

Workshop II: Basic Image Segmentation

Assignment Workshop 03

1. ให้นักศึกษาตอบคำถามต่อไปนี้

1.1 อธิบายเบื้องต้นเกี่ยวกับโมเดลที่เป็นโครงสร้างแบบ Autoencoder ว่าคืออะไร มีโครงสร้าง อะไรบ้าง

Autoencoder ถูกใช้ในงานต่างๆ เช่น การลดมิติข้อมูล (Dimensionality Reduction), การลบสัญญาณรบกวน (Denoising), และการสร้างข้อมูลใหม่ (Generative Modeling) เป็น ANN (Unsupervised learning)

ใช้เรียนรู้การบีบอัดข้อมูลและการสร้างข้อมูลกลับมาใหม่ในลักษณะใกล้เคียงต้นฉบับ ประกอบด้วย 3 ส่วนหลัก คือ

- Encoder:

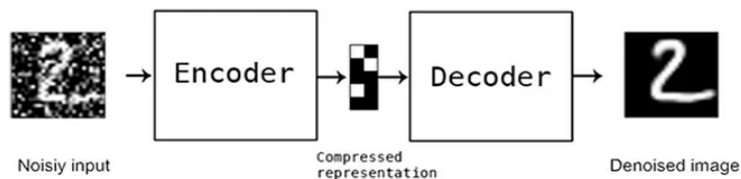
ทำหน้าที่แปลงข้อมูลต้นฉบับ (Input) ให้อยู่ในรูปแบบที่มีมิติที่ต่ำกว่า (Compressed Representation) โดยใช้ชั้นของโครงข่ายประสาทเทียม เช่น Convolutional Layers หรือ Fully Connected Layers

- Latent Space หรือ Code:

บริเวณที่เก็บข้อมูลในรูปแบบที่ถูกบีบอัดไว้มากที่สุด (Bottleneck) ซึ่งเป็นการแสดงผลข้อมูลต้นฉบับในลักษณะเชิงนามธรรม (Abstract Features)

- Decoder:

ทำหน้าที่สร้างข้อมูลกลับมาจาก Latent Space ให้อยู่ในรูปแบบที่ใกล้เคียงกับข้อมูลต้นฉบับมากที่สุด



1.2 Latent Space มีหน้าที่อะไรในโครงสร้างของ Autoencoder

Latent Space มีหน้าที่เก็บข้อมูลในรูปแบบที่ถูกบีบอัด (Compressed Features) และเป็นตัวแทนของข้อมูลต้นฉบับในลักษณะที่ลดมิติลง แต่ยังคงเก็บลักษณะเด่นที่สำคัญ (Key Features) ไว้ ตรงส่วนของ Latent layer ต้องมีขนาดเล็กกว่าข้อมูลต้นฉบับ ต้องมีลักษณะถูกบีบอัดให้เป็น bottleneck เพราะจะเป็นการบังคับให้ Model เรียนรู้ที่จะดึงเอาสาระสำคัญออกมาเท่านั้น

1.3 ในโครงสร้างสถาปัตยกรรมแบบ U-Net จะมีส่วนสำคัญที่เรียกว่า Skip Connection ให้ อธิบายว่าคืออะไร และช่วยอะไรบ้างใน U-Net

Skip Connection ใน U-Net เป็นการเชื่อมโยงข้อมูลจากชั้น Encoder (Downsampling) ไปยังชั้น Decoder (Upsampling) ที่มีระดับชั้นลึกเท่ากัน โดยการ "Copy and Crop" ข้อมูลจาก Encoder แล้วรวมเข้ากับข้อมูลใน Decoder

ประโยชน์ของ **Skip Connection** ใน U-Net:

- รักษาข้อมูลรายละเอียด(Spatial Information) ช่วยเก็บรักษาคุณสมบัติของข้อมูลที่อยู่สูงหายระหว่างกระบวนการลดมิติใน Encoder
- ปรับปรุงประสิทธิภาพของการเรียนรู้ ช่วยให้โมเดลสร้างข้อมูลกลับมาได้ใกล้เคียงต้นฉบับมากขึ้น โดยเฉพาะในงาน Segmentation ที่ต้องการข้อมูลเชิงตำแหน่งที่แม่นยำ
- แก้ปัญหา Gradient Vanishing ช่วยให้ Gradient ถูกส่งผ่านไปยังชั้นลึกๆ ได้ดีขึ้น ทำให้การฝึกโมเดลมีเสถียรภาพมากขึ้น
- ช่วยเพิ่มประสิทธิภาพในการแยกแยะรายละเอียดของข้อมูล เช่น การแบ่งส่วนภาพ(Image Segmentation)

Skip Connection เป็นส่วนสำคัญที่ทำให้ U-Net เหมาะสำหรับงานที่ต้องการความแม่นยำสูง

2. ให้นักศึกษาทำโจทย์ของไฟล์Workshop_II-03_ImageSegmentation.ipynb ให้เรียบร้อย พร้อมทั้งมาตอบคำถามต่อไปนี้

2.1 อธิบายวิธีการสร้างข้อมูลสังเคราะห์ให้มีการสร้างรูปแบบทั้งหมด 4 รูปแบบสำหรับการสร้างข้อมูลเพื่อนำมาเทรนโมเดล

รูปแบบที่สร้าง: วงกลม (Circle), สี่เหลี่ยม (Rectangle), ห้าเหลี่ยม (Pentagon), วงรี (Oval)

1. เริ่มต้น Input Image
 - สร้าง Noise แบบสุ่มด้วย `torch.rand` ขนาด `[1, image_size, image_size]` เพื่อเริ่มต้น Input Image
2. สร้าง Target (4 ช่อง)
 - สร้างข้อมูลสำหรับแต่ละรูปทรงใน 4 ช่องของ `target`:
 - Circle: ใช้สมการ $(x-x_c)^2 + (y-y_c)^2 \leq r^2$
 - Rectangle: ระบุนิพจน์ `(x, y)` และ `(x2, y2)` แล้วเติมค่าภายใน
 - Pentagon: ใช้ฟังก์ชัน 5 จุดสร้างขอบเขตด้วย `polygon`
 - Oval: ใช้สมการ $((x-x_c)/a)^2 + ((y-y_c)/b)^2 \leq 1$
3. รวม Target กับ Input Image
 - รวมรูปทรงทั้งหมดจาก `target` เข้ากับ Noise ของ Input Image ด้วยน้ำหนัก 0.8 และตัดค่าให้อยู่ในช่วง `[0, 1]`
4. บันทึกข้อมูล
 - หากกำหนด `save_path`:
 - บันทึก Input Image เป็น `.png`
 - บันทึก Target เป็น `.npy`

2.2 คำสั่ง Conv2d กับ ConvTranspose2d ใน pytorch ใช้งานต่างกันอย่างไร

- Conv2d (Convolution Layer)
ใช้สำหรับ ลดขนาดเชิงพื้นที่ (spatial size) ของข้อมูล หรือ ดึงคุณลักษณะ (feature extraction) จากข้อมูลอินพุต เช่น ภาพ จะคำนวณค่าผลลัพธ์โดยเลื่อน

Kernel/Filter ไปตามอินพุต เพื่อให้ได้ Feature Map ขนาดเล็กลง ใช้ใน Encoder หรือ Downsampling ในโครงสร้าง CNN/U-Net

```
nn.Conv2d(in_channels, out_channels, kernel_size=3, stride=1, padding=1)
```

- ConvTranspose2d (Transposed Convolution Layer)

ใช้สำหรับ เพิ่มขนาดเชิงพื้นที่ (spatial size) ของข้อมูล เช่น การ Upsampling หรือ Reconstruct ภาพจาก Feature Map จะกระจาย Kernel/Filter ย้อนกลับ เพื่อเพิ่มขนาดของข้อมูลผลลัพธ์

ใช้ใน Decoder หรือ Upsampling ในโครงสร้าง U-Net

```
nn.ConvTranspose2d(in_channels, out_channels, kernel_size=2, stride=2)
```

2.3 แสดงผลการทำนายของการทำโมเดล U-Net ที่เทรนด้วยข้อมูลสังเคราะห์แบบมี 4 วัตถุ

