Course outline

About NPTEL

How does an NPTEL online course work?

Week 0

Week 1

Week 2

Week 3

Week 4

Week 5

Week 6

Week 7

Week 8

Week 9

Week 10

Week 11

Week 12

● Contrastive Learning and History in Face Understanding - Part 1

○ Contrastive Learning and History in Face Understanding - Part 2

○ Self-Supervised Learning: SimCLR

○ Vision Language Models: Introduction and History

○ CLIP: The Anchoring Inflection Point

○ Beyond CLIP: BLIP, BLIP-2 and CoCA

○ From VLMs to Multimodal LLMs

○ Course Conclusion

○ Lecture materials

○ Practice: Week 12 : Assignment 12(Non-Graded)

● **Quiz: Week 12: Assignment 12**

○ Week 12 Feedback Form : Deep Learning for Computer Vision

Download Videos

Live Session

Text Transcripts

Books

Problem Solving Session - July 2024

Week 12: Assignment 12

The due date for submitting this assignment has passed.

Due on 2024-10-16, 23:59 IST.

Assignment submitted on 2024-10-16, 00:34 IST

1) What is the effect of increasing the guidance scale in classifier-free guidance? ***2 points***

○ The generated images become more random
● The generated images become more realistic
○ The generated images become more stylized
○ The generated images become less diverse

No, the answer is incorrect.
Score: 0

Accepted Answers:
The generated images become less diverse

2) In the forward process of a diffusion model, the data is incrementally corrupted by adding noise across T timesteps. If the variance of the noise added at each step t is β_t , which decreases linearly from 0.03 at the first step to 0.02 at the last step over 150 timesteps, calculate the total variance of the noise added over 150 timesteps. (Hint: Calculate average variance per step to calculate total variance) ***2 points***

● 3.25
○ 4.5
○ 7.5
○ 3.75

No, the answer is incorrect.
Score: 0

Accepted Answers:
3.75

3) Consider a reverse process in a diffusion model where the goal is to reconstruct the original data from the noise. If the model correctly reduces the variance of the noise by 0.02 in each reverse step, and starts with a noise variance of 1.0 at timestep $T = 50$, how many steps are required to reduce the noise variance to 0.1? ***2 points***

○ 45
○ 50
● 40
○ 30

No, the answer is incorrect.
Score: 0

Accepted Answers:
45

4) Classifier-free guidance is a technique used in diffusion models to improve sample quality without the explicit use of a classifier. This technique involves modifying the sampling process based on a control parameter. If the control parameter, denoted as γ , is set to zero, what effect does this have on the generation process? ***2 points***

☐ It fails to generate realistic samples
☑ It removes all guidance, effectively making the process equivalent to the unconditional generation
☐ It maximizes the influence of the classifier, leading to highly detailed generations
☐ It can lead to more diverse samples compared to higher values of γ , as the generation process is less constrained by the conditional information

Partially Correct.
Score: 1

Accepted Answers:
It removes all guidance, effectively making the process equivalent to the unconditional generation
It can lead to more diverse samples compared to higher values of γ , as the generation process is less constrained by the conditional information

5) Which of the following are **FALSE** for self-supervised learning (SSL) techniques? (Select ALL possible correct options) ***0 points***

☐ Bootstrap Your Own Latent (BYOL) method does not depend on negative samples to achieve state-of-the-art results
☑ MoCo maintains the dictionary as a stack of data samples, this enabling use of encoded keys from the immediately preceding mini-batches
☐ In SimCLR, the number of negative samples is limited by batch size
☑ In image rotation-based SSL, the task typically involves generating the correct image for the given rotated input image
☐ In the image inpainting task, the goal is to fill the gaps of an image based on surrounding information

Partially Correct.
Score: 0

Accepted Answers:
MoCo maintains the dictionary as a stack of data samples, this enabling use of encoded keys from the immediately preceding mini-batches
In image rotation-based SSL, the task typically involves generating the correct image for the given rotated input image
In the image inpainting task, the goal is to fill the gaps of an image based on surrounding information

6) What is the purpose of the reverse process in DDPMs? ***2 points***

○ To add noise to the image
● To remove noise from the image
○ To generate new images

Yes, the answer is correct.
Score: 2

Accepted Answers:
To remove noise from the image

7) What is the purpose of the forward process in DDPMs? ***2 points***

● To add noise to the image
○ To remove noise from the image
○ To generate new images

Yes, the answer is correct.
Score: 2

Accepted Answers:
To add noise to the image

8) What is the diffusion process in a diffusion model? ***2 points***

● A stochastic process that gradually adds noise to an image
○ A deterministic process that gradually removes noise from an image

○ A neural network-based process that generates images
○ A generative adversarial network-based process that generates images

Yes, the answer is correct.
Score: 2
Accepted Answers:
A stochastic process that gradually adds noise to an image

9) What is the primary goal of CLIP? ***2 points***

○ To generate images from text descriptions
○ To translate text into different languages
● To learn a joint embedding space for text and images
○ To perform image classification

Yes, the answer is correct.
Score: 2
Accepted Answers:
To learn a joint embedding space for text and images

10) Which technique does CLIP use to learn a joint embedding space ***2 points***

○ Reinforcement learning
○ Supervised learning
● Contrastive learning
○ Unsupervised learning

Yes, the answer is correct.
Score: 2
Accepted Answers:
Contrastive learning

11) Which of the following is a common self-supervised learning task? ***2 points***

○ Image inpainting
○ Image colorization
○ Image denoising
● All of the above

Yes, the answer is correct.
Score: 2
Accepted Answers:
All of the above

12) What task does BLIP primarily excel at? ***2 points***

● Image captioning
○ Image classification
○ Text-to-image generation
○ Object detection

Yes, the answer is correct.
Score: 2
Accepted Answers:
Image captioning