามทะเบียนบ้าน 228 ซอยลาดพร้าว 64 แยก 2 ถนนลาดพร้าว แขวงวังทองหลาง เขตวังทองหลาง กรุงเทพมหานคร 10310

สถานที่ติดต่อ 203 ถนนเสือป่า แขวงป้อมปราบ เขตป้อมปราบศัตรูพ่าย กรุงเทพมหานคร 10100

โทรศัพท์ 02-223-2692 **มือถือ** 083-298-9275 **โทรสาร** - **e-mail** thananop.k@gmail.com

ลงชื่อ.....

อาจารย์ที่ปรึกษาโครงการ

ชื่อ-นามสกุล ผศ.ดร. ธนารัตน์ ชลิดาพงศ์

สังกัด/สถาบัน ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

สถานที่ติดต่อ ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

โทรศัพท์ 02-218-6993 **มือถือ** - **โทรสาร** - **e-mail** thanarat.c@chula.ac.th

คำรับรอง “โครงการนี้เป็นความคิดริเริ่มของนักพัฒนาโครงการและไม่ได้ลอกเลียนแบบมาจากผู้อื่นผู้ใด ข้าพเจ้าขอรับรองว่าจะให้คำแนะนำและ สนับสนุนให้นัก พัฒนาในความดูแลของข้าพเจ้าดำเนินการศึกษา/วิจัย/พัฒนาตามหัวข้อที่เสนอและจะทำหน้าที่ประเมินผลงานดังกล่าวให้กับโครงการฯ ด้วย”

ลงชื่อ.....

หัวหน้าสถาบัน (อธิการบดี/คณบดี/หัวหน้าภาควิชา/ผู้อำนวยการ/อาจารย์ใหญ่/หัวหน้าหมวด)

ชื่อ-นามสกุล ผศ. ณัฐวุฒิ หนูไพโรจน์

สังกัด/สถาบัน ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

สถานที่ติดต่อ ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

โทรศัพท์ 02-218-6956 **มือถือ** - **โทรสาร** - **e-mail** natawut.n@eng.chula.ac.th

คำรับรอง “ข้าพเจ้าขอรับรองว่าผู้พัฒนามีสิทธิ์ขอรับทุนสนับสนุนตามเงื่อนไขที่โครงการฯกำหนดและอนุญาตให้ดำเนินการศึกษา/วิจัย/พัฒนาตามหัวข้อที่ได้เสนอ มานี้ในสถาบันได้ภายใต้การบังคับบัญชาของข้าพเจ้า”

ลงชื่อ.....

2. สาระสำคัญของโครงการ

โครงการ ข้อความเอย เจ้าอ่านว่าอย่างไร, ระบบระบุตำแหน่งและรู้จำข้อความบนภาพถ่ายทั่วไป มีจุดประสงค์เพื่อสร้างระบบที่สามารถระบุตำแหน่งข้อความและรู้จำข้อความจากภาพถ่าย โดยมุ่งเน้นที่จะพัฒนาขั้นตอนวิธีที่จะทำให้การระบุตำแหน่งข้อความและรู้จำข้อความจากภาพถ่าย ให้มีความแม่นยำมากขึ้น สำหรับโครงการนี้จะเป็นการสร้างขั้นตอนวิธีในการระบุตำแหน่งข้อความจากภาพถ่ายและรู้จำตัวเลขจากบริเวณข้อความที่สามารถระบุตำแหน่งได้ โดยจะทำการระบุตำแหน่งและสร้างกรอบให้กับบริเวณที่มีข้อความปรากฏอยู่ในภาพถ่ายนั้น หลังจากนั้นจะทำการรู้จำตัวเลขจากข้อความที่สามารถระบุตำแหน่งได้ และทำการทดสอบประสิทธิภาพในการระบุตำแหน่งข้อความและรู้จำตัวเลขจากภาพถ่าย โดยใช้ชุดข้อมูล (dataset) ที่ถูกจัดเตรียมเพื่อใช้ในการแข่งขัน BEST 2016 : การแข่งขันสุดยอดการหาตำแหน่งและรู้จำข้อความในภาพถ่าย ซึ่งในปีนี้จะเป็นภาพถ่ายของผู้เข้าแข่งขันในการวิ่งมาราธอน โดยต้องทำการระบุตำแหน่งเลขประจำตัวของผู้เข้าแข่งขัน และทำการรู้จำตัวเลขที่ระบุตำแหน่งได้ ซึ่งชุดข้อมูลดังกล่าว ทางหน่วยปฏิบัติการวิจัยเทคโนโลยีภาพ ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ (NECTEC) จะเป็นผู้จัดเตรียมเลือกชุดข้อมูลและจัดทำผลเฉลย

ในการจัดทำโครงการนี้ทางผู้วิจัยได้มีแนวคิดจาก ความต้องการในการอ่านข้อความที่ปรากฏอยู่ในภาพทั่วไป ซึ่งจะต้องผ่านกระบวนการในการระบุตำแหน่งข้อความในภาพ (Text Localization) แล้วจึงนำไปทำการรู้จำตัวอักษร (Text Recognition) ดังนั้นการพัฒนาขั้นตอนวิธีในการระบุตำแหน่งและรู้จำข้อความในภาพถ่ายทั่วไป ซึ่งมีความซับซ้อนของภาพสูงกว่าภาพเอกสาร ให้มีความแม่นยำจึงเป็นส่วนสำคัญที่ทำให้การอ่านข้อความในภาพทั่วไปนั้นมีความแม่นยำมากขึ้น

3. หลักการและเหตุผล

การสื่อสารเป็นสิ่งจำเป็นสิ่งหนึ่งสำหรับการดำรงชีวิตของมนุษย์ในปัจจุบัน ซึ่งรูปแบบของการสื่อสารที่พบในชีวิตประจำวันนอกเหนือจากการสื่อสารด้วยเสียงแล้ว ยังมีการสื่อสารจากภาพ และข้อความต่างๆ ซึ่งเราสามารถพบข้อความปรากฏอยู่ทั่วไป ทั้งบนท้องถนน ภายในอาคาร และสื่อต่างๆ ดังแสดงในรูปที่ 1 ซึ่งข้อความเหล่านี้เป็นสิ่งสำคัญที่จะสื่อสาร และให้ข้อมูลกับผู้อ่าน ซึ่งถ้าเราสามารถสกัดและบอกได้ว่าข้อความในภาพประกอบด้วยตัวอักษรตัวใดได้ จะสามารถนำไปใช้ประโยชน์กับงานวิจัยทางด้านอื่นๆ ได้อีกมาก

รูปที่ 1 ตัวอย่างของข้อความที่พบในชีวิตประจำวัน

การแปลภาษาอัตโนมัติจากภาพ เป็นงานหนึ่งที่สามารถใช้ประโยชน์จากการอ่านข้อความในภาพได้ ตัวอย่างเช่น การช่วยแปลภาษาของป้ายในสถานที่ท่องเที่ยวหรือตามเส้นทางการเดินทาง ให้เป็นภาษาที่นักท่องเที่ยวสามารถเข้าใจได้ซึ่งเพิ่มความสะดวกและช่วยเหลือให้นักท่องเที่ยวเดินทางได้ง่ายขึ้น

การวิจัยทางด้านหุ่นยนต์ก็เป็นอีกด้านหนึ่ง ที่สามารถใช้ข้อมูลจากการอ่านข้อความในภาพที่หุ่นยนต์เห็นนั้นมาช่วยให้หุ่นยนต์สามารถตัดสินใจโต้ตอบกับสภาพแวดล้อมได้หลากหลายยิ่งขึ้น ยกตัวอย่างเช่น งานวิจัยทางด้านการสร้างระบบขับรถอัตโนมัติแบบไร้คนขับ ซึ่งต้องอาศัยความสามารถในการอ่านป้าย และเครื่องหมายจราจร เพื่อตัดสินใจว่าจะควบควบคุมทิศทางและความเร็วในการควบคุมรถอย่างไร

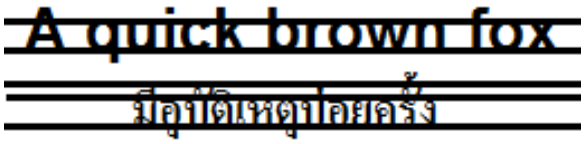
งานวิจัยด้านความเป็นจริงเสมือน (augmented reality) ก็สามารถใช้ข้อมูลที่ได้จากอ่านข้อความนั้น มาช่วยเพิ่มข้อมูลของสิ่งต่างๆที่ปรากฏภายในความจริงเสมือนที่สร้างขึ้น ยกตัวอย่างเช่น Google Glass ซึ่งเป็นผลิตภัณฑ์จากงานวิจัยด้านดังกล่าว ที่สามารถใช้ข้อมูลภายนอก เช่นข้อความที่พบจากการมองด้วย Google Glass ร่วมกับการโปรแกรมต่างๆเพื่อสร้างความเป็นจริงเสมือนที่อ้างอิงจากข้อความที่พบจากการมองเห็นได้ เป็นต้น

เทคโนโลยีด้านการช่วยเหลือคนพิการ ก็สามารถใช้ประโยชน์จากการอ่านข้อความในภาพได้เช่นกัน ซึ่งในปัจจุบันมีผู้พิการทางด้านการมองเห็นเป็นจำนวนมาก ยกตัวอย่างเช่น ผู้มีสายตาเลือนราง ที่สามารถมองเห็นได้บ้างแต่ไม่ค่อยชัดเจน ทำให้การใช้ชีวิตประจำวันที่ต้องอาศัยการอ่านข้อความจากสิ่งต่างๆนั้นทำได้ค่อนข้างลำบาก การสร้างเครื่องช่วยอ่านที่สามารถอ่านข้อความที่พบในชีวิตประจำวันได้ จะช่วยพัฒนาคุณภาพชีวิตของผู้พิการกลุ่มนี้ให้ดีขึ้น

แต่ในปัจจุบัน งานวิจัยด้านการรู้จำตัวอักษรภาษาไทยนั้นมุ่งเน้นในด้านการรู้จำตัวอักษรจากเอกสารที่ได้จากเครื่องสแกนภาพและการรู้จำตัวอักษรภาษาไทยที่ได้จากการเขียน ซึ่งภาพเอกสารที่ได้จากการสแกนภาพนั้นมีความแตกต่างกับภาพของตัวอักษรและข้อความที่ปรากฏอยู่ในภาพถ่าย ทั้งในแง่สภาพของแสง ตำแหน่งของข้อความ รูปแบบของตัวอักษร การเบลอของตัวอักษร ฉากหลังที่มีความซับซ้อนและมีการบดบังเกิดขึ้น ซึ่งได้แสดงตัวอย่างของภาพเหล่านี้เทียบกับภาพที่ได้จากการสแกนเอกสารในรูปที่ 2 อีกทั้งโครงสร้างของภาษาไทยที่ประกอบด้วยสระและวรรณยุกต์ที่มีระดับแตกต่างจากภาษาอื่นๆ ปัญหาเหล่านี้เป็นอุปสรรคที่ทำให้วิธีการรู้จำตัวอักษรที่ใช้กับเอกสารที่ได้จากเครื่องสแกนภาพนั้น ทำงานได้ผลไม่ดีนักกับอักษรที่ได้จากภาพถ่าย

รูปที่ 2 ความแตกต่างระหว่างข้อความบนเอกสาร เปรียบเทียบกับข้อความที่ปรากฏในภาพถ่ายทั่วไป

จากปัญหาเหล่านี้ เมื่อทำการสืบค้นงานวิจัยที่เกี่ยวข้องกับการรู้จำข้อความบนภาพถ่ายข้อความภาษาอังกฤษและทำการเปรียบเทียบระหว่างภาพเอกสารที่ได้จากการสแกนกับภาพข้อความในภาพถ่ายฉากธรรมชาติแล้วจะเห็นได้ว่า ข้อความในภาพถ่ายนั้นจะอยู่ในฉาก (scene) ที่มีความซับซ้อนสูงกว่ามาก ทั้งในแง่องค์ประกอบและการรบกวนต่างๆ และการระบุตำแหน่งข้อความในภาพสำหรับภาษาไทยนั้น มีความยากมากกว่าภาษาอังกฤษเนื่องจากโครงสร้างของภาษาไทยที่ประกอบด้วยสระและวรรณยุกต์หลายระดับดังแสดงในรูปที่ 3 ทำให้วิธีการระบุตำแหน่งข้อความบนภาพทั่วถ่ายไปสำหรับภาษาอังกฤษนั้นทำงานได้ไม่นักกับภาษาไทยเนื่องจากอาจจะสามารถระบุตำแหน่งของตัวอักษรได้ดี แต่การระบุตำแหน่งของสระและวรรณยุกต์ที่ไม่ได้อยู่ในระดับเดียวกับตัวอักษรอาจจะทำได้ไม่ดีนัก



รูปที่ 3 ความแตกต่างของระดับของอักษร สระและวรรณยุกต์ ในภาษาไทยและภาษาอังกฤษ

สำหรับในโครงการนี้ ซึ่งเป็นโครงการที่ส่งเข้าแข่งขันในประเภท BEST 2016 : การแข่งขันสุดยอดการหาตำแหน่งและรู้จำข้อความในภาพถ่าย (Text Localization and Recognition Contest) ซึ่งมีหัวข้อในการแข่งขันคือ การระบุตำแหน่งและรู้จำข้อความจากภาพถ่าย โดยการระบุตำแหน่งข้อความบนภาพถ่ายการแข่งวิ่งมาราธอน และทำการรู้จำตัวเลขจากบริเวณที่ระบุตำแหน่งได้ดังแสดงตัวอย่างของภาพทดสอบในรูปที่ 4 ซึ่งลักษณะงานดังกล่าวมีความคล้ายคลึงกับการรู้จำตัวอักษรจากภาพถ่าย เนื่องด้วยตัวเลขที่ปรากฏในภาพถ่ายทั่วไปก็มีความหลายหลากเช่นเดียวกับตัวอักษร

รูปที่ 4 ตัวอย่างภาพทดสอบ

ที่มา : www.shutterrunning.com/index.php?option=com_kunena&view=topic&catid=4&id=21123&Itemid=110

4. วัตถุประสงค์

- เพื่อสร้างระบบที่สามารถระบุตำแหน่งข้อความและรู้จำตัวเลขจากภาพถ่ายที่เห็นข้อความได้ชัดเจนและมีความคมชัด โดยระบบที่สร้างจะสามารถทำงานได้โดยอัตโนมัติ

- เพื่อเป็นต้นแบบในการสร้างและพัฒนาขั้นตอนวิธีในการระบุตำแหน่งข้อความภาษาไทยและรู้จำตัวเลขในภาพถ่ายให้มีความแม่นยำมากขึ้น

- เพื่อส่งเสริมให้วงการนักวิจัยทางด้านการประมวลผลภาพในประเทศไทย ให้มีการคิดค้น และพัฒนาขั้นตอนวิธีที่เหมาะสมในการระบุตำแหน่งและรู้จำข้อความจากภาพถ่าย

5. ปัญหาหรือประโยชน์ที่เป็นเหตุผลให้ควรพัฒนาโปรแกรม

ในปัจจุบันปัญหาทางด้านการรู้จำข้อความจากภาพถ่ายทั่วไปนั้น ได้รับความสนใจจากนักวิจัยจำนวนมาก เนื่องจากสามารถนำไปพัฒนาใช้ในโปรแกรมประยุกต์ได้ในหลายรูปแบบเช่น การแปลภาษาจากภาพแบบอัตโนมัติ การอ่านภาพของป้ายจราจร เส้นทางต่างๆ ซึ่งสามารถนำไปใช้ในระบบขับรถยนต์อัตโนมัติ เป็นต้น แต่ในปัจจุบันยังไม่มีขั้นตอนวิธี ที่ออกแบบและทำงานได้อย่างมีประสิทธิภาพสำหรับภาษาไทย ซึ่งโครงการนี้มุ่งเน้นที่จะพัฒนาขั้นตอนวิธีในการระบุตำแหน่งข้อความภาษาไทยที่สามารถทำงานได้อย่างมีประสิทธิภาพ และทำการรู้จำตัวเลขจากบริเวณที่ระบุตำแหน่ง ซึ่งสามารถพัฒนาไปสู่การรู้จำตัวอักษรภาษาไทยซึ่งมีความซับซ้อนสูงกว่าได้เช่นกัน อีกทั้งยังช่วยส่งเสริมวงการวิจัยและการพัฒนาในประเทศไทย ให้มีความตื่นตัวในการพัฒนาขั้นตอนวิธีที่สามารถทำงานได้อย่างมีประสิทธิภาพกับภาษาไทยอีกด้วย

6. เป้าหมายและขอบเขตของโครงการ

6.1 เป้าหมายของโครงการ

เพื่อพัฒนาระบบและขั้นตอนวิธีในการระบุตำแหน่งข้อความและรู้จำตัวเลขจากภาพถ่าย โดยมุ่งเน้นที่จะเพิ่มความแม่นยำในการระบุตำแหน่งข้อความภาษาไทยและภาษาอื่นๆ รวมถึงการสร้างขั้นตอนวิธีรู้จำตัวเลขที่พบในภาพถ่ายที่มีความแม่นยำ ซึ่งการพัฒนาในทิศทางดังกล่าวนั้น จะช่วยให้การพัฒนาระบบรู้จำตัวอักษรจากภาพถ่ายในอนาคตมีความแม่นยำมากขึ้น

6.2 ขอบเขตของโครงการ

ทำการพัฒนาระบบและขั้นตอนในการระบุตำแหน่งข้อความและรู้จำตัวเลขจากภาพถ่าย โดยภาพถ่ายนำเข้าต้องมีขนาดไม่ต่ำกว่า 480x320 จุดภาพ ข้อความในภาพนั้นจะต้องสามารถมองเห็นได้ชัดเจน และมีขนาดความสูงมากกว่า 32 จุดภาพ โดยขั้นตอนวิธีที่ได้ออกแบบจะสามารถระบุตำแหน่งข้อความ สกัดส่วนภาพที่มีข้อความ และสามารถรู้จำตัวเลขจากบริเวณที่สกัดได้

7. รายละเอียดของการพัฒนา

7.1 เนื้อเรื่องย่อ

แผนภาพการทำงาน

รูปที่ 5 แผนภาพการทำงานของโปรแกรมที่เสนอ

งานวิจัยที่เกี่ยวข้อง

จากงานวิจัยของ (Ye & Doermann, 2014) ได้แบ่งวิธีการระบุตำแหน่งข้อความในภาพถ่ายโดยพิจารณาองค์ประกอบที่เชื่อมต่อกัน (Connected Component Based Approach) และวิธีการระบุตำแหน่งข้อความบนภาพถ่ายโดยพิจารณาส่วนภาพ (Region Based Approach) ซึ่งสามารถสรุปขั้นตอนวิธีที่ใช้ของทั้ง 2 วิธีได้ดังตารางที่ 1 และ 2

ตารางที่ 1 ขั้นตอนวิธีที่ใช้ในการระบุตำแหน่งข้อความบนภาพถ่ายโดยพิจารณาองค์ประกอบที่เชื่อมต่อกัน

	เสนอโดย	**คุณลักษณะสำคัญ**	**ตัวจำแนกประเภท**
1	(Subramanian, Natarajan, Decerbo, & Castanon, 2007)	stroke and color features	Feature Analysis
2	(Jirattitichareon & Chalidabhongse, 2006)	Edge and Color	CCA and Gaussian Mixture Model
3	(Epshtein, Ofek, & Wexler, 2010)	stroke-width transform	CCA (Connected Component Analysis)
4	(Karaoglu, Fernando, & Trémeau, 2010)	geometric, shape regularity and corner based features	Random Forest
5	(Huang & Ma, 2010)	coarseness and statistic features	CCA
6	(Neumann & Matas, 2010, 2012, 2013)	MSER and scale invariant features	SVM
7	(Yin, Yin, Hao, & Iqbal, 2012)	MSER , statistic, histogram and edge-linking features	Adaboost
8	(Woraratpanya, Boonchukusol, Kuroki, & Kato, 2013)	Edge , position and colors	K-Mean Clustering and CCA
9	(Woraratpanya et al., 2014)	Edge , position and colors	K-Mean Clustering and CCA

ตารางที่ 2 ขั้นตอนวิธีที่ใช้ในการระบุตำแหน่งข้อความบนภาพถ่ายโดยพิจารณาส่วนภาพ

	เสนอโดย	คุณลักษณะสำคัญ	ตัวจำแนกประเภท	วิธีในการรวมองค์ประกอบ
1	(Chen & Yuille, 2004)	statistic, histogram and edge-linking features	Adaboost	CCA and Niblack Binarization
2	(Gllavata, Ewerth, & Freisleben, 2004)	Histrogram SD of high frequency wavelet coefficient	K-mean Clustering	CCA
3	(Hanif, Prevost, & Negri, 2008)	Mean difference, SD and HOG	Adaboost	CCA
4	(Pan, Hou, & Liu, 2008)	Multiscaled Local Binary Pattern (mLBP) HOG	Adaboost	Relaxation Labelling and CCA based on Markov RF
5	(Hanif & Prevost, 2009)	Mean difference, SD and HOG	Complexity Adaboost	CCA + Neural Network
6	(Pan, Hou, & Liu, 2009, 2011)	HOG	Waldboost	CCA based on CRF
7	(Pan, Liu, & Hou, 2010)	Gradient,edge features, HOG, DCT and LBP	Waldboost and Polynomial	CCA
8	(Meng & Song, 2012)	Salient features, aspect ratio, occupy ratio, roughness, compactness and stroke	SVM	CCA + CRF
9	(Mishra, Alahari, & Jawahar, 2012)	HOG	SVM	NMS
10	(Kobchaisawat & Chalidabhongse, 2014)	Learnable Feature Extractors	Convolutional Neural Networks	-

11	(Kobchaisawat & Chalidabhongse, 2015)	Learnable Feature Extractors	Convolutional Neural Networks	NMS and CCA

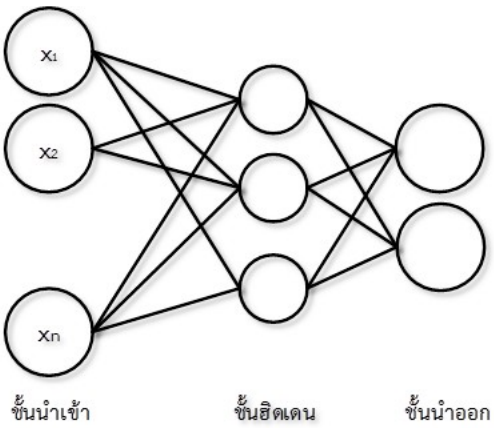
สำหรับส่วนของการรู้จำตัวเลขจากภาพทั่วไปนั้น เมื่อทำการสืบค้นงานวิจัยที่เกี่ยวข้อง จะพบงานวิจัยที่มีความน่าสนใจ ดังแสดงในตารางที่ 3

ตารางที่ 3 ขั้นตอนวิธีที่ใช้ในการรู้จำตัวเลขจากภาพถ่ายทั่วไป

	เสนอโดย	**คุณลักษณะสำคัญ**	**ตัวจำแนกประเภท**
1	(Goodfellow, Bulatov, Ibarz, Arnoud, & Shet, 2013)	Learnable Feature Extractors	Convolutional Neural Networks
2	(Sermanet, Chintala, & Lecun, 2012)	Learnable Feature Extractors	Convolutional Neural Networks
3	(Zeiler & Fergus, 2013)	Learnable Feature Extractors	Convolutional Neural Networks (Stochastic Pooling)
4	(Netzer et al., 2011)	HOG, Learnable Feature Extractors (Stacked Auto Encoder)	Neural Networks, K-Mean Clustering, Stacked Auto Encoder
5	(Goodfellow, Warde-Farley, Mirza, Courville, & Bengio, 2013)	Learnable Feature Extractors	Convolutional Neural Networks (Maxout Through Feature Maps)
6	(Lee, Xie, Gallagher, Zhang, & Tu, 2014)	Learnable Feature Extractors	Deeply-supervised Net (DSN)
7	(Liang & Hu, 2015)	Learnable Feature Extractors	Recurrent Convolutional Neural Networks

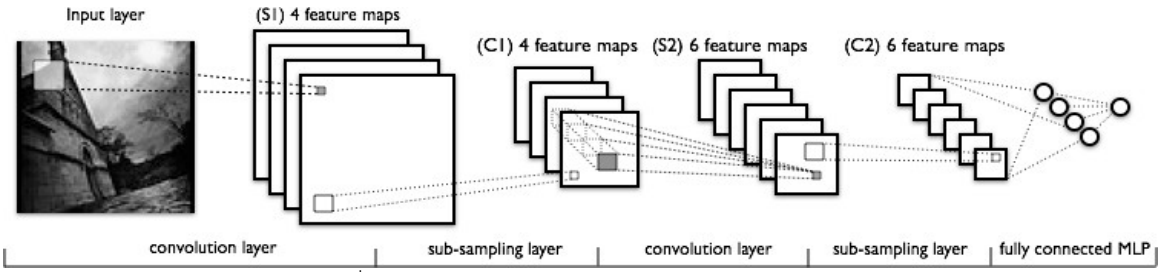
7.2 เทคนิคหรือเทคโนโลยีที่ใช้

นิวรอลเน็ตเวิร์ค (Neural Network) จะประกอบด้วย 3 ส่วนได้แก่ ชั้นนำเข้า (Input Layer) ซึ่งข้อมูลนำเข้าเป็นเวคเตอร์ 1 มิติ ของคุณลักษณะสำคัญ (feature) ที่ได้จากตัวสกัดคุณลักลักษณะสำคัญที่เลือกใช้ ชั้นฮิดเด่น(Hidden Layer) และชั้นนำออก (Output Layer) เชื่อมต่อกัน โดยมีโครงสร้างดังรูปที่ 6



รูปที่ 6 โครงสร้างของนิวรอลเน็ตเวิร์ค

คอนโวลูชันนอล นิวรอลเน็ตเวิร์ค (Convolutional Neural Network) มีความแตกต่างจากนิวรอลเน็ตเวิร์ค คือ คอนโวลูชันนอล นิวรอลเน็ตเวิร์คนั้น มีข้อมูลนำเข้าเป็นภาพ ซึ่งข้อมูลชนิด 2 มิติ และมีโครงสร้างซึ่งประกอบด้วย 2 ส่วนได้แก่ ชั้นของตัวสกัดคุณลักษณะสำคัญที่เรียนรู้ได้ (Trainable Feature Extractor Layer) และ ชั้นของการจำแนก(Fully-Connected Layer) ที่มีโครงสร้างเหมือนนิวรอลเน็ตเวิร์คทั่วไป คอนโวลูชั่นนอลนิวรอลเน็ตเวิร์ค มีโครงสร้างดังรูปที่ 7



รูปที่ 7 โครงสร้างของคอนโวลูชันนอล นิวรอลเน็ตเวิร์ค

ชั้นของตัวสกัดคุณลักษณะสำคัญที่เรียนรู้ได้ (Trainable Feature Extractor Layer) ประกอบไปด้วย 3 ชั้นหลักคือ ชั้นคอนโวลูชัน (Convolution), ชั้นของฟังก์ชันไม่เชิงเส้น (Nonlinear Function) และชั้นการสุ่มตัวอย่าง (Subsampling) และยังมีชั้นอื่นๆเช่น ชั้นดรอปเอ้าท์ (Dropout)

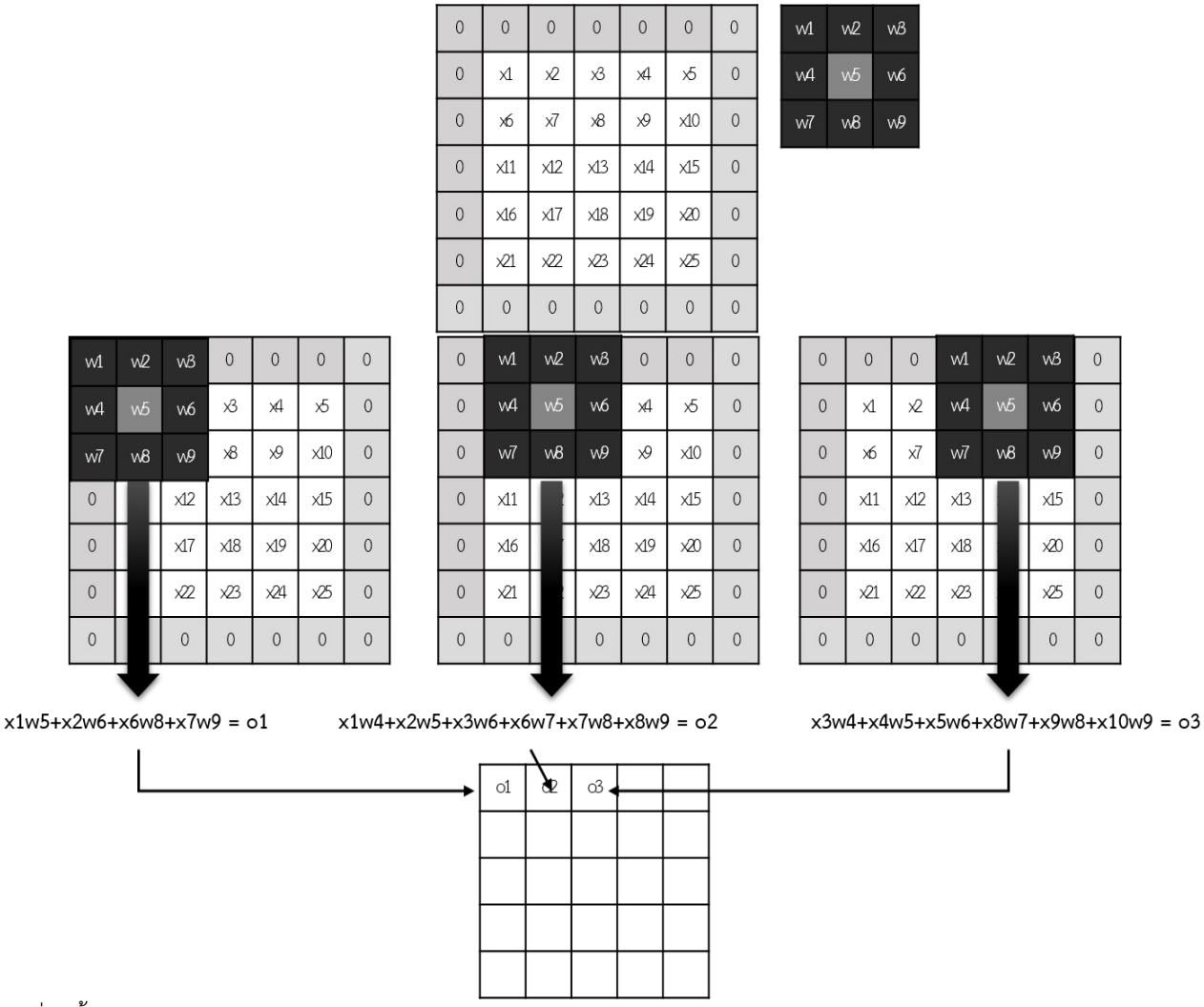
คอนโวลูชันนอล นิวรอลเน็ตเวิร์คจะมีการจัดเรียงภายในชั้นของตัวสกัดคุณลักษณะสำคัญได้ในหลายรูปแบบ โดยปกติจะมีการจัดเรียงชุดตัวสกัดคุณลักษณะสำคัญ ประกอบด้วยชั้นคอนโวลูชัน ชั้นของฟังก์ชันไม่เชิงเส้นและชั้นการสุ่มตัวอย่าง ดังแสดงในรูปที่ 8 โดยแต่ละชุดตัวสกัดคุณลักษณะสำคัญ สามารถเรียงต่อกันได้ขึ้นอยู่กับการเลือกใช้ของผู้ใช้งาน

(ก)

(ข)

รูปที่ 8 (ก) ตัวอย่างการจัดเรียงภายในชุดตัวสกัดคุณลักษณะสำคัญ
(ข) ตัวอย่างการจัดเรียงภายในชั้นของตัวสกัดคุณลักษณะสำคัญที่เรียนรู้ได้ ที่มีชุดตัวสกัดคุณลักษณะสำคัญจำนวน n ชุด

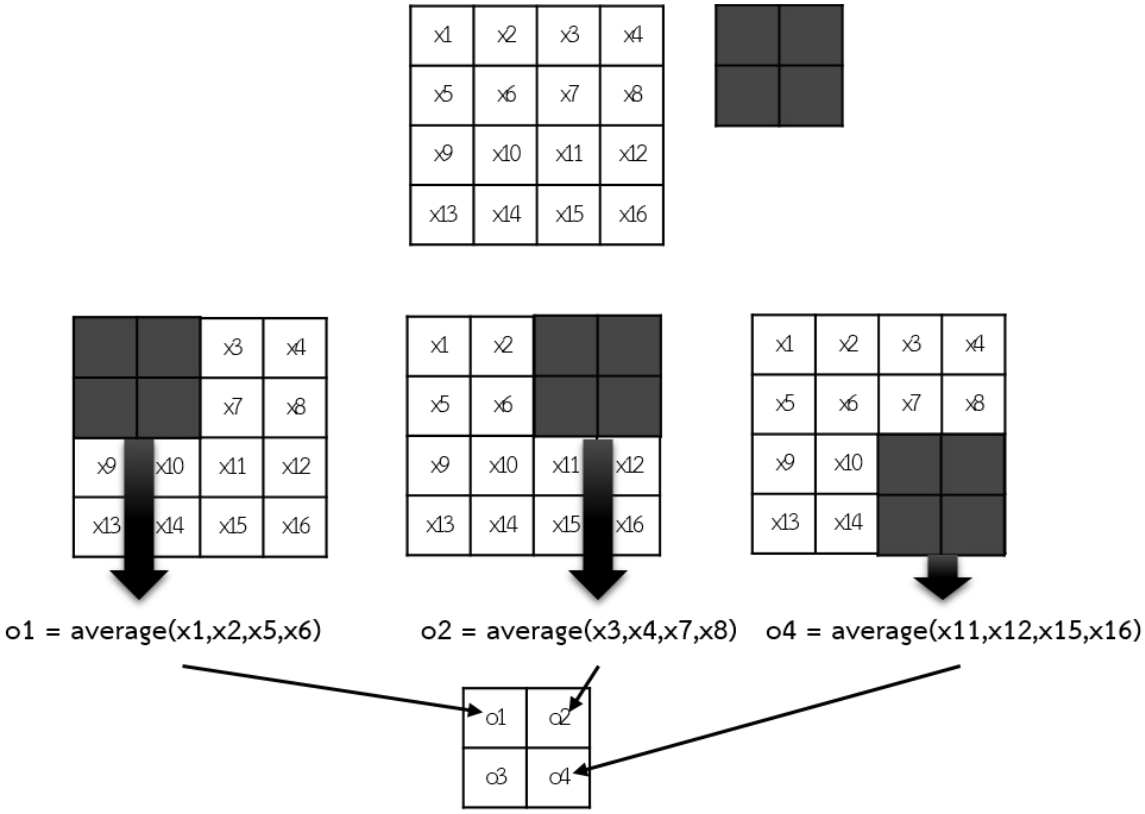
ในชั้นคอนโวลูชันนั่น จะทำการคอนโวลูชันแบบ valid ซึ่งแสดงวิธีการในรูปที่ 9 กับเวคเตอร์ 2 มิตินำเข้า โดยจะเรียกผลของชั้นคอนโวลูชั่นว่า feature maps แล้วจึงนำไปผ่านฟังก์ชั่นแบบไม่เชิงเส้น (non-linearity function) ในตอนเริ่มต้น ค่าน้ำหนักในแต่ละเคอร์เนลนั้นจะถูกสุ่ม และจะถูกปรับให้มีความเหมาะสมกับชุดข้อมูลสอนโดยขั้นตอนวิธีแบคพรอพพาเกชั่น (backpropagation)



รูปที่ 9 ขั้นตอนการคอนโวลูชันบนภาพขนาด 5x5 จุดภาพ โดยใช้เคอร์เนลขนาด 3x3 จุดภาพและให้ผลลัพธ์ขนาดเท่ากับภาพนำเข้า

เคอร์เนลที่ใช้ในของชั้นคอนโวลูชัน จะถูกสุ่มในตอนตั้งต้นและจะถูกปรับให้มีความเหมาะสมกับชุดข้อมูลสอน โดยขั้นตอนวิธีแบคพรอพพาเกชั่น (backpropagation) และผลที่ได้จากชั้นคอนโวลูชันนั้น จะเรียกว่า feature maps ซึ่งจะมีจำนวนเท่ากับเคอร์เนล

การทำงานของชั้นการสุ่มตัวอย่างนั้น เพื่อลดจำนวนตัวแปรและลดความซ้ำซ้อนของข้อมูล เนื่องจากผลที่ได้จากชั้นคอนโวลูชัน ซึ่งสำหรับคอนโวลูชัน นิวรอลเน็ตเวิร์คนั้น จะมีวิธีการสุ่มตัวอย่างที่เป็นที่นิยมใช้กัน 2 แบบคือ การหาค่าสูงสุด และการหาค่าเฉลี่ย โดยชั้นของการสุ่มตัวอย่างนั้น จะมีวิธีการเลื่อน window แตกต่างจากชั้นคอนโวลูชัน คือ โดยส่วนใหญ่จะไม่มีการทับกันของ window เกิดขึ้น วิธีการทำงานของชั้นการสุ่มตัวอย่างนั้นดังแสดงในรูปที่ 10



รูปที่ 10 ขั้นตอนการสุ่มตัวอย่างบนภาพขนาด 4x4 จุดภาพ โดยใช้เคอร์เนลการสุ่มตัวอย่างแบบหาค่าเฉลี่ยขนาด 2x2 จุดภาพ

เมื่อผ่านชั้นของตัวสกัดคุณลักษณะสำคัญที่เรียนรู้ได้แล้ว คุณลักษณะสำคัญทั้งหมดที่ได้จะถูกนำไปเป็นข้อมูลนำเข้าของชั้นการจำแนก ซึ่งสามารถเลือกได้ว่าจะใช้ขั้นตอนวิธีใดในการจำแนกกลุ่มของคุณลักษณะสำคัญที่กัดได้ โดยอาจจะเป็นนิวรอลเน็ตเวิร์คทั่วไปที่ประกอบด้วย ชั้นนำเข้า ชั้นฮิดเด่น และชั้นนำออก เพื่อให้ผลลัพธ์การจำแนกของคอนโวลูชันนอล นิวรอลเน็ตเวิร์ค ซึ่งอาจจะเป็นชั้นของการจำแนกเป็น ชั้นซอฟต์แมกซ์ (Softmax Layer) โดยไปเป็นตามสมการ

$$f(x)=\frac{e^{x}}{\sum e^{x}}$$

เมื่อ x เป็นเวคเตอร์ของคุณลักษณะสำคัญ ที่ได้จากชั้นของตัวสกัดคุณลักษณะสำคัญ โดยจะให้ผลลัพธ์ที่อยู่ในช่วง [0,1] ซึ่งเป็นค่าความน่าจะเป็นของแต่ละคุณลักษณะสำคัญที่สกัดได้

7.3 เครื่องมือที่ใช้ในการพัฒนา

- ภาษาที่ใช้ในการพัฒนา C++ และ MATLAB

- เครื่องมือและไลบรารี่ที่ใช้ในการพัฒนา

1. CodeLite และ gcc 5.2
2. Octave
3. OpenCV 3.0
4. CUDA Toolkit 7.5
5. matio
6. cuDNN
7. Boost
8. Matconvnet

7.4 รายละเอียดโปรแกรมที่จะพัฒนา

7.4.1 Input / Output Specification

Input : ภาพถ่ายที่ต้องการระบุตำแหน่งข้อความและรู้จำตัวเลข

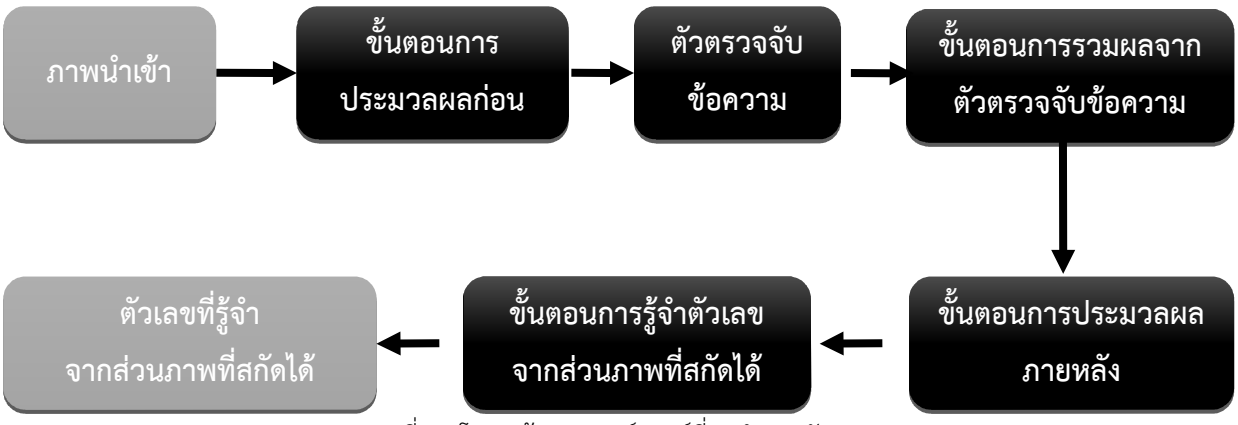
Output : ตัวเลขที่สามารถรู้จำได้จากบริเวณที่ระบุตำแหน่งข้อความได้

7.4.2 Functional Specification

- โปรแกรมสามารถระบุตำแหน่งข้อความที่ปรากฏบนภาพถ่ายนำเข้าได้
- โปรแกรมสามารถรู้จำตัวเลขจากบริเวณข้อความที่ระบุตำแหน่งได้

7.4.3 โครงสร้างของซอร์ฟแวร์ (Software Design)

โปรแกรมที่จะทำการพัฒนา สามารถแบ่งการทำงานได้ดังรูปที่ 11 และมีรายละเอียดในแต่ละส่วนดังต่อไปนี้



รูปที่ 11 โครงสร้างของซอร์ฟแวร์ที่จะทำการพัฒนา

- ขั้นตอนการประมวลผลก่อน

เพื่อปรับปรุงคุณภาพของภาพที่ต้องการระบุตำแหน่งข้อความให้มีความเหมาะสมกับการนำไปประมวลผลต่อไป

- ตัวตรวจจับข้อความ

มีหน้าที่จำแนกระหว่างส่วนภาพที่เป็นข้อความและไม่ใช่ข้อความ

- การรวมผลจากตัวตรวจจับข้อความ

เพื่อสร้างสมมุติฐานของบริเวณที่น่าจะมีข้อความจากผลของตัวตรวจจับข้อความ

- การประมวลผลภายหลัง

เพื่อกำจัดส่วนภาพที่ไม่มีข้อความที่ได้จากขั้นตอนการรวมผลจากตัวตรวจจับข้อความ

- การรู้จำตัวเลขจากบริเวณข้อความที่ตรวจจับได้

เพื่อรู้จำตัวเลขจากภาพบริเวณข้อความที่ตรวจจับได้จากขั้นตอนการตรวจจับข้อความ

7.5 ขอบเขตและข้อจำกัดของโปรแกรมที่จะพัฒนา

- ภาพนำเข้าต้องมีขนาดอย่างน้อย 480x320 จุดภาพ
- ข้อความที่ปรากฏในภาพมีความเปรียบต่างระหว่างฉากหลังและตัวอักษรชัดเจน
- ข้อความที่ปรากฏในภาพไม่มีการเบลอที่เกิดจากการเคลื่อนไหว
- ข้อความที่ปรากฏในภาพต้องมีขนาดอย่างน้อย 32 จุดภาพ
- การรู้จำข้อความนั้นจะทำการรู้จำเฉพาะตัวเลขเท่านั้น (เนื่องจากหัวข้อการแข่งขัน)

8. บรรณานุกรม

Chen, X., & Yuille, A. L. (2004). Detecting and reading text in natural scenes. In *IEEE Conference on Computer Vision and Pattern Recognition, 2004. (CVPR2004)* (Vol. 2, pp. II–366–II–373 Vol.2). http://doi.org/10.1109/CVPR.2004.1315187

Epshtein, B., Ofek, E., & Wexler, Y. (2010). Detecting text in natural scenes with stroke width transform. In *IEEE Conference on Computer Vision and Pattern Recognition, 2012. (CVPR2012)* (pp. 2963–2970). Ieee. http://doi.org/10.1109/CVPR.2010.5540041

Gllavata, J., Ewerth, R., & Freisleben, B. (2004). Text detection in images based on unsupervised classification of high-frequency wavelet coefficients. In *17th International Conference on Pattern Recognition, 2004. (ICPR2004)* (Vol. 1, pp. 425–428 Vol.1). http://doi.org/10.1109/ICPR.2004.1334146

Goodfellow, I. J., Bulatov, Y., Ibarz, J., Arnoud, S., & Shet, V. (2013). Multi-digit Number Recognition from Street View Imagery using Deep Convolutional Neural Networks. In *2nd International Conference on Learning Representations, 2014. (ICLR2014)* (pp. 1–13). Retrieved from http://arxiv.org/abs/1312.6082

Goodfellow, I. J., Warde-Farley, D., Mirza, M., Courville, A., & Bengio, Y. (2013). Maxout networks. *arXiv Preprint arXiv:1302.4389*.

Hanif, S. M., & Prevost, L. (2009). Text Detection and Localization in Complex Scene Images using Constrained AdaBoost Algorithm. In *10th International Conference on Document Analysis and Recognition, 2009. (ICDAR2009)* (pp. 1–5). http://doi.org/10.1109/ICDAR.2009.172

Hanif, S. M., Prevost, L., & Negri, P. A. (2008). A cascade detector for text detection in natural scene images. In *19th International Conference on Pattern Recognition,2008 (ICPR2008)* (pp. 1–4). http://doi.org/10.1109/ICPR.2008.4761536

Huang, X., & Ma, H. (2010). Automatic Detection and Localization of Natural Scene Text in Video. In *20th International Conference on Pattern Recognition,2010 (ICPR2010)* (pp. 3216–3219). http://doi.org/10.1109/ICPR.2010.786

Jirattitichareon, W., & Chalidabhongse, T. H. (2006). Automatic Detection and Segmentation of Text in Low Quality Thai Sign Images. In *IEEE Asia-Pacific Conference on Circuits and Systems,*

2006. (APCCAS2006) (pp. 1000–1003). http://doi.org/10.1109/APCCAS.2006.342256

Karaoglu, S., Fernando, B., & Trémeau, A. (2010). A Novel Algorithm for Text Detection and Localization in Natural Scene Images. In *International Conference on Digital Image Computing: Techniques and Applications, 2010. (DICTA2010)* (pp. 635–642). http://doi.org/10.1109/DICTA.2010.115

Kobchaisawat, T., & Chalidabhongse, T. H. (2014). Thai text localization in natural scene images using Convolutional Neural Network. In *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, 2014. (APSIPA2014)* (pp. 1–7). http://doi.org/10.1109/APSIPA.2014.7041775

Kobchaisawat, T., & Chalidabhongse, T. H. (2015). A Method for Multi-Oriented Thai Text Localization in Natural Scene Images using Convolutional Neural Network. In *4th IEEE International Conference on Signal and Image Processing Applications, 2015. (ICSIPA 2015)*.

Lee, C.-Y., Xie, S., Gallagher, P., Zhang, Z., & Tu, Z. (2014). Deeply-supervised nets. *arXiv Preprint arXiv:1409.5185*.

Liang, M., & Hu, X. (2015). Recurrent convolutional neural network for object recognition. In *IEEE Conference on Computer Vision and Pattern Recognition, 2015. (CVPR 2015)* (pp. 3367–3375). http://doi.org/10.1109/CVPR.2015.7298958

Meng, Q., & Song, Y. (2012). Text Detection in Natural Scenes with Salient Region. In *10th IAPR International Workshop on Document Analysis Systems, 2012. (DAS2012)* (pp. 384–388). http://doi.org/10.1109/DAS.2012.85

Mishra, A., Alahari, K., & Jawahar, C. V. (2012). Top-down and bottom-up cues for scene text recognition. In *IEEE Conference on Computer Vision and Pattern Recognition, 2012. (CVPR2012)* (pp. 2687–2694). http://doi.org/10.1109/CVPR.2012.6247990

Netzer, Y., Wang, T., Coates, A., Bissacco, A., Wu, B., & Ng, A. Y. (2011). Reading digits in natural images with unsupervised feature learning. *NIPS Workshop on Deep Learning and Unsupervised Feature Learning*, *2011*.

Neumann, L., & Matas, J. (2010). A Method for Text Localization and Recognition in Real-world Images. In *Computer Vision – ACCV 2010* (pp. 770–783). http://doi.org/10.1007/978-3-642-

19318-7_60

Neumann, L., & Matas, J. (2012). Real-time scene text localization and recognition. In *IEEE Conference on Computer Vision and Pattern Recognition, 2012. (CVPR2012)* (pp. 3538–3545). http://doi.org/10.1109/CVPR.2012.6248097

Neumann, L., & Matas, J. (2013). Scene Text Localization and Recognition with Oriented Stroke Detection. In *IEEE International Conference on Computer Vision, 2013. (ICCV2013)* (pp. 97–104). http://doi.org/10.1109/ICCV.2013.19

Pan, Y.-F., Hou, X., & Liu, C.-L. (2008). A Robust System to Detect and Localize Texts in Natural Scene Images. In *8th IAPR International Workshop on Document Analysis Systems, 2008. (DAS2008)* (pp. 35–42). http://doi.org/10.1109/DAS.2008.42

Pan, Y.-F., Hou, X., & Liu, C.-L. (2009). Text Localization in Natural Scene Images Based on Conditional Random Field. In *10th International Conference on Document Analysis and Recognition, 2009. (ICDAR2009)* (pp. 6–10). Ieee. http://doi.org/10.1109/ICDAR.2009.97

Pan, Y.-F., Hou, X., & Liu, C.-L. (2011). A Hybrid Approach to Detect and Localize Texts in Natural Scene Images. In *IEEE Transactions on Image Processing* (Vol. 20, pp. 800–813). http://doi.org/10.1109/TIP.2010.2070803

Pan, Y.-F., Liu, C.-L., & Hou, X. (2010). Fast scene text localization by learning-based filtering and verification. In *17th IEEE International Conference on Image Processing, 2010. (ICIP2010)* (pp. 2269–2272). http://doi.org/10.1109/ICIP.2010.5651862

Sermanet, P., Chintala, S., & Lecun, Y. (2012). Convolutional neural networks applied to house numbers digit classification. In *21st International Conference on Pattern Recognition,2012 (ICPR2012)* (pp. 3288–3291).

Subramanian, K., Natarajan, P., Decerbo, M., & Castanon, D. (2007). Character-Stroke Detection for Text-Localization and Extraction. In *9th International Conference on Document Analysis and Recognition, 2007. (ICDAR2007)* (Vol. 1, pp. 33–37). http://doi.org/10.1109/ICDAR.2007.4378671

Woraratpanya, K., Boonchukusol, P., Kuroki, Y., & Kato, Y. (2013). Improved Thai text detection from natural scenes. In *International Conference on Information Technology and Electrical*

Engineering, 2013. (ICITEE2013) (pp. 137–142). http://doi.org/10.1109/ICITEED.2013.6676227

Woraratpanya, K., Pasupa, K., Suttapakti, U., Boonchukusol, P., Titijaroonroj, T., Hokking, R., … Kato, Y. (2014). Text-background decomposition for thai text localization and recognition in natural scenes. In *Information Technology and Electrical Engineering, 2014. (ICITEE 2014)* (pp. 1–6). http://doi.org/10.1109/ICITEED.2014.7007914

Ye, Q., & Doermann, D. (2014). Text Detection and Recognition in Imagery: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1–1. http://doi.org/10.1109/TPAMI.2014.2366765

Yin, X., Yin, X.-C., Hao, H.-W., & Iqbal, K. (2012). Effective text localization in natural scene images with MSER, geometry-based grouping and AdaBoost. In *21st International Conference on Pattern Recognition,2012 (ICPR2012)* (pp. 725–728).

Zeiler, M. D., & Fergus, R. (2013). Stochastic pooling for regularization of deep convolutional neural networks. *arXiv Preprint arXiv:1301.3557*.

9. ประวัติและผลงานวิจัยดีเด่นของผู้พัฒนา

ประวัติของผู้พัฒนา

นาย ธนานพ กอบชัยสวัสดิ์ ปัจจุบันเป็นนิสิตสาขาวิศวกรรมคอมพิวเตอร์ ระดับปริญญาเอก จากห้องปฏิบัติการวิจัยคอมพิวเตอร์กราฟฟิกและเทคโนโลยีทางภาพ (Computer Graphic and Computer Imaging : CGCI) ภาควิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย โดยมี ผศ. ดร. ธนารัตน์ ชลิดาพงศ์ เป็นอาจารย์ที่ปรึกษา

ผลงานและผลงานวิจัยดีเด่นของผู้พัฒนา

- เข้าร่วมโครงการโอลิมปิกวิชาการ สาขาวิชาคอมพิวเตอร์ ปี พ.ศ. 2545 – 2547
- ส่งผลงานเข้าประกวดในงาน NSC 2005 และได้ผ่านเข้าสู่รอบชิงชนะเลิศ หมวดโปรแกรมส่งเสริมการเรียนรู้ในหัวข้อ กลับคืนสู่วันวาน
- ส่งผลงานเข้าประกวดในงาน NSC 2010 และได้ผ่านเข้าสู่รอบชิงชนะเลิศ หมวดโปรแกรมช่วยเหลือคนพิการ ในหัวข้อ เสียงพูดสู่ภาษามือ
- ได้รับรางวัลชนะเลิศจากการส่งผลงานในงาน NSC 2015 ประเภทหัวข้อพิเศษ BEST 2015 : การแข่งขันสุดยอดการหาตำแหน่งข้อความบนภาพถ่าย
- ได้รับรางวัลรองชนะเลิศจากโครงการ Thai Sign Language Finger Spelling System for Bridging Communication between Thai People and Thai Deaf ในการประกวด 2nd IPTV Application Challenge ประเภท Value to Society ที่จัดโดย International Telecommunication Union (ITU)
- ได้รับการตอบรับการตีพิมพ์ผลงานทางวิชาการในหัวข้อ Thai Text Localization in Natural Scene Images using Convolutional Neural Network ในการประชุมทางวิชาการ Asia-Pacific Signal and Information Processing Association Annual Summit and Conference 2014 (APSIPA 2014)
- ได้รับการตอบรับการตีพิมพ์ผลงานทางวิชาการในหัวข้อ A Method for Multi-Oriented Thai Text Localization in Natural Scene Images using Convolutional Neural Network ในการประชุมทางวิชาการ IEEE International Conference on Signal and Image Processing Applications (ICSIPA 2015)