Topics
My posts
More
CATEGORIES
General
Site Feedback
All categories

2025 USA-NA-AIO Round 2, Problem 3, Part 5

USAAIO May 2025

Part 5 (5 points, non-coding task)

Note that our final goal is to build a CLIP neural network. For the image data, we will use Vision Transformers (ViT) to extract image embeddings.

With the above high level information, please explain the reasons behind the following things that you did in Part 4.

1. Why the channel dimension is ahead of the height and width dimensions?
2. Why the sizes of all images are normalized to `(224,224)`?
3. Why each pixel value is normalized between -1 and 1?

USAAIO May 2025

Misplaced '#'

1. The input of ViT requires the channel dimension to go ahead of the height and width dimensions.
2. ViT model requires this dimension.
3. ViT model requires data to fall into this range.

""" END OF THIS PART """

Skip to main content

Related topics

Topic	Replies	Activity
2025 USA-NA-AIO Round 2, Problem 3, Part 10	1	May 2025
2025 USA-NA-AIO Round 2, Problem 3, Part 11	1	May 2025
2025 USA-NA-AIO Round 2, Problem 3, Part 13	1	May 2025
2025 USA-NA-AIO Round 2, Problem 3, Part 4	1	May 2025
2025 USA-NA-AIO Round 2, Problem 3, Part 1	1	May 2025