# Classifying Textual Data with pretrained Vision Models through Transfer Learning and Data Transformations

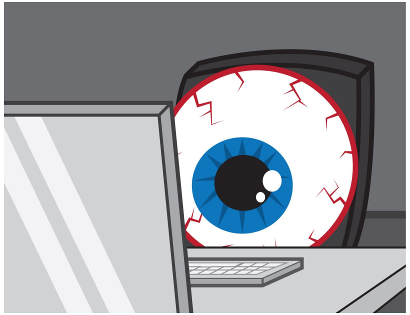
TeIAS | Tehran Institute for Advanced Studies

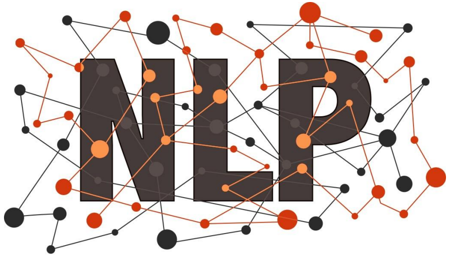
Melika Nobakhtian

# Introduction

# Introduction

- Natural Language Processing
- Computer Vision
- Transfer Learning
- Make Closer Langauge and Vision?





# In this work...

- Using knowledge from vision models to train text classifiers
- Creating image dataset from text dataset
- The main parts are:
  - Using BERT Embeddings from IMDB-text dataset to create IMDB-image dataset
  - Analyzing domain shifts between source and target datasets and avoid them
  - Using early layers of benchmark vision models as feature extractor
  - Training different models on these features

# Preliminaries

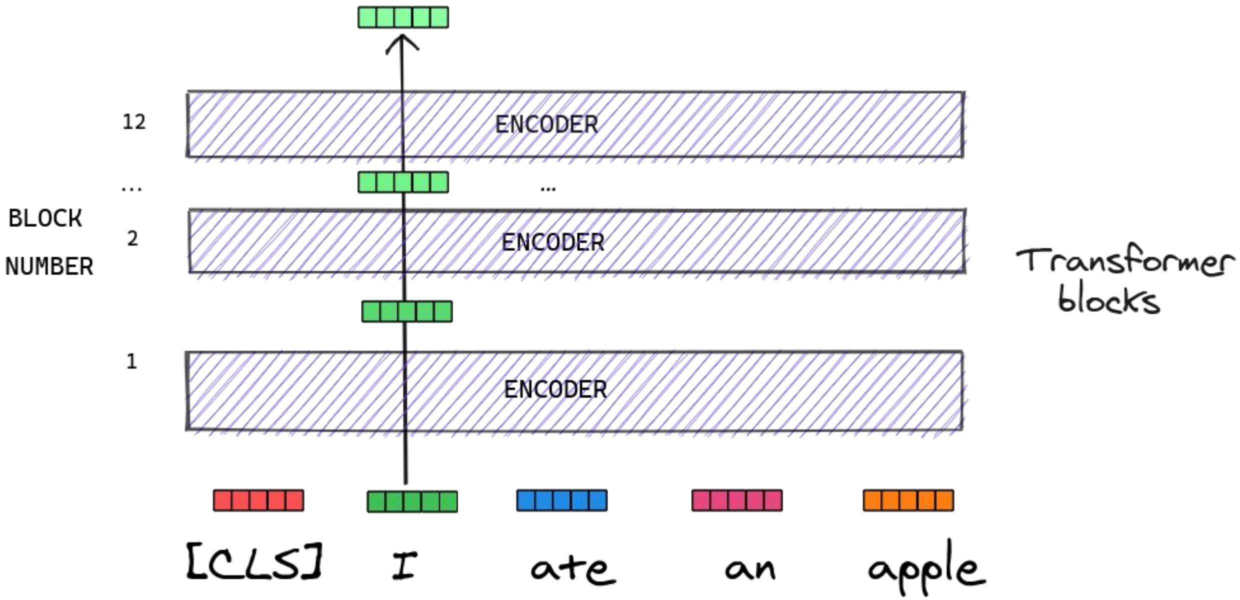
# IMDB Dataset

- For sentiment analysis or text classification
- 50000 examples in two classes: “Positive” or “Negative”
- Small and balanced

# BERT

- BERT (Bidirectional Encoder Representations from Transformers)
- Pre-trained on large unlabeled dataset
- Contextualized word embeddings (word representations vary based on context)
- Transferring across different NLP tasks

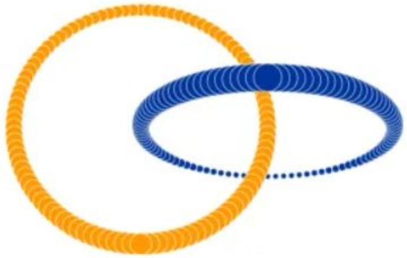
# BERT



12
ENCODER
...
...
BLOCK
NUMBER
2
ENCODER
Transformer
blocks
1
ENCODER
[CLS]
I
ate
an
apple

# t-SNE and DeepInsight Method

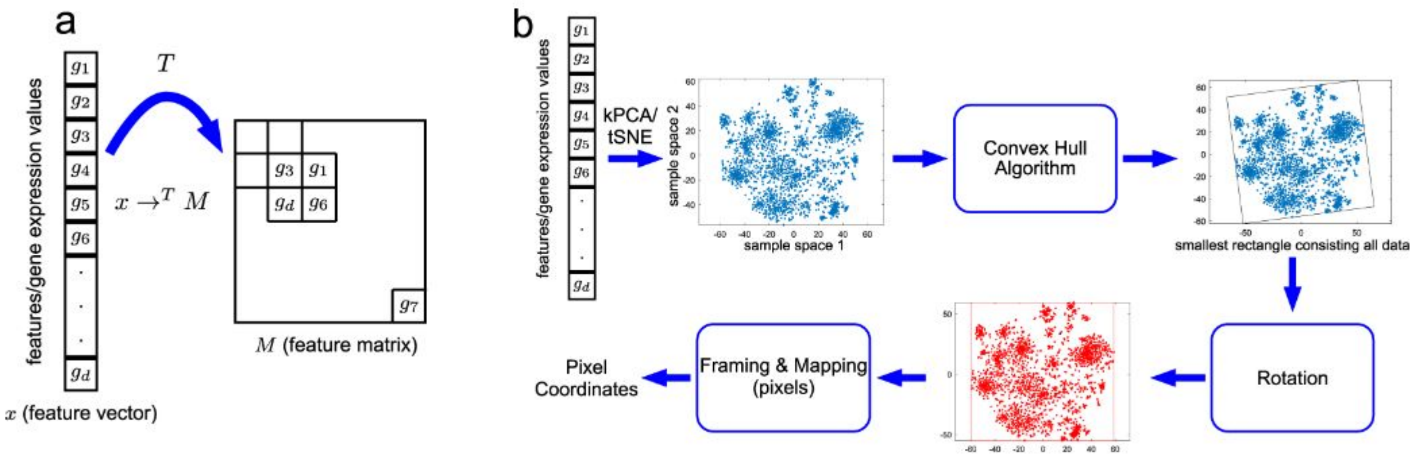
- t-SNE (t-distributed Stochastic Neighbor Embedding)
  - A dimensionality reduction method like PCA
  - Non-linear, unlike PCA
  - focuses on maintaining the pairwise similarities between data points in both the original and the reduced space



# t-SNE and DeepInsight Method

- DeepInsight
  - A methodology to transform non-image data to an image for CNN architecture
  - Useful where the data is often non-image, like DNA sequences
  - Leverage the power of CNNs
  - Using t-SNE or K-PCA  to project the transpose of dataset
  - Creating a set of related features on a 2D plane according to their similarity
  - Re-transposing to get the new samples with size [NxNx3]

# t-SNE and DeepInsight Method



# Method

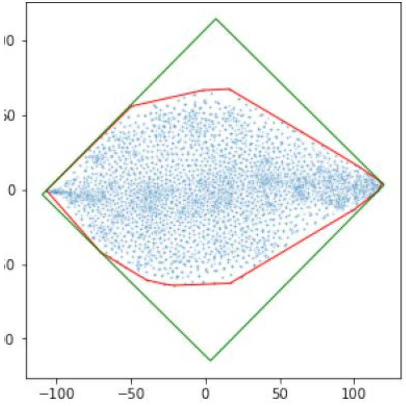
# Generating the IMDB-Image Dataset

# BERT Embedding Generation

- A very special feature of BERT
  - [CLS] token
  - Added at the beginning of a sentence embedding, at each layer output
  - A sentence beginning and a unique representation for classification purpose
- A high dimensional space to apply t-SNE
  - The semantic nature of higher layers of BERT
  - From the last six layers
  - [6x768] vector for each input sample

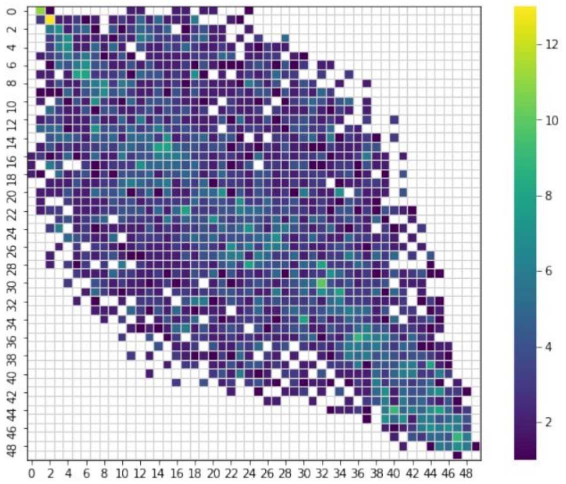
# Transforming BERT Embeddings to Images

- After that we obtained Embeddings, our dataset has this shape: [50000, 4608]:
  - Transpose the matrix to apply t-SNE
  - t-SNE projection will give us a 2D plane
  - Convex-Hull algorithm, to improve image quality
    - Isolating the rectangle containing all points
  - Rotating rectangle to get horizontal matrix of pixels



# Transforming BERT Embeddings to Images

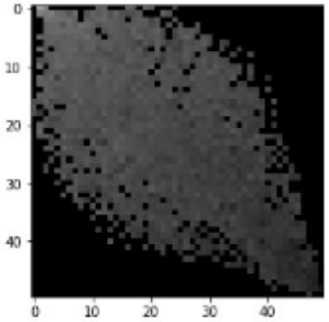
- Feature values are continuous:
    - Multiple features may be associated with a single location
    - Features are averaged to generate a discrete space of pixels
- Chosen pixel space: [50x50]



# Data Domains and Transfer Learning

# Domain Adoption

- Generated IMDB-Image dataset and ImageNet Dataset are extremely different
- Distribution mismatch and domain shift problems
- Solve this problem: **Domain Adoption**




(a)

(b)

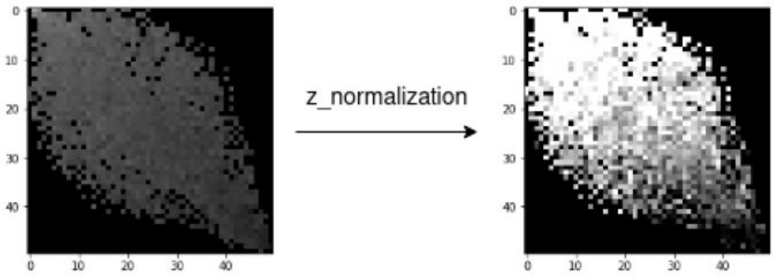
# Transfer Learning

- A successful Transfer Learning: an architecture able to adapt the Target Domain to Source Domain
- Here we focus on the shared features in both datasets:
  - Instead of forcing to learn domains
  - Our dataset is small-in-size and it will cause overfitting
  - Focus on geometric features like edges, curves, and blobs
  - Z-normalization: adjust image contrast and improve pixel space clarity

# Transfer Learning

- ***X*** is a set of input vectors, $\mu$ and $\Sigma$ are the mean and standard deviation of the entire image pixel space, $\epsilon$ is a small value to prevent dividing by zero. In the following, you can see an image before and after applying Z-normalization:

$$\mu = \mathbb{E}_x \in \mathcal{X}[x],$$
$$\sigma^2 = \mathbb{E}_x \in \mathcal{X}[(x-\mu)^2],$$
$$\hat{x}_i = \frac{x_i - \mu}{\sigma + \epsilon}$$

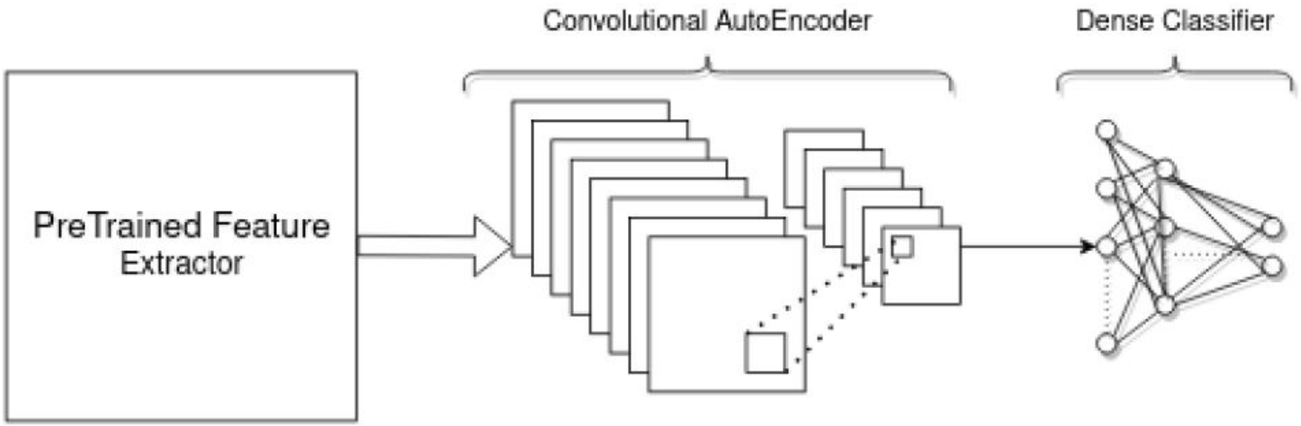


# Architecture

# Architecture

- The IMDB-Image dataset is too small compared to ImageNet
- Danger of overfitting in the process of training!
- Sliced Convolutional Feature Extractors from their original pretrained models
- Stacked to a Convolutional Auto-Encoder
    - With randomly initialized parameters
    - Followed by a Dense Classifier
- To avoid overfitting problem: freezing the feature extractors

# Architecture



# Architecture

- **AlexNet**: The first two pretrained Convolutional Layers, outputs 192 feature maps
- **ResNet:** The first downsampling Convolutional layer and the first residual layer
- **ResNext**: The first Convolutional layer and the first Residual layer
- **ShuffleNet V2**: The first Convolutional layer followed by Batch normalization and stage2
- **VGG16**: only the first 12 layers, containing 4 Convolutional layers

# Results

# Experiments

- 50000 samples: 40000 for training, 10000 for validation
- ReLU, Batch Normalization and Adam optimizer
- Batch size: 32
- Different Learning rates based on the model

# Original Results vs. Reproduced Results

| Feature Ext | Nbr of FM's | CAE LR | LC LR | Val Acc |
|---|---|---|---|---|
| AlexNet [14] | 192 | 0.00001 | 0.0005 | 0.87 ($\pm$0.01) |
| ResNet [9] | 256 | 0.00005 | 0.0001 | 0.85 ($\pm$0.01) |
| ResNext [26] | 256 | 0.00005 | 0.001 | 0.85 ($\pm$0.01) |
| ShuffleNet [15] | 116 | 0.0005 | 0.001 | 0.86 ($\pm$0.01) |
| VGG16 [20] | 256 | 0.00005 | 0.001 | 0.86 ($\pm$0.01) |

| Feature Ext | Val Acc |
|---|---|
| AlexNet | 0.801 |
| ResNet | 0.824 |
| ResNext | 0.819 |
| ShuffleNet | 0.812 |
| VGG16 | 0.813 |

# Original Results vs. Reproduced Results

**We have different results, but why?**

- Validation set a part of main dataset: We don't know the exact samples which are in it
- Text preprocessing techniques
- Normalization or scaling before t-SNE
- Random weights for Conv-AE


THANKS FOR YOUR ATTENTION
NOW
GIVE A 10 TO THIS TEAM