

การขยายความสามารถตัวต้นคิดของแกนส์สำหรับการสร้างภาพที่หลากหลายจากภาพร่าง



วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

สาขาวิชาวิทยาศาสตร์คอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2562

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

Enhancement of GANs' Generator for Diverse Image Generation from Sketch



A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science in Computer Science

Department of Computer Engineering

FACULTY OF ENGINEERING

Chulalongkorn University

Academic Year 2019

Copyright of Chulalongkorn University

หัวข้อวิทยานิพนธ์	การขยายความสามารถตัวต้นคิดของแกนส์สำหรับการสร้าง ภาพที่หลากหลายจากภาพร่าง
โดย	น.ส.วรินธร ภูสมสาย
สาขาวิชา	วิทยาศาสตร์คอมพิวเตอร์
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	รองศาสตราจารย์ ดร.ญาใจ ลีมีปิยะกรณ์

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้หัวข้อวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

.....	คณบดีคณะวิศวกรรมศาสตร์
(ศาสตราจารย์ ดร.สุพจน์ เตชวรสินสกุล)	
คณะกรรมการสอบวิทยานิพนธ์	ประธานกรรมการ
.....	
(ผู้ช่วยศาสตราจารย์ ดร.สุกรี สิ้นธุภิณโณ)	อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก
.....	
(รองศาสตราจารย์ ดร.ญาใจ ลีมีปิยะกรณ์)	กรรมการภายนอกมหาวิทยาลัย
.....	
(อาจารย์ ดร.ภาสกร อภิรักษ์วรพินิต)	

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

วรินทร์ ภูมสมสาย : การขยายความสามารถตัวต้นคิดของแกนส์สำหรับการสร้างภาพที่หลากหลายจากภาพร่าง. (Enhancement of GANs' Generator for Diverse Image Generation from Sketch) อ.ที่ปรึกษาหลัก : รศ. ดร.ญาใจ ลี้มปิยะกรณ์

การเปลี่ยนแปลงลักษณะกายภาพใบหน้าอาจเป็นผลมาจากวัยที่เพิ่มขึ้น การทำศัลยกรรม หรือการปลอมแปลง ผู้ต้องสงสัยหรืออาชญากรปกปิดอัตลักษณ์ของตนด้วยภาพลักษณ์ที่ผิดไปจากเดิม เช่น ใส่วิกผม สวมแว่นตา หรือติดหนวดเครา งานวิจัยนี้เป็นประโยชน์ต่อการสร้างลักษณะหลากหลายที่ประดิษฐ์ขึ้นให้ผิดไปจากเดิมของผู้ต้องสงสัย หรือใบหน้าที่เปลี่ยนไปของบุคคลสูญหาย เทคนิคแกนส์ได้ถูกประยุกต์ใช้เพื่อสังเคราะห์ภาพสีจากภาพร่าง โดยภาพผลลัพธ์ที่ได้สามารถสร้างความหลากหลายของใบหน้าที่เกิดจากความแตกต่างของคุณลักษณะบนใบหน้าสลับไปมาได้ 5 แบบ ประกอบด้วย มี/ไม่มีผม มี/ไม่มีการแต่งหน้า ผมตรง/ผมลอน ใส่/ไม่ใส่แว่นตา และมี/ไม่มีหนวดเครา แนวทางที่นำเสนอได้ขยายความสามารถตัวต้นคิดของสตาร์แกน2ด้วยสถาปัตยกรรมยูเน็ต การประเมินสมรรถนะของแบบจำลองที่นำเสนอกระทำการทดลองเปรียบเทียบกับสมรรถนะของสตาร์แกน2 โดยใช้ค่าคะแนนเอฟไอดีสำหรับวัดคุณภาพของภาพที่สร้างขึ้น พบว่า ค่าคะแนนตัววัดเอฟไอดีบนชุดข้อมูลทดสอบให้ผลน้อยกว่า 40% ของค่าที่ได้จากแบบจำลองเส้นฐานสตาร์แกน2 อีกทั้งผลลัพธ์ภาพสังเคราะห์ที่สร้างขึ้นจากคุณลักษณะหลากหลายบนใบหน้าดูเป็นธรรมชาติและสมจริง

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

สาขาวิชา วิทยาศาสตร์คอมพิวเตอร์
ปีการศึกษา 2562

ลายมือชื่อนิสิต
ลายมือชื่อ อ.ที่ปรึกษาหลัก

6170958521 : MAJOR COMPUTER SCIENCE

KEYWORD: Generative Adversarial Networks, Image synthesis, Facial attribute editing, Sketch-to-image translation, U-Net architecture

Warintorn Phusomsai : Enhancement of GANs' Generator for Diverse Image Generation from Sketch. Advisor: Assoc. Prof. Yachai Limpiyakorn

Face physical changes may result from aging, surgery, or disguise. The criminal suspects conceal their identity with false appearances such as wearing a wig, glasses, beard and mustache. This research benefits the generation of various fictitious appearances of the suspects or facial changes of lost persons. The technique of GANs is applied for synthesizing a color image from a sketch. The output image can be varied in five facial attributes with a toggle: haired/ bald wear/ not wear makeup straight/ wavy hair wear/ not wear glasses beard and mustache/ no beard and mustache. The approach enhances the Generator of StarGan2 with the U-Net architecture. The experiments were carried out to evaluate the performance of the proposed model compared to that of StarGan2. FID scores are used for measuring the quality of the generated images. The FID scores measured on the test data reported about 40% less than that of the baselined StarGan2 model and the synthesized images with varied facial attributes look natural and realistic.

วิทยาลัย
CHULALONGKORN UNIVERSITY

Field of Study: Computer Science

Student's Signature

Academic Year: 2019

Advisor's Signature

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงเป็นอย่างดีได้ด้วยความอนุเคราะห์จากรองศาสตราจารย์ ดร. ญาใจ ลิ้มปิยะกรณ์ อาจารย์ที่ปรึกษาวิทยานิพนธ์ ที่ได้สละเวลาอันมีค่ามาให้ความรู้ แนวคิด คำปรึกษา ตลอดจนตรวจสอบ และแก้ไขปัญหาข้อผิดพลาดต่าง ๆ ผู้วิจัยขอกราบขอบพระคุณเป็นอย่างสูงไว้ ณ ที่นี้ด้วย

ขอขอบพระคุณ ผู้ช่วยศาสตราจารย์ ดร.สุกรี สิ้นธุภิณฺโญ ประธานกรรมการสอบวิทยานิพนธ์ และอาจารย์ ดร.ภาสกร อภิรักษ์วรพิณิต กรรมการสอบวิทยานิพนธ์ ที่กรุณาเสียสละเวลาอันมีค่า ให้คำแนะนำที่เป็นประโยชน์ในการทำวิทยานิพนธ์ในครั้งนี้

ขอขอบพระคุณบิดา มารดา ที่ให้การสนับสนุน ความรัก ความเป็นห่วง และเป็นกำลังใจที่ดีเสมอมา

ขอขอบคุณเพื่อน ๆ พี่ ๆ น้อง ๆ ภาควิชาวิศวกรรมคอมพิวเตอร์ที่คอยช่วยเหลือ ให้คำปรึกษา แลกเปลี่ยนความรู้ ความคิดเห็นในด้านต่าง ๆ และมีมิตรภาพที่ดีให้กันตลอดระยะเวลาที่ผ่านมา

สุดท้ายนี้ขอขอบพระคุณผู้ที่เกี่ยวข้องทุกท่านที่ไม่ได้กล่าวมาข้างต้นที่คอยให้ความช่วยเหลือ ซึ่งทำให้วิทยานิพนธ์สำเร็จลุล่วงไปได้ด้วยดี ผู้วิจัยหวังเป็นอย่างยิ่งว่าวิทยานิพนธ์ฉบับนี้จะเป็นประโยชน์ต่อผู้ที่สนใจไม่มากนักน้อย

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

วรินธร ภูสมสาย

สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	ค
บทคัดย่อภาษาอังกฤษ.....	ง
กิตติกรรมประกาศ.....	จ
สารบัญ.....	ฉ
สารบัญตาราง.....	ฉ
สารบัญภาพ.....	ช
บทที่ 1 บทนำ.....	1
1.1 ที่มาและความสำคัญ.....	1
1.2 วัตถุประสงค์การวิจัย.....	2
1.3 ขอบเขตงานวิจัย.....	2
1.4 ขั้นตอนการดำเนินงาน.....	2
1.5 ประโยชน์ที่คาดว่าจะได้รับ.....	3
1.6 ลำดับการจัดเรียงเนื้อหาในวิทยานิพนธ์.....	3
1.7 ผลงานที่ได้รับการตีพิมพ์จากวิทยานิพนธ์.....	3
บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง.....	5
2.1 ทฤษฎีที่เกี่ยวข้อง.....	5
2.1.1 โครงข่ายการสร้างเชิงประปักษ์ หรือ แจนส์ (Generative Adversarial networks: GANs).....	5
2.1.2 การแปลงภาพจากภาพ (Image to image translation).....	6
2.1.4 สถาปัตยกรรมยูเน็ต (U-net Architecture).....	7
2.2. งานวิจัยที่เกี่ยวข้อง.....	8

2.2.1 Image-to-Image Translation with Conditional Adversarial Networks.....	8
2.2.2 Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks.....	9
2.2.3 Interactive Sketch & Fill: Multiclass Sketch-to-Image Translation	10
2.2.4 StarGAN v2: Diverse Image Synthesis for Multiple Domains	10
2.2.5 Image Generation from Sketch Constraint Using Contextual GAN	11
2.2.6 GD-StarGAN: Multi-domain image-to-image translation in garment design	11
2.2.7 AttGAN: Facial Attribute Editing by Only Changing What You Want	12
2.2.8 MU-GAN: Facial Attribute Editing based on Multi-attention Mechanism...	13
2.2.9 SSCGAN: Facial Attribute Editing via Style Skip Connections.....	14
บทที่ 3 แนวคิดและวิธีวิจัย.....	15
บทที่ 4 การออกแบบและพัฒนาแบบจำลอง	21
4.1 โครงสร้างเน็ตเวิร์กของแบบจำลอง.....	21
4.2 สภาพแวดล้อมและเครื่องมือที่ใช้ในการพัฒนา.....	22
4.3 ขั้นตอนการเรียนรู้แบบจำลอง.....	23
4.3 ขั้นตอนการตรวจสอบความสมเหตุสมผล.....	25
4.4 ปัญหาที่พบของการเรียนรู้แบบจำลองและวิธีการแก้ไข.....	26
บทที่ 5 การวิเคราะห์และประเมินผล	27
บทที่ 6 สรุปผลการวิจัย.....	32
6.1 สรุปผลการวิจัย.....	32
6.2 ข้อจำกัดในงานวิจัย.....	32
6.2 งานวิจัยในอนาคต	33
บรรณานุกรม.....	34

ประวัติผู้เขียน..... 37



จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

สารบัญตาราง

	หน้า
ตาราง 1 เปรียบเทียบงานวิจัยที่ประยุกต์ใช้แกนส์กับการสังเคราะห์ภาพกับงานวิจัยนี้	15
ตาราง 2 การเก็บค่าเอฟไอดีเพื่อหาค่าเอฟไอดีแอบแฝง 1 ครั้ง.....	20
ตาราง 3 การเก็บค่าเอฟไอดีเพื่อหาค่าเอฟไอดีอ้างอิง 1 ครั้ง	20
ตาราง 4 โครงสร้างสถาปัตยกรรมยูเน็ตของตัวต้นคิด	21
ตาราง 5 ค่าเอฟไอดีเปรียบเทียบระหว่างเทคนิคสตาร์แกน2 กับเทคนิคสตาร์แกน2ยูเน็ต.....	27



จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

สารบัญภาพ

	หน้า
ภาพที่ 1 โครงสร้างของโครงข่ายการสร้างเชิงปรปักษ์ [7].....	6
ภาพที่ 2 โครงสร้างของโครงข่ายคอนโวลูชัน [9]	7
ภาพที่ 3 โครงสร้างของยูเน็ต [8].....	8
ภาพที่ 4 ข้อมูลนำเข้าและผลลัพธ์จากการสร้างภาพจากภาพต้นแบบด้วยโครงข่ายการสร้างเชิง ปรปักษ์แบบมีเงื่อนไข [3].....	9
ภาพที่ 5 ค่าผลลัพธ์จากตัววัดเอเอ็มที่เปรียบเทียบระหว่างไซเคิลแกนและเทคนิคอื่น [4]	9
ภาพที่ 6 แนวทางสองขั้นตอนของระบบที่พัฒนาใน [5].....	10
ภาพที่ 7 ตัวต้นคิด ตัวโครงข่ายการแปลง ตัวเข้ารหัสสโตล์ และตัวแบ่งแยกของสตาร์แกน2 [6]	11
ภาพที่ 8 โครงสร้างของเทคนิคซ่อมแซมภาพและคอนเท็กซ์แกน [10].....	11
ภาพที่ 9 โครงสร้างของเทคนิคจีดีสตาร์แกน [11].....	12
ภาพที่ 10 โครงสร้างของเทคนิคเอทีทีแกน [12].....	13
ภาพที่ 11 โครงสร้างของเทคนิคเอ็มยูแกนส์ [13].....	13
ภาพที่ 12 โครงสร้างของเทคนิคเอสเอสซีแกนส์ [14].....	14
ภาพที่ 13 ภาพร่างที่แปลงโดยคลังโปรแกรมเอ็กดีโอจี.....	16
ภาพที่ 14 ภาพรวมของงานวิจัย	17
ภาพที่ 15 โครงสร้างเน็ตเวิร์กของ 1) ตัวแบ่งแยก 2) ตัวโครงข่ายการแปลง 3) ตัวเข้ารหัสรูปแบบ .	22
ภาพที่ 16 ค่าความสูญเสียระหว่างการเรียนรู้ของแบบจำลอง	24
ภาพที่ 17 ตัวอย่างรูปภาพระหว่างการเรียนรู้ของแบบจำลอง	25
ภาพที่ 18 ค่าเอฟไอดีของการตรวจสอบความสมเหตุสมผล.....	26
ภาพที่ 19 ภาพเสมือนที่ได้จากการทดสอบโดยแสดงในแต่ละคู่ของคุณลักษณะของเพศหญิง.....	29
ภาพที่ 20 ภาพเสมือนที่ได้จากการทดสอบโดยแสดงในแต่ละคู่ของคุณลักษณะของเพศชาย.....	30

ภาพที่ 21 ผลลัพธ์การสังเคราะห์ภาพเสมือนจากภาพร่างเดียวกันในทุกคุณลักษณะแบ่งตามเพศหญิง	31
ภาพที่ 22 ผลลัพธ์การสังเคราะห์ภาพเสมือนจากภาพร่างเดียวกันในทุกคุณลักษณะแบ่งตามเพศชาย	31



บทที่ 1

บทนำ

1.1 ที่มาและความสำคัญ

หนึ่งในวิธีการชี้ตัวคนร้ายที่ใช้ในปัจจุบันที่นำไปสู่การค้นหาและจับกุมคนร้ายได้เป็นอย่างดี คือ การวาดภาพร่าง ซึ่งถูกร่างมาจากการให้ปากคำของพยานที่อยู่ในเหตุการณ์ โดยอาศัยการจดจำ ลักษณะหรือจุดเด่นของอวัยวะบนใบหน้า รวมไปถึงวิธีการร่างจากการรวบรวมอวัยวะต่าง ๆ ของ ใบหน้า ต่อมา มีการพัฒนาแอปพลิเคชันเพื่อช่วยในการร่างภาพคนร้ายให้ง่ายขึ้น โดยสร้างเป็นภาพ ลายเส้นการ์ตูนที่มีลักษณะของอวัยวะบนใบหน้าที่หลากหลายและสามารถเลือกแต่ละอวัยวะมา รวมกัน แต่ยังคงมีความไม่ครอบคลุมในทุกลักษณะของอวัยวะและส่งผลให้การร่างภาพคนร้ายมีความ คลาดเคลื่อนจากความเป็นจริง ในปัจจุบันจึงได้มีการนำเอาปัญญาประดิษฐ์หรือการเรียนรู้เชิงลึกมา ช่วยในการสร้างภาพเพื่อให้ได้ข้อมูลที่มีความเหมือนจริงและรวดเร็วมากยิ่งขึ้น

Ian J. G. และคณะ [1] ได้นำเสนอเทคนิคการสร้างข้อมูลรูปภาพด้วยโครงข่ายการสร้างเชิง ประปักษ์ หรือ แกนส์ (Generative Adversarial Networks: GANs) ซึ่งสามารถนำมาประยุกต์ใช้ในการ สร้างภาพเหมือนบุคคลได้ โดยวิธีการทำงานจะแบ่งออกเป็นสองฝ่าย ประกอบด้วย ตัวต้นคิด (Generator) และตัวแบ่งแยก (Discriminator) เน็ตเวิร์กทั้งสองจะทำงานร่วมกัน โดยตัวต้นคิดจะ สร้างข้อมูลเหมือนขึ้นมาและตัวแบ่งแยกจะตัดสินว่าใกล้เคียงกับข้อมูลจริงหรือไม่ หากยังไม่ใกล้เคียง ตัวต้นคิดจะเรียนรู้และปรับค่าพารามิเตอร์ด้วยค่าความผิดพลาดจนกว่าจะได้ข้อมูลที่เหมือนจริง ต่อมา ในปี ค.ศ. 2014 Radford A. และคณะ [2] ได้นำโครงข่ายประสาทคอนโวลูชัน หรือ ซีเอ็นเอ็น (Convolution neuron network: CNN) มาประยุกต์ใช้กับโครงข่ายการสร้างเชิงประปักษ์เพื่อสกัด คุณลักษณะเด่นของข้อมูลภาพ โดยนำเสนอเทคนิคดีซีแกน (Deep Convolutional Generative Adversarial Network : DCGAN) ซึ่งใช้ข้อมูลนำเข้าเป็นสัญญาณรบกวนที่สุ่มขึ้นมาโดยผ่านชั้นของ โครงข่ายประสาทคอนโวลูชันจนกลายเป็นภาพเหมือนภาพต้นแบบ ต่อมา Phillip I. และคณะ [3] ได้ พัฒนาการสร้างข้อมูลรูปภาพเหมือนโดยมีภาพพิมพ์เขียวของวัตถุนั้นเป็นข้อมูลนำเข้า และมีการ ปรับเปลี่ยนพารามิเตอร์โดยเรียนรู้จากค่าความผิดพลาด ซึ่งรูปภาพที่ได้มีความเหมือนจริงมากกว่าการ สร้างจากสัญญาณรบกวน แต่ข้อจำกัดของเทคนิคนี้คือ ต้องมีภาพพิมพ์เขียวของวัตถุก่อน จึงจะ สามารถนำไปปรับและสร้างเป็นภาพจริงได้ นอกจากนี้ ภาพที่ได้จะไม่มีความหลากหลายมากนัก เพราะมีภาพต้นแบบที่คล้าย ๆ กัน ต่อมา มีการประยุกต์เทคนิคการสร้างภาพโดยอาศัยภาพต้นแบบ แปลงเป็นภาพในกลุ่มอื่นได้ ยกตัวอย่างเช่น เปลี่ยนภาพม้าให้เป็นม้าลาย เปลี่ยนภาพถ่ายตุ๊กตาร้อนให้

เป็นฤดูหนาว ซึ่งนำเสนอโดย Jun-Yan Z. และคณะ [4] ในปี ค.ศ. 2019 Arnab G. และคณะ [5] ประยุกต์ใช้โครงข่ายการสร้างเชิงประปักษ์ในการสร้างภาพจากภาพร่างให้มีผลลัพธ์หลากหลายกลุ่ม เช่น ภาพร่างวงกลม เมื่อนำไปเรียนรู้สามารถสร้างภาพผลลัพธ์ออกมาในกลุ่มต่าง ๆ ที่มีลักษณะเป็นวงกลม เช่น ลูกบอล ลูกบาศก์ แดงโม ดาวจันทร์ เป็นต้น นอกจากการเปลี่ยนกลุ่มของสิ่งของแล้วยังมีงานวิจัยที่พัฒนาเกี่ยวกับการเปลี่ยนลักษณะของคนได้ เช่น การเปลี่ยนสีผม การเปลี่ยนอายุ นำเสนอโดย Yunjey C. และคณะ [6] โดยเสนอเทคนิคสตาร์แกน2 (StarGan2) ซึ่งรับข้อมูลนำเข้าเป็นภาพบุคคล และสามารถสร้างความหลากหลายของภาพเสมือนของบุคคลนั้นได้มากถึง 40 คุณลักษณะที่แตกต่างกัน ทั้งนี้ ผู้วิจัยเห็นว่าคนร้ายสามารถเปลี่ยนแปลงคุณลักษณะของใบหน้าจากเดิมได้และอาจทำให้ต่างจากคำให้การ ด้วยความสามารถในการสร้างภาพเสมือนของคนในลักษณะต่าง ๆ ของเทคนิคสตาร์แกน2 ผู้วิจัยจึงได้นำเสนอแนวทางในการสร้างภาพเสมือนที่มีคุณลักษณะบนใบหน้าที่หลากหลายของบุคคลจากภาพร่าง ซึ่งผลของการเปลี่ยนแปลงคุณลักษณะต่างๆของใบหน้าสามารถนำไปใช้ประโยชน์ในการสืบหาตัวบุคคลหรือคนร้ายได้

1.2 วัตถุประสงค์การวิจัย

เพื่อขยายความสามารถของโครงข่ายการสร้างเชิงประปักษ์สำหรับสร้างภาพบุคคลเสมือนจริงจากภาพร่าง โดยสามารถสร้างความหลากหลายของภาพผลลัพธ์จากการกำหนดค่าคุณลักษณะของใบหน้าในระบบรองรับ

1.3 ขอบเขตงานวิจัย

1. ใช้ชุดข้อมูลภาพใบหน้าบุคคลชนิดเจเพ็ก (JPEG) จากฐานข้อมูลเซเลบเอ จำนวน 350 ภาพ ขนาด 128×128 พิกเซล
2. ความหลากหลายคุณลักษณะของใบหน้าที่กำหนดได้ครั้งละ 1 คุณลักษณะ
3. ประเมินผลแบบจำลองผลลัพธ์ด้วยตัววัดเอฟไอดี (Frechet Inception Distance: FID) ซึ่งเป็นตัววัดค่าความแตกต่างระหว่างภาพจริงและภาพเสมือน โดยเปรียบเทียบค่าเอฟไอดีระหว่างเทคนิคสตาร์แกน2 และเทคนิคที่นำเสนอ

1.4 ขั้นตอนการดำเนินงาน

1. ศึกษางานวิจัยเกี่ยวกับเทคนิคการสร้างข้อมูลจากโครงข่ายการสร้างเชิงประปักษ์
2. วิเคราะห์ข้อดี/ ข้อจำกัดของงานวิจัยที่เกี่ยวข้อง เพื่อประยุกต์ใช้กับงานวิจัย

3. กำหนดขอบเขตและวัตถุประสงค์ของงานวิจัย
4. ออกแบบสถาปัตยกรรมของงานวิจัย
5. พัฒนางานวิจัย
6. ทดสอบและวัดผลความคล้ายของภาพผลลัพธ์ที่สร้างกับภาพร่าง
7. เผยแพร่ผลงานวิชาการ
8. สรุปผลการวิจัย
9. เรียบเรียงวิทยานิพนธ์

1.5 ประโยชน์ที่คาดว่าจะได้รับ

1. ได้เทคนิคและแบบจำลองการสร้างข้อมูลด้วยโครงข่ายการสร้างเชิงประปักษ์สำหรับสร้างภาพใบหน้าบุคคลที่มีความหลากหลายของคุณลักษณะต่าง ๆ บนใบหน้าที่ระบรองรับจากภาพร่าง
2. ได้ระบบต้นแบบสำหรับการสร้างภาพเสมือนใบหน้าที่หลากหลายจากภาพร่างที่อาจเป็นประโยชน์ต่อการสร้างภาพถ่ายผู้ต้องสงสัยหรือบุคคลสูญหาย

1.6 ลำดับการจัดเรียงเนื้อหาในวิทยานิพนธ์

เนื้อหาของวิทยานิพนธ์ฉบับนี้แบ่งออกเป็น 6 บท ได้แก่ บทที่ 1 บทนำ ประกอบด้วย ความสำคัญของปัญหา วัตถุประสงค์ของงานวิจัย ขอบเขตงานวิจัย ประโยชน์ที่คาดว่าจะได้รับ และผลงานที่ได้รับการตีพิมพ์จากวิทยานิพนธ์ บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง บทที่ 3 แนวคิดและวิธีวิจัย บทที่ 4 การออกแบบและพัฒนาแบบจำลอง บทที่ 5 การวิเคราะห์และประเมินผล และบทที่ 6 สรุปผลการวิจัย ข้อจำกัดในการทำงานวิจัย และงานวิจัยในอนาคต

1.7 ผลงานที่ได้รับการตีพิมพ์จากวิทยานิพนธ์

ส่วนหนึ่งของวิทยานิพนธ์ฉบับนี้ได้รับการตีพิมพ์ผลงานทางวิชาการจำนวน 2 บทความ:

1. Phusomsai, W. & Limpiyakorn, Y. (2020). Applying GANs for Generating Image with Varied Facial Attributes from Sketch. In 13th International Conference on Computer and Electrical Engineering. Beijing, China.

2. Phusomsai, W. & Limpiyakorn, Y. (2020). Toward U-Net-based GANs for Diverse Facial Image Synthesis from Sketch. In 13th International Conference on Advanced computer theory and Engineering. Hangzhou, China.



บทที่ 2

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

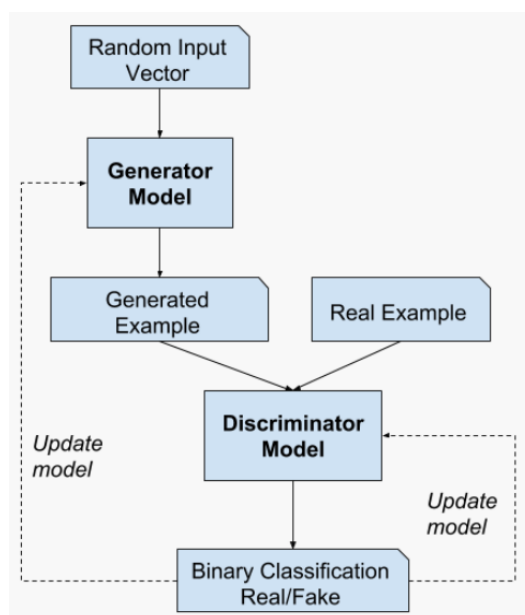
2.1 ทฤษฎีที่เกี่ยวข้อง

2.1.1 โครงข่ายการสร้างเชิงประปักษ์ หรือ แจนส์ (Generative Adversarial networks: GANs)

การสร้างข้อมูลจากโครงข่ายการสร้างเชิงประปักษ์ เป็นเทคโนโลยีใหม่ที่พัฒนาขึ้นสำหรับการจำลองชุดข้อมูลใหม่ให้เหมือนข้อมูลจริงมากที่สุด เพื่อเป็นประโยชน์กับงานที่ต้องการข้อมูลจำนวนมาก เช่น การจำแนกประเภท เพราะการเก็บข้อมูลหรือการติดฉลากข้อมูลอาจใช้เวลาและค่าใช้จ่ายที่สูง ซึ่งถ้ามีการจำลองหรือสร้างข้อมูลเหมือนจริงขึ้นมาจะช่วยลดปัญหาในส่วนนี้ได้ โครงสร้างของเทคนิคการสร้างข้อมูลจากโครงข่ายการสร้างเชิงประปักษ์ จะมีการทำงานแบบแข่งขันกันระหว่างคู่ของเน็ตเวิร์ก (peers of network) [7] ประกอบด้วย

1. ตัวต้นคิด (G) ทำหน้าที่สร้างข้อมูลเหมือนให้เหมือนข้อมูลจริงมากที่สุด
2. ตัวแบ่งแยก (D) ทำหน้าที่รับข้อมูลนำเข้าได้แก่ข้อมูลเหมือนและข้อมูลจริงและจำแนกว่าข้อมูลที่รับว่าเป็นข้อมูลจริงหรือไม่

จากรูปที่ 1 ตัวต้นคิดจะรับสัญญาณรบกวนที่ได้จากการสุ่มเข้ามาเพื่อสร้างเป็นข้อมูลเหมือนและส่งไปยังตัวแบ่งแยก โดยตัวแบ่งแยกจะรับข้อมูลมาสองชุด ได้แก่ข้อมูลจริงและข้อมูลเหมือน จากนั้นตัวแบ่งแยกจะมีการคำนวณหาค่าความน่าจะเป็น ถ้าหากค่าความน่าจะเป็นเข้าใกล้ค่า 1 หมายความว่า ข้อมูลเหมือนมีความคล้ายข้อมูลจริง ถ้าหากค่าความจริงเข้าใกล้ค่า 0 หมายความว่า ข้อมูลเหมือนไม่มีความคล้ายข้อมูลจริง แต่ถ้าความน่าจะเป็นเป็นเท่ากับ 0.5 หมายความว่าตัวแบ่งแยกไม่สามารถแบ่งแยกความแตกต่างของข้อมูลชุดนั้นได้ เพราะฉะนั้นจะถือได้ว่าข้อมูลเหมือนมีความคล้ายกับข้อมูลจริงจนแยกไม่ออก สังเกตได้ว่าตัวต้นคิดไม่มีการรับรู้ถึงข้อมูลของจริงใด ๆ แต่ตัวต้นคิดจะอาศัยผลลัพธ์ของการเรียนรู้จากตัวแบ่งแยกเพื่อนำมาปรับปรุงวิธีการเรียนรู้ของตัวเองให้สามารถสร้างข้อมูลที่มีความเหมือนจริงได้ โดยมีเป้าหมายคือการทำให้ค่าความน่าจะเป็นของข้อมูลเหมือนมีค่าความน่าจะเป็นใกล้เคียงกับค่าความน่าจะเป็นของข้อมูลจริงมากที่สุด



ภาพที่ 1 โครงสร้างของโครงข่ายการสร้างเชิงประปรักษ์ [7]

เทคนิคการสร้างข้อมูลจากโครงข่ายการสร้างเชิงประปรักษ์สามารถแบ่งตามการประยุกต์และพัฒนาได้ 3 ประเภท

1. การประยุกต์และพัฒนาในส่วนของตัวต้นคิด เช่น ดับเบิ้ลยูแกน (W-GAN) บีแกน (BE-GAN) และเอซีแกน (AC-GAN) เป็นต้น
2. การประยุกต์และพัฒนาในส่วนของตัวแบ่งแยก เช่น ซีแกน (cGAN) พิกซ์ทูพิกซ์ (Pix2Pix) และ ไซเคิลแกน (CycleGAN) เป็นต้น
3. การประยุกต์และพัฒนาในเชิงโครงสร้าง เช่น ดีซีแกน (DCGAN) สตาร์แกน (StarGAN) และสตาร์แกน2 (StarGAN2) เป็นต้น

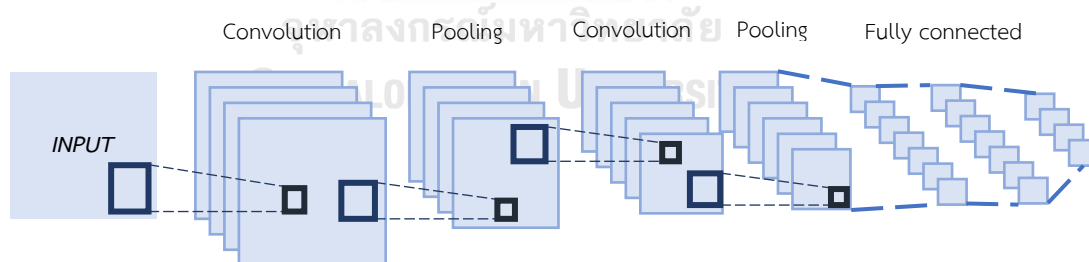
2.1.2 การแปลงภาพจากภาพ (Image to image translation)

การแปลงภาพจากภาพเป็นเทคนิคในการหาความสัมพันธ์ระหว่างภาพต้นแบบและภาพเป้าหมาย โดยสามารถสร้างภาพที่มีลักษณะคล้ายกันหรือต่างกันได้ การสร้างภาพที่มีลักษณะคล้ายกันหรือจัดอยู่ในกลุ่มเดียวกันจะต้องเตรียมคู่ของภาพต้นแบบและภาพเป้าหมาย เพื่อใช้หาความสัมพันธ์ของภาพและใช้ในการเรียนรู้เพื่อสร้างแบบจำลอง Phillip I. และคณะ [3] นำเสนอเทคนิคพิกซ์ทูพิกซ์ โดยใช้ภาพโครงสร้างของวัตถุเป็นภาพต้นแบบและสร้างเป็นภาพเสมือนให้คล้ายกับภาพเป้าหมาย โดยการสร้างภาพที่ต้องอาศัยคู่ของข้อมูลจะมีความหลากหลายไม่มากนัก เพราะว่ามีข้อมูลที่ได้จะมีความใกล้เคียงกับภาพต้นแบบเท่านั้น รวมไปถึงเวลาของการเตรียมคู่ของข้อมูลใช้

เวลานาน ต่อมา จึงมีการพัฒนางานวิจัยที่สามารถสร้างภาพที่ไม่จำเป็นต้องเป็นคู่กันขึ้น โดยสร้างแบบจำลองที่สามารถหาความสัมพันธ์ของทั้งสองภาพและสร้างออกมาเป็นภาพเสมือนตามที่ต้องการได้ เช่น งานวิจัยของ Jun-Yan Z. และคณะ [4] นำเสนอการสร้างภาพจากต่างโดเมนด้วยเทคนิคหาความสอดคล้องของภาพทั้งสองเพื่อนำไปปรับโครงสร้างของภาพ และแปลงเป็นภาพเสมือนตามที่ตั้งเป้าไว้ โดยหัวใจสำคัญของการแปลงภาพจากภาพต้นแบบ คือ การคงไว้ของคุณลักษณะหลักของภาพต้นแบบและแทนที่ส่วนที่เหลือด้วยคุณลักษณะเด่นของภาพเป้าหมาย

2.1.4 สถาปัตยกรรมยูเน็ต (U-net Architecture)

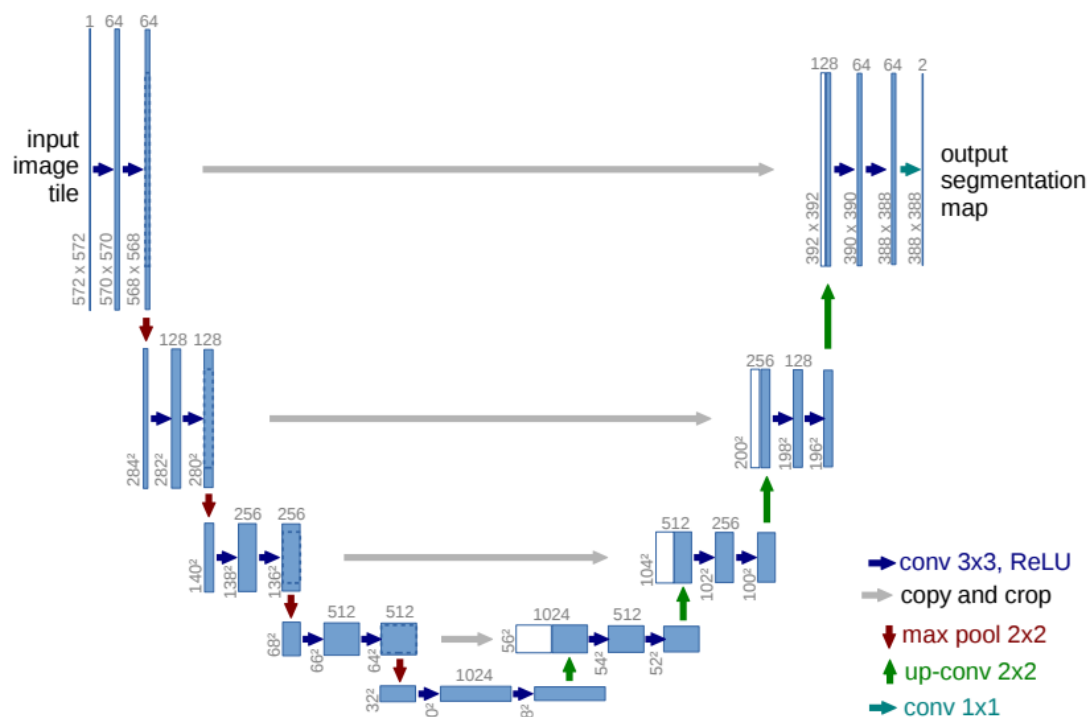
สถาปัตยกรรมยูเน็ตมีพื้นฐานมาจากโครงข่ายคอนโวลูชัน (Convolution network) เสนอโดย Olaf R. และคณะ [8] โดยถูกนำมาใช้แก้ปัญหาการแบ่งส่วนของภาพทางชีววิทยา ทั้งนี้ ได้มีการปรับปรุงโครงสร้างของโครงข่ายคอนโวลูชันมาเชื่อมต่อกันอย่างเต็มรูปแบบ เบื้องต้นโครงข่ายคอนโวลูชันประกอบด้วยชั้นการทำงาน 4 ชั้น ดังภาพที่ 2 ได้แก่ 1) Convolution Layer เป็นชั้นที่ทำการสกัดคุณลักษณะของข้อมูลนำเข้าโดยใช้ตัวกรอง (Filter) หรือเคอร์เนล (Kernel) ผลลัพธ์ที่ได้เรียกว่า ฟีเจอร์แมป (Feature map) [8] ซึ่งเป็นคุณลักษณะที่บ่งบอกลักษณะเด่นของข้อมูลชุดนั้นๆ 2) Non- Linearity Layer เป็นชั้นที่ใช้ปรับเปลี่ยนหรือกำหนดข้อมูลขาออกจากข้อมูลจากชั้นคอนโวลูชัน 3) Pooling Layer เป็นชั้นคัดกรองคุณลักษณะเฉพาะของข้อมูลที่มีความสำคัญเพื่อลดความซับซ้อนของข้อมูล และ 4) Fully Connected Layer เป็นโครงข่ายประสาทเทียมที่เชื่อมต่อและประมวลผลของแต่ละโหนดของข้อมูล เพื่อให้ได้ชุดข้อมูลสำหรับนำไปใช้ต่อไป



ภาพที่ 2 โครงสร้างของโครงข่ายคอนโวลูชัน [9]

สถาปัตยกรรมยูเน็ตประกอบด้วยสองส่วน คือ 1) ส่วนเข้ารหัส (Encoder) และ 2) ส่วนถอดรหัส (Decoder) โดยทั้งสองส่วนจะทำงานตรงกันข้าม กล่าวคือ ส่วนเข้ารหัสจะทำงานตามโครงข่ายคอนโวลูชันโดยทำให้ข้อมูลต้นแบบมีขนาดเล็กลงและสกัดคุณลักษณะเด่นของข้อมูลออกมาเรียกว่า Latent space ในขณะที่ส่วนถอดรหัสจะทำงานตรงกันข้ามกับ Convolution Layer หรือที่

เรียกว่า Convolution transpose layer และ Pooling layer เพื่อที่จะขยาย Latent space ไปอยู่ในรูปแบบที่ต้องการและมีขนาดเท่ากับข้อมูลต้นแบบ สถาปัตยกรรมยูเน็ตมีการเชื่อมต่อแบบก้าวกระโดด (Skip connection) ระหว่างส่วนเข้ารหัสและส่วนถอดรหัสเพื่อส่งต่อคุณลักษณะของภาพต้นแบบให้คงอยู่ในภาพผลลัพธ์ ดังภาพที่ 3

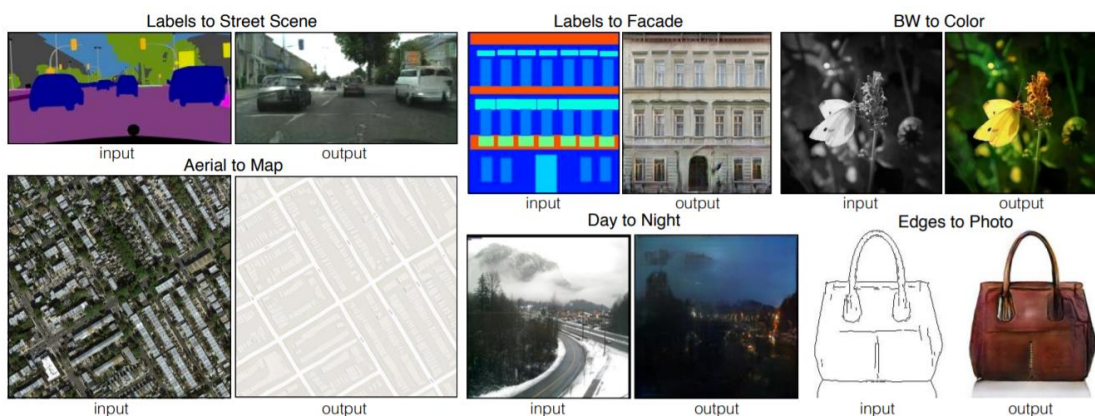


ภาพที่ 3 โครงสร้างของยูเน็ต [8]

2.2. งานวิจัยที่เกี่ยวข้อง

2.2.1 Image-to-Image Translation with Conditional Adversarial Networks

Phillip I. และคณะ [3] เสนองานวิจัยเกี่ยวกับการสร้างภาพเสมือนจากภาพต้นแบบด้วยโครงข่ายการสร้างเชิงประปักษ์แบบมีเงื่อนไข โดยตัวต้นคิดรับข้อมูลนำเข้าและเงื่อนไขเพื่อสร้างภาพเสมือนให้ใกล้เคียงตามเงื่อนไขนั้นๆ ชื่อว่า ซีแกน (cGAN) ซึ่งมีพื้นฐานมาจากเทคนิคฟิซซุทิกซ์ ประเด็นที่นำมาเสนอในงานวิจัยนี้ คือ การปรับโครงสร้างของตัวต้นคิด เปลี่ยนฟังก์ชันการคิดค่าความผิดพลาดของแบบจำลอง และเสนอผลงานในรูปแบบของแอปพลิเคชัน แบบจำลองที่นำเสนอสามารถประยุกต์ใช้กับฐานข้อมูลหลายชนิด เช่น ภาพดอกไม้ ภาพกระเป๋ ภาพแผนที่ เป็นต้น ดังภาพที่ 4



ภาพที่ 4 ข้อมูลนำเข้าและผลลัพธ์จากการสร้างภาพจากภาพต้นแบบด้วยโครงข่ายการสร้างเชิง
ปรปักษ์แบบมีเงื่อนไข [3]

2.2.2 Unpaired Image-to-Image Translation using Cycle-Consistent

Adversarial Networks

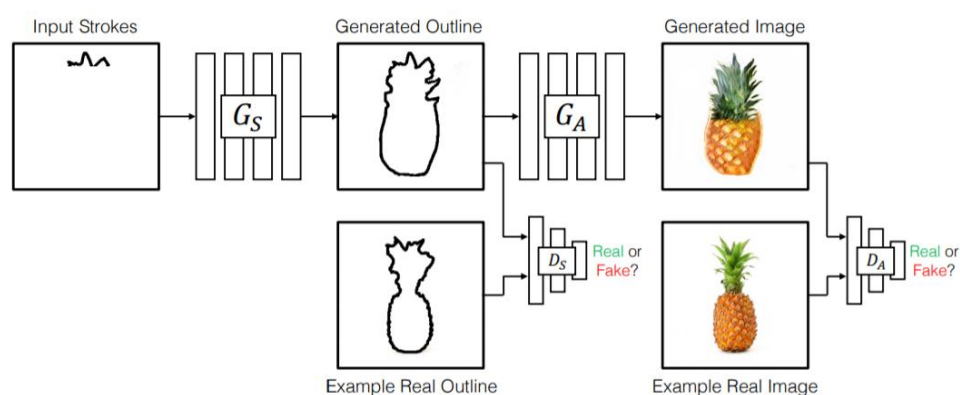
งานวิจัย [4] นำเสนอโดย Jun-Yan Z. และคณะ มีแนวคิดมาจากการสร้างภาพเสมือนจากภาพต้นแบบ ซึ่งปกติมาจากภาพที่เป็นคู่ เช่น ภาพร่างโครงสร้างตึกนำมาสร้างให้เป็นภาพตึก แต่ไม่ใช่ทุกข้อมูลที่มีการเก็บข้อมูลแบบเป็นคู่ จึงเป็นแนวคิดของงานวิจัยในการสร้างภาพที่ไม่เป็นคู่หรือมาจากข้อมูลคนละกลุ่ม เช่น การสร้างภาพม้าจากภาพม้าลาย การสร้างภาพถ่ายหน้าหนาวจากภาพถ่ายหน้าร้อน เป็นต้น การแปลงโครงสร้างของภาพต้นแบบและภาพผลลัพธ์ในลักษณะนี้จะใช้คุณลักษณะของวัตถุด้วยกันแต่ต่างกันที่รูปแบบขึ้นอยู่กับเงื่อนไขที่กำหนดไว้ เทคนิคที่เสนอเรียกว่า ไซเคิลแกน (CycleGAN) วัดผลโดยเปรียบเทียบกับโครงข่ายการสร้างเชิงปรปักษ์อื่น ๆ โดยใช้ตัววัดเอเอ็มที (Amazon Mechanical Turk: AMT) ซึ่งเป็นเว็บไซต์ที่มีกลุ่มคนมาตัดสินใจหรือวิเคราะห์ข้อมูลให้ โดยไซเคิลแกนให้ผลลัพธ์ที่ดีและมีค่าเอเอ็มทีมากกว่าเทคนิคอื่น ดังภาพที่ 5

Loss	Map → Photo	Photo → Map
	% Turkers labeled real	% Turkers labeled real
CoGAN [32]	0.6% ± 0.5%	0.9% ± 0.5%
BiGAN/ALI [9, 7]	2.1% ± 1.0%	1.9% ± 0.9%
SimGAN [46]	0.7% ± 0.5%	2.6% ± 1.1%
Feature loss + GAN	1.2% ± 0.6%	0.3% ± 0.2%
CycleGAN (ours)	26.8% ± 2.8%	23.2% ± 3.4%

ภาพที่ 5 ค่าผลลัพธ์จากตัววัดเอเอ็มทีเปรียบเทียบระหว่างไซเคิลแกนและเทคนิคอื่น [4]

2.2.3 Interactive Sketch & Fill: Multiclass Sketch-to-Image Translation

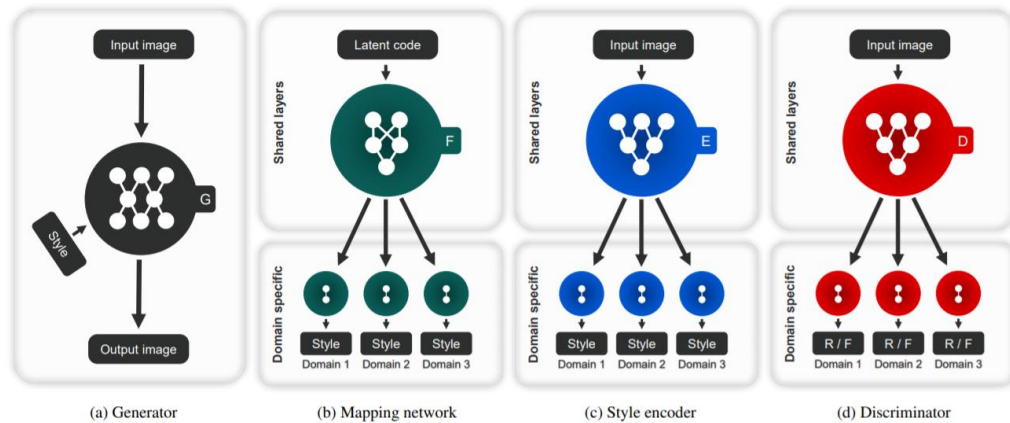
Arnab G. และคณะ [5] ได้เสนอวิธีการสร้างภาพเสมือนจากภาพร่าง โดยเริ่มจากภาพร่างแค่บางส่วนเพื่อต้องการให้มีผลลัพธ์ที่หลากหลาย มีการกำหนดตัวต้นคิดขึ้นมาสองตัว ตัวแรกทำหน้าที่ต่อเติมขอบของภาพร่างจากลายเส้นบางส่วน จากนั้นตัวต้นคิดตัวที่สองต่อเติมภาพร่างด้วยสี และพื้นผิวของผลลัพธ์ที่ต้องการ ตัวต้นคิดทั้งสองตัวจะมีตัวแบ่งแยกของตัวเองเพื่อทดสอบภาพที่สร้างขึ้นมากับภาพจริงและใช้ค่าความผิดพลาดปรับปรุงแบบจำลองต่อไป ดังภาพที่ 6 โดยนำเสนอโครงสร้างสกินี่เรสเนสที่ให้ผลลัพธ์ที่เพิ่มขึ้น



ภาพที่ 6 แนวทางสองขั้นตอนของระบบที่พัฒนาใน [5]

2.2.4 StarGAN v2: Diverse Image Synthesis for Multiple Domains

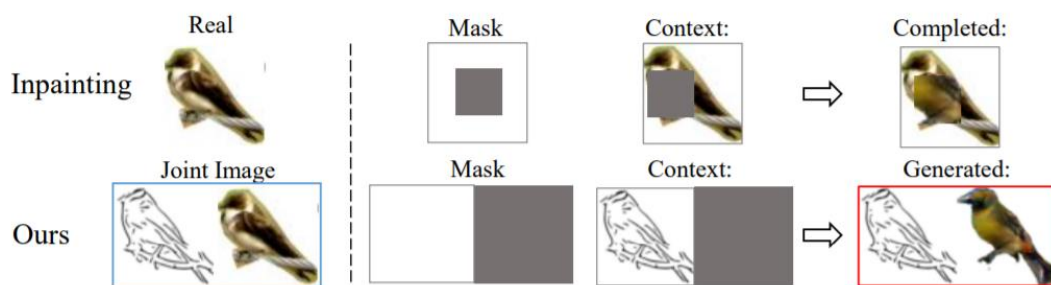
Yunjey C. และคณะ [6] มีแนวคิดในการสร้างภาพเสมือนที่แตกต่างกันในผลลัพธ์กลุ่มนั้น ตัวอย่างเช่น การสร้างภาพเสมือนที่เป็นผู้หญิง ผลลัพธ์ที่ได้จะได้ผู้หญิงในลักษณะที่ต่างกัน โดยใช้ภาพนำเข้าเพียงภาพเดียว คณะวิจัยเสนอการปรับโครงสร้างของโครงข่ายการสร้างเชิงประปักษ์จากเดิมมีเพียงตัวต้นคิดและตัวแบ่งแยก โดยมีการเพิ่มตัวโครงข่ายการแปลง (Mapping Network) ที่เป็นตัวสร้างคุณลักษณะแบบสุ่มขึ้นมาเพื่อให้คุณลักษณะมีความหลากหลาย และตัวเข้ารหัสสไตล์ (Style encoder) ที่ทำหน้าที่สกัดค่าคุณลักษณะของภาพจริงออกมา ผลลัพธ์ที่ได้มีค่าสูงกว่าเมื่อเปรียบเทียบกับโครงข่ายการสร้างเชิงประปักษ์แบบอื่น ภาพที่ 7 แสดงโครงสร้างตัวต้นคิด โครงข่ายการเชื่อมโยง ตัวเข้ารหัสสไตล์ และตัวแบ่งแยก ของสตาร์แกน2 ในงานวิจัย [6]



ภาพที่ 7 ตัวต้นคิด ตัวโครงข่ายการแปลง ตัวเข้ารหัสสไตล์ และตัวแบ่งแยกของสตาร์แกน2 [6]

2.2.5 Image Generation from Sketch Constraint Using Contextual GAN

งานวิจัย [10] เสนอโดย Yongyi L. และคณะ มีเป้าหมายคือการสร้างภาพเสมือนจริงจากภาพร่าง แม้ว่าภาพร่างนั้นจะไม่ได้มาจากศิลปินวาดก็ตาม โดยอาศัยหลักการของเทคนิคการซ่อมแซมภาพ (Image inpainting) เรียกว่า คอนเท็กซ์แกน (ContextGAN) จากภาพที่ 8 เห็นได้ว่าเทคนิคซ่อมแซมภาพจะพยายามเพิ่มเติมส่วนของที่ขาดหายไปของภาพให้มีความใกล้เคียงกับส่วนอื่น งานวิจัยนี้กำหนดภาพร่างเป็นข้อมูลนำเข้าแล้วกำหนดให้ภาพผลลัพธ์เป็นส่วนที่ขาดหายไป และนำไปเรียนรู้และเพิ่มเติมส่วนที่ขาดหายไปจนกลายเป็นภาพที่ใกล้เคียงกับผลลัพธ์ที่ต้องการ

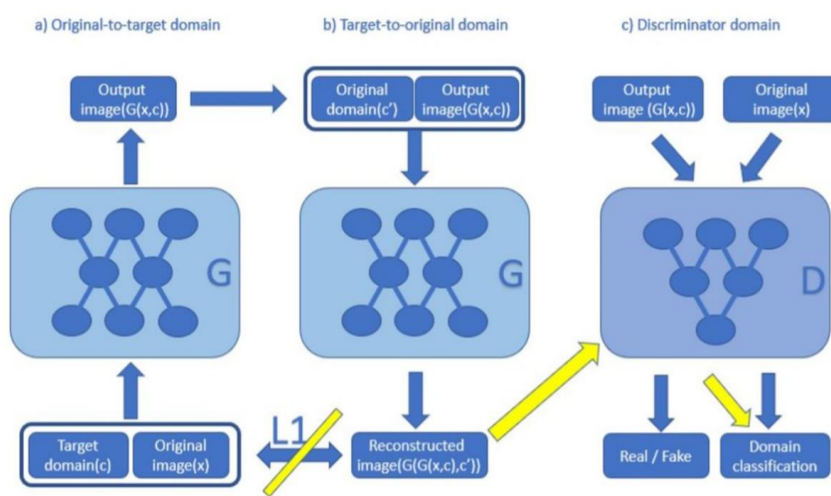


ภาพที่ 8 โครงสร้างของเทคนิคซ่อมแซมภาพและคอนเท็กซ์แกน [10]

2.2.6 GD-StarGAN: Multi-domain image-to-image translation in garment design

งานวิจัย [11] เสนอโดย Yangyun S. และคณะ ได้ประยุกต์เทคนิคสตาร์แกนโดยปรับเปลี่ยนโครงสร้างของตัวต้นคิดจากเดิมใช้โครงข่ายคอนโวลูชันที่คงเหลือเป็นสถาปัตยกรรมยูเน็ต และทดลองกับรูปภาพเสื้อผ้าเพื่อช่วยประหยัดเวลาและค่าใช้จ่ายของนักออกแบบเสื้อผ้าในการ

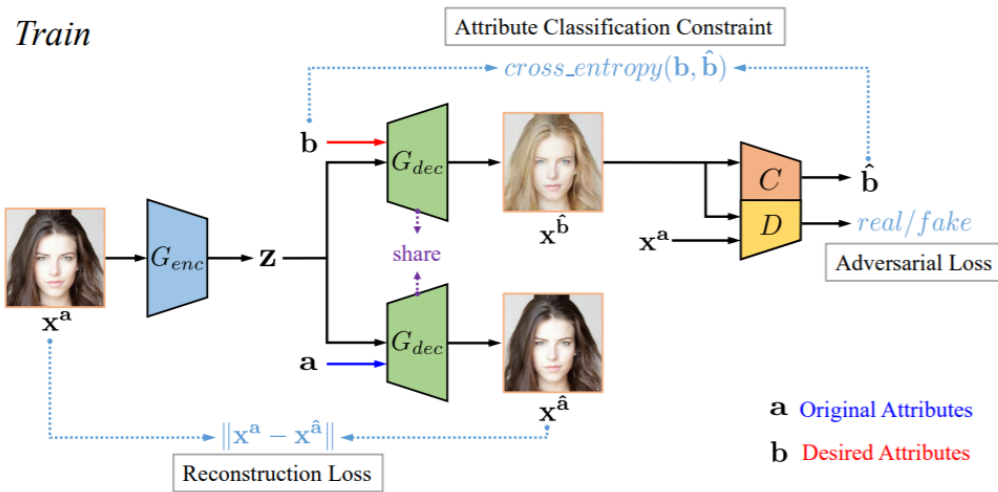
ออกแบบเสื้อผ้าให้มีความหลากหลาย นอกจากนี้ยังมีการปรับเปลี่ยนวิธีการคำนวณค่าความสูญเสียของเทคนิคสตาร์แกนเพื่อให้แบบจำลองเหมาะสมกับการสร้างภาพของเสื้อผ้าได้ดียิ่งขึ้น ดังแสดงในภาพที่ 9 ซึ่งมีการละเว้นการหาค่าความสูญเสียของรูปแบบระหว่างภาพต้นแบบและภาพเสมือนที่สร้างมาจากภาพเสมือนและเพิ่มการหาค่าความสูญเสียของตัวแบ่งแยกโดยใช้เปรียบเทียบระหว่างภาพเป้าหมายและภาพเสมือนที่สร้างมาจากภาพเสมือน โดยงานวิจัยนี้ให้ประสิทธิภาพในการสร้างภาพเสื้อผ้าที่ดีกว่าเมื่อเปรียบเทียบกับเทคนิคสตาร์แกน



ภาพที่ 9 โครงสร้างของเทคนิคจีดีสตาร์แกน [11]

2.2.7 AttGAN: Facial Attribute Editing by Only Changing What You Want

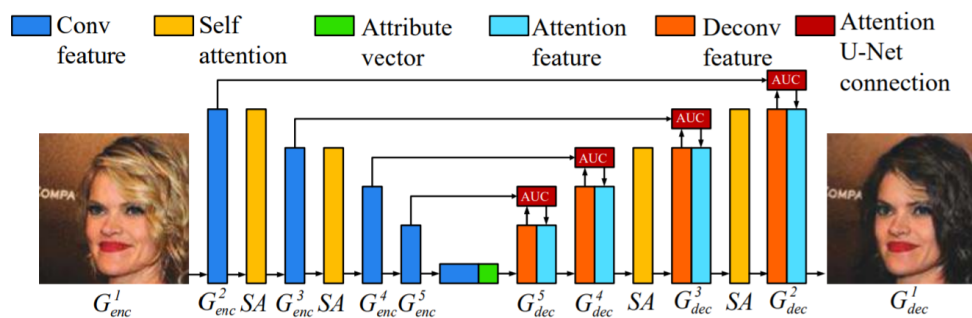
Zhenliang H. และคณะ [12] เสนอเทคนิคที่ชื่อว่าเอทีทีแกน ซึ่งเป็นเทคนิคที่ประยุกต์ใช้แกนสีในการสร้างภาพใบหน้าที่มีคุณลักษณะที่หลากหลาย มีการใช้สถาปัตยกรรมตัวเข้ารหัสและตัวถอดรหัสในตัวต้นคิดโดยมีการนำข้อมูลของคุณลักษณะใส่เข้าไปในส่วนของการถอดรหัสทำให้ภาพที่ได้มีคุณลักษณะตามที่ต้องการ โดยมีโครงสร้างดังภาพที่ 10 เอทีทีแกนมีการวัดค่าความสูญเสียทั้งหมดสามค่า ประกอบด้วยค่าความสูญเสียประปักษ์โดยเปรียบเทียบความน่าจะเป็นระหว่างภาพเสมือนและภาพเป้าหมายที่ได้จากการคำนวณของตัวแบ่งแยก ค่าความสูญเสียของรูปแบบโดยเปรียบเทียบระหว่างภาพต้นแบบและภาพเสมือนที่สร้างจากตัวต้นคิด และค่าความสูญเสียของการจำแนกประเภทซึ่งได้มาจากการจำแนกประเภทของภาพเสมือนที่มีคุณลักษณะที่ต้องการ



ภาพที่ 10 โครงสร้างของเทคนิคเอทีทีแกน [12]

2.2.8 MU-GAN: Facial Attribute Editing based on Multi-attention Mechanism

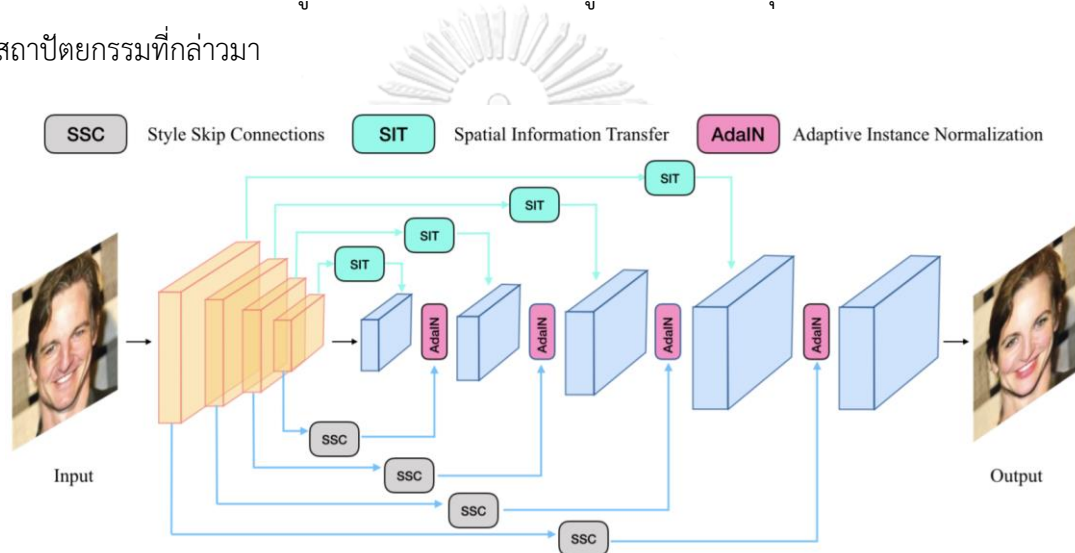
เทคนิคเอ็มยูแกนส์เป็นเทคนิคการสร้างภาพเสมือนใบหน้าเสนอโดย Ke Z. และคณะ [13] มีวัตถุประสงค์หลักอยู่สองอย่างคือการเปลี่ยนโดเมนของภาพต้นแบบและเปลี่ยนคุณลักษณะของภาพต้นแบบให้เหมือนคุณลักษณะของภาพเป้าหมายโดยเค้าโครงเดิมของใบหน้ายังคงอยู่ โดยมีการพัฒนาสถาปัตยกรรมของตัวต้นคิดจากเดิมใช้โครงข่ายคอนโวลูชันแบบเข้ารหัสและถอดรหัสมาเป็นสถาปัตยกรรมยูเน็ต นอกจากนี้ยังปรับปรุงในส่วนของการเชื่อมต่อแบบก้าวกระโดดของสถาปัตยกรรมยูเน็ตโดยการแทนที่ด้วยการเชื่อมต่อยูเน็ตแบบใส่ใจ (Attention U-Net connection : AUCs) การเชื่อมต่อยูเน็ตแบบใส่ใจมีการคำนวณค่าสัมประสิทธิ์ที่มีขนาดเท่ากับข้อมูลผลลัพธ์และนำมาต่อท้ายข้อมูลผลลัพธ์ในแต่ละชั้นของการถอดรหัส ดังภาพที่ 11 การประยุกต์ใช้การเชื่อมต่อยูเน็ตแบบใส่ใจทำให้ภาพที่สร้างยังคงคุณลักษณะของภาพต้นแบบได้ ทำให้ภาพผลลัพธ์มีคุณภาพ สามารถคงใบหน้าเดิมขณะที่เปลี่ยนแปลงคุณลักษณะอื่นๆบนใบหน้า



ภาพที่ 11 โครงสร้างของเทคนิคเอ็มยูแกนส์ [13]

2.2.9 SSCGAN: Facial Attribute Editing via Style Skip Connections

Wenqing C. และคณะ [14] ได้เสนอวิธีการสร้างภาพเหมือนใบหน้าบุคคลโดยใช้แกนส์ที่มีโครงสร้างของสถาปัตยกรรมยูเน็ต โดยเพิ่มส่วนการส่งต่อข้อมูลตามพื้นที่ (Spatial Information Transfer: SIT) เพื่อส่งต่อคุณลักษณะของภาพต้นแบบให้ปรากฏในภาพที่ผลลัพธ์ โดยเพิ่มการส่งต่อข้อมูลตามพื้นที่ในทุก ๆ ชั้นของตัวถอดรหัส และเปรียบเทียบคุณภาพของภาพเหมือนระหว่างสถาปัตยกรรมยูเน็ต สถาปัตยกรรมยูเน็ตที่ใช้การส่งต่อข้อมูลตามพื้นที่แทนการเชื่อมต่อแบบก้าวกระโดดและสถาปัตยกรรมยูเน็ตที่เพิ่มการส่งต่อข้อมูลตามพื้นที่ ตามโครงสร้างที่แสดงใน ภาพที่ 12 ผลลัพธ์คือสถาปัตยกรรมยูเน็ตที่เพิ่มการส่งต่อข้อมูลตามพื้นที่ให้คุณภาพของภาพที่ดีกว่าสองสถาปัตยกรรมที่กล่าวมา



ภาพที่ 12 โครงสร้างของเทคนิคเอสเอสซีแกนส์ [14]

บทที่ 3

แนวคิดและวิธีวิจัย

เทคนิคสตาร์แกน2 เดิมใช้งานกับข้อมูลนำเข้าที่เป็นภาพสีโดยมีจุดประสงค์เพื่อเปลี่ยนแปลงคุณลักษณะบนในหน้าเช่น เปลี่ยนสีผม เปลี่ยนการแต่งหน้า เป็นต้น และจากปัญหาของงานวิจัย ผู้วิจัยได้นำเอาเทคนิคสตาร์แกน2 มาประยุกต์ใช้โดยเปลี่ยนข้อมูลนำเข้าจากภาพสีเป็นภาพร่างและสร้างภาพผลลัพธ์ให้มีคุณลักษณะตามที่ต้องการ และเห็นว่าคุณภาพของภาพผลลัพธ์ที่ได้ยังไม่ดีมาก จากนั้นผู้วิจัยได้ศึกษางานวิจัยที่เกี่ยวข้องและสรุปผลแสดงในตารางที่ 1

ตาราง 1 เปรียบเทียบงานวิจัยที่ประยุกต์ใช้แกนส์กับการสังเคราะห์ภาพกับงานวิจัยนี้

Ref. Propose	Propose Method	G/ D Architecture	Measure
[3] Image to Image translation and Application	cGAN	U-Net/ PacthGAN	AMT, IS
[4] Image to Image translation and Application	CycleGAN	CNN/ PacthGAN	AMT
[5] Sketch to multiclass Image translation	N/A	SkinnyResNet on both	FID
[6] Facial attribute editing in different visual domains	StarGan -> StarGan2	Encoder-decoder on both	FID
[10] Sketch to Image translation	Contextual GAN	CNN on both	SSIM
[11] Image to Image translation	GD StarGan	U-Net/CNN	IS
[12] Facial attribute editing different visual domains	AttGan	Encoder-Decoder / CNN	Attribute editing accuracy
[13] Facial attribute editing in different visual domains	MU-GAN	U-Net with AUCs	Attribute editing accuracy
[14] Facial attribute editing in different visual domains	SSCGAN	U-Net with SIT	FID
[Thesis] Facial attribute editing in different visual domain from sketch	StarGan2- Unet	U-Net/ Residual convolution block	FID

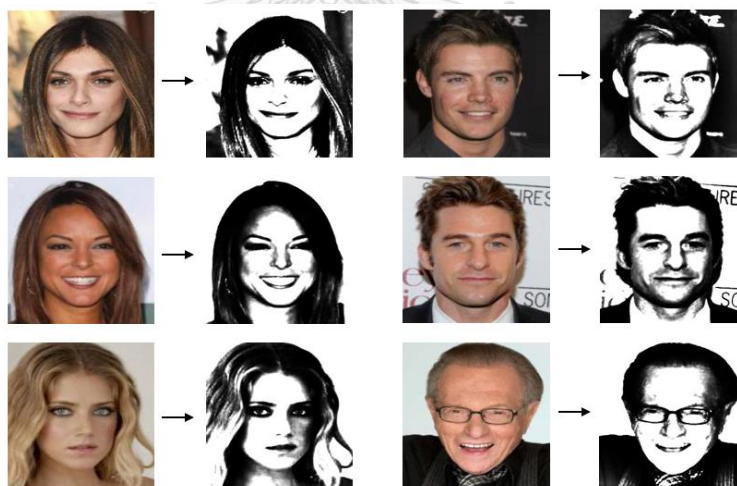
จากตารางที่ 1 พบว่า สถาปัตยกรรมยูเน็ตเป็นเทคนิคนิยมนำมาประยุกต์ใช้กับโครงข่ายการสร้างเชิงปรบักษ์เพราะว่าภายในของสถาปัตยกรรมยูเน็ตมีการเชื่อมโยงแบบก้าวกระโดดระหว่างส่วนเข้ารหัสและส่วนถอดรหัสทำให้คุณลักษณะของภาพต้นแบบยังคงปรากฏในภาพผลลัพธ์ที่ได้ ด้วยเหตุนี้งานวิจัยจึงประยุกต์ใช้สถาปัตยกรรมยูเน็ตในส่วนของตัวต้นคิดของสตาร์แกน2 วิธีการสำหรับระเบียบวิธีวิจัยที่น่าเสนอ ประกอบด้วยขั้นตอนหลัก ๆ อธิบายได้ดังนี้

1. จัดเตรียมภาพร่างของบุคคลจากฐานข้อมูลจากฐานข้อมูลเซเลบเอ (CelebA) จำนวน 350 รูป โดยจัดลักษณะตามเพศ โดยแบ่งได้ดังนี้

1.1 เพศหญิง ประกอบด้วย ทรงผม $\in \{ \text{ผมตรง, ผมหยัก} \}$ การแต่งหน้า $\in \{ \text{แต่งหน้า, ไม่แต่งหน้า} \}$ และ แว่นตา $\in \{ \text{ใส่แว่น, ไม่ใส่แว่น} \}$

1.2 เพศชาย ประกอบด้วย ทรงผม $\in \{ \text{มีผม, ไม่มีผม} \}$ ขนบนในหน้า $\in \{ \text{มีหนวดและเครา, ไม่มีหนวดและเครา} \}$ และ แว่นตา $\in \{ \text{ใส่แว่น, ไม่ใส่แว่น} \}$

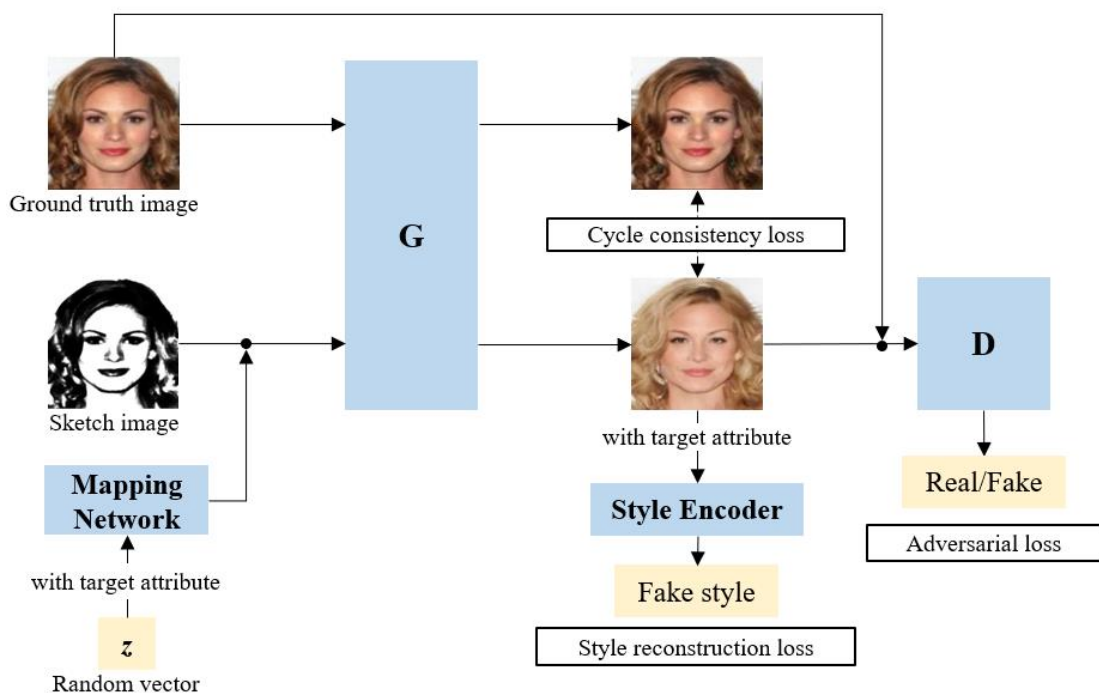
จัดเตรียมภาพร่างใช้คลังโปรแกรมเอ็กซ์ดีโอจี (XDoG) ซึ่งเป็นคลังโปรแกรมสำหรับการเปลี่ยนภาพสีเป็นภาพร่าง โดยตรวจหาขอบของวัตถุในภาพและสร้างภาพร่างตามขอบของวัตถุ ยกตัวอย่างภาพร่างที่ได้ ดังภาพที่ 13



ภาพที่ 13 ภาพร่างที่แปลงโดยคลังโปรแกรมเอ็กซ์ดีโอจี

2. แบ่งข้อมูลรูปภาพใบหน้าออกเป็น 3 ชุด ประกอบด้วย ชุดข้อมูลสอน (training set) ชุดข้อมูลตรวจสอบความสมเหตุสมผล (validation set) และชุดข้อมูลทดสอบ (test set) ในอัตราส่วน 80 ต่อ 10 ต่อ 10 ตามลำดับ

- 2.1 ใช้ชุดข้อมูลสอนสำหรับการเรียนรู้ของแบบจำลองด้วยเทคนิคที่นำเสนอ และปรับค่าพารามิเตอร์โดยใช้ค่าความสูญเสียของแบบจำลอง
- 2.2 ใช้ชุดข้อมูลตรวจสอบความสมเหตุสมผล เพื่อวัดสมรรถนะของแบบจำลอง กำหนดค่าพารามิเตอร์และจำนวนรอบที่ใช้ในการเรียนรู้
- 2.3 ทดสอบประเมินผลสมรรถนะแบบจำลองที่ได้ด้วยชุดข้อมูลทดสอบ
3. จัดเตรียมสภาพแวดล้อมและเครื่องมือที่ใช้ในการดำเนินการของแบบจำลอง
4. ออกแบบโครงสร้างงานวิจัยโดยมีโครงสร้างพื้นฐานมาจากเทคนิคสตาตาร์แกน2 ดังภาพที่ 14 ซึ่งแตกต่างกับเทคนิคสตาตาร์แกน2 คือกำหนดให้ภาพนำเข้าเป็นภาพร่างร่วมกับคุณลักษณะที่ต้องการผ่านตัวต้นคิดเพื่อให้ได้ภาพผลลัพธ์ที่เป็นภาพสีและมีคุณลักษณะที่ต้องการ จากนั้นนำภาพผลลัพธ์ที่ได้ไปคำนวณความน่าจะเป็นของภาพผลลัพธ์ว่ามีค่าใกล้เคียงกับภาพต้นแบบหรือไม่โดยผ่านตัวแบ่งแยก



ภาพที่ 14 ภาพรวมของงานวิจัย

5. ออกแบบวิธีการคำนวณค่าความสูญเสียของแบบจำลองของตัวต้นคิดและตัวแบ่งแยก โดยโครงสร้างพื้นฐานมาจากเทคนิคสตาตาร์แกน2

5.1 ค่าความสูญเสียของตัวต้นคิดมาจากผลรวมของค่า 3 ค่า ประกอบด้วย

1) ความสูญเสียปรปักษ์ (Adversarial loss) คำนวณจากค่าความคลาดเคลื่อน (Binary Cross Entropy) ระหว่างความน่าจะเป็นของภาพผลลัพธ์ที่ได้จากตัวแบ่งแยก $P(x')$ และความน่าจะเป็นของภาพเป้าหมาย โดยค่าความน่าจะเป็นของภาพเป้าหมายจะกำหนดให้เป็น 1 เพราะคาดหวังให้ภาพผลลัพธ์มีความใกล้เคียงกับภาพเป้าหมายมากที่สุด สามารถคำนวณได้จากสมการที่ 1

$$\text{Generator adversarial loss} = -\log(P(x')) \quad (1)$$

2) ค่าความสูญเสียการหมุนเวียน (Cycle consistency loss) คำนวณจากค่าเฉลี่ยของอาร์เรย์ที่ได้มาจากความแตกต่างระหว่างภาพผลลัพธ์ที่สร้างจากตัวต้นคิด (x'_y) และภาพต้นแบบในคุณลักษณะที่ต้องการ (x_y) ดังสมการที่ 2

$$\text{Cycle consistency loss} = \overline{|x'_y - x_y|} \quad (2)$$

3) ค่าความสูญเสียรูปแบบ (Style reconstruction loss) คำนวณจากค่าเฉลี่ยของอาร์เรย์ที่ได้มาจากความแตกต่างระหว่างรูปแบบที่สร้างจากตัวโครงข่ายการแปลง (S_y) กับรูปแบบของคุณลักษณะที่ต้องการ (S_t) ดังสมการที่ 3

$$\text{Style reconstruction loss} = \overline{|S_y - S_t|} \quad (3)$$

5.2 ค่าความสูญเสียของตัวแบ่งแยก คำนวณมาจากค่าความคลาดเคลื่อนระหว่างความน่าจะเป็นของภาพผลลัพธ์ที่ได้จากตัวแบ่งแยก $P(x')$ และความน่าจะเป็นของภาพเป้าหมาย โดยกำหนดค่าความน่าจะเป็นของภาพเป้าหมายเป็น 0 เพราะว่าตัวแบ่งแยกเรียนรู้เพื่อแยกแยะภาพผลลัพธ์และภาพเป้าหมายโดยกำหนดให้เป้าหมายของภาพผลลัพธ์เป็น 0 สามารถคำนวณได้จากสมการที่ 4

$$\text{Discriminator adversarial loss} = -\log(1 - P(x')) \quad (4)$$

6. พัฒนาแบบจำลองโดยขยายความสามารถของโครงข่ายการสร้างเชิงปรปักษ์ เพื่อให้สามารถสร้างภาพบุคคลจากภาพร่างที่เตรียมไว้ โดยประยุกต์ใช้เทคนิคสตาร์แกน2 เพื่อสร้างภาพเสมือนที่มีความหลากหลายของคุณลักษณะที่ระบบรองรับ

7. กำหนดตัววัดผลและวิธีประเมินผลของงานวิจัย โดยตัววัดที่ใช้ในงานวิจัยคือค่าเอฟโอดี แบ่งวิธีการวัดเป็น 2 วิธี ประกอบด้วย

7.1 ค่าเอฟไอดีแอบแฝง (FID latent) วัดความคล้ายระหว่างภาพเป้าหมายกับภาพผลลัพธ์ที่สร้างมาจากภาพร่าง โดยทั้งคู่มีคุณลักษณะที่เหมือนกัน ซึ่งแสดงให้เห็นถึงความสามารถของแบบจำลองในการสร้างภาพเสมือนที่มีคุณลักษณะเดิม การวัดค่าเอฟไอดีแอบแฝงแต่ละครั้งมาจากค่าเฉลี่ยของผลรวมของค่าเอฟไอดีทั้งหมด ตามตารางที่ 2

7.2 ค่าเอฟไอดีอ้างอิง (FID reference) วัดจากภาพเป้าหมายกับภาพผลลัพธ์ที่สร้างมาจากภาพร่าง โดยทั้งคู่มีคุณลักษณะที่แตกต่างกัน ซึ่งแสดงให้เห็นถึงความสามารถของแบบจำลองในการสร้างภาพผลลัพธ์ที่มีคุณลักษณะที่แตกต่างจากเดิม การวัดค่าเอฟไอดีอ้างอิงมาจากการหาค่าเฉลี่ยของผลรวมของค่าเอฟไอดีทั้งหมด ตามตารางที่ 3



บทที่ 4

การออกแบบและพัฒนาแบบจำลอง

4.1 โครงสร้างเน็ตเวิร์กของแบบจำลอง

องค์ประกอบของโครงข่ายการสร้างเชิงประปักษ์ประกอบด้วย ตัวต้นคิด ตัวแบ่งแยก ตัวโครงข่ายการแปลง (Mapping network) และตัวเข้ารหัสรูปแบบ (Style encoder) งานวิจัยนี้ได้ปรับปรุงตัวต้นคิดจากเดิมใช้สถาปัตยกรรมโครงข่ายคอนโวลูชันที่คงเหลือเป็นสถาปัตยกรรมยูเน็ต ซึ่งแบ่งออกเป็นสองส่วนได้แก่ 1) ส่วนเข้ารหัส ประกอบด้วยชั้นคอนโวลูชันจำนวน 8 ชั้น โดยชั้นที่ 2 ถึง 8 มีการใช้ฟังก์ชันการกระตุ้น lrelu (Leaky Rectified Linear Activation) ก่อนเข้าสู่ชั้นคอนโวลูชัน 2) ส่วนถอดรหัส ประกอบด้วยชั้นคอนโวลูชันแบบสลับจำนวน 8 ชั้น โดยชั้นที่ 1 ถึง 7 มีการใช้ฟังก์ชันการกระตุ้น lrelu ก่อนเข้าสู่ชั้นคอนโวลูชันแบบสลับ ซึ่งผลลัพธ์ของแต่ละชั้นจะมาจากการรวมผลลัพธ์ของชั้นคอนโวลูชันแบบสลับกับชั้นคอนโวลูชันจากส่วนเข้ารหัสชั้นจากชั้นล่างสุดตามลำดับ ดังแสดงในตารางที่ 4

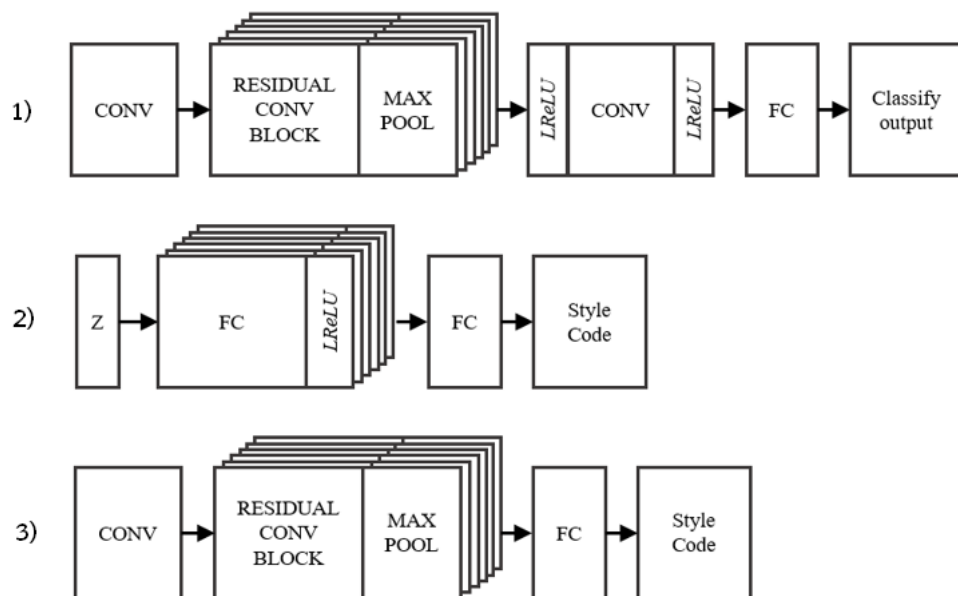
ตาราง 4 โครงสร้างสถาปัตยกรรมยูเน็ตของตัวต้นคิด

Generator – Encoder	Generator – Decoder
Conv (64, 3, 1)	<i>lrelu</i> , DeConv (512, 3, 1) + Encoder layer 7
<i>lrelu</i> , Conv (128, 4, 2)	<i>lrelu</i> , DeConv (512, 3, 1) + Encoder layer 6
<i>lrelu</i> , Conv (256, 4, 2)	<i>lrelu</i> , DeConv (512, 3, 1) + Encoder layer 5
<i>lrelu</i> , Conv (512, 4, 2)	<i>lrelu</i> , DeConv (512, 3, 1) + Encoder layer 4
<i>lrelu</i> , Conv (512, 3, 1)	<i>lrelu</i> , DeConv (512, 4, 2) + Encoder layer 3
<i>lrelu</i> , Conv (512, 3, 1)	<i>lrelu</i> , DeConv (256, 4, 2) + Encoder layer 2
<i>lrelu</i> , Conv (512, 3, 1)	<i>lrelu</i> , DeConv (128, 4, 2) + Encoder layer 1
<i>lrelu</i> , Conv (512, 3, 1)	DeConv (64, 3, 1)

โครงสร้างเน็ตเวิร์กของตัวแบ่งแยก ตัวโครงข่ายการแปลง และตัวเข้ารหัสรูปแบบ อ้างอิงจากเทคนิคสตาร์แกน2 ดังภาพที่ 15 ตัวแบ่งแยกประกอบด้วย 4 ส่วนที่ทำงานต่อเนื่องกันตามลำดับ ได้แก่ 1) ชั้นคอนโวลูชัน 1 ชั้น 2) สถาปัตยกรรมโครงข่ายคอนโวลูชันที่คงเหลือ 6 ชั้น ภายในประกอบด้วยชั้นคอนโวลูชัน 2 ชั้นโดยมีฟังก์ชันการกระตุ้น lrelu คั่นระหว่างชั้นทั้งสอง จากนั้นมีการลดขนาดของข้อมูลลง (Down sampling) โดยใช้ตัวกรองขนาด 3×3 3) ชั้นนี้มีการใช้ฟังก์ชันการ

กระตุ้น lrelu จากนั้นส่งข้อมูลผ่านชั้นคอนโวลูชัน 1 ชั้นและใช้ฟังก์ชันการกระตุ้น lrelu กับผลลัพธ์ที่ได้อีกครั้ง 4) เป็นชั้นรวบรวมการเชื่อมต่ออย่างเต็มรูปแบบ (Fully Connected Layer) และได้ผลลัพธ์สุดท้ายเป็นความน่าจะเป็นของภาพนำเข้า

ตัวโครงข่ายการแปลงและตัวเข้ารหัสรูปแบบมีจุดประสงค์เดียวกันคือสร้างรูปแบบของคุณลักษณะที่ต้องการ แต่แตกต่างกันตรงที่ตัวโครงข่ายการแปลงจะรับข้อมูลนำเข้าเป็นค่าที่ได้จากการสุ่มร่วมกับฉลากของคุณลักษณะที่ต้องการ ส่วนตัวเข้ารหัสรูปแบบจะรับข้อมูลนำเข้าเป็นรูปภาพและฉลากของคุณลักษณะของรูปภาพนั้นและแปลงออกมาเป็นรูปแบบที่แสดงถึงคุณลักษณะที่ต้องการ ตัวโครงข่ายการแปลงประกอบด้วยชั้นรวบรวมการเชื่อมต่ออย่างเต็มรูปแบบจำนวน 7 ชั้น โดยชั้นที่ 1 ถึง 6 มีการใช้ฟังก์ชันการกระตุ้น lrelu ตัวเข้ารหัสรูปแบบประกอบด้วย 3 ส่วน ได้แก่ 1) ชั้นคอนโวลูชัน 1 ชั้น 2) สถาปัตยกรรมโครงข่ายคอนโวลูชันที่คงเหลือ 6 ชั้น ภายในประกอบด้วย ชั้นคอนโวลูชัน 2 ชั้นโดยมีฟังก์ชันการกระตุ้น lrelu คั่นระหว่างชั้นทั้งสอง จากนั้นมีการลดขนาดของข้อมูลลง (Down sampling) โดยใช้ตัวกรองขนาด 3×3 3) ชั้นรวบรวมการเชื่อมต่ออย่างเต็มรูปแบบ



ภาพที่ 15 โครงสร้างเน็ตเวิร์กของ 1) ตัวแบ่งแยก 2) ตัวโครงข่ายการแปลง 3) ตัวเข้ารหัสรูปแบบ

4.2 สภาพแวดล้อมและเครื่องมือที่ใช้ในการพัฒนา

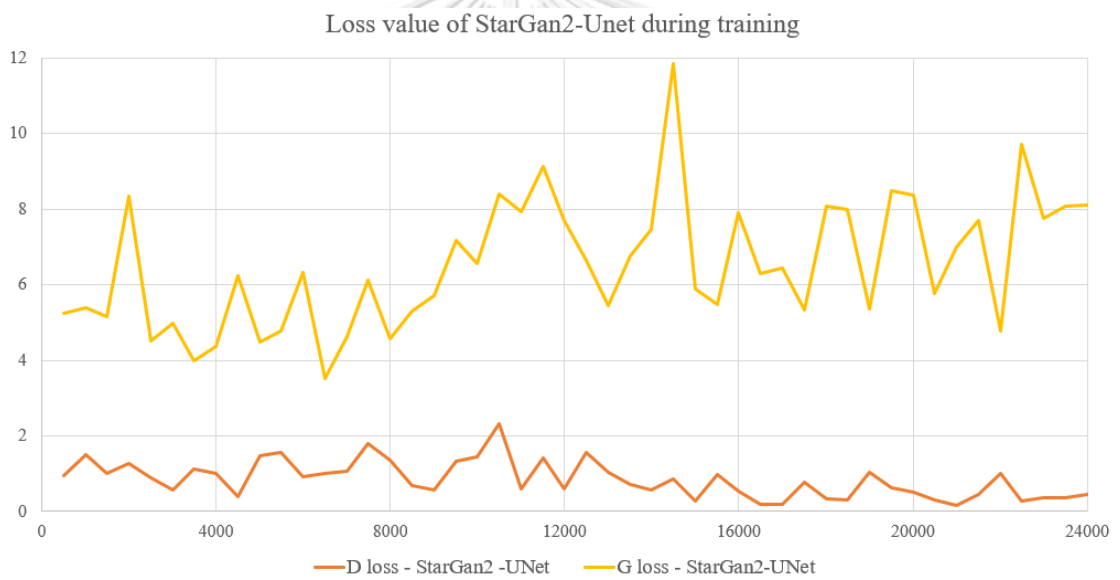
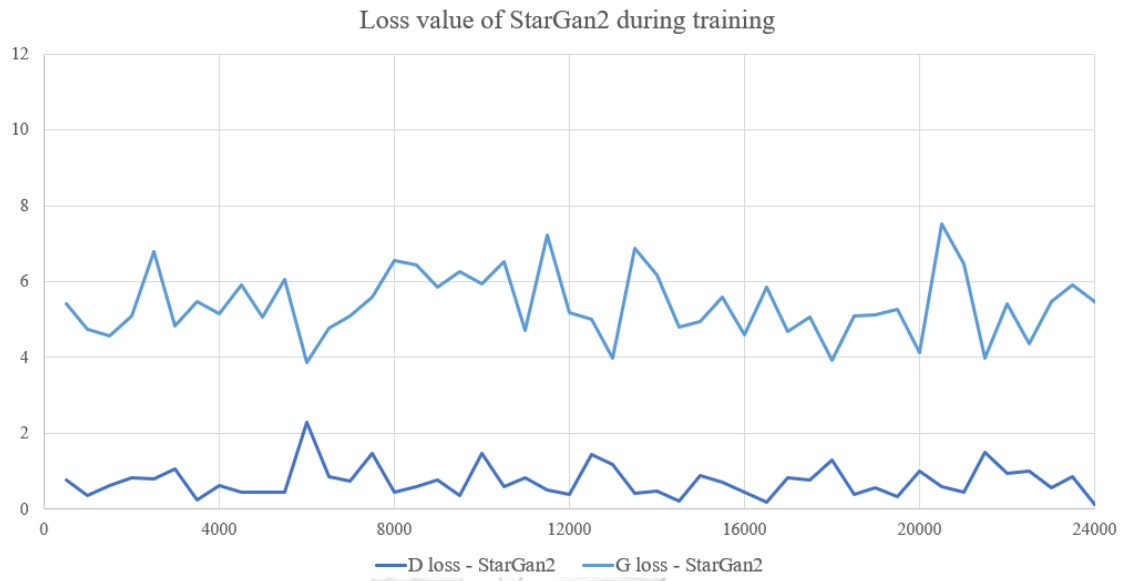
งานวิจัยดำเนินการบนกูเกิลคลาวด์ (Google cloud) ผ่านโฮสต์โปรแกรม โคลแล็บ (Colab) มีหน่วยประมวลผลภาพ (Graphics Processing unit) 1 ตัว รุ่น Tesla P100-PCIE ขนาด 16

กิกะไบต์ และดำเนินการแปลรหัสโดยภาษาไพทอน (Python) เวอร์ชัน 2.0 ประกอบกับคลังโปรแกรมไพทอน (Pytorch library)

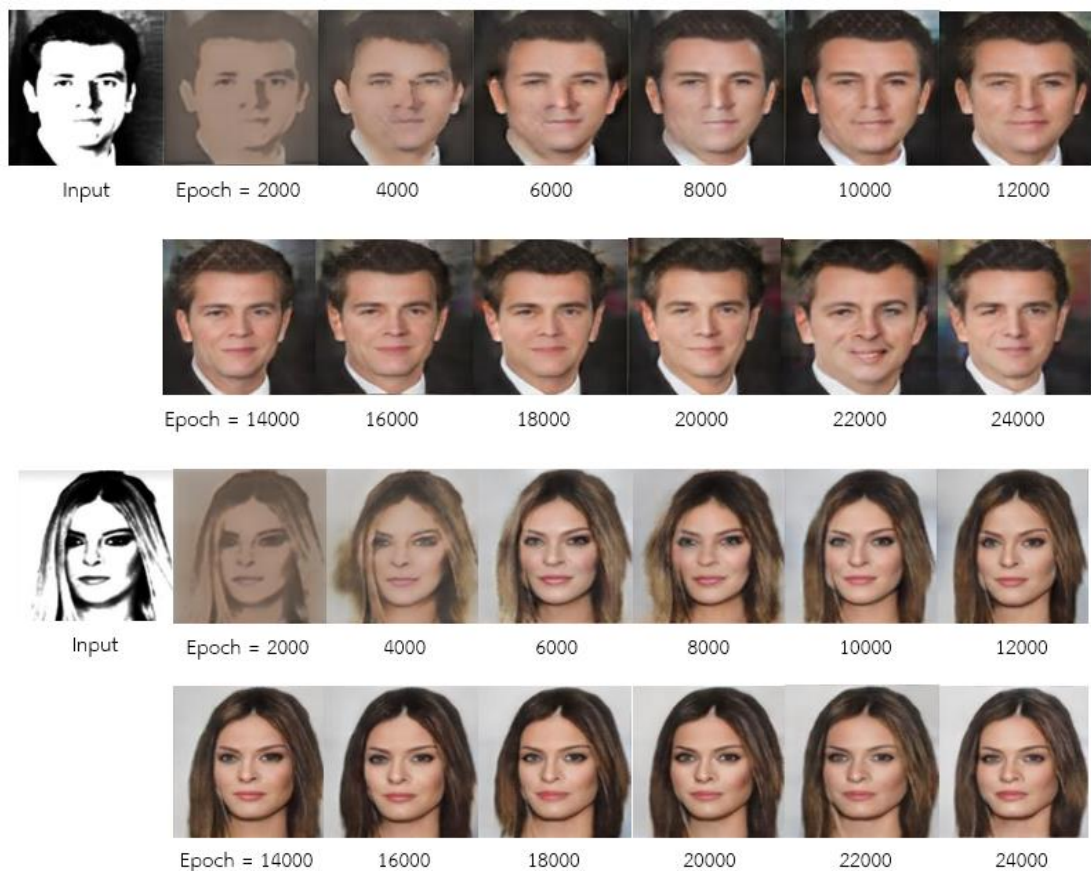
4.3 ขั้นตอนการเรียนรู้แบบจำลอง

ในแต่ละรอบของการเรียนรู้แบบจำลอง เริ่มจาก ตัวต้นคิดรับข้อมูลนำเข้าเป็นเวกเตอร์ของภาพร่างขนาด 128×128 พิกเซล ร่วมกับรูปแบบของคุณลักษณะที่ต้องการที่สร้างมาจากตัวโครงข่ายการแปลง ผลลัพธ์ที่ได้จากตัวต้นคิดคือภาพเสมือนที่เป็นภาพสีและมีคุณลักษณะตามที่กำหนด และส่งต่อไปยังตัวแบ่งแยกเพื่อหาความน่าจะเป็นของภาพเสมือนและคำนวณค่าความสูญเสียประจักษ์โดยเปรียบเทียบกับภาพเป้าหมาย เพื่อให้การเรียนรู้ของตัวต้นคิดมีประสิทธิภาพมากยิ่งขึ้นจึงมีการเรียนรู้ของตัวต้นคิดโดยรับภาพข้อมูลนำเข้าที่เป็นภาพเป้าหมายและได้ภาพเสมือนที่สร้างมาจากภาพเป้าหมาย บนสมมติฐานว่าภาพเสมือนที่สร้างจากภาพร่างและภาพเป้าหมายจะต้องมีความใกล้เคียงกัน จึงมีการวัดค่าความสูญเสียการหมุนเวียนจากภาพทั้งสองด้วย เพื่อให้การเรียนรู้สร้างรูปแบบที่แสดงถึงคุณลักษณะที่กำหนดให้ดียิ่งขึ้น จึงมีการเรียนรู้ของตัวเข้ารหัสรูปแบบและหาค่าความสูญเสียของรูปแบบ ซึ่งคำนวณจากรูปแบบของคุณลักษณะที่สร้างจากภาพเสมือนที่ได้จากตัวต้นคิดร่วมกับคุณลักษณะที่กำหนด เปรียบเทียบกับรูปแบบของคุณลักษณะที่สร้างมาจากภาพเป้าหมายจริง

งานวิจัยนี้ได้ทำการฝึกทั้งสองแบบจำลอง (สตาร์แกน2 และสตาร์แกน2ยูเน็ต) ทั้งหมด 24000 รอบ ระหว่างการเรียนรู้มีการเก็บค่าความสูญเสียของตัวต้นคิดและตัวแบ่งแยกโดยแสดงให้เห็นในทุก ๆ 500 รอบ ดังภาพที่ 16 โดยทฤษฎีของแกนส์ค่าความสูญเสียของตัวต้นคิดและตัวแบ่งแยกจะมีความสัมพันธ์กันเชิงตรงกันข้าม จากภาพที่ 16 ค่าความสูญเสียของตัวต้นคิดของเทคนิคสตาร์แกน2ยูเน็ตมีค่ามากขึ้นและค่าความสูญเสียของตัวแบ่งแยกมีค่าน้อยลง เป็นเพราะว่าความสามารถของตัวต้นคิดในการเรียนรู้เริ่มคงที่หลังจากรอบที่ 10000 เป็นต้นไป และทำให้ค่าความสูญเสียมีค่าเพิ่มมากขึ้น นอกจากนี้ ยังส่งผลให้ความสามารถในการจำแนกภาพเสมือนของตัวแบ่งแยกเพิ่มขึ้นอีกด้วย อย่างไรก็ตาม ค่าความสูญเสียของตัวต้นคิดเพิ่มขึ้นไม่มาก ผู้วิจัยจึงยังคงฝึกแบบจำลองต่อไปและใช้การตรวจสอบความสมเหตุสมผล (train-validation) เป็นปัจจัยต่อไปในการตัดสินใจเลือกแบบจำลองที่ดีที่สุด แต่ละรอบของการเรียนรู้แบบจำลองมีการบันทึกรูปภาพที่สร้างเอาไว้ สำหรับตรวจสอบคุณภาพของรูปภาพและตรวจสอบการเปลี่ยนแปลงในแต่ละรอบ เพื่อประกอบการตัดสินใจในการเลือกแบบจำลองที่ดีที่สุด ตัวอย่างการเปลี่ยนแปลงของรูปภาพที่ถูกสร้างในทุก ๆ 2000 รอบของแบบจำลองแสดง ดังภาพที่ 17



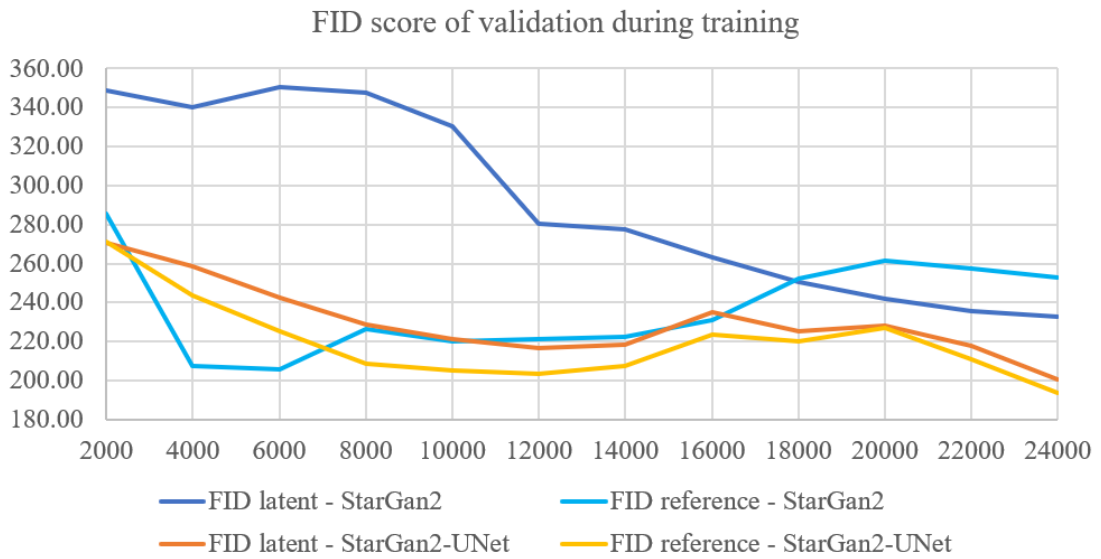
ภาพที่ 16 ค่าความสูญเสียระหว่างการเรียนรู้ของแบบจำลอง



ภาพที่ 17 ตัวอย่างรูปภาพระหว่างการเรียนรู้ของแบบจำลอง

4.3 ขั้นตอนการตรวจสอบความสมเหตุสมผล

แบบจำลองมีการตรวจสอบความสมเหตุสมผลในทุก ๆ 2000 รอบ โดยวัดค่าเอฟไอดี 2 แบบ ได้แก่ 1) ค่าเอฟไอดีแอบแฝง (FID - latent) วัดความคล้ายของภาพเป้าหมายกับภาพเสมือนที่สร้างมาจากภาพร่างที่มีคุณลักษณะเหมือนภาพเป้าหมาย ซึ่งจะแสดงให้เห็นถึงความสามารถในการสร้างภาพเสมือนจากภาพร่าง 2) ค่าเอฟไอดีของการอ้างอิง (FID reference) วัดจากภาพเป้าหมายกับภาพเสมือนที่สร้างมาจากภาพร่างที่มีคุณลักษณะที่แตกต่างจากภาพเป้าหมาย ซึ่งจะแสดงให้เห็นถึงความสามารถในการสร้างภาพเสมือนและเปลี่ยนแปลงภาพเสมือนนั้นให้อยู่ในคุณลักษณะตามที่กำหนด



ภาพที่ 18 ค่าเอฟไอดีของการตรวจสอบความสมเหตุสมผล

4.4 ปัญหาที่พบของการเรียนรู้แบบจำลองและวิธีการแก้ไข

1. ปัญหาที่พบบ่อยระหว่างการเรียนรู้แบบจำลองเกิดขึ้นเมื่อทดสอบแบบจำลองและพบว่าแบบจำลองไม่สามารถคงคุณลักษณะของภาพต้นแบบให้ปรากฏในภาพผลลัพธ์ได้ เนื่องจากเดิมข้อมูลการเรียนรู้ถูกแบ่งกลุ่มตามคุณลักษณะอย่างชัดเจน โดยมีการปะปนภาพของคุณลักษณะอื่นเข้ามาเพียงเล็กน้อยหรือไม่มีเลย จึงทำให้ภาพตัวอย่างในการเรียนรู้ไม่ครอบคลุมในทุกคุณลักษณะที่ต้องการ ยกตัวอย่างเช่นการเตรียมภาพของกลุ่มใส่แว่นตาแล้วข้อมูลชุดนี้ไม่มีคนที่ไม่มีแว่นตาที่ใส่แว่นตา จึงทำให้ตอนเรียนรู้และทดสอบแบบจำลองไม่สามารถสร้างภาพคนจากเดิมมีแว่นตาให้ใส่แว่นตาได้ ผู้จัดทำได้แก้ไขโดยการจัดกลุ่มภาพที่ใช้ในการทดสอบใหม่ โดยกำหนดให้ทุกกลุ่มของคุณลักษณะจะต้องมีภาพที่ประกอบด้วยคุณลักษณะอื่น ๆ ด้วย

2. ปัญหาสภาพแวดล้อมที่ใช้ประมวลผลของการเรียนรู้แบบจำลอง เนื่องจากการเรียนรู้ของแกนส์เป็นการเรียนรู้ที่ใช้เวลานานหากใช้หน่วยประมวลผลกลาง (Central Processing Unit: CPU) จึงต้องปรับเปลี่ยนวิธีการเขียนโปรแกรมเพื่อประยุกต์ใช้กับหน่วยประมวลผลภาพ (Graphic Processing Unit: GPU) และต้องเป็นโปรแกรมที่สามารถทำงานแบบขนานได้ ผู้จัดทำจึงได้ปรับปรุงโปรแกรมและอาศัยการประมวลผลบนคลาวด์ เพื่อลดระยะเวลาในการเรียนรู้แบบจำลอง

บทที่ 5

การวิเคราะห์และประเมินผล

งานวิจัยนี้ใช้ตัววัดเอฟไอดีซึ่งเป็นตัววัดที่บอกถึงความเหมือนของภาพ 2 ภาพที่นำมาเปรียบเทียบ ซึ่งค่ายิ่งน้อยแสดงว่าภาพทั้งสองมีความใกล้เคียงกันมาก จากค่าเอฟไอดีที่แสดงในภาพที่ 12 ผู้วิจัยจึงเลือกแบบจำลองที่มีค่าเอฟไอดีต่ำที่สุดจากรอบที่เรียนรู้ทั้งหมด ได้แก่ รอบที่ 14000 กับ 24000 ของเทคนิคสตาร์แกน2 และสตาร์แกน2ยูเน็ต ตามลำดับ มาเป็นแบบจำลองที่ดีที่สุดของทั้ง 2 เทคนิค งานวิจัยนี้ได้้นำข้อมูลชุดทดสอบมาวัดผลกับแบบจำลองที่ดีที่สุดของทั้ง 2 เทคนิค และได้ค่าเอฟไอดี สรุปดังแสดงในตารางที่ 5

ตาราง 5 ค่าเอฟไอดีเปรียบเทียบระหว่างเทคนิคสตาร์แกน2 กับเทคนิคสตาร์แกน2ยูเน็ต

Model	FID Latent	FID reference
StarGan2	219.04	181.91
StarGan2-Unet	118.81	119.93

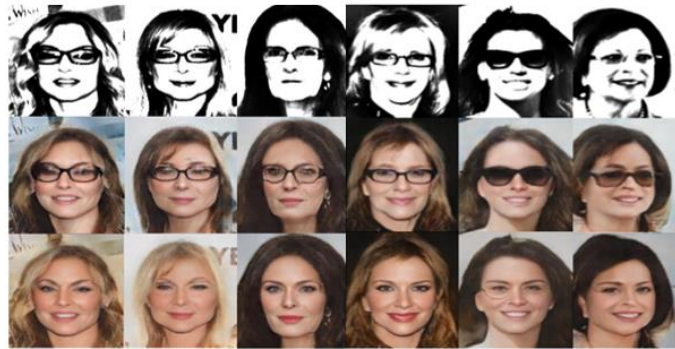
ภาพที่ 19 แสดงผลลัพธ์ที่ได้จากการทดสอบแบบจำลอง โดยแสดงให้เห็นเป็นคู่ของคุณลักษณะแบบสลับคู่ (toggle) ของภาพร่างต้นแบบเพศหญิง ประกอบด้วย

+ ใส่แว่นตา → ไม่ใส่แว่นตา, ไม่ใส่แว่นตา → ใส่แว่นตา;
 แต่งหน้า → ไม่แต่งหน้า, ไม่แต่งหน้า → แต่งหน้า;
 ผมลอน → ผมตรง, ผมตรง → ผมลอน

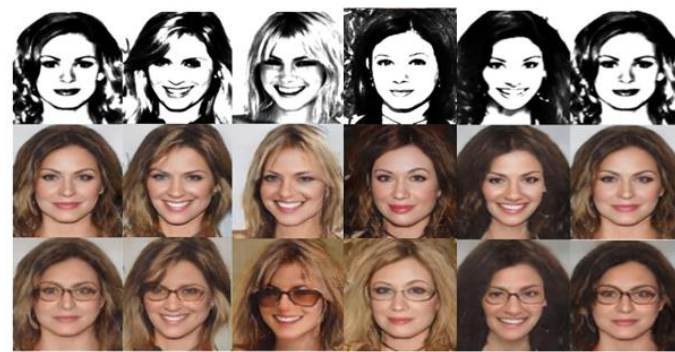
ภาพที่ 20 แสดงผลลัพธ์ที่ได้จากการทดสอบแบบจำลอง โดยแสดงให้เห็นเป็นคู่ของคุณลักษณะแบบสลับคู่ (toggle) ของภาพร่างต้นแบบเพศชาย ประกอบด้วย

+ ใส่แว่นตา → ไม่ใส่แว่นตา, ไม่ใส่แว่นตา → ใส่แว่นตา;
 มีผม → ไม่มีผม, ไม่มีผม → มีผม;
 มีหนวดเครา → ไม่มีหนวดเครา, ไม่มีหนวดเครา → มีหนวดเครา

ผลลัพธ์การสังเคราะห์ภาพเสมือนจากภาพร่างเดียวกันในทุกคุณลักษณะแบ่งตามเพศหญิงและเพศชายแสดงดังภาพที่ 21 และภาพที่ 22 ตามลำดับ



Glasses → No Glasses



No Glasses → Glasses



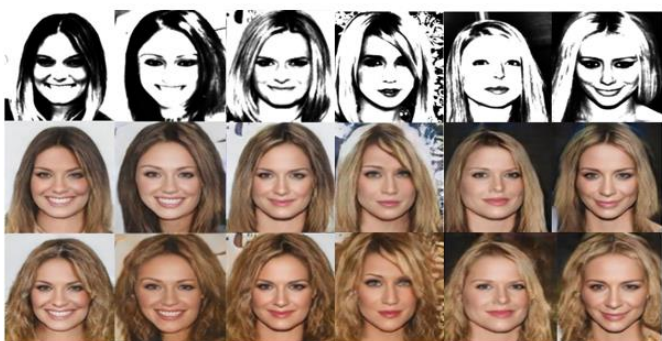
Makeup → No Makeup



No Makeup → Makeup

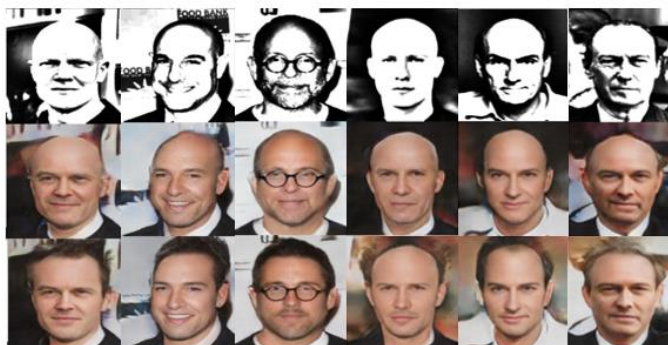


Wavy hair → Straight hair



Straight hair → Wavy hair

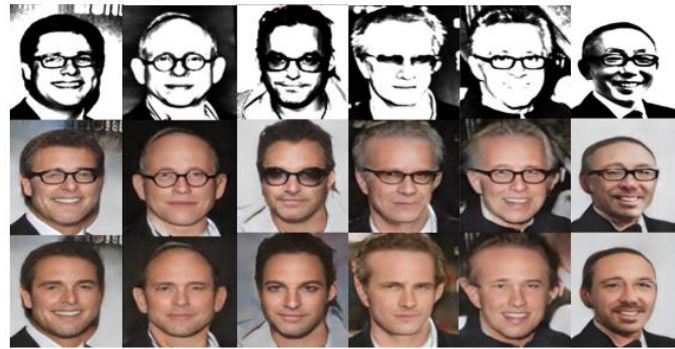
ภาพที่ 19 ภาพเสมือนที่ได้จากการทดสอบโดยแสดงในแต่ละคู่ของคุณลักษณะของเพศหญิง



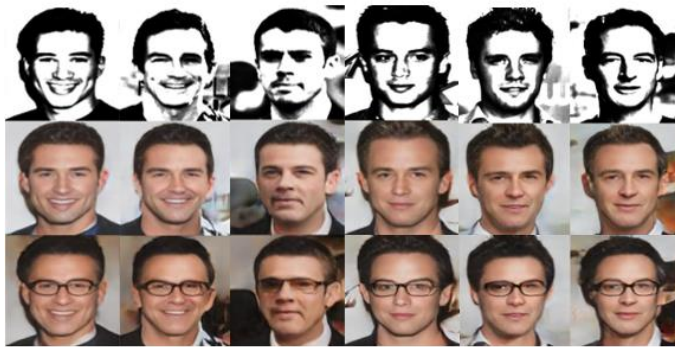
Bald → Not Bald



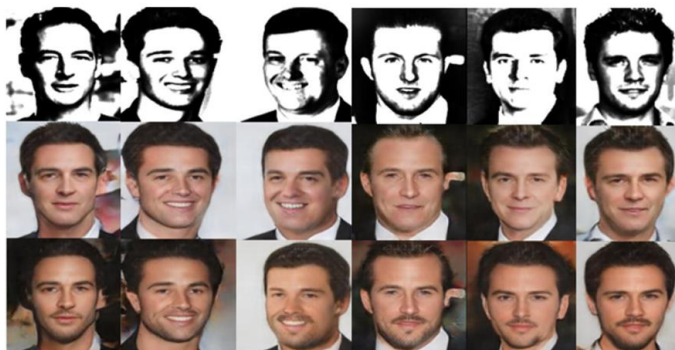
Not Bald → Bald



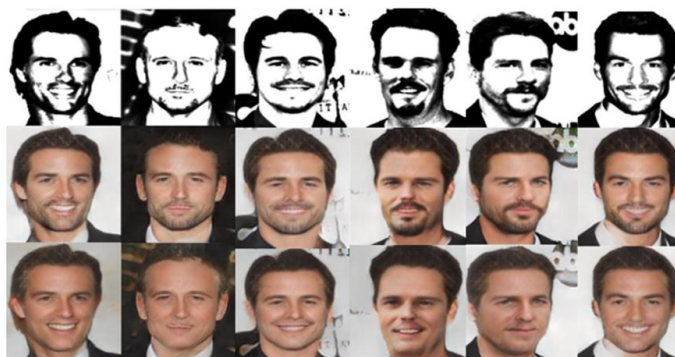
Glasses → No Glasses



No Glasses → Glasses



No Beard and Mustache → Beard and Mustache

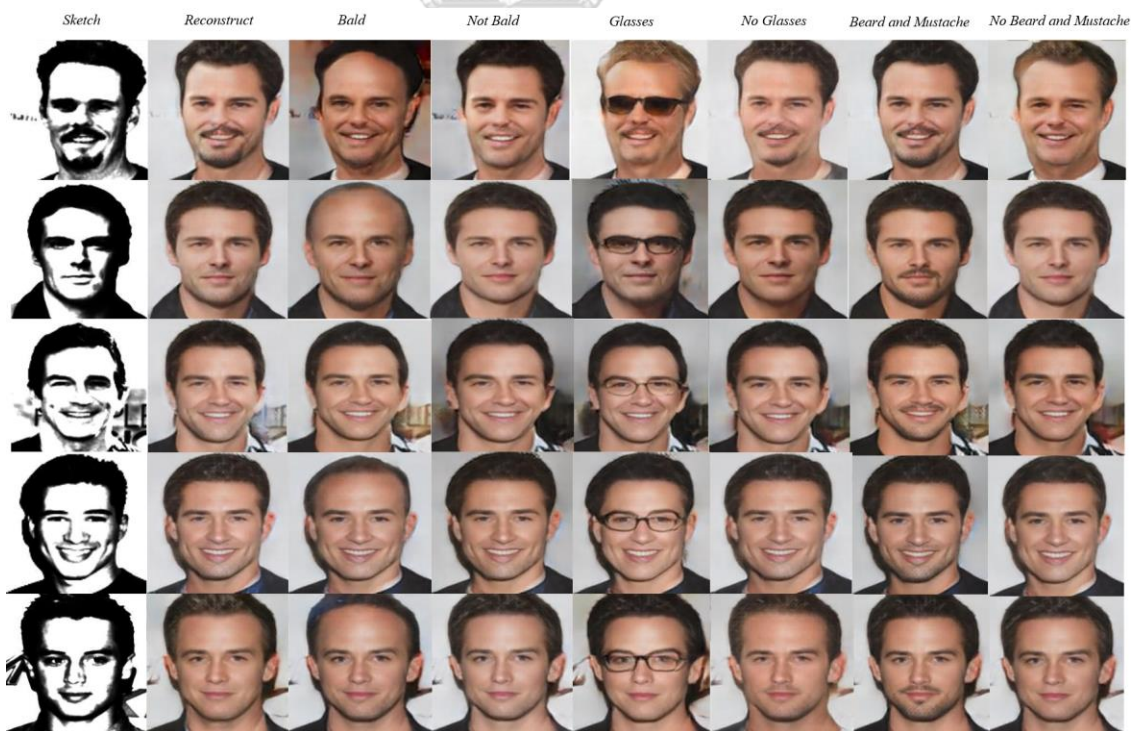


Beard and Mustache → No Beard and Mustache

ภาพที่ 20 ภาพเสมือนที่ได้จากการทดสอบโดยแสดงในแต่ละคู่ของคุณลักษณะของเพศชาย



ภาพที่ 21 ผลลัพธ์การสังเคราะห์ภาพเสมือนจากภาพร่างเดียวกันในทุกคุณลักษณะแบ่งตามเพศหญิง



ภาพที่ 22 ผลลัพธ์การสังเคราะห์ภาพเสมือนจากภาพร่างเดียวกันในทุกคุณลักษณะแบ่งตามเพศชาย

บทที่ 6

สรุปผลการวิจัย

6.1 สรุปผลการวิจัย

วิทยานิพนธ์ฉบับนี้ได้นำเสนอเทคนิคการสร้างภาพเสมือนที่มีคุณลักษณะบนใบหน้าที่หลากหลายของบุคคลจากภาพร่าง เพื่อนำไปใช้ประโยชน์ในการสืบหาตัวบุคคลหรือคนร้าย ซึ่งสามารถสร้างภาพเสมือนที่มีคุณลักษณะทั้งหมด 5 คู่ของคุณลักษณะโดยแบ่งตามเพศ ได้แก่ 1) เพศหญิง ประกอบด้วย ทรงผม \in { ผมตรง, ผมหยัก } การแต่งหน้า \in { แต่งหน้า, ไม่แต่งหน้า } และ แว่นตา \in { ใส่แว่น, ไม่ใส่แว่น } 2) เพศชาย ประกอบด้วย ทรงผม \in { มีผม, ไม่มีผม } ขนบนใบหน้า \in { มีหนวดเครา, ไม่มีหนวดเครา } และ แว่นตา \in { ใส่แว่น, ไม่ใส่แว่น } งานวิจัยนี้ได้นำเอาเทคนิคสตาร์แกน2 จากเดิมที่มีจุดประสงค์ในการเปลี่ยนแปลงคุณลักษณะบนใบหน้าของภาพสี มาประยุกต์ใช้กับภาพร่างและเปลี่ยนแปลงโครงสร้างของตัวต้นคิดของสตาร์แกน2 เพื่อเพิ่มประสิทธิภาพของการสร้างภาพเสมือนจากภาพร่างให้ดียิ่งขึ้น โดยมีการเรียนรู้และตรวจสอบความสมเหตุสมผลควบคู่กันและเมื่อครบกำหนดรอบที่เรียนรู้ ผู้จัดทำได้เลือกแบบจำลองในรอบที่มีค่าเอฟไอดีทีที่ต่ำสุดที่ได้จากการตรวจสอบความสมเหตุสมผลเพื่อนำแบบจำลองในรอบนั้นมาวัดผล การวัดผลของงานวิจัยนี้จะเปรียบเทียบระหว่างเทคนิคพื้นฐานคือเทคนิคสตาร์แกนส์2 และเทคนิคที่ได้นำเสนอคือเทคนิคสตาร์แกน2ยูเน็ต เทคนิคที่นำเสนอให้ผลลัพธ์ที่ดีกว่าเทคนิคพื้นฐานโดยวัดจากค่าเอฟไอดีทีถึงร้อยละ 40 นั้นแสดงให้เห็นถึงความสามารถของการปรับปรุงแบบจำลองที่ดีขึ้น นอกจากนั้น ภาพที่ได้จากการสร้างภาพเสมือนมีคุณภาพที่ดีหากตัดสินด้วยตาเปล่า โดยภาพยังคงคุณสมบัติของใบหน้าของภาพร่างและเปลี่ยนแปลงไปตามคุณลักษณะที่ต้องการได้เป็นอย่างดี

6.2 ข้อจำกัดในงานวิจัย

งานวิจัยนี้มีข้อจำกัดของการสร้างภาพเสมือนอยู่ 2 อย่าง ได้แก่ ความหลากหลายคุณลักษณะของใบหน้าที่กำหนดได้ครั้งละ 1 คุณลักษณะเท่านั้น และแบบจำลองผลลัพธ์ไม่สามารถสร้างภาพเสมือนที่มีคุณลักษณะนอกเหนือจากที่เรียนรู้ได้

6.2 งานวิจัยในอนาคต

จากข้อจำกัดของงานวิจัยที่ว่าความหลากหลายของคุณลักษณะของใบหน้าสามารถสร้างได้ครั้งละ 1 คุณลักษณะเท่านั้น จึงสามารถต่อยอดในงานวิจัยในอนาคตได้ โดยสามารถปรับปรุงวิธีการเรียนรู้ของตัวโครงข่ายการแปลงและตัวเข้ารหัสการแปลงเพิ่มให้สามารถสร้างรูปแบบของคุณลักษณะที่แสดงถึงลักษณะพิเศษได้มากกว่า 1 คุณลักษณะ เพื่อเพิ่มสมรรถนะของความหลากหลายให้มากยิ่งขึ้น



บรรณานุกรม



จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

- [1] Ian J. G., Jean P., Mehdi M., Bing X., David W., Sherjil O., Aaron C. and Yoshua B. (2014, 1 June 2014). “Generative Adversarial Networks.”. Retrieved 12 Dec 2019, from <https://ui.adsabs.harvard.edu/abs/2014arXiv1406.2661G>.
- [2] Alec R., Luke M. and Soumith C. (2015, 1 Nov 2015). “Unsupervised Representation Learning With Deep Convolutional Generative Adversarial Networks.”. Retrieved 12 Dec 2019, from <https://ui.adsabs.harvard.edu/abs/2015arXiv151106434R>.
- [3] Phillip I., Jun-Yan Z., Tinghui Z. and Alexei A. E. (2016, 1 Nov 2016). “Image-to-Image Translation with Conditional Adversarial Networks.”. Retrieved 12 Dec 2019, from <https://ui.adsabs.harvard.edu/abs/2016arXiv161107004I>.
- [4] Jun-Yan Z., Taesung P., Phillip I., and Alexei A. E. (2018, 15 Nov 2018). “Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks.”. Retrieved 02 Feb 2020, from <https://arxiv.org/abs/1703.10593>.
- [5] Jason B. (2019, 17 June. 2019). “A Gentle Introduction to Generative Adversarial Networks (GANs).” . Retrieved 12 Dec 2019, from <https://machinelearningmastery.com/what-are-generative-adversarial-networks-gans/>.
- [6] Arnab G., Richard Z., Puneet K. D., Oliver W., Alexei A. E., Philip H. S. T. and Eli S. (2019, 15 Sep 2019). “Interactive Sketch & Fill: Multiclass Sketch-to-Image Translation.”. Retrieved 02 Feb 2020, from <https://arxiv.org/abs/1909.11081>.
- [7] Yongyi L., Shangzhe W., Yu-Wing T. and Chi-Keung T. (2018, 26 Jul 2018). “Image Generation from Sketch Constraint Using Contextual GAN.”. Retrieved 02 Feb 2020, from <https://arxiv.org/abs/1711.08972>.
- [8] Yunjey C., Youngjung U., Jaejun Y. and Jung-Woo H. (2019, 4 Dec 2019) “StarGAN v2: Diverse Image Synthesis for Multiple Domains.”. Retrieved 12 Dec 2019, from <https://arxiv.org/abs/1912.01865>.
- [9] Olaf R., Philipp F. and Thomas B. (2015, 18 May 2018). “U-Net: Convolutional Networks for Biomedical Image Segmentation.”. 11 Mar 2020, from <https://arxiv.org/pdf/1505.04597.pdf>.

- [10] Cornelisse D. (2018). “An intuitive guide to Convolutional Neural Networks.”. Retrieved 12 Oct 2020, from <https://medium.freecodecamp.org/an-intuitive-guide-to-convolutional-neural-networks-260c2de0a050>.
- [11] Yangyun S., Runnan H, Wenkai H. (2020). “GD-StarGAN: Multi-domain image- to-image translation in garment design.”. Retrieved 30 Aug 2020, from <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0231719>.
- [12] Zhenliang H., Wangmeng Z., Meina K., Shiguang S. and Xilin C. (2018). “AttGAN: Facial Attribute Editing by Only Changing What You Want.”. Retrieved 30 Aug 2020, from <https://arxiv.org/pdf/1711.10678.pdf>.
- [13] Ke Z., Yukun S., Xiwang G., Liang Q. and Zhenbing Z. (2020). MU-GAN: Facial Attribute Editing based on Multi-attention Mechanism. Retrieved 30 Aug 2020, from <http://www.ieee-jas.org/article/id/62788e30-59dd-43b8-a670-ba9b91bfe68f?pageType=en>.
- [14] Wengqing C., Ying T., Chengjie W., Jilin L., Feiyue H. and Rongrong J. (2020). SSCGAN: Facial Attribute Editing via Style Skip Connections. European Computer Vision Association. Retrieved 30 Aug 2020, from https://www.ecva.net/papers/eccv_2020/papers_ECCV/papers/123600409.pdf.

ประวัติผู้เขียน

ชื่อ-สกุล	Warintorn Phusomsai
วัน เดือน ปี เกิด	22 June 1994
สถานที่เกิด	Bangkok, Thailand
วุฒิการศึกษา	Bachelor of Science, Khon Kaen University
ที่อยู่ปัจจุบัน	105 M.7 Renenaknon, Nakhonphanom province, Thailand, 48170
ผลงานตีพิมพ์	<p>Phusomsai, W. & Limpiyakorn, Y. (2020). Applying GANs for Generating Image with Varied Facial Attributes from Sketch. In 13th International Conference on Computer and Electrical Engineering. Beijing, China.</p> <p>Phusomsai, W. & Limpiyakorn, Y. (2020). Toward U-Net-based GANs for Diverse Facial Image Synthesis from Sketch. In 13th International Conference on Advanced computer theory and Engineering. Hangzhou, China.</p>