第25回ディスカッション

実験状況

ER20038 小林亮太

担当：鈴木雅★，福井，張

MPRG

MACHINE PERCEPTION AND ROBOTICS GROUP

はじめに

- 実験概要
- 実験条件
- 実験状況
- 評価結果の検証

実験概要

- 3モーダルを二段階で学習実験
  - 3パターンを実験
    - AV_T：AとVで学習した後からTを追加
    - VT_A：VとTで学習して後からAを追加
    - AT_V：AとTで学習して後からVを追加

- HowTo100Mデータセットを用いて学習
- YouCook2，MSR-VTTの2つのデータセットを用いてゼロショットで評価
- テキストからビデオの検索タスクで評価
  - テキストによるビデオ内の該当箇所の検索

- エポック数を増加させた追加の実験

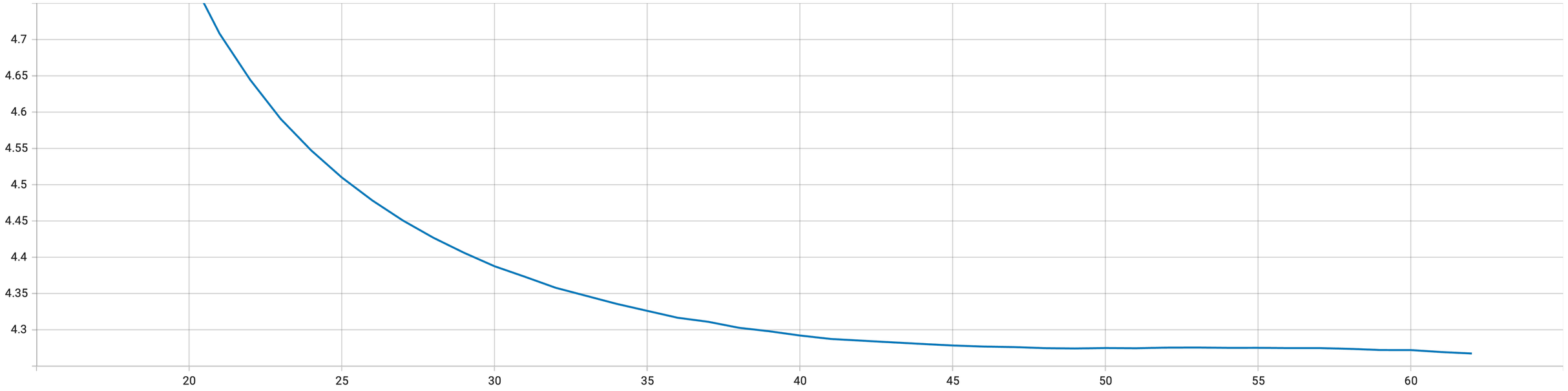
実験条件

- Feature Extractor
  - ビデオ　　：ResNet152
  - オーディオ　　：DaveNet [D Harwath+, ECCV’18]
  - テキスト　　：Word2vec
- バッチサイズ　　：128
- 学習率　　：0.0001
- 最適化手法　　：Adam
- GPU　　：A100 × 4

- エポック数　　：未定
  - Epoch数を多めに設定し，lossの推移を確認しepoch数を決定

実験状況

- VT_Aのバターンの一段階目で実験
- 45エポックまでlossが減少を確認



評価結果の検証

- 検索結果のカテゴリの分布を確認
  - 一部のカテゴリが検索結果上位を占領している可能性

- R@k
  - R：再現率 (Recall)
  - k：検索結果の上位k個内に正解が含まれている確率

6

$$\mathrm{R@k} = \frac{(\text{上位k個内の正解数})}{(\text{上位k個内の正解数}) + (\text{上位k個以外の正解数})}$$

$$= \frac{(\text{上位}k\text{個内の正解数})}{\text{検索回数}}$$

おわりに

- 実験　　：必要なエポック数を確認

- 今後の予定：
  - 他の組み合わせでの追加の実験

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