Week 10: Assignment 10

The due date for submitting this assignment has passed.

Due on 2024-10-02, 23:59 IST.

Assignment submitted on 2024-10-02, 15:56 IST

1) Why might Segment Anything (SAM) be particularly useful in data annotation tasks compared to traditional segmentation models? *1 point*

- It produces perfect segmentation masks.
- It can adapt to segment any object, even those not seen during training
- It requires less computational resources
- It automatically labels all objects in an image without user input

Yes, the answer is correct.
Score: 1
Accepted Answers:
It can adapt to segment any object, even those not seen during training

2) What is the primary advantage of using DETR (Detection Transformer) over traditional object detection methods? *1 point*

- DETR uses an anchor-based approach, which simplifies the object localization process.
- DETR eliminates the need for region proposals and anchor boxes, simplifying the object detection pipeline.
- DETR requires significantly fewer training epochs compared to traditional methods.
- DETR can only detect objects in high-resolution images due to its reliance on self-attention mechanisms.

Yes, the answer is correct.
Score: 1
Accepted Answers:
DETR eliminates the need for region proposals and anchor boxes, simplifying the object detection pipeline.

3) How does the patch size in a Vision Transformer impact performance? *1 point*

- Smaller patch sizes lead to better local feature extraction but increase computational cost.
- Larger patch sizes always improve model performance.
- Smaller patch sizes are computationally cheaper but may miss global context.
- Patch size has no significant impact on model performance.

Yes, the answer is correct.
Score: 1
Accepted Answers:
Smaller patch sizes lead to better local feature extraction but increase computational cost.

4) What is a key characteristic of the Swin Transformer that differentiates it from the standard Vision Transformer (ViT)? *1 point*

- Swin Transformer uses global attention throughout the entire image for every layer.
- Swin Transformer employs a hierarchical structure with shifted windows for local attention, allowing it to scale to larger images.
- Swin Transformer is designed exclusively for small image resolutions.
- Swin Transformer eliminates the use of multi-head self-attention in favor of convolutional operations

Yes, the answer is correct.
Score: 1
Accepted Answers:
Swin Transformer employs a hierarchical structure with shifted windows for local attention, allowing it to scale to larger images.

5) What is the purpose of the class token in a Vision Transformer? *1 point*

- It encodes the position of each image patch.
- It serves as the representation of the entire image, which is used for classification.
- It performs the same function as a softmax layer in traditional neural networks.
- It stores the output of each transformer layer.

Yes, the answer is correct.
Score: 1
Accepted Answers:
It serves as the representation of the entire image, which is used for classification.

6) Why do Vision Transformers often require large datasets for effective training? *1 point*

- They are inherently more data-efficient than CNNs.
- They lack the inductive biases of convolutions, making them more reliant on data to learn structure.
- Their self-attention mechanism directly reduces the need for large datasets.
- They can overfit more easily without large datasets.

Yes, the answer is correct.
Score: 1
Accepted Answers:
They lack the inductive biases of convolutions, making them more reliant on data to learn structure.

7) What is the primary challenge when training GANs? *1 point*

- Maximizing the discriminator loss.
- Ensuring the generator and discriminator learn in balance.
- Training the generator faster than the discriminator.
- Reducing the number of parameters in the generator.

Yes, the answer is correct.
Score: 1
Accepted Answers:
Ensuring the generator and discriminator learn in balance.

8) Which of the following best describes "mode collapse" in GANs? *1 point*

- The discriminator becoming too powerful.
- The generator producing a limited variety of outputs.
- The loss function of the discriminator diverging.
- The generator generating random noise instead of real-like data.

Yes, the answer is correct.
Score: 1
Accepted Answers:
The generator producing a limited variety of outputs.

9) What is the role of the latent space in a VAE? *1 point*

- It stores the compressed data.
- t stores real-valued outputs of the decoder.
- It represents the error between the input and output.
- It captures a distribution of latent variables for data generation.

Yes, the answer is correct.
Score: 1
Accepted Answers:
It captures a distribution of latent variables for data generation.

10) Which of the following statements are false? (Select all that apply) *1 point*

- [ ] Generative adversarial networks (GANs) generate sharper images compared to Variational AutoEncoders (VAE)
- [ ] GAN is an example of an implicit density estimation model
- [x] Fully connected layers in mapping network of Style-GAN do not change the dimension of its input
- [x] The generator and discriminator are always trained together in a GAN

Yes, the answer is correct.
Score: 1
Accepted Answers:
Fully connected layers in mapping network of Style-GAN do not change the dimension of its input
The generator and discriminator are always trained together in a GAN

11) What are the capabilities of these models? *1 point*

1) Discriminative model	i) Assigns labels to data; Performs supervised feature learning
2) Generative model	ii) Assigns labels while rejecting outliers; Generates new data conditioned on input labels
3) Conditional generative model	iii) Detects outliers; Performs unsupervised feature learning; Samples to generate new data

- 1→iii, 2→i, 3→ii

1→i, 2→iii, 3→ii

1→ii, 2→iii, 3→i

1→i, 2→ii, 3→iii

Yes, the answer is correct.
Score: 1

Accepted Answers:
1→i, 2→iii, 3→ii

In a VAE, if the encoder outputs $\mu = [0.3, 0.1, 0.2, 0.4]$ and $\sigma = [0.1, 0.4, 0.2, 0.3]$, and ϵ sampled from $N(0, I)$ is $[0.6, 0.2, 0.4, 0.1]$, then the latent value z given to the decoder is?

12) Element 1:_____

0.36

Yes, the answer is correct.
Score: 0.25

Accepted Answers:
(Type: Numeric) 0.36

0.25 points

13) Element 2:_____

0.18

Yes, the answer is correct.
Score: 0.25

Accepted Answers:
(Type: Numeric) 0.18

0.25 points

14) Element 3:_____

0.28

Yes, the answer is correct.
Score: 0.25

Accepted Answers:
(Type: Numeric) 0.28

0.25 points

15) Element 4:_____

0.43

Yes, the answer is correct.
Score: 0.25

Accepted Answers:
(Type: Numeric) 0.43

0.25 points