HW1: AutoEncoder & VAE

314581006 林鼎翔

Homework Question:

Q1: Describe your Autoencoder model, including details about the Encoder and Decoder architecture, the activation functions used, any regularization techniques applied, and other relevant components.

A1:

- Encoder: self.encoder = nn.Sequential(

```
        nn.Conv2d(3, 64, kernel_size=4, stride=2, padding=1), # 64->32
        nn.ReLU(inplace=True),
        nn.Conv2d(64, 128, kernel_size=4, stride=2, padding=1), # 32->16
        nn.BatchNorm2d(128),
        nn.ReLU(inplace=True),
        nn.Conv2d(128, 256, kernel_size=4, stride=2, padding=1), # 16->8
        nn.BatchNorm2d(256),
        nn.ReLU(inplace=True)
)
```

  輸入影像大小為 3*64*64，經過三個stride=2的卷積層後會變成 256*8*8，中間加入ReLU當作activation function，還有加BatchNorm去做正規化。

- Decoder: self.decoder = nn.Sequential(

```
        nn.ConvTranspose2d(256, 128, kernel_size=4, stride=2, padding=1), # 8->16
        nn.BatchNorm2d(128),
        nn.ReLU(inplace=True),
        nn.ConvTranspose2d(128, 64, kernel_size=4, stride=2, padding=1), # 16->32
        nn.BatchNorm2d(64),
        nn.ReLU(inplace=True),
        nn.ConvTranspose2d(64, 3, kernel_size=4, stride=2, padding=1), # 32->64
        nn.Sigmoid()
)
```

  從256*8*8逐漸轉換回3*64*64，隱藏層使用ReLU當activation function和使用BatchNorm正規化，最後輸出層使用Sigmoid，讓重建後的像素範圍介於[0,1]。

Q2: Explain how you trained the model, specifying the loss function, optimization method, learning rate, and any other training-related details.

A2:

- AE: Loss Function使用MSE，Optimizer用Adam(沒有做weight decay或scheduling)，學習率稍微調高到0.002，並提高epoch到150，幫助模型收斂。
- VAE: Loss Function使用 restruction loss+KL divergence，Optimizer一樣用Adam(沒有做weight decay或scheduling)，學習率提高到0.002，epoch提高到150。

Q3: Qualitative Analysis: Display the inference results for a random sample of 10 images, showing the Ground Truth, the predictions from the AE, and the predictions from the VAE.

A3:

Q4: Introduce the qualitative evaluation metrics: PSNR (Peak Signal-to-Noise Ratio), SSIM (Structural Similarity Index Measure), and LPIPS (Learned Perceptual Image Patch Similarity).

A4:

- PSNR (Peak Signal-to-Noise Ratio): 看整體的像素誤差是多少，PSNR分數越高表示跟標準答案越接近、雜訊比較小。
- SSIM (Structural Similarity Index Measure): 不只有看像素誤差，還會看亮度、對比、結構有沒有對齊，數值範圍是0到1，數值越接近1越好。
- LPIPS (Learned Perceptual Image Patch Similarity): 去比較兩張圖在hyper特徵上面的差異，像是紋理和風格之類的。數值範圍是0到1，數值越接近0越好。

Q5: Present the qualitative analysis results for both the AE and VAE models, including the PSNR, SSIM, and LPIPS values.

A5:

	Metric	AE	VAE
0	PSNR	26.554852	19.827398
1	SSIM	0.846709	0.574641
2	LPIPS	0.127692	0.291075

Q6: Conclusion: Based on the training and testing results of the AE and VAE models,provide a summary and final conclusion for this assignment.

A6: AE 在三個評估指標下，都表現得比 VAE 好。我認為原因是 VAE 的處理過程中，會 map 到常態分布，導致有些細節被消去掉。而且 VAE 的 loss function 是 reconstruction loss 和 KL Divergence，可能會失衡。如果 KL 太重，就可能輸出比較糊和缺少細節；如果 reconstruction loss 比較重，生成圖的品質就會比較差。