HW1-ImageBind

Google Colab Links

Google Colaboratory

https://colab.research.google.com/drive/1shMvudeqw4A-ECFbC9qjWQYuMm1p2mv4

Google Drive Folder

https://drive.google.com/drive/folders/1WWqnQKCLWTzpLAPoy2eAz-vZfXb1PinW?usp=drive_link

Results and Discussions

1. **Provide the three 3x3 matrices obtained in Task 1 (as described in the “Fill up the TODO blocks” section), and perform a comparative analysis. Explore the differences within and between these matrices and describe your observations in detail.**

   - Text data — “*Dog*”, “*Car*”, “*Bird*”
   - Image data



   - Dot Product Matrix

```
Dot Product Matrix:
             Dog         Car        Bird
Dog  37.04533005 22.74183273 20.08076859
Car  17.24113083 27.90612793 22.92059326
Bird 22.98768997 21.50801849 26.77712440
```

> The dot product value between this dog image and the dog text is the highest (37.0), indicating a greater similarity between them.

> The dot product value between Car-Car is higher than other combinations containing car data.

> The dot product value between Bird-Bird is relatively lower (26.8), indicating a low degree of similarity between them.

- Softmax Matrix

```
Softmax Matrix:
            Dog        Car       Bird
Dog  0.99999940 0.00000061 0.00000004
Car  0.00002319 0.99318731 0.00678955
Bird 0.02199780 0.00500917 0.97299308
```

> Compared with the Dot Product Matrix above, application of the Softmax function widened the gap between the values in the matrix.

- Cosine Product Matrix

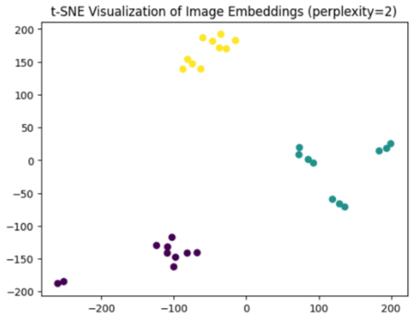
```
Cosine Product Matrix:
            Dog        Car       Bird
Dog  0.24359924 0.16094951 0.12362319
Car  0.09137049 0.18363532 0.08193187
Bird 0.10504004 0.10618439 0.24519134
```

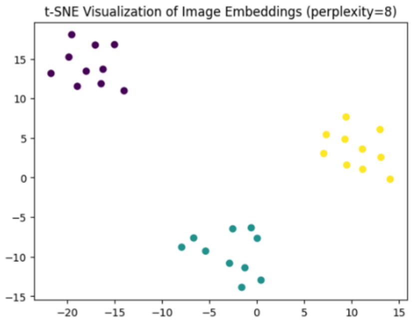
> Cosine similarity measures how similar or dissimilar two vectors are by looking at the cosine of the angle between them. It helps us understand the relationships between embedding vectors, identify similar pairs, and is commonly used in clustering, classification, and retrieval tasks.

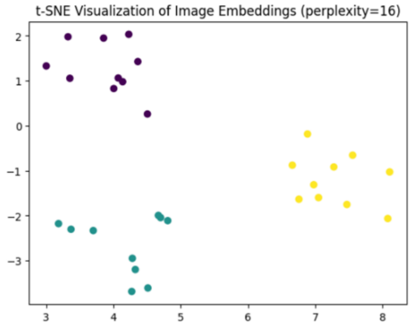
2. **Why do we need to use softmax after calculating the inner product?**

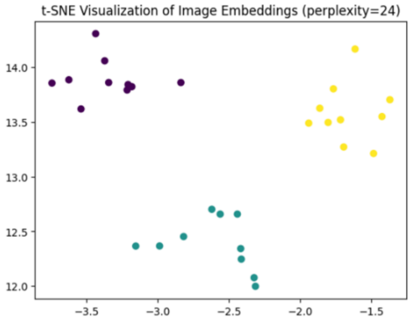
> Softmax normalizes inner product outputs into a probability distribution, crucial for applications like multi-class classification. It also enhances stability during training by reducing disparities in raw values, aiding optimization.

3. **Include the figures and discuss the outputs of t-SNE when using different numbers for perplexity.**



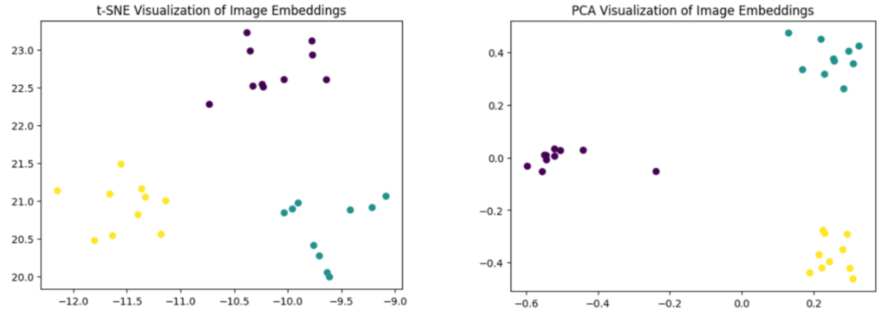






> Lower perplexity values tend to result in tighter, more focused clusters in the t-SNE plots, with data points closer together. On the other hand, higher perplexity values yield more spread-out data points with a tendency to form larger, less distinct clusters.

4. **Include the figure generated by t-SNE and PCA, compare the result and discuss the differeces.**



> In the t-SNE plot, the relative distances between data points are better preserved. This facilitates capturing complex nonlinear relationships in the data, making t-SNE a preferred choice for visualizing high-dimensional data.

> In the PCA plot, data point distribution relies more on principal components, which typically represent linear relationships in the data. Therefore, PCA tends to capture linear relationships during dimensionality reduction and has limited capacity for handling nonlinear relationships.

5. **Include the three pictures and text you selected and described. Record the result.**

- Text data — “*A cat with rabbit headwear*”, "*A standing cat*", "*Stacking Cats*”
- Image data

- Results matrices

```
Dot Product Matrix:
24.359924 16.094948 12.362318
9.137048  18.363531  8.193187
10.504002 10.618438 24.519131

Softmax Matrix:
0.999999 0.000001 0.000000
0.000023 0.993187 0.006790
0.021998 0.005009 0.972993
```

> The higher dot product value between *“A cat with rabbit headwear”* can be attributed to the presence of more contextually relevant keywords such as '*cat*', '*rabbit*', and '*headwear*' in the text.

> The presence of the letter *'s'* in *'Stacking Cats'* text likely contributed to its association with images containing multiple cats.

6. **Anything you find out interesting about the embedding space.**

> In the past, when training neural networks, models were primarily designed based on the format of the data. Therefore, I am delighted with the introduction of *ImageBind*, which allows various types of data to be embedded into a unified

> embedding space, opening up new possibilities for the future of AI development.