


รายงานโครงงานวิศวกรรมไฟฟ้า วิชา 2102499

การพัฒนารูปภาพความละเอียดสูงยิ่งยวดโดยใช้
แบบจำลองการเพิ่มประสิทธิภาพของรูปภาพ
Image Super Resolution Improvement Using
Image Enhancement Model

นายราชันย์ ปัญญาเกียรติคุณ เลขประจำตัว 6230458121
อาจารย์ที่ปรึกษา รศ. ดร.สุภาวดี อร่ามวิทย์

ภาควิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์
จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2565

ลงชื่ออาจารย์ที่ปรึกษาหลัก  (รศ.ดร. สุภาวดี อร่ามวิทย์) วันที่ 12 พค 2566	ลงชื่ออาจารย์ที่ปรึกษาร่วม (ถ้ามี) _____ (_____) วันที่ _____	ลงชื่อตัวแทนบริษัท (เฉพาะนิสิตใน โปรแกรมความเชื่อมโยง อุตสาหกรรม) _____ (_____) วันที่ _____
---	--	---

บทคัดย่อ

เทคโนโลยีโครงข่ายประสาทเทียมแบบสังวัตนาการเป็นเทคโนโลยีที่ถูกนำไปใช้อย่างหลากหลายกับงานต่าง ๆ ซึ่งหนึ่งในงานเหล่านั้นคือ การสร้างคืนภาพความละเอียดสูงยิ่งยวด และในปัจจุบันเทคโนโลยีโครงข่ายประสาทเทียมแบบสังวัตนาการก็ได้กลายเป็นส่วนหนึ่งของแบบจำลองที่ถูกพัฒนาขึ้นใหม่มากมาย เช่น แบบจำลองเชิงความน่าจะเป็นการแพร่กระจายการลดทอนของสัญญาณรบกวน ซึ่งแบบจำลองนี้เป็นแบบจำลองที่ถูกนำไปประยุกต์ใช้กับงานทางด้านอื่นมากมาย เช่น งานด้านการสร้างภาพหรือวิดีโอ โดยในโครงการนี้ได้ทดลองนำแบบจำลองดังกล่าวมาประยุกต์ใช้กับงานด้านการสร้างคืนภาพความละเอียดสูงยิ่งยวด โดยนำไปใช้กับผลลัพธ์ที่ออกมาจากแบบจำลองการแปลงความสนใจแบบผสมผสานในอัตราส่วนของการขยายรูปภาพแบบ 2 เท่า, 3 เท่า และ 4 เท่า เพื่อทดสอบว่าแบบจำลองนี้สามารถกำจัดสัญญาณรบกวนที่เกิดจากแบบจำลองการแปลงความสนใจแบบผสมผสาน ซึ่งผลลัพธ์ที่ได้พบว่ายังไม่สามารถทำได้ดีพอ เมื่อพิจารณาค่าอัตราส่วนต่อสัญญาณสูงสุด (PSNR) เป็นตัววัดประสิทธิภาพ โดยในชุดข้อมูลทดสอบเซตห้าที่มีอัตราส่วนของการขยายแบบ 2 เท่า ได้ค่าเฉลี่ย PSNR เท่ากับ 19.28 dB เมื่อใช้แบบจำลองเชิงความน่าจะเป็นการแพร่กระจายการลดทอนของสัญญาณรบกวน แต่สำหรับแบบจำลองการแปลงความสนใจแบบผสมผสานได้ค่าเฉลี่ยมากถึง 45.96 dB

คำสำคัญ: การสร้างคืนภาพ, ภาพความละเอียดสูงยิ่งยวด, การเรียนรู้เชิงลึก, การลดสัญญาณรบกวนของภาพ

Abstract

Convolution neural network technology is used in a wide range of applications. One of those jobs is Super-Resolution Image. Convolution neural network technology has become part of many newly developed models, like the Denoising Diffusion Model, which is a model with many applications, including the images and video generation. In this project, the denoising diffusion model has been tried and applied to reconstruct super-resolution images. It was used to the results from the Hybrid Attention Transformer (HAT) model at 2x, 3x, and 4x scales to test whether the model could eliminate the noise generated by the Hybrid Attention Transformer model. From the result, it was found that Denoising Diffusion Model performed poorly compared to Hybrid Attention Transformer. When Peak Signal Noise Ratio (PSNR) was considered a metric and Set 5 2x scale was a test set, the average PSNR using Denoising Diffusion Model was 19.28 dB, but the average PSNR using Hybrid Attention Transformer was 45.96 dB.

Keywords: Image Reconstruction, Super-Resolution Image, Deep Learning, Image Denoising

สารบัญ

บทคัดย่อ.....	ก
Abstract.....	ก
สารบัญ.....	ข
1. บทนำ.....	1
1.1 ที่มาและความสำคัญของโครงการ.....	1
1.2 วัตถุประสงค์ของโครงการ.....	2
1.3 ขอบเขตของโครงการ.....	2
1.4 ผลลัพธ์ที่คาดหวังจากโครงการ.....	2
2. หลักการและทฤษฎีที่เกี่ยวข้อง.....	3
2.1 วิธีการสร้างคุณภาพความละเอียดสูงยิ่งยวด.....	3
2.1.1 วิธีการประมาณค่าในช่วงของรูปภาพ.....	3
2.1.2 แบบจำลองการแปลงความสนใจแบบผสมผสาน.....	4
2.2 แบบจำลองเชิงความน่าจะเป็นการแพร่กระจายการลดทอนของสัญญาณรบกวน.....	5
2.2.1 กระบวนการไปข้างหน้า (Forward Process).....	5
2.2.2 กระบวนการย้อนกลับ (Reverse Process).....	5
2.2.3 ฟังก์ชันสูญเสีย (Loss function).....	6
2.2.4 ระเบียบวิธีของการเรียนรู้และกระบวนการย้อนกลับของแบบจำลอง.....	8
2.2.5 โครงข่ายประสาทเทียมแบบสังวัตนาการ (Convolution Neural Network).....	9
2.2.6 สถาปัตยกรรมโครงข่ายประสาทเทียมแบบยูเน็ต (U-net).....	12
2.3 เครื่องมือวัดประสิทธิภาพของแบบจำลอง.....	13
2.3.1 ค่าอัตราส่วนต่อสัญญาณสูงสุด (Peak Signal to Noise Ratio; PSNR).....	13
2.3.2 ค่าอัตราความคล้ายของโครงสร้าง (SSIM).....	14
2.3.3 อัตราคะแนนความโง่เขลา (Fool Score Rate; FSR).....	15
3. ผลลัพธ์ของโครงการและการอภิปรายผล.....	15

3.1 ชุดข้อมูลสำหรับการเรียนรู้ของแบบจำลอง.....	16
3.2.1 ดีไอวีทูเค (DIV2K).....	17
3.2.2 ฟลิคเกอร์ (Flickr2K).....	17
3.2.3 เซเลบเอ เฮชคิว (CelebA HQ).....	17
3.2 ชุดข้อมูลสำหรับการทดสอบ.....	17
3.2.1 เซตห้า (Set5).....	17
3.2.2 เซตสิบสี่ (Set14).....	17
3.2.3 บีเอสดีเอสหนึ่งร้อย (BSDS100).....	17
3.2.4 เออเบิร์นหนึ่งร้อย (Urban100).....	17
3.3 รายละเอียดการฝึกฝนแบบจำลอง.....	17
3.3.1 แบบจำลองการแปลงแบบผสมผสาน.....	17
3.3.2 แบบจำลองเชิงความน่าจะเป็นการแพร่กระจายการลดทอนของสัญญาณรบกวน.....	18
3.4 ตารางผลลัพธ์.....	18
3.5 รูปภาพตัวอย่างผลลัพธ์.....	20
4. บทสรุป.....	22
4.1 สรุปผลการดำเนินการ.....	22
4.2 ปัญหา อุปสรรค และแนวทางแก้ไข.....	23
4.3 ข้อเสนอแนะ.....	23
5. กิตติกรรมประกาศ.....	24
6. เอกสารอ้างอิง.....	25

1. บทนำ

1.1 ที่มาและความสำคัญของโครงการ

การสร้างคุณภาพความละเอียดสูงยิ่งยวดคือการเพิ่มความละเอียดของภาพจากภาพที่มีความละเอียดต่ำไปเป็นภาพที่มีความละเอียดสูง ประโยชน์ของการสร้างคุณภาพความละเอียดสูงยิ่งยวดสามารถใช้ได้ในระบบเฝ้าระวัง เนื่องจากกล้องวงจรปิดในระบบเฝ้าระวังไม่มีความละเอียดที่สูงมาก จึงทำให้รูปภาพใบหน้าของคนในกล้องของระบบเฝ้าระวังไม่มีความคมชัดหรือมีความละเอียดต่ำ ส่งผลให้ระบุตัวตนได้ยาก การสร้างคุณภาพความละเอียดสูงยิ่งยวดจึงเป็นส่วนสำคัญในการที่ทำให้สามารถระบุตัวตนของบุคคลในภาพจากกล้องวงจรปิดได้ อีกทั้งในงานทางด้านการแพทย์ รูปภาพจากการสร้างคุณภาพด้วยแม่เหล็ก (MRI) ยังมีความละเอียดที่ไม่สูงพอเนื่องจากปัญหาขีดจำกัดของเครื่องสแกน หรือการที่มีสัญญาณรบกวนเกิดขึ้นในกรณีที่เกิดการขยับของผู้ป่วยที่อยู่ในเครื่องสแกน จากปัญหานี้จึงสามารถนำการสร้างคุณภาพความละเอียดสูงยิ่งยวดมาช่วยในการแก้ปัญหาได้ หรือในปัญหาด้านพื้นที่การจัดเก็บข้อมูล จากรูปภาพที่มีความละเอียดสูงทำให้ขนาดของข้อมูลมีค่ามาก ซึ่งส่งผลให้มีค่าใช้จ่ายที่สูงในการจัดเก็บข้อมูล การแก้ปัญหาคือเก็บรูปภาพที่ความละเอียดที่ต่ำลง และเมื่อต้องการนำรูปภาพที่มีความละเอียดต่ำออกมาใช้ จึงสามารถใช้การสร้างคุณภาพความละเอียดสูงยิ่งยวดมาช่วยทำให้ภาพมีความละเอียดที่สูงขึ้นทำให้เหมาะสมแก่การใช้งาน

โครงการนี้มีประเด็นสำคัญคือ ต้องการทดสอบประสิทธิภาพในการสร้างคุณภาพความละเอียดสูงยิ่งยวดจากวรรณกรรม [1] ซึ่งทดสอบประสิทธิภาพของรูปภาพเดิมโดยการนำรูปภาพที่มีความละเอียดสูงยิ่งยวดที่เป็นผลลัพธ์มาจากวรรณกรรม [1] มาทำการประมวลผลอีกครั้งหนึ่ง โดยใช้แบบจำลองเชิงความน่าจะเป็นการแพร่กระจายของการลดทอนของสัญญาณรบกวน [2]

สิ่งที่คาดหวังเมื่อจบโครงการคือ การนำแบบจำลองใหม่มานำมาใช้ในการสร้างคุณภาพความละเอียดสูงยิ่งยวด หลังจากนั้นจะลองเปรียบเทียบคุณภาพของการสร้างคุณภาพความละเอียดสูงยิ่งยวดโดยมีแนวทางการดำเนินงานโดยย่อดังนี้

1. จัดเตรียมชุดข้อมูล และแยกชุดข้อมูลที่จะนำมาใช้ฝึกและใช้ทดสอบแบบจำลอง
2. ทำให้ชุดข้อมูลมีความละเอียดต่ำ
3. นำชุดข้อมูลไปผ่านการสร้างคุณภาพความละเอียดสูงยิ่งยวด
4. สร้างแบบจำลองเชิงความน่าจะเป็นการแพร่กระจายของการลดทอนของสัญญาณรบกวน
5. ทดสอบและเปรียบเทียบประสิทธิภาพที่ได้จากข้อ 4 และรูปจากการสร้างคุณภาพความละเอียดสูงยิ่งยวดโดยวรรณกรรม [1] และรูปต้นฉบับ

1.2 วัตถุประสงค์ของโครงการ

1. เพื่อพัฒนาองค์ความรู้ด้านการสร้างค่านภาพละเอียดสูงยิ่งยวดโดยใช้แบบจำลองการเพิ่มประสิทธิภาพของรูปภาพ
2. เพื่อเปรียบเทียบประสิทธิภาพที่ได้จากแบบจำลองการเพิ่มประสิทธิภาพของรูปภาพ ที่ได้จากวิธีการสร้างค่านภาพความละเอียดสูงยิ่งยวดและรูปต้นฉบับ

1.3 ขอบเขตของโครงการ

1. โครงการนี้ได้ใช้ชุดข้อมูลในการเรียนรู้ 2 ชุด ได้แก่ ดีไอวีทูเคและฟลิคเกอร์ทูเค
2. โครงการนี้จะใช้รูปภาพที่มีความละเอียดต่ำที่ถูกลดขนาดลงมาจากรูปต้นฉบับทั้งสามแบบ ได้แก่ ลดขนาดลงมา 2 เท่า, 3 เท่า และ 4 เท่า
3. โครงการนี้จะใช้ชุดข้อมูลในการทดสอบประสิทธิภาพทั้งหมด 4 ชุด ได้แก่ เซตห้า, เซตสิบสี่, บีเอสดีหนึ่งร้อย, เออร์เบินหนึ่งร้อย

1.4 ผลลัพธ์ที่คาดหวังจากโครงการ

สามารถพัฒนาแบบจำลองการแพร่กระจายของการลดทอนของสัญญาณรบกวนที่สามารถเพิ่มคุณภาพให้กับภาพความละเอียดสูงยิ่งยวดได้

1.5 ขั้นตอนการดำเนินงาน

1. ศึกษาทฤษฎีการสร้างค่านภาพความละเอียดสูงยิ่งยวดโดยใช้การเรียนรู้เชิงลึก (แบบจำลองการแปลงความสนใจแบบผสมผสาน) และแบบจำลองการแพร่กระจายของการลดทอนของสัญญาณรบกวน
2. ศึกษาข้อมูลชุดในการเรียนรู้และชุดข้อมูลที่ใช้ในการทดสอบ
3. ทดสอบแบบจำลองการแปลงความสนใจแบบผสมผสานและวัดประสิทธิภาพของรูปภาพ
4. ทดสอบแบบจำลองการแพร่กระจายของการลดทอนของสัญญาณรบกวนและวัดประสิทธิภาพของรูปภาพ
5. เปรียบเทียบประสิทธิภาพของแบบจำลองและสรุปผล
6. เขียนรายงาน

2. หลักการและทฤษฎีที่เกี่ยวข้อง

เนื่องจากแบบจำลองการเพิ่มประสิทธิภาพของรูปภาพจะต้องใช้องค์ความรู้หลายๆอย่าง เพื่อทำความเข้าใจตั้งแต่การลดคุณภาพของรูปภาพ วิธีการสร้างคืนภาพความละเอียดสูงยิ่งยวด และแบบจำลองในการเพิ่มประสิทธิภาพของรูปภาพซึ่งจะใช้แบบจำลองเชิงความน่าจะเป็นการแพร่กระจายการลดทอนของสัญญาณรบกวน ซึ่งจะมีรายละเอียดดังนี้

2.1 วิธีการสร้างคืนภาพความละเอียดสูงยิ่งยวด

วิธีการสร้างคืนภาพความละเอียดสูงแบบดั้งเดิมจะใช้วิธีการประมาณค่าในช่วงของรูปภาพ โดยวิธีการเป็นวิธีการดั้งเดิมจึงยังไม่มีมีการใช้การเรียนรู้เชิงลึก แต่ในปัจจุบันที่การเรียนรู้เชิงลึกมีความก้าวหน้ามากขึ้น ทำให้การเรียนรู้เชิงลึกถูกนำไปประยุกต์ใช้ในงานด้านต่างๆ โดยการคืนสร้างภาพความละเอียดสูงก็เป็นหนึ่งในงานที่ถูกนำไปใช้ และในโครงงานนี้จะใช้แบบจำลองการแปลงความสนใจแบบผสมผสาน รายละเอียดของวิธีต่างๆจะถูกอธิบายดังนี้

2.1.1 วิธีการการประมาณค่าในช่วงของรูปภาพ

เป็นวิธีการที่ใช้เพื่อเปลี่ยนขนาดของรูปภาพให้มีขนาดใหญ่ขึ้นหรือเล็กลง หรือใช้เพื่อบิดเบือนรูปภาพ ตัวอย่างเช่นการหมุนรูปภาพ หรือการเปลี่ยนมุมมองของรูปภาพ โดยหลักการในการประมาณค่าช่วงในรูปภาพ จะใช้พิกเซล (Pixel) ของรูปภาพที่มีอยู่ ผ่านการประมาณทางคณิตศาสตร์เพื่อให้ได้พิกเซลใหม่ และนำไปวางในรูปภาพใหม่เพื่อให้มีขนาดหรือการบิดเบือนที่ต้องการ โดยทั่วไปแล้วจะแบ่งวิธีการประมาณค่าช่วงออกเป็นสองกลุ่ม ได้แก่ การประมาณค่าช่วงแบบปรับตัวได้และการประมาณค่าช่วงแบบปรับตัวไม่ได้ ซึ่งในโครงงานนี้จะสนใจเพียงแค่การเปลี่ยนขนาดของรูปภาพโดยใช้ขั้นตอนวิธีการประมาณค่าช่วงแบบปรับตัวได้ทั้งหมด 3 วิธี

1. วิธีการประมาณค่าช่วงจากตำแหน่งใกล้สุด (Nearest Neighbor Interpolation)

วิธีการประมาณค่าช่วงจากตำแหน่งใกล้สุดมีหลักการคือจะเลือกจุดพิกเซลที่อยู่ใกล้จุดที่สนใจที่สุด แล้วประมาณจุดที่สนใจด้วยจุดพิกเซลนั้น จากวิธีการนี้จะทำให้เป็นวิธีที่เรียบง่ายและเร็วที่สุด แต่จะมีข้อเสียคือจะได้ภาพที่หยابหรือไม่มีความคมชัด

2. วิธีการประมาณแบบเส้นคู่ (Bilinear Interpolation)

วิธีการประมาณแบบเส้นคู่มีหลักการคือจะเลือก 4 จุดรอบๆ จุดที่สนใจแล้วนำค่าของทั้ง 4 จุดนี้มาเฉลี่ยแบบถ่วงน้ำหนักกับค่าระยะห่างระหว่างจุดที่สนใจกับทั้ง 4 จุดรอบ ๆ โดยจุดที่อยู่ใกล้กับจุดที่สนใจจะมีค่าน้ำหนักที่สูงกว่าจุดที่อยู่ไกลกว่า ซึ่งวิธีการนี้จะได้รูปภาพที่มีความต่อเนื่อง (Smooth) กว่าวิธีการประมาณค่าช่วงจากตำแหน่งใกล้สุด แต่จะสร้างภาพได้ช้ากว่าเนื่องจากการคำนวณมากกว่า

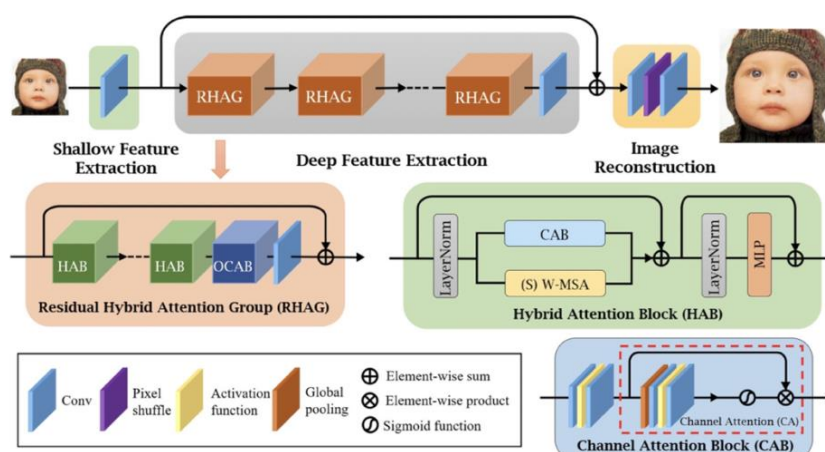
3. วิธีการประมาณค่าแบบประสาณเชิงลูกบาศก์ (Bicubic Interpolation)

วิธีการประมาณค่าแบบประสาณเชิงลูกบาศก์จะใช้หลักการเดียวกันกับการประมาณแบบเส้นคู่ แต่จะเลือกจุดรอบๆ ที่มากกว่านั้นคือเลือกทั้งหมด 16 จุดพิกเซลรอบจุดที่สนใจ แล้วนำค่าของทั้ง 16 จุดมาเฉลี่ยแบบถ่วงน้ำหนักกับค่าระยะห่างระหว่างจุดที่สนใจกับทั้ง 16 จุดรอบๆ โดยจุดที่อยู่ใกล้กับจุดที่สนใจจะมีค่าน้ำหนักที่สูงกว่าจุดที่อยู่ไกลกว่า จากการที่วิธีการนี้ใช้จำนวนพิกเซลมากกว่าวิธีการประมาณแบบเส้นคู่ ดังนั้นจะได้รูปภาพที่มีความต่อเนื่องมากกว่า

2.1.2 แบบจำลองการแปลงความสนใจแบบผสมผสาน (Hybrid Attention Transformer; HAT)

แบบจำลองการแปลงความสนใจแบบผสมผสานเป็นวิธีการที่นำรูปภาพที่มีความละเอียดต่ำมาทำให้มีความละเอียดสูงขึ้น โดยการใช้วิธีการเรียนรู้เชิงลึกเป็นส่วนสำคัญของวิธีการนี้ ซึ่งโครงสร้างของแบบจำลองการแปลงความสนใจแบบผสมผสานแสดงดังรูปที่ 1

จากรูปที่ 1 จะเห็นว่าโครงสร้างนี้แบ่งออกเป็นสามส่วนได้แก่ การสกัดคุณลักษณะตื้น (Shallow feature extraction) การสกัดคุณลักษณะลึก (Deep feature extraction) และ การสร้างคืนรูปภาพ (Image reconstruction) ในส่วนแรก การสกัดคุณลักษณะตื้นจะประกอบไปด้วยชั้นสังวัตนาการ 1 ชั้น เพื่อสกัดคุณลักษณะในส่วนแรกก่อนและช่วยให้การหาค่าที่เหมาะสมที่สุดมีเสถียรภาพ หลังจากนั้นจะทำการสกัดคุณลักษณะเชิงลึก ในส่วนนี้จะประกอบไปด้วย กลุ่มความสนใจแบบผสมผสานตกค้าง (Residual Hybrid Attention Group; RHAG) ทั้งหมด N กลุ่มและจะมีชั้นสังวัตนาการอีก 1 ชั้นที่ปลายของส่วนนี้ เพื่อช่วยในการสะสมข้อมูลของการสกัดคุณลักษณะเชิงลึก หลังจากนั้นจะทำการรวมแบบค่าต่อค่า ระหว่างค่าที่ออกมาจากการสกัดคุณลักษณะตื้นและการสกัดคุณลักษณะลึก เพื่อนำเข้าไปในส่วนสุดท้ายคือการสร้างคืนรูปภาพโดยภาพที่ได้จะเป็นภาพความละเอียดสูงยิ่งยวด



2.2 แบบจำลองเชิงความน่าจะเป็นการแพร่กระจายการลดทอนของสัญญาณรบกวน

แบบจำลองเชิงความน่าจะเป็นการแพร่กระจายการลดทอนของสัญญาณรบกวน (Denoising Diffusion Probabilistic Model) เป็นแบบจำลองที่ถูกพัฒนา มาจากแบบจำลองเชิงความน่าจะเป็นการแพร่กระจาย (Diffusion Probabilistic Model) [4] โดยแบบจำลองนี้ยังมีความสามารถในการสร้างรูปภาพที่มีคุณภาพสูงไม่ด้อย จึงเกิดการพัฒนารูปแบบจำลองเชิงความน่าจะเป็นการแพร่กระจายการลดทอนของสัญญาณรบกวน ภาพรวมของแบบจำลองนี้จะทำการรับรูปภาพที่เป็นข้อมูลขาเข้ามาทำการเพิ่มสัญญาณรบกวนหลังจากนั้นจะให้แบบจำลองทำการเรียนรู้เพื่อที่จะสร้างภาพคืนกลับมา ซึ่งหลักการของแบบจำลองนี้จะแบ่งออกเป็นสองขั้นตอนได้แก่ กระบวนการไปข้างหน้าและกระบวนการย้อนกลับ นอกจากนี้ยังมีรายละเอียดสำคัญได้แก่ การหาฟังก์ชันสูญเสีย เพื่อนำไปใช้ในการหาค่าที่เหมาะสมที่สุดของพารามิเตอร์ โดยจะมีการนำสถาปัตยกรรมโครงข่ายประสาทเทียมแบบยูเน็ต ซึ่งภายในโครงข่ายนี้มีการใช้โครงข่ายประสาทเทียมแบบสังวัตนาการ โดยแต่ละขั้นตอนมีรายละเอียดที่สำคัญดังนี้

2.2.1 กระบวนการไปข้างหน้า (Forward process)

ในขั้นตอนนี้จะมีรูปภาพเป็นข้อมูลขาเข้าของแบบจำลองและทำการเพิ่มสัญญาณรบกวนให้แก่รูปภาพทั้งหมด T ขั้นตอน จนรูปภาพมีการกระจายตัวแบบไอโซโทรปิกเกาส์เซียน (Isotropic gaussian) ดังรูปที่ 2 โดยในแต่ละขั้นตอนที่มีการเพิ่มสัญญาณรบกวน จะมีสมมติฐานที่ว่ากระบวนการนี้เป็นกระบวนการลูกโซ่มาร์คอฟ (Markov chain) และในแต่ละขั้นตอนมีการกระจายตัวแบบเกาส์เซียน ดังในสมการที่ 1

$$q(x_t|x_{t-1}) = \mathcal{N}(x_t; \sqrt{1 - \beta_t}x_{t-1}, \beta_t I) \quad (1)$$

โดย $q(x_t|x_{t-1})$ คือการกระจายตัวของ $x_t|x_{t-1}$ และจากสมการที่ 1 มีความหมายคือการกระจายตัวของ $q(x_t|x_{t-1})$ เป็นการกระจายแบบเกาส์เซียนที่มีค่าเฉลี่ยคือ $\sqrt{1 - \beta_t}x_{t-1}$ และมีเมทริกซ์ของความแปรปรวนคือ $\beta_t I$ โดย β_t เป็นค่าคงที่ที่ถูกกำหนดไว้ซึ่งมีความเกี่ยวข้องกับการเพิ่มสัญญาณรบกวนให้กับรูปภาพ จากสมการที่ 1 จะสามารถเขียนให้อยู่ในรูปสมการที่ 2 ได้ดังนี้

$$x_t = \sqrt{\alpha_t}x_0 + \sqrt{1 - \alpha_t}\epsilon \quad (2)$$

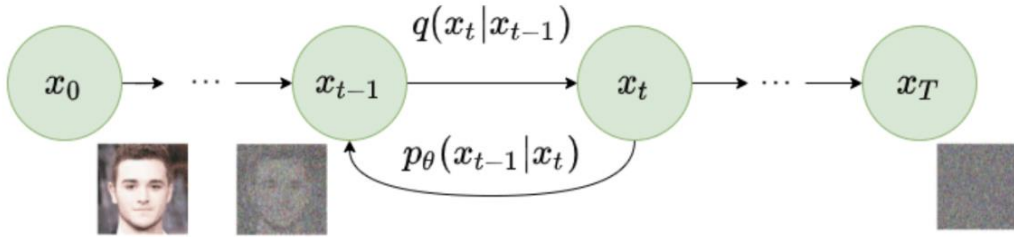
โดยที่

$\epsilon \sim \mathcal{N}(0, I)$ เป็นค่าสัญญาณรบกวนของ x_t

$$\alpha_t = \prod_{s=0}^t 1 - \beta_s$$

จากสมการที่ 2 สามารถสรุปเป็นสมการที่ 3 ได้ดังนี้ โดยสมการที่ 3 นี้จะถูกนำไปใช้เพื่อช่วยในการจัดรูปทางคณิตศาสตร์ต่อในขั้นตอนของการสร้างฟังก์ชันสูญเสีย

$$x_t \sim q(x_t|x_0) = \mathcal{N}(x_t; \sqrt{\bar{\alpha}_t}x_0, (1 - \bar{\alpha}_t)I) \quad (3)$$



รูปที่ 2 แสดงขั้นตอนของกระบวนการไปข้างหน้าและกระบวนการย้อนกลับ [2]

2.2.2 กระบวนการย้อนกลับ (Reverse process)

เมื่อให้จำนวนขั้นตอนการเพิ่มสัญญาณรบกวน มีค่าสูงขึ้นจะทำให้ x_t มีการกระจายตัวที่เข้าใกล้การเป็นไอโซโทโรปิกเกาส์เซียน หลังจากนั้นจะนำ x_t ในขั้นตอนสุดท้ายซึ่งก็คือ x_T ไปผ่านกระบวนการย้อนกลับแสดงในรูปที่ 2 ในกระบวนการนี้จะทำการประมาณค่า $q(x_{t-1}|x_t)$ เนื่องจากการที่เลือกค่า β มีค่าน้อยๆและให้ค่าเพิ่มขึ้นเมื่ออยู่ในขั้นตอนที่สูงขึ้น ทำให้ค่า $q(x_{t-1}|x_t)$ มีการกระจายตัวแบบเกาส์เซียน โดยจะใช้พารามิเตอร์ $p_\theta(x_{t-1}|x_t)$ ที่มีการกระจายตัวแบบเกาส์เซียนแทน $q(x_{t-1}|x_t)$ เพื่อแสดงว่าค่านี้เป็นค่าที่ได้รับมาจากการเรียนรู้ของแบบจำลอง และมีค่าดังสมการที่ 4

$$p_\theta(x_{t-1}|x_t) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \beta_t I) \quad (4)$$

จากสมการที่ 4 จะเห็นว่าค่าเฉลี่ยของ $p_\theta(x_{t-1}|x_t)$ คือ $\mu_\theta(x_t, t)$ ซึ่งมีความหมายว่าค่านี้เกิดจากการที่แบบจำลองเรียนรู้ และพยายามทำนายค่าเฉลี่ยในแต่ละขั้นตอน นอกจากนี้เพื่อความง่ายของแบบจำลอง จะให้แบบจำลองเรียนรู้และทำนายแค่ค่าเฉลี่ย เนื่องจากจะกำหนดให้เมทริกซ์ของความแปรปรวนในแต่ละขั้นตอนในกระบวนการย้อนกลับ มีค่าเท่ากับเมทริกซ์ของความแปรปรวนในกระบวนการไปข้างหน้า จึงสามารถสรุปได้ว่าในขั้นตอนการย้อนกลับจะทำการประมาณค่าเฉลี่ยจากการเรียนรู้ของแบบจำลองเพื่อที่จะได้ค่าการกระจายตัวของรูปภาพ โดยการหาค่า x_{t-1} ดังได้ในตารางที่ 2

ในส่วนต่อไปจะทำการอธิบายเกี่ยวกับฟังก์ชันสูญเสียซึ่งในท้ายที่สุดแล้วจากการประมาณค่าเฉลี่ยในกระบวนการย้อนกลับจะเปลี่ยนเป็นปัญหาการทำนายสัญญาณรบกวนของรูปภาพที่เวลานั้นแทน

2.2.3 ฟังก์ชันการสูญเสีย (Loss function)

เนื่องจากแบบจำลองนี้เป็นการประมาณค่าการกระจายตัวของข้อมูลจึงเลือกใช้ค่าลบของลอการิทึมของภาวะน่าจะเป็น (Negative log likelihood) ดังสมการที่ 5 แต่จากการหาค่า $p_\theta(x_0)$ เป็นกระบวนการที่ใช้เวลาคำนวณมาก จึงใช้ทฤษฎีขอบเขตการเปลี่ยนแปลง (Variational bound) ทำให้ได้สมการที่ 6

$$L = \mathbb{E}[-\log p_\theta(x_0)] \quad (5)$$

$$L = \mathbb{E}_q[D_{KL}(q(x_t|x_{t-1}) \parallel p(x_T)) + \sum_{t>1} D_{KL}(q(x_{t-1}|x_t) \parallel p_\theta(x_{t-1}|x_t)) - \log p_\theta(x_0|x_1)] \quad (6)$$

จากการที่ β เป็นค่าที่ถูกกำหนดไว้แล้วทำให้ $\mathbb{E}_q[D_{KL}(q(x_t|x_{t-1}) \parallel p(x_T))]$ เป็นค่าคงที่จึงจะไม่พิจารณาในฟังก์ชันการสูญเสีย แต่จะพิจารณาค่า $\mathbb{E}_q[\sum_{t>1} D_{KL}(q(x_{t-1}|x_t) \parallel p_\theta(x_{t-1}|x_t))]$ จากพจน์นี้มีการใช้ ค่าการลู่ออกของคัลล์แบก-ลึบเลอว์ (KL divergence) มีหมายความว่ายิ่งค่า $q(x_{t-1}|x_t)$ และ $p_\theta(x_{t-1}|x_t)$ มีการกระจายตัวที่เหมือนกันจะทำให้ค่าการลู่ออกของคัลล์แบก-ลึบเลอว์มีค่าลดลง จึงทำให้ฟังก์ชันการสูญเสียมีค่าลดลง และจากการกระจายตัวของแบบจำลองนี้เป็นการกระจายแบบเกาส์เซียน โดยมีพารามิเตอร์แค่ค่าเฉลี่ย และเมทริกซ์ของความแปรปรวนซึ่งรู้ค่าอยู่แล้วจากการกำหนดให้มีค่าเท่ากันของทั้งกระบวนการไปข้างหน้าและกระบวนการย้อนกลับ จึงทำให้สามารถลดรูปฟังก์ชันของการสูญเสียได้ดังสมการที่ 7

$$L = \mathbb{E}_q \|\tilde{\mu}_t(x_t, x_0) - \mu_\theta(x_t, t)\|^2 \quad (7)$$

โดยที่

$$\tilde{\mu}_t(x_t, x_0) = \frac{\sqrt{\alpha_t}\beta_t}{1-\bar{\alpha}_t}x_0 + \frac{\sqrt{1-\beta_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}x_t \quad (8)$$

และจากสมการที่ 2 และสมการที่ 8 จึงสามารถเขียนการประมาณของค่าเฉลี่ยได้ดังสมการ

$$\tilde{\mu}_t(x_t, x_0) = \frac{1}{\sqrt{\alpha_t}} \left(x_t - \frac{\beta_t}{(1-\bar{\alpha}_t)} \epsilon \right) \quad (9)$$

จากสมการ 9 จึงทำให้สามารถคิดการประมาณของค่าเฉลี่ยได้ในรูปแบบเดียวกันดังสมการ 11

$$\mu_\theta(x_t, x_0) = \frac{1}{\sqrt{\alpha_t}} \left(x_t - \frac{\beta_t}{(1-\bar{\alpha}_t)} \epsilon_\theta(x_t, t) \right) \quad (10)$$

จากสมการ 8, 9 และ 10 จึงได้ฟังก์ชันการสูญเสียดังสมการ 11

$$L = \mathbb{E}_{t, x_0, \epsilon} \|\epsilon - \epsilon_\theta(x_t, t)\|^2 \quad (11)$$

จากฟังก์ชันการสูญเสียในสมการที่ 11 หมายความว่า จากปัญหาของแบบจำลองนี้ที่ต้องประมาณค่าเฉลี่ยของการกระจายตัวของรูปภาพ จะกลายเป็นปัญหาของการทำนายค่าสัญญาณรบกวนของรูปภาพในแต่ละขั้นตอนแทนซึ่งจะทำให้ใช้เวลาในการคำนวณมีค่าน้อยลงจึงมีประสิทธิภาพการคำนวณที่ดีขึ้น และจากค่าสุดท้ายในสมการที่ 6 คือ $\mathbb{E}_q[\log p_\theta(x_0|x_1)]$ จะส่งผลต่อขั้นตอนสุดท้ายในการหา x_0 ของกระบวนการย้อนกลับสามารถดูได้ในตารางที่ 2

2.2.4 ระเบียบวิธีของการเรียนรู้และกระบวนการย้อนกลับของแบบจำลอง

ขั้นตอนวิธีที่ 1 ระเบียบวิธีการเรียนรู้ของแบบจำลอง

ระเบียบวิธีการเรียนรู้ของแบบจำลอง

1. วนลูป

2. $x_0 \sim q(x_0)$

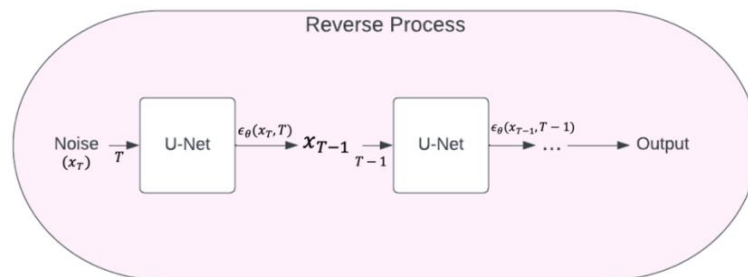
3. $t \sim \text{Uniform}(\{1, \dots, T\})$

4. ใช้วิธีการเคลื่อนที่ลงของความชัน (Gradient descent) กับ

$$\nabla_{\theta} \|\epsilon - \epsilon_{\theta}(x_t, t)\|^2$$

5. จนกระทั่ง ลู่เข้า

จากขั้นตอนวิธีที่ 1 จะเห็นรายละเอียดของการเรียนรู้ของแบบจำลองที่สำคัญคือแบบจำลองจะทำการสุ่มขั้นตอนที่เวลาใดๆขึ้นมาโดยให้ที่เวลาหนึ่ง ๆ มีการแจกแจงแบบเอกรูป และทำการหาค่าที่เหมาะสมที่สุดโดยใช้วิธีการเคลื่อนที่ลงของความชันทำซ้ำจนกระทั่งเกิดการลู่เข้า นอกจากนี้สิ่งที่สำคัญที่สุดของแบบจำลองคือค่าของสัญญาณรบกวนที่ถูกประมาณ ซึ่งค่านี้จะเกิดจากการนำ x_t และ t ไปผ่านโครงข่ายประสาทเทียมแบบยูเน็ต เพื่อที่จะได้ผลลัพธ์ $\epsilon_{\theta}(x_t, t)$ ดังรูปที่ 3 ข้อดีของโครงข่ายประสาทเทียมแบบยูเน็ตคือสามารถกำหนดให้ข้อมูลขาเข้าและข้อมูลขาออกมีขนาดที่เท่ากันได้ จึงมีความเหมาะสมกับงานนี้เนื่องจากเราต้องให้ขนาดของสัญญาณรบกวนมีค่าเท่ากับขนาดของรูปภาพนั้นๆ



รูปที่ 3 แผนภาพกระบวนการย้อนกลับ

ขั้นตอนวิธีที่ 2 ระเบียบวิธีกระบวนการย้อนกลับ

ระเบียบวิธีกระบวนการย้อนกลับ

1. $x_T \sim \mathcal{N}(0, I)$
2. สำหรับ $t = T, \dots, 1$ ทำ
3. $z \sim \mathcal{N}(0, I)$ เมื่อ $t > 1$ นอกจากนี้ $z = 0$
4. $x_{t-1} = \frac{1}{\sqrt{\alpha_t}} [x_t + \frac{1-\alpha_t}{\sqrt{1-\alpha_t}} \epsilon_\theta(x_t, t)] + \sqrt{\beta_t} z$
5. จบ
6. คืนค่า x_0

จากตารางที่ 2 จะเป็นการนำสัญญาณรบกวนที่ได้การเรียนรู้ของแบบจำลองมาใช้เพื่อให้ได้รูปภาพดั้งเดิมหรือ x_0

2.2.5 โครงข่ายประสาทเทียมแบบสังวัตนาการ (Convolution Neural Network)

โครงข่ายประสาทเทียมแบบสังวัตนาการได้ถูกใช้อย่างแพร่หลายในวิธีการเรียนรู้เชิงลึก (Deep learning) ซึ่งโครงข่ายนี้ได้รับแนวคิดมาจากการมองเห็นของสัตว์ โดยในช่วงแรกโครงข่ายนี้ได้ถูกใช้ในงานด้านการจดจำวัตถุ แต่ในปัจจุบันโครงข่ายนี้ได้ถูกนำไปประยุกต์ใช้ในงานด้านอื่น ๆ อีก เช่น การติดตามวัตถุ การประมาณท่าทางของร่างกาย หรือการตรวจจับความผิดปกติของวัตถุที่เคลื่อนที่

โครงข่ายประสาทเทียมแบบสังวัตนาการจะประกอบไปด้วย ชั้นขาเข้า (Input layer), ชั้นซ่อน (Hidden layer) และ ชั้นขาออก (Output layer) ภายในชั้นซ่อนจะมีการสังวัตนาการอยู่ภายในชั้นนี้ และแต่ละชั้นในชั้นซ่อนก็จะมีวิธีการคำนวณแบบต่าง ๆ ดังนี้

1. ชั้นสังวัตนาการ (Convolution layer)

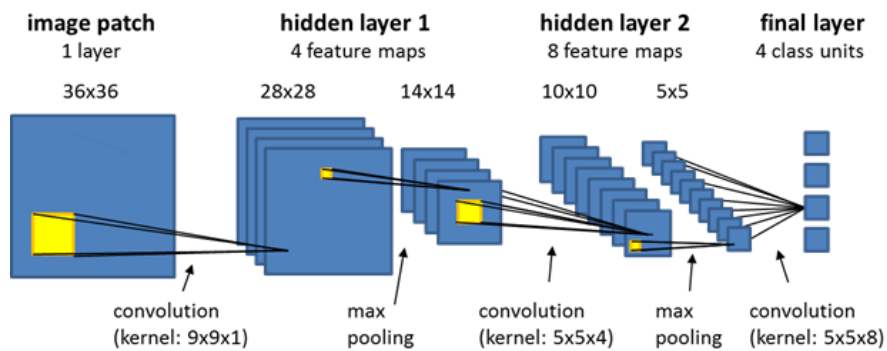
ชั้นนี้จะเป็นชั้นที่พบมากที่สุดในการโครงข่ายประสาทเทียมแบบสังวัตนาการ และเป็นชั้นที่มีการคำนวณมากที่สุด ซึ่งพารามิเตอร์ (Parameter) ที่สามารถเรียนรู้ได้ในชั้นนี้ คือกลุ่มของฟิลเตอร์ (Filter) หรือ เคอร์เนล (Kernel) โดยฟิลเตอร์เหล่านี้จะทำการสังวัตนาการกับรูปภาพขาเข้า เพื่อที่จะได้ผลลัพธ์เป็นฟีเจอร์แมพ (Feature map) ที่มีลักษณะเป็นสองมิติหลาย ๆ อันซึ่งจะถูกต่อกันจนมีลักษณะคล้ายรูปกล่องสามมิติที่มีความกว้าง ความยาว และความลึก และฟีเจอร์แมพเหล่านี้ก็จะทำการสังวัตนาการกับฟิลเตอร์ของชั้นต่อไปดังรูปที่ [1] นอกจากนี้ยังมีไฮเปอร์พารามิเตอร์ (Hyperparameter) ที่ส่งผลต่อขนาดของผลลัพธ์ที่ได้จากการสังวัตนาการได้แก่ จำนวนของฟิลเตอร์จะเป็นตัวกำหนดความลึกของฟีเจอร์แมพ การก้าวข้าม (Striding) เพื่อบอกถึงลักษณะการเคลื่อนที่ของฟิลเตอร์ในขณะทำการสังวัตนาการ และการเสริมเติม (Padding) จะทำให้ขนาดของผลลัพธ์มีขนาดเท่ากับก่อนการสังวัตนาการ

โดยขนาดความกว้าง ขนาดความยาว และขนาดความลึกของพีเจอร์แมพที่ได้หลังจากการสังวัตนาการ
เป็นดังสมการที่ 12 และสามารถดูได้จากรูปที่ 4

$$(n_H^l, n_W^l, n_d^l) = \left(\left\lfloor \frac{n_H^{l-1} + 2 \times p^l - f^l}{s^l} + 1 \right\rfloor, \left\lfloor \frac{n_W^{l-1} + 2 \times p^l - f^l}{s^l} + 1 \right\rfloor, n_f^l \right) \quad (12)$$

โดยที่

- n_H^l คือความยาวของพีเจอร์แมพในชั้นซ่อนที่ l
- n_W^l คือความกว้างของพีเจอร์แมพในชั้นซ่อนที่ l
- n_d^l คือความลึกของพีเจอร์แมพในชั้นซ่อนที่ l
- p^l คือการเสริมเติมในชั้นที่ l
- f^l คือขนาดของฟิลเตอร์ในชั้นที่ l
- s^l คือจำนวนการก้าวข้ามในชั้นที่ l
- n_f^l คือจำนวนของฟิลเตอร์ในชั้นที่ l



รูปที่ 4 โครงสร้างของโครงข่ายประสาทเทียมแบบสังวัตนาการ

ที่มาของรูป: https://docs.ecognition.com/eCognition_documentation/User%20Guide%20Developer/8%20Classification%20-%20Deep%20Learning.htm

การคำนวณค่าของพีเจอร์แมพจะมีวิธีการคือ นำแต่ละฟิลเตอร์มาวางทับแต่ละบริเวณของรูป โดยในแต่ละบริเวณจะทำการคูณแบบค่าต่อค่า แล้วจะนำค่าที่ได้ในหนึ่งบริเวณมารวมกัน หลังจากนั้นจะนำผลลัพธ์ที่ได้ไปเป็นค่าพิกเซลของพีเจอร์แมพ โดยจะนำค่าพิกเซลที่ได้จากการคูณแบบค่าต่อค่าและนำรวมกันของแต่ละบริเวณมาสร้างเป็นเมทริกซ์ มีการคำนวณเป็นไปดังสมการที่ 13 สามารถดูได้ในรูปที่ 5 และเมื่อทำการสังวัตนาการสำหรับทุก ๆ ฟิลเตอร์เสร็จแล้วจะได้เป็นผลลัพธ์ที่มีขนาดสามมิติตามสมการที่ 12

$$G[m, n] = \sum_j \sum_k h[j, k] f[m - j, n - k] \quad (13)$$

โดยที่

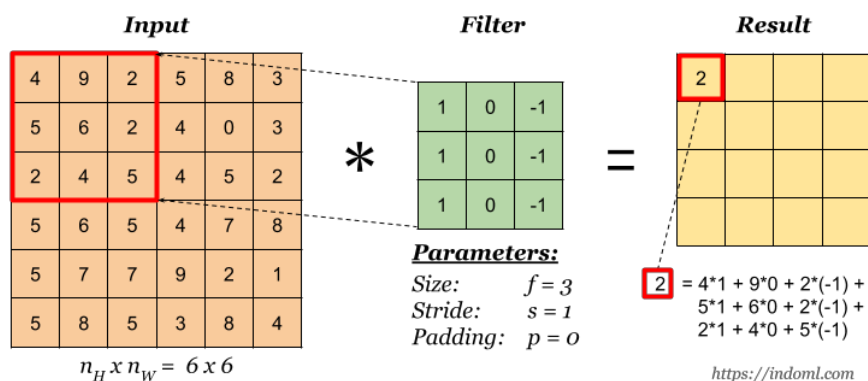
$G[m, n]$ คือค่าของพิกเซลที่ได้หลังจากการรวมผลของการคูณแบบค่าต่อค่าโดยจะอยู่ในคอลัมน์ที่ m แถวที่ n

$h[j, k]$ คือค่าพิกเซลของรูปภาพขาเข้าโดยจะอยู่ในคอลัมน์ที่ j แถวที่ k

f คือฟิลเตอร์

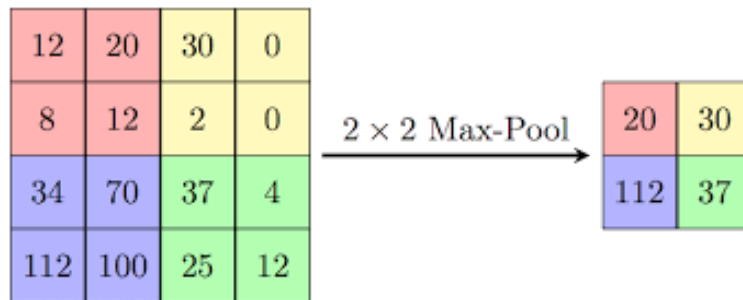
2. ขั้นตอนการรวมกลุ่ม

ขั้นตอนการรวมกลุ่มจะมีหน้าที่ในการลดขนาดของฟังก์ชันลักษณะและความพิเศษของขั้นตอนการรวมกลุ่มนี้คือจะไม่มีพารามิเตอร์ที่สามารถเรียนรู้ได้ ซึ่งจะทำให้การคำนวณมีความซับซ้อนน้อยลง จึงส่งผลให้เวลาในการคำนวณลดลง นอกจากนี้การที่แบบจำลองมีความซับซ้อนน้อยลงจะช่วยลดปัญหาการโอเวอร์ฟิต (Overfitting) โดยขั้นตอนการรวมกลุ่มที่จะใช้ในโครงงานนี้คือการรวมกลุ่มแบบสูงสุด (Max pooling) จะมีหลักการคือจะนำฟิลเตอร์ที่มีขนาดเป็นสี่เหลี่ยมจัตุรัส ทำการสังวัตนาการกับพีเจอร์แมพ แต่แทนที่จะคำนวณตามปกติการรวมกลุ่มแบบสูงสุดจะเลือกค่าที่มากที่สุดที่อยู่ในบริเวณฟิลเตอร์นั้นแทน แสดงในรูปที่ 6



รูปที่ 5 ตัวอย่างการสังวัตนาการ

ที่มาของรูป: <https://www.projectpro.io/article/introduction-to-convolutional-neural-networks-algorithm-architecture/560>



รูปที่ 6 ตัวอย่างการรวมกลุ่มแบบสูงสุด

ที่มาของรูป: <https://computersciencewiki.org/index.php/File:MaxpoolSample2.png>

3.ชั้นเชื่อมโยงแบบสมบูรณ์ (Fully Connected Layer)

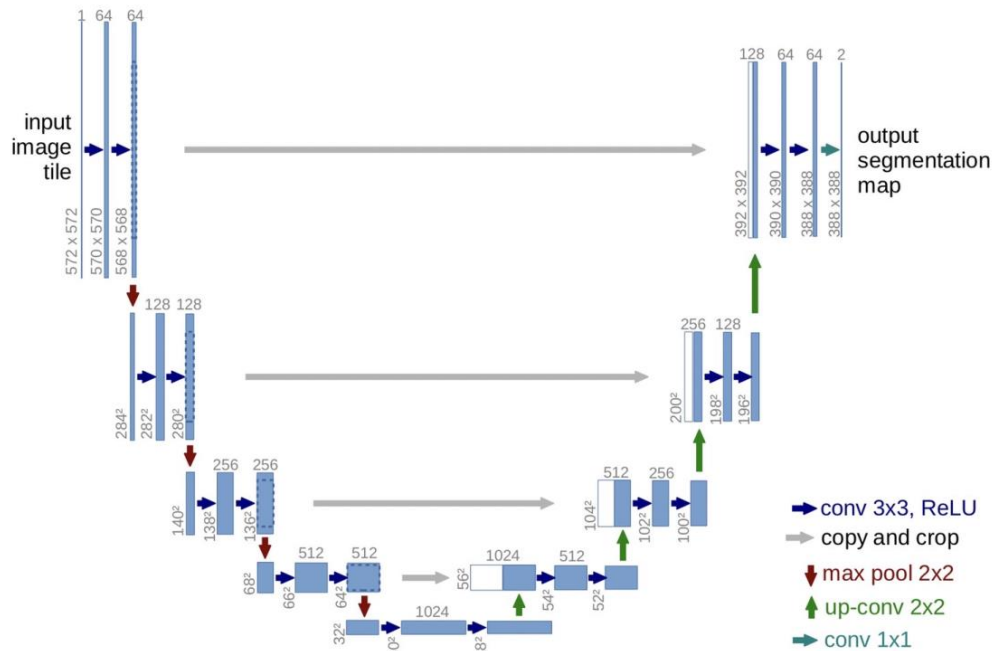
ชั้นเชื่อมโยงแบบสมบูรณ์จะเป็นชั้นที่เปลี่ยนฟีเจอร์แมพ จากการที่มีลักษณะเป็น 3 มิติให้เหลือเพียง 1 มิติ โดยจากผลลัพธ์ที่ได้ในชั้นนี้จะทำให้สามารถนำไปใช้ในงานประเภท การจำแนกประเภทได้ (Classification)

2.2.6 สถาปัตยกรรมโครงข่ายประสาทเทียมแบบยูเน็ต (U-Net)

สถาปัตยกรรมโครงข่ายประสาทเทียมแบบยูเน็ต คือโครงข่ายที่ใช้โครงข่ายประสาทเทียมแบบสังวัตนาการอยู่ภายในชั้นต่าง ๆ พัฒนาต่อยอดขึ้นมาจากสถาปัตยกรรมโครงข่ายสังวัตนาการแบบสมบูรณ์ (Fully convolution network) มีจุดประสงค์เพื่อใช้สำหรับการแบ่งส่วนภาพทางการแพทย์ (Biomedical image segmentation)

โครงสร้างของสถาปัตยกรรมนี้จะแสดงในรูปที่ 7 จะเห็นได้ว่าโครงสร้างนี้จะประกอบไปด้วยขั้นตอนการหดซึ่งแสดงอยู่ทางฝั่งซ้ายของรูปที่ 7 และส่วนขั้นตอนการขยายซึ่งแสดงทางฝั่งขวาของรูปที่ 7 และเมื่อดูโครงสร้างนี้โดยรวมแล้วจะพบว่า มีรูปร่างเป็นตัวยู (U) จึงเรียกระบบสถาปัตยกรรมนี้ว่า ยูเน็ต โดยในทางฝั่งขั้นตอนการหดจะพบว่า ในชั้นต่างๆจะประกอบไปด้วย ชั้นสังวัตนาการที่มีฟิลเตอร์ขนาด 3×3 ($f = 3$) และไม่มีการเสริมเติมค่าศูนย์ทั้งหมด 2 ชั้น ในแต่ละชั้นจะต้องไปผ่านฟังก์ชันแรกทีโพดเชิงเส้น (ReLU) เป็นฟังก์ชันกระตุ้น (Activation function) หลังจากนั้นจะตามด้วยขั้นตอนการรวมกลุ่มแบบสูงสุดที่มีฟิลเตอร์ขนาด 2×2 ($f = 2$) ที่มีค่าการก้าวข้ามเท่ากับ 2 นอกจากนี้ หลังจากผ่านขั้นตอนการรวมกลุ่มแบบสูงสุดในแต่ละครั้ง จะทำการเพิ่มจำนวนฟิลเตอร์ (n_f) เป็นสองเท่าจากรูปที่ 4 จะเห็นว่าจำนวนฟิลเตอร์จะถูกเพิ่มเป็น 64, 128, 256, 512 และ 1024 ตามลำดับ และในขั้นตอนการขยายทางฝั่งขวา จะยังมีชั้นสังวัตนาการที่มีฟิลเตอร์ขนาด 3×3 และไม่มีการเสริมเติมค่าศูนย์เหมือนกับขั้นตอนการหด แต่ในขั้นตอนนี้จะมีการขยาย จากการใช้ขั้นตอนการเพิ่มขั้นของการสังวัตนาการ (Up-convolution) ซึ่งทำให้ความกว้างและความยาวของฟีเจอร์แมพมีขนาดเพิ่มขึ้นสองเท่า นอกจากนี้จะทำการลดความลึกของฟีเจอร์แมพลงสองเท่าเมื่อผ่านขั้นตอนการเพิ่มขั้นของการสังวัตนาการ ในขั้นตอนการขยายจะมีความ

พิเศษคือ มีการนำฟีเจอร์แมพจากฝั่งขั้นตอนการหดมาต่อเข้ากับฟีเจอร์แมพในขั้นตอนการขยาย เพื่อลดการสูญเสียของพิกเซลในการสังวนการ และในขั้นสุดท้ายจะใช้ชั้นสังวนการที่มีฟิลเตอร์ขนาด 1×1 ($f = 1$) อีกทั้งทำให้ความลึกของฟีเจอร์แมพลดลงจนเหลือสองดังแสดงในรูปที่ 7



รูปที่ 7 โครงสร้างสถาปัตยกรรมยูเน็ต [4]

2.3 เครื่องมือวัดประสิทธิภาพของแบบจำลอง

ในโครงการนี้จะใช้เครื่องมือวัดประสิทธิภาพของแบบจำลองสองแบบได้แก่ค่าอัตราส่วนต่อสัญญาณรบกวนสูงสุด (PSNR) และค่าการวัดตำแหน่งความคล้ายของโครงสร้าง (SSIM)

2.3.1 ค่าอัตราส่วนต่อสัญญาณรบกวนสูงสุด (Peak Signal to Noise Ratio; PSNR)

คือค่าอัตราส่วนระหว่างกำลังสูงที่สุดที่เป็นไปได้ของสัญญาณและกำลังของสัญญาณรบกวนโดยค่านี้นี้จะเป็นเครื่องมือที่ถูกใช้กันอย่างแพร่หลายในงานด้านการสร้างคืนของรูปภาพ หรือ วิดีทัศน์โดยมีการคำนวณดังสมการที่ 14 ดังนี้

$$PSNR = 10\log_{10} \left(\frac{R^2}{MSE} \right) \quad (14)$$

โดยที่

R คือค่าสูงสุดที่เป็นไปได้ของค่าพิกเซลในรูปภาพได้แก่ 255

MSE คือค่าเฉลี่ยของความผิดพลาดกำลังสอง คัดจากสมการที่ 15

$$MSE = \frac{\sum_{m,n} [I_1(m,n) - I_2(m,n)]^2}{m \times n} \quad (15)$$

โดยที่

I_1 คือรูปภาพดั้งเดิม

I_2 คือรูปภาพที่ผ่านการสร้างขึ้น

m คือความกว้างของรูปภาพ

n คือความยาวของรูปภาพ

N คือพื้นที่ของรูปภาพ

2.3.2 ค่าการวัดตำแหน่งความคล้ายของโครงสร้าง (structural similarity index measure; SSIM)

เป็นค่าที่วัดคุณภาพของรูปภาพโดยอ้างอิงกับรูปภาพต้นฉบับที่ไม่มีการบีบอัดหรือไม่มีการสูญเสียโดยมีสมการการคำนวณดังสมการที่ 16

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (16)$$

โดยที่

x, y คือรูปภาพที่นำมาเปรียบเทียบกัน

μ_x, μ_y คือค่าเฉลี่ยของรูปภาพ x, y ตามลำดับ

σ_{xy} คือค่ารากของความแปรปรวนข้ามของ x กับ y

σ_x^2, σ_y^2 คือค่าความแปรปรวนของ x และ y ตามลำดับ

C_1, C_2 เป็นค่าคงที่

2.3.3 อัตราคะแนนความโง่เขลา (Fool Score Rate; FRS) [3]

เป็นค่าที่ใช้เปรียบเทียบผลในเชิงคุณภาพมีจุดประสงค์เพื่อเปรียบเทียบประสิทธิภาพระหว่าง แบบจำลองการแปลงความสนใจแบบผสมผสาน (Hybrid Attention Transformer; n_1) และแบบจำลองเชิงความน่าจะเป็นการแพร่กระจายการลดทอนของสัญญาณรบกวน (Denoising Diffusion Model; n_2)

มีวิธีการดังนี้

1. ให้คนดูรูปภาพความละเอียดต่ำพร้อมกับรูปภาพต้นฉบับและรูปภาพผลลัพธ์ที่ออกมาจากแบบจำลอง
2. ให้เลือกว่ารูปภาพไหนเป็นรูปภาพความละเอียดสูงของรูปภาพความละเอียดต่ำ
3. จากนั้นมาคำนวณเป็นเปอร์เซ็นต์ของคนที่เลือก มีสมการดังนี้

$$FRS = \frac{n_i}{N} \times 100 \quad (17)$$

โดยที่

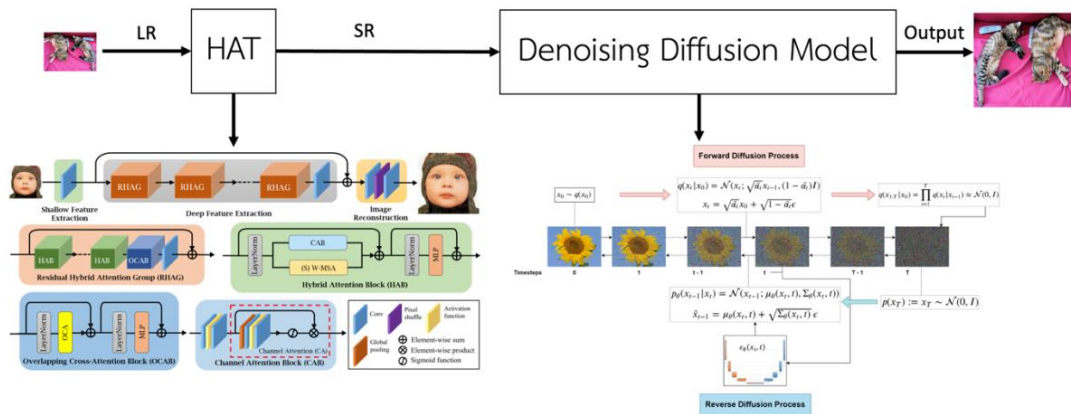
n_i คือจำนวนของคนที่เลือกรูปภาพที่มาจากแบบจำลองที่ i

N คือจำนวนของคนที่เลือกรูปภาพทั้งหมด

จากสมการที่ 17 จะเห็นว่าค่า FRS มีค่าอยู่ระหว่าง 0 ถึง 100 และยิ่งค่านี้มีค่ามากแสดงว่าคนที่ดูรูปภาพที่เป็นผลลัพธ์จากแบบจำลองที่ i นั้นเชื่อว่ารูปภาพนี้เป็นรูปภาพจริง

3. ผลลัพธ์ของโครงงานและการอภิปรายผล

จากความรู้และทฤษฎีที่กล่าวไปในหัวข้อที่ 2 จะนำทุกอย่างมาเชื่อมโยงกันเพื่อให้ได้แบบจำลองที่ใช้ในการพัฒนาภาพความละเอียดสูงยิ่งยวด โดยจะเริ่มจากการนำรูปภาพความละเอียดต่ำ (LR) หลังจากนั้นจะนำไปผ่านแบบจำลองการแปลงแบบผสมผสานเพื่อสร้างคืนภาพความละเอียดสูงยิ่งยวด (SR) ซึ่งจะเพิ่มความละเอียดทั้งหมด 3 แบบได้แก่ เพิ่มความละเอียดขึ้น 2, 3 และ 4 เท่า เมื่อได้ภาพความละเอียดสูงยิ่งยวดแล้วก็จะนำไปผ่านแบบจำลองสุดท้ายซึ่งก็คือแบบจำลองเชิงความน่าจะเป็นการแพร่กระจายการลดทอนของสัญญาณรบกวน ซึ่งในแบบจำลองนี้จะประกอบไปด้วยสองขั้นตอนได้แก่ กระบวนการไปข้างหน้าจะมีหน้าที่เพิ่มสัญญาณรบกวนให้แก่รูปภาพ และกระบวนการย้อนกลับจะมีหน้าที่ในการทำนายสัญญาณรบกวนของรูปภาพที่เวลาต่างๆ เพื่อนำไปใช้ในการสร้างภาพ ณ เวลานั้นๆ และในท้ายที่สุดก็จะได้รูปภาพขาออกเป็นรูปภาพ (Output) ดังรูปที่ 8



รูปที่ 8 แบบจำลองที่เสนอเพื่อใช้ในการพัฒนาภาพความละเอียดสูงยิ่งยวด

3.1 ชุดข้อมูลสำหรับการเรียนรู้ของแบบจำลอง

จากรูปภาพที่ใช้ก่อนผ่านแบบจำลองวิธีการแปลงความสนใจแบบผสมผสาน มีการลดขนาดลงทั้งหมดสามแบบ ได้แก่ ลดขนาดลงสองเท่า สามเท่า และสี่เท่า ทำให้ในการเรียนรู้ของแบบจำลองจะถูกแบ่งเป็นการเรียนรู้ทั้งหมด 3 แบบจำลอง และจะใช้ชุดข้อมูลทั้งหมดสองชุดนำมารวมกันในการเรียนรู้ของแบบจำลองได้แก่

3.1.1 ดีเอฟทูเค (DF2K)

เป็นชุดข้อมูลที่ประกอบไปด้วยรูปภาพทั้งหมด 1000 รูปภาพและความละเอียดของรูปภาพคือ 2048x1080 พิกเซล โดยชนิดของรูปภาพจะมีความหลากหลาย อาทิเช่น ภาพใบหน้าคน ภาพอาคาร

3.1.2 ฟลิคเกอร์ทูเค (Flickr2K)

เป็นชุดข้อมูลที่ถูกเก็บมาจากเว็บไซต์ฟลิคเกอร์โดยรูปภาพจะมีความละเอียดคือ 2048x1080 พิกเซล โดยข้อมูลชุดนี้จะถูกใช้อย่างแพร่หลายสำหรับงานที่เกี่ยวข้องกับรูปภาพ

3.1.3 เซเลบเอ เฮชคิว (CelebA HQ)

เป็นชุดข้อมูลที่ประกอบไปด้วย รูปภาพใบหน้าคนที่มีความละเอียดสูงจำนวน 30,000 รูปภาพ และมีความละเอียด 1024x1024 พิกเซล

3.2 ชุดข้อมูลสำหรับการทดสอบ

สำหรับชุดข้อมูลที่ใช้ในการทดสอบแบบจำลองจะเลือกใช้ชุดข้อมูลการทดสอบเหมือนวรรณกรรม [1] ซึ่งชุดข้อมูลการทดสอบที่เลือกใช้นี้จะเป็นชุดข้อมูลที่เป็นมาตรฐานที่ใช้กันอย่างแพร่หลายสำหรับงานสร้างคืนภาพความละเอียดสูงยิ่งยวด

3.2.1 เซตห้า (Set 5)

เป็นชุดข้อมูลที่ประกอบไปด้วยรูปภาพทั้งหมด 5 รูปภาพได้แก่ รูปเด็กทารก รูปผีเสื้อ รูปนก รูปหัวคน และรูปผู้หญิง

3.2.2 เซตสิบสี่ (Set 14)

เป็นชุดข้อมูลที่ประกอบไปด้วยรูปภาพทั้งหมด 14 รูป และมีความหลากหลายของรูปภาพเช่น คน สัตว์ หรือดอกไม้

3.2.3 บีเอสดีหนึ่งร้อย (BSD 100)

เป็นชุดข้อมูลที่ประกอบไปด้วยรูปภาพ 100 รูป ซึ่งมีความหลากหลายของรูปภาพตั้งแต่ คน วิถีธรรมชาติ ต้นไม้ หรืออาหาร

3.2.4 เออร์เบินหนึ่งร้อย (Urban 100)

เป็นชุดข้อมูลที่ประกอบไปด้วยรูปภาพอาคารต่างๆทั้งหมด 100 รูปภาพ

3.3 รายละเอียดการฝึกฝนแบบจำลอง

3.3.1 แบบจำลองการแปลงแบบผสมผสาน

ข้อมูลที่ใช้ในการฝึกฝนได้แก่ชุดข้อมูล ดีไอวีทูเคและฟลิคเกอร์ทูเค นอกจากนี้จะใช้จำนวนของชั้นกลุ่มความสนใจแบบผสมผสานตึกค้าง (RHAG) ทั้งหมด 6 ชั้น ขนาดแพทช์ (patch size) 64x64 จำนวนการทำซ้ำ (iteration) 500,000 รอบ อัตราการเรียนรู้ (learning rate) 0.0002 และอัตราการเรียนรู้จะลดลงครึ่งหนึ่งทุกๆการทำซ้ำที่ 250,000, 400,000, 450,000, 475,000 รอบ และใช้ตัวหาค่าเหมาะสม (optimizer) คือ อัดัม (Adam)

3.3.2 แบบจำลองเชิงความน่าจะเป็นการแพร่กระจายการลดทอนของสัญญาณรบกวน

ข้อมูลที่ใช้ในการฝึกฝนได้แก่ชุดข้อมูล ชุดข้อมูลดีไอวีทูเคและ เซเลบเอ นอกจากนี้ลำดับขั้นเวลา (timestep) สำหรับแบบจำลองที่จะถูกตั้งไว้ที่ 2000 ขนาดแบตช์ (batch size) 256 จำนวนการทำซ้ำ 1 ล้านรอบ อัตราการเรียนรู้ (learning rate) 0.0001 และใช้ตัวหาค่าที่เหมาะสม (optimizer) คือ อัดัม (Adam)

3.4 ตารางผลลัพธ์

ตารางที่ 1 ตารางการเปรียบเทียบคุณภาพเชิงปริมาณของชุดข้อมูลทดสอบ

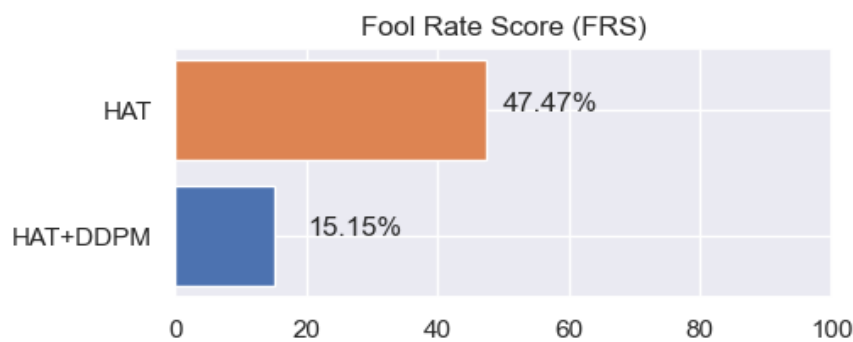
วิธีการ	ชุดข้อมูล	มาตราส่วน	PSNR (dB)	SSIM
HAT	Set5	X2	45.96	0.9891
HAT+DDPM	Set5	X2	19.28	0.4887
HAT	Set5	X3	42.20	0.9731
HAT+DDPM	Set5	X3	20.36	0.4573
HAT	Set5	X4	39.31	0.9552
HAT+DDPM	Set5	X4	20.95	0.5196
HAT	Set14	X2	27.34	0.7867
HAT+DDPM	Set14	X2	19.38	0.4226
HAT	Set14	X3	34.47	0.9098
HAT+DDPM	Set14	X3	18.69	0.4226
HAT	Set14	X4	31.94	0.8606
HAT+DDPM	Set14	X4	19.86	0.4872
HAT	BSDS100	X2	38.80	0.9673
HAT+DDPM	BSDS100	X2	19.15	0.3940
HAT	BSDS100	X3	32.28	0.9217
HAT+DDPM	BSDS100	X3	19.54	0.4223
HAT	BSDS100	X4	31.77	0.8495
HAT+DDPM	BSDS100	X4	19.27	0.4300
HAT	Urban100	X2	37.35	0.9697

วิธีการ	ชุดข้อมูล	มาตราส่วน	PSNR (dB)	SSIM
HAT+DDPM	Urban100	X2	17.98	0.4155
HAT	Urban100	X3	34.17	0.9078
HAT+DDPM	Urban100	X3	18.41	0.4236
HAT	Urban100	X4	29.43	0.8670
HAT+DDPM	Urban100	X4	19.36	0.4723

HAT คือ แบบจำลองการแปลงความสนใจแบบผสมผสาน

HAT+DDPM คือ การนำผลลัพธ์ที่ได้จาก HAT ไปผ่านแบบจำลองเชิงความน่าจะเป็นการแพร่กระจายการลดทอนของสัญญาณรบกวน

จากผลลัพธ์ของตารางที่ 1 จะเห็นว่าสำหรับทุกๆชุดข้อมูลการทดสอบ แบบจำลองการแปลงความสนใจแบบผสมผสานจะมีประสิทธิภาพสูงกว่า การนำผลลัพธ์ที่ได้จาก HAT ไปผ่านแบบจำลองเชิงความน่าจะเป็นการแพร่กระจายการลดทอนของสัญญาณรบกวน (HAT+DDPM) ซึ่งทั้งค่า PSNR หรือ SSIM ยังมีค่าสูงจะมีความหมายว่ารูปที่ออกมานั้นมีคุณภาพที่ใกล้เคียงกับรูปภาพต้นฉบับ

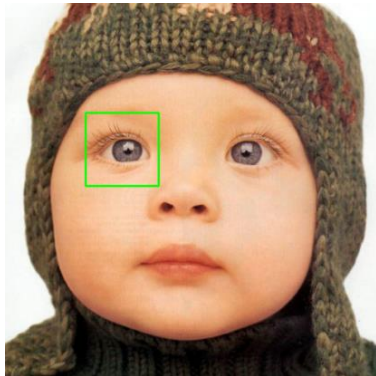


รูปที่ 9 รูปแผนภูมิแท่งของอัตราคะแนนความโง่เขลา


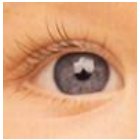

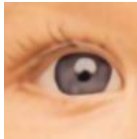



จากการสุ่มรูปภาพมาทั้งหมด 9 รูป จากชุดข้อมูลทดสอบในแบบจำลองการแปลงความสนใจแบบผสมผสาน และเลือกเพิ่มอีก 9 รูปภาพจากผลลัพธ์จากแบบจำลองที่ 2 หลังจากนั้นให้ผู้ทำแบบทดสอบทั้งหมด 11 คน เลือกรูปภาพที่คิดว่าเป็นรูปภาพจริง จึงนำผลลัพธ์มาแสดงเป็นแผนภูมิแท่งได้ดังรูปที่ 9 ซึ่งจากผลลัพธ์ที่ได้จะสรุปได้ว่า ในแบบจำลองการแปลงความสนใจแบบผสมผสาน (HAT) ผู้เข้าทดสอบทั้งหมดตอบผิดว่ารูปภาพที่เกิดจากแบบจำลองนี้เป็นรูปภาพจริงถึง 47.47% เท่านั้น แต่สำหรับแบบจำลองที่ 2 (HAT+DDPM) พบว่าผู้เข้าทดสอบตอบผิดเพียง 15.15% แสดง

ให้เห็นว่าแบบจำลองที่ 2 นี้มีประสิทธิภาพที่ไม่ดี เมื่อเทียบกับแบบจำลองการแปลงความสนใจแบบผสมผสาน

3.5 รูปภาพตัวอย่างผลลัพธ์



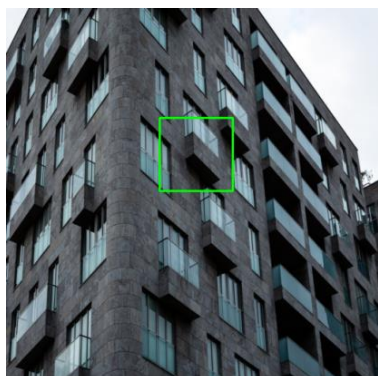
รูปที่ 10 รูปภาพ Baby จาก Set5

			
HR	HATx2	HATx3	HATx4
PSNR/SSIM	37.62/0.9621	37.62/0.8112	32.65/0.8770
			
	DDPMx2	DDPMx3	DDPMx4
	19.46/0.5508	23.79/0.5638	24.70/0.6747



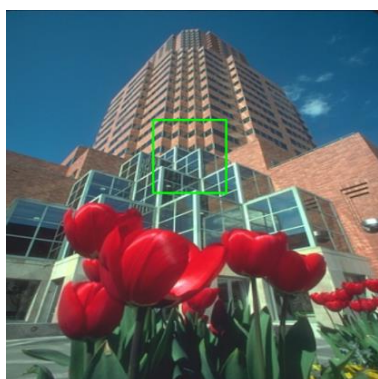
รูปที่ 11 รูปภาพ Zebra จาก Set14

			
HR	HATx2	HATx3	HATx4
PSNR/SSIM	37.62/0.9720	20.53/0.6780	29.56/0.8421
			
	DDPMx2	DDPMx3	DDPMx4
	17.43/0.4260	17.04/0.3768	17.04/0.3768



รูปที่ 12 รูปภาพ img001 จาก Urban100

HR	HATx2	HATx3	HATx4
PSNR/SSIM	36.82/0.9648	23.45/0.7407	29.62/0.8668
	DDPMx2	DDPMx3	DDPMx4
	16.27/0.2625	15.22/0.2201	18.18/0.3253



รูปที่ 13 รูปภาพ 86000 จาก BSDS100

HR	HATx2	HATx3	HATx4
PSNR/SSIM	39.23/0.9822	24.82/0.8010	32.05/0.9184
	DDPMx2	DDPMx3	DDPMx4
	21.40/0.5493	19.47/0.3742	22.62/0.6334

จากผลลัพธ์รูปภาพที่ 10, 11, 12 และ 13 ซึ่งเป็นตัวอย่างที่นำมาจากข้อมูลชุดทดสอบ และผลลัพธ์ที่เกิดจากแบบจำลองทั้งสองแบบจำลอง จะพบว่าผลลัพธ์จากแบบจำลองการแปลงความสนใจแบบผสมผสานมีค่า PSNR และ SSIM สูงกว่าผลลัพธ์จากแบบจำลองที่เกิดจากแบบจำลองเชิงความน่าจะเป็นการแพร่กระจายการลดทอนของสัญญาณรบกวนอย่างมาก ซึ่งจากตัวอย่างรูปภาพจะเห็นว่ามีความแตกต่างอย่างชัดเจน จากผลลัพธ์ตัวอย่าง ก็จะสะท้อนถึงค่า FSR ที่แบบจำลองการแปลงความสนใจแบบผสมผสานมีค่าสูงกว่าแบบจำลองเชิงความน่าจะเป็นการแพร่กระจายการลดทอนของสัญญาณรบกวนอย่างเห็นได้ชัด

4. บทสรุป

4.1 สรุปผลการดำเนินการ

จากการทำงานตลอดภาคการศึกษาที่ผ่านมาได้ศึกษาวรรณกรรมต่างๆที่เกี่ยวข้องกับการสร้างคุณภาพความละเอียดสูงยิ่งยวด จึงได้เลือกแบบจำลองการแปลงความสนใจแบบผสมผสานมาใช้ เนื่องจากแบบจำลองนี้มีประสิทธิภาพในการสร้างคุณภาพความละเอียดสูงยิ่งยวดที่ดีที่สุด และศึกษาแบบจำลองการลดทอนเชิงความน่าจะเป็นการแพร่กระจายการลดทอนของสัญญาณรบกวน เพื่อนำมาประยุกต์ใช้กับแบบจำลองการแปลงความสนใจแบบผสมผสาน หลังจากศึกษาวรรณกรรมต่างๆเรียบร้อยแล้วจึงได้ทดลองนำแบบจำลองการลดทอนเชิงความน่าจะเป็นการแพร่กระจายการลดทอนของสัญญาณรบกวนมาใช้กับผลลัพธ์จากแบบจำลองการแปลงความสนใจแบบผสมผสาน

จากผลลัพธ์ในการสร้างคุณภาพความละเอียดสูงยิ่งยวดในส่วนแรกได้แก่ ผลลัพธ์ที่เกิดจากแบบจำลองการแปลงความสนใจแบบผสมผสาน (HAT) พบว่าผลลัพธ์ที่ได้มีผลลัพธ์ในเชิงปริมาณที่น่าพอใจเนื่องจากมีผลลัพธ์ใกล้เคียงกับในวรรณกรรมที่ [1] โดยอ้างอิงจากค่าอัตราส่วนต่อสัญญาณสูงสุด (PSNR) และ ค่าการวัดตำแหน่งความคล้ายของโครงสร้าง (SSIM) จะเห็นว่าค่า PSNR และ SSIM มีค่าสูงมากสำหรับทุกข้อมูลในชุดทดสอบ ซึ่งสองค่านี้มีค่าสูงแสดงว่ารูปที่สร้างขึ้นกลับมานั้นมีคุณภาพที่ดี และผลลัพธ์ในเชิงคุณภาพโดยอ้างอิงจากอัตราคะแนนความเียงเผลา (FRS) ในรูปที่ 9 พบว่าจากผู้เข้าทดสอบทั้งหมดสามารถแยกภาพที่เกิดจากแบบจำลอง HAT ได้ประมาณครึ่งหนึ่ง และในส่วนที่สองผลลัพธ์ที่ได้จากแบบจำลองเชิงความน่าจะเป็นการแพร่กระจายการลดทอนของสัญญาณรบกวน พบว่าผลลัพธ์ในเชิงปริมาณยังไม่น่าพอใจเนื่องจาก จะเห็นว่าค่า PSNR และ SSIM มีค่าน้อยมากเมื่อเปรียบเทียบกับแบบจำลองการแปลงแบบผสมผสาน ซึ่งเมื่อสองค่านี้น้อยแสดงว่ารูปภาพที่ถูกสร้างขึ้นกลับมานั้นมีคุณภาพที่ไม่ดี ซึ่งสามารถสังเกตได้จากในรูปที่ 10, 11, 12 และ 13 ซึ่งรูปภาพเหล่านี้เป็นรูปที่มาจากข้อมูลชุดทดสอบทั้ง 4 ชุด เหตุผลอย่างหนึ่งที่ทำให้ผลลัพธ์ที่ออกมาไม่น่าพอใจมีประสิทธิภาพน้อยกว่าแบบจำลองการแปลงความสนใจแบบผสมผสานอย่างเห็นได้ชัดเป็นเพราะว่า แบบจำลองเชิงความน่าจะเป็นการแพร่กระจายการลดทอนของสัญญาณรบกวนนี้ เป็นแบบจำลองที่เพิ่งถูกพัฒนาได้ไม่นาน ทำให้ประสิทธิภาพไม่ดีพอทั้งในเชิงปริมาณและในเชิงคุณภาพเมื่อเทียบกับแบบจำลองการแปลงความสนใจแบบผสมผสาน ซึ่งเป็นแบบจำลองนี้มีโครงสร้างมาจากกระบวนการความสนใจของหม้อแปลง (transformer attention mechanism) ที่ได้ถูกพัฒนามานานแล้ว แต่ข้อดีของแบบจำลองเชิงความน่าจะเป็นการแพร่กระจายการลดทอนของสัญญาณรบกวนคือมีระยะเวลาในการฝึกฝนแบบจำลองที่เร็วกว่าทำให้สามารถนำไปฝึกฝนกับชุดข้อมูลหลายๆชุดได้มากกว่า

4.2 ปัญหา อุปสรรค และแนวทางแก้ไข

1. ข้อจำกัดด้านทรัพยากรของเครื่องทำให้ไม่สามารถใช้แบบจำลองการสร้างคุณภาพความละเอียดสูงยิ่งยวดกับรูปภาพที่มีขนาดมากกว่า 1024x1024 ได้จึงทำให้ต้องมีการปรับขนาดของรูปก่อนการใช้

4.3 ข้อเสนอแนะ

1. ทดลองปรับโครงสร้างของยูนิต์ในแบบจำลองเชิงความน่าจะเป็นการแพร่กระจายการลดทอนของสัญญาณรบกวน ให้มีความซับซ้อนมากขึ้นเพื่อที่จะได้ผลลัพธ์ที่มีประสิทธิภาพมากขึ้น

5. กิตติกรรมประกาศ

ทางผู้จัดทำขอขอบคุณ อาจารย์ที่ปรึกษา รศ.ดร. สุภาวดี อร่ามวิทย์ นายวัชร เรืองสังข์ และนิสิตในกลุ่มวิจัยเทคโนโลยีวิทัศน์ หน่วยปฏิบัติการวิจัยการวิเคราะห์และประมวลสื่อประสม ที่คอยให้คำปรึกษาและช่วยเหลือตลอดการทำโครงการ

6. เอกสารอ้างอิง

- [1] Chen, Xiangyu and Wang, Xintao and Zhou, Jiantao and Dong, Chao, "Activating More Pixels in Image Super-Resolution Transformer," *arXiv preprint arXiv:2205.04437*, 2022.
- [2] Ho, Jonathan and Jain, Ajay and Abbeel, Pieter, "Denoising diffusion probabilistic models," *Advances in Neural Information Processing Systems*, volume 33, pp. 6840--6851, 2020.
- [3] Saharia, Chitwan and Ho, Jonathan and Chan, William and Salimans, Tim and Fleet, David J and Norouzi, Mohammad, "Image super-resolution via iterative refinement," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.
- [4] Ronneberger, Olaf and Fischer, Philipp and Brox, Thomas, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, Springer, 2015, pp. 234--241.
- [5] Sohl-Dickstein, Jascha and Weiss, Eric and Maheswaranathan, Niru and Ganguli, Surya, "Deep unsupervised learning using nonequilibrium thermodynamics," in *International Conference on Machine Learning*, PMLR, 2015, pp. 2256--2265.
- [6] Aloysius, Neena and Geetha, M, "A review on deep convolutional neural networks," in *2017 international conference on communication and signal processing (ICCSP)*, IEEE, 2017, pp. 0588-0592.