Doc-2

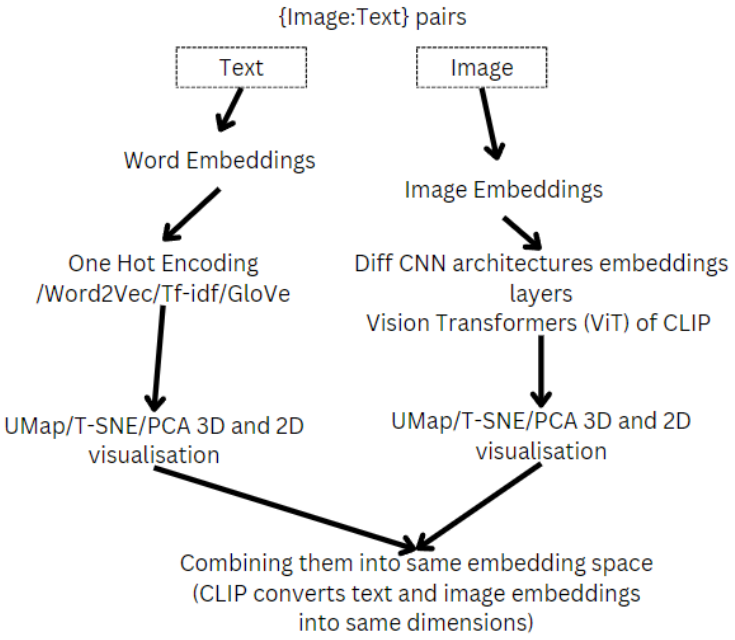
Text-Image Embeddings & Knowledge Graph Embeddings

The project will be divided into 2 parts:

1. Text-Image Vector Embeddings
2. Knowledge Graph Embeddings (KGEs)

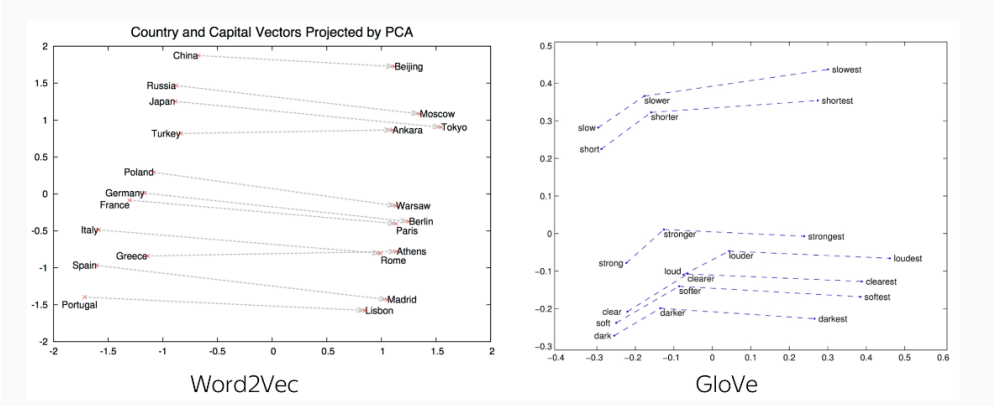
PART-1

Text/Image Embeddings

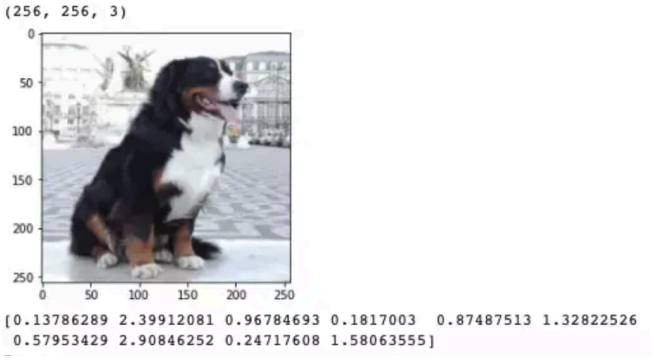


Prerequisites (for Part 1):

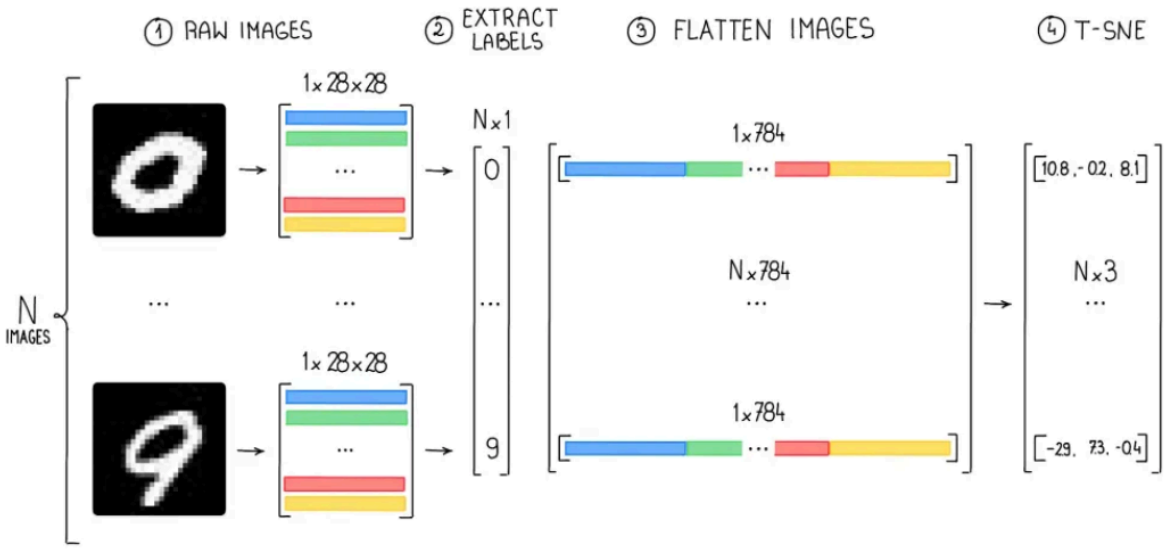
- **Understanding Word and Image embeddings + Basics of NLP**
  https://lena-voita.github.io/nlp_course/word_embeddings.html -Word Embeddings
  1. Understand Context Windows,
  2. Word2Vec ka Log loss function,
  3. Embedding dimensions.
  4. Read analysis and interpretability section
  5. Linear Structure



- **Image Embeddings of CLIP model and CNNs:** https://www.activeloop.ai/resources/generate-image-embeddings-using-a-pre-trained-cnn-and-store-them-in-hub/



Each CNN model has different embedding layers for Image, we can use them for visualization.



This is any CNN model embedding

- What is an Image Embedding?
- Leveraging Embeddings and Clustering Techniques in Computer Vision

This RoboFlow blog discusses about CLIP and Large Multimodal Model's Image embeddings

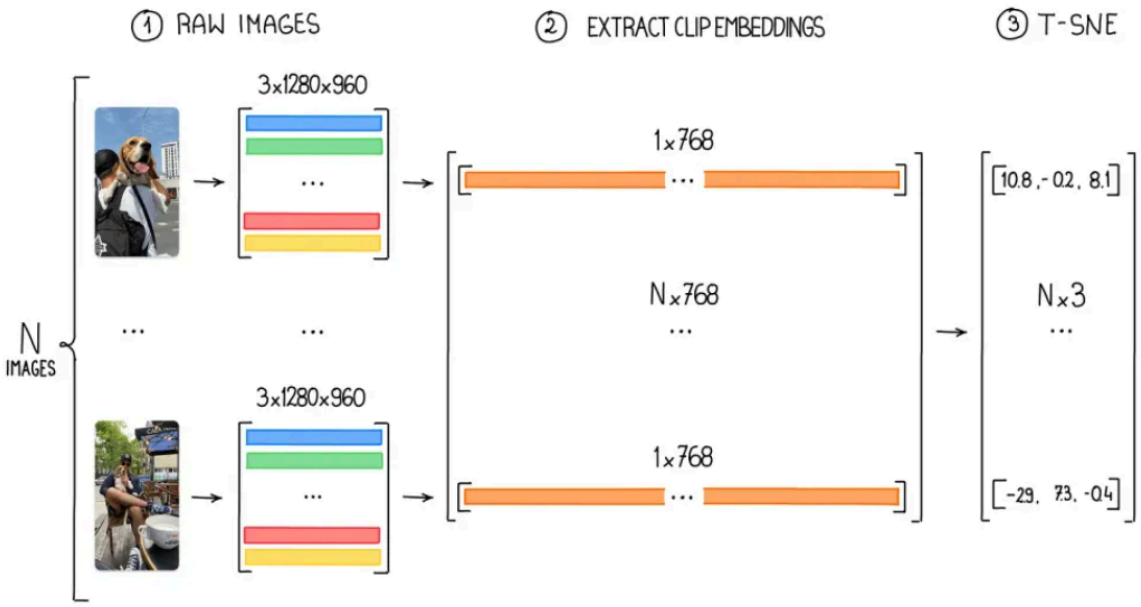


Diagram depicting the dimensions of the tensors at each stage of processing when using CLIP embeddings.

CLIP embeddings address this issue by providing a more abstract and compact representation of images, effectively encoding high-level visual and semantic information. By using CLIP embeddings, we can efficiently work with high-resolution images while preserving their essential visual and semantic characteristics for various computer vision tasks

An image embedding will encode this information. We could then compare the image embedding to a text embedding like "fruit" to see how similar the concept of "fruit" is to the contents of the image. We could take two prompts, such as "fruit" and "vegetable", and see how similar each one is. The most similar prompt is considered the most representative of the image.

You can save embeddings in a special database called a vector database (used to store embedding) to run efficient searches between image and text embeddings. This efficient searching enables you to build a "semantic" image search engine.

- **Dimensionality Reduction Techniques**
  https://programminghistorian.org/en/lessons/clustering-visualizing-word-embeddings - Read Dimension Reduction section only
  PCA: https://perma.cc/XSG8-NLU7 , https://www.datacamp.com/tutorial/pca-analysis-r
  T-SNE- https://www.datacamp.com/tutorial/introduction-t-sne
  UMap- https://pair-code.github.io/understanding-umap/

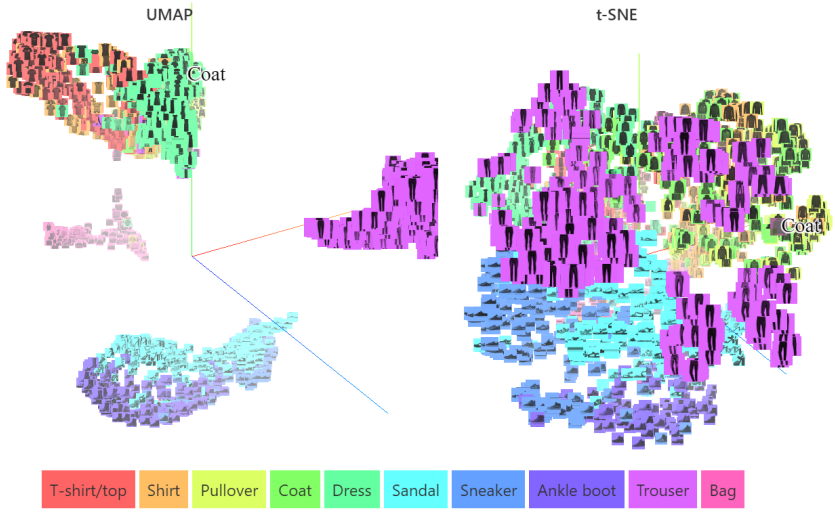


Figure 2: *Dimensionality reduction applied to the Fashion MNIST dataset. 28x28 images of clothing items in 10 categories are encoded as 784-dimensional vectors and then projected to 3 using UMAP and t-SNE.*

- Embedding projector - **COOL THING**