第12回ディスカッション

論文調査と実験

ER20038 小林亮太

担当：鈴木雅★，福井，張

MPRG

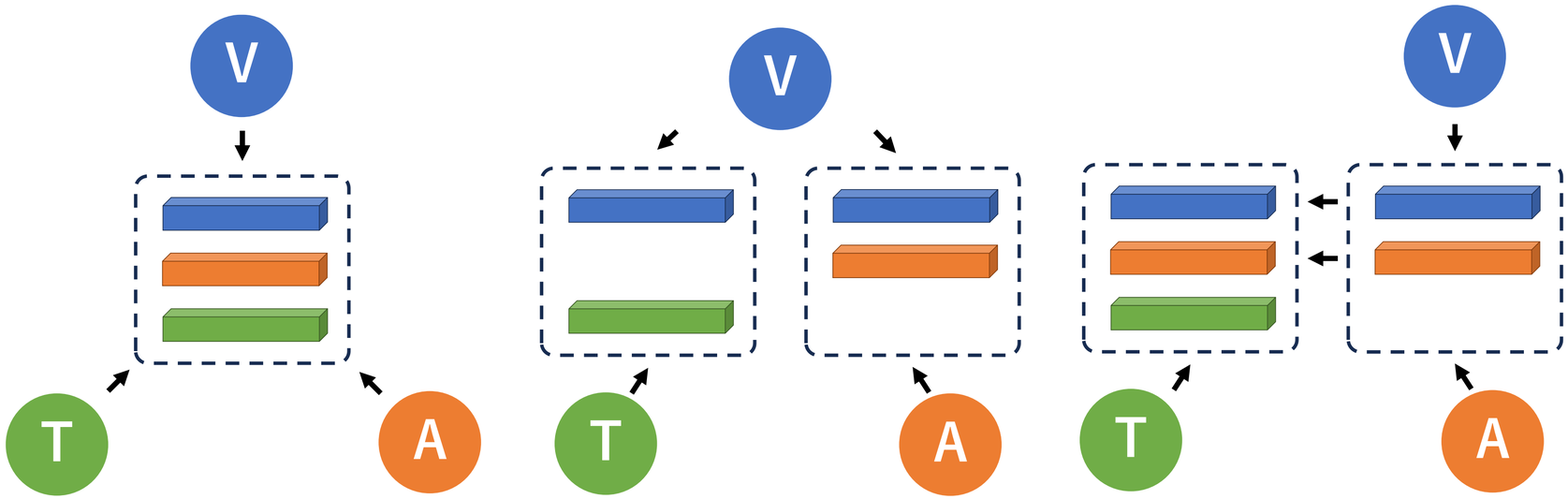
MACHINE PERCEPTION AND ROBOTICS GROUP

はじめに

- 研究テーマ
- CLIP^2: Contrastive Language-Image-Point Pertaining from Real-World Point Cloud Data
- 応用
- 実験状況

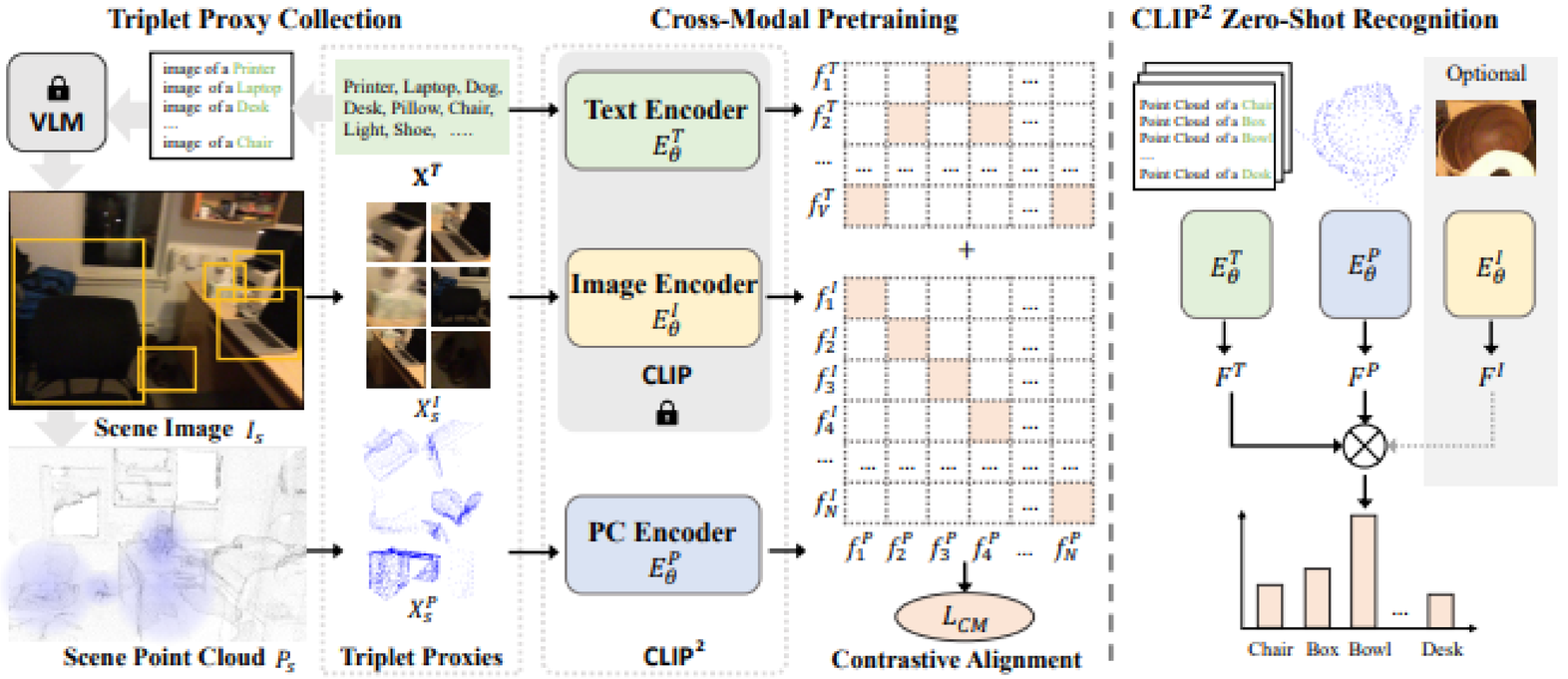
研究テーマ

- 3モーダル（ビデオ，オーディオ，テキスト）のマルチモーダル自己教師あり学習
- テキストに比べビデオやオーディオにはノイズが多く存在
  - 各モーダルの組み合わせでノイズを抽出せずに学習ができる可能性
    - 近づけるモーダルの組み合わせによる学習効果への影響について調査



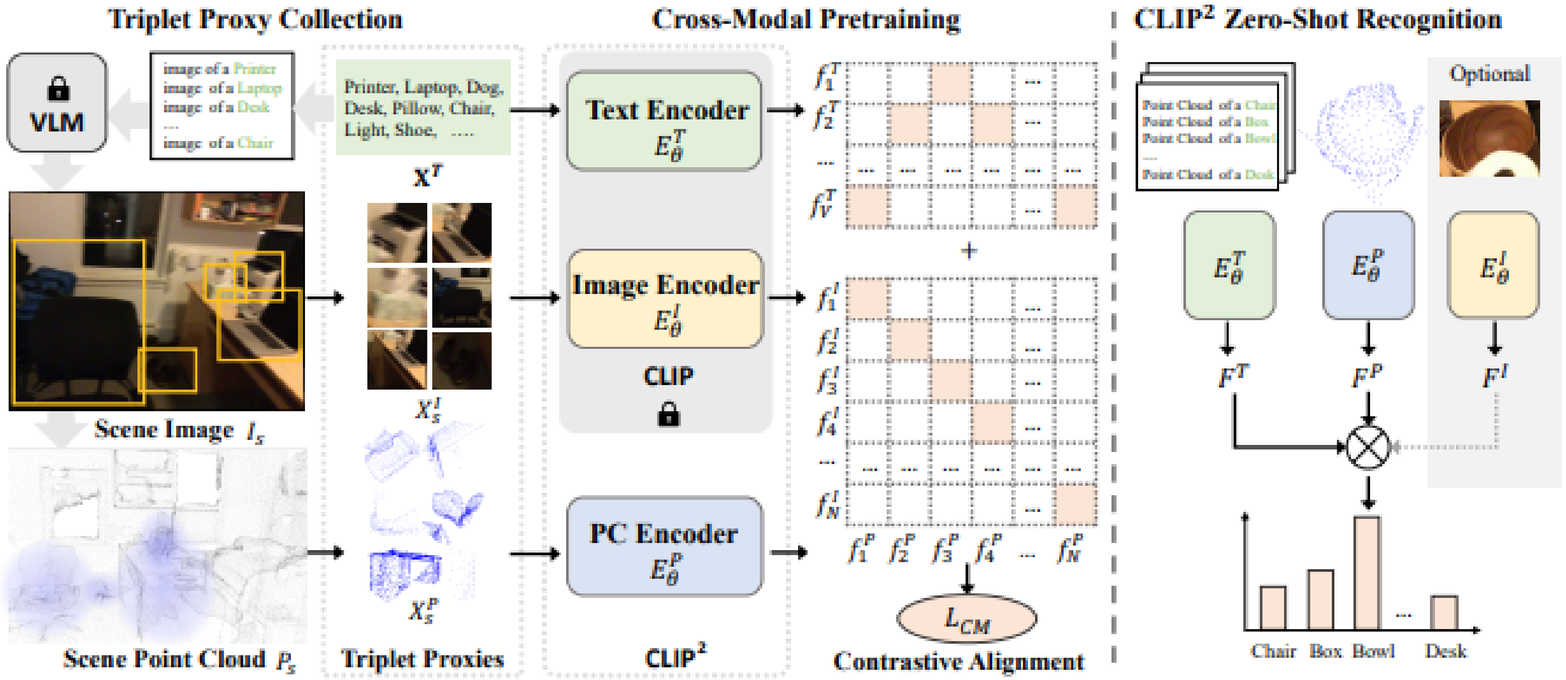
$CLIP^2$: Contrastive Language-Image-Point Pertaining from Real-World Point Cloud Data [Y. Zeng+, CVPR'23]

- CLIPのテキストと画像に点群を追加
- 事前学習済みのCLIPを利用
  - 新たに点群のエンコーダを学習



応用

- 入力をテキスト，オーディオ，ビデオの組み合わせに適用
  - 画像の領域　→動画のクリップのフレーム
  - 画像の領域に対応した点群　→動画のクリップに対応したオーディオ
  - 画像の領域の物体の単語　→動画のクリップの字幕



実験状況

- 事前学習の実行時設定ミス
  - オーディオとビデオの2モダールで学習
- 再度3モーダルで事前学習
  - メモリ関連のエラー
    - バッチサイズやnum_workerの値の変更で一時的に対処
  - LossがNaNになる問題
    - 対処中

おわりに

- CLIP^2　　：研究への応用
- 実験　　：エラー対応中

- 今後の予定：CLIP2の論文調査

実験結果

- 事前学習が完了
- hoge