**Introduction / Background**

Contrastive Learning has been widely explored in computer vision to improve visual representations of objects. This project aims to apply contrastive learning methods to improve the quality of sentence-level embeddings, as an alternative to masked language modeling (MLM). MLM is a token-level objective which does not perform too well on topic prediction tasks. This project demonstrates that a wider span-level contrastive objective can perform better on topic prediction. I explore a variety of views, expanding on the DeCLUTR paper.

**Methods**

SimCLR Objective Function:

\[
\text{sim}(u, v) = \frac{u^T v}{\|u\|\|v\|}
\]

\[
\ell_{ij} = -\log \sum_{k \neq j} \exp(\text{sim}(z_i, z_j)/\tau)
\]

\(z_1, z_2\) are views

Longformer Model:

![Longformer Model Diagram]

**View Selection**

Example: I went to the bakery to buy a loaf of bread. It was very crispy and tasty. The loaf cost $5.00.

**Slice views:**

View 1: I <MASK> the bakery <MASK> loaf of bread. View 2: I went to <MASK> to buy a <MASK>.

**Neighboring views:**

View 1: I went to the bakery to buy a loaf of bread. View 2: It was very crispy and tasty.

**Neighboring Slice views:**

Example: I went to the bakery to buy a loaf of bread. It was very crispy and tasty. The loaf cost $5.00.

View 1: I <MASK> the bakery <MASK> loaf of bread. View 2: It was <MASK> crispy and <MASK>.

**Negative views:**

Negative views are different document spans within the same batch, which was used for calculating SimCLR loss.

**Experiment**

1. Fine-tune 4 models on Wikipedia Dataset
   a. Slice-longformer, neighbor-longformer, neighbor-slice-longformer, roberta-longformer MLM

2. Train Logistic Regression Model on longformers

3. Evaluate on Transfer Tasks:
   a. Sentiment Analysis: SentEval SST2 binary classification
   b. Topic Prediction: 20 Newsgroup dataset

**Results & Analysis**

- Models performed slightly above chance (50%) for SST2
  - Wikipedia Dataset written in unbiased tone
- The slice-views and neighbor-views outperformed roberta-longformer MLM on topic prediction
  - Contrasting masking and neighboring spans results in better sentence-embeddings for topic prediction
- The wide span views allow the contrastive model to pick up on longer range topics, inter-sentence context, and latent semantics.

**Future Work**

- Explore more transfer tasks to see if contrastive learning models are generating better representations for wider range of tasks
- Scale up the size of the models