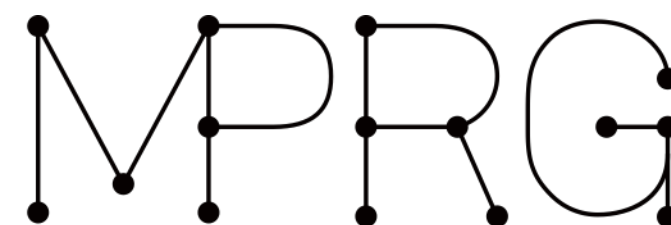


第18回ディスカッション

バッチサイズとメモリエラー

ER20038 小林亮太

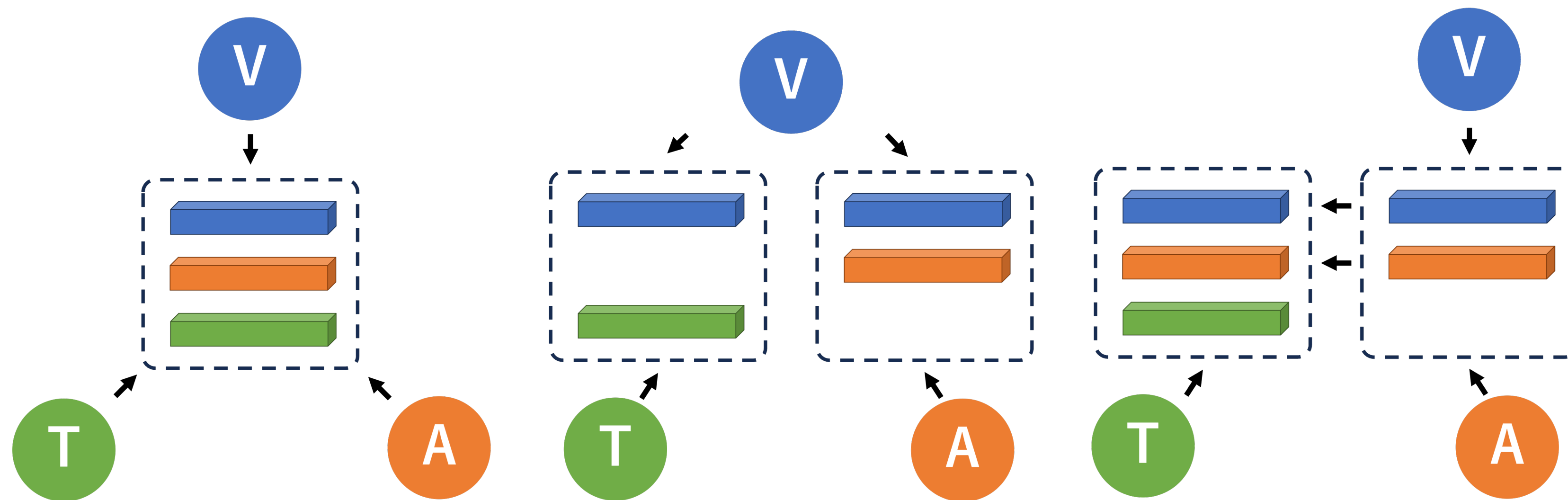
担当：鈴木雅★，福井，張



MACHINE PERCEPTION AND ROBOTICS GROUP

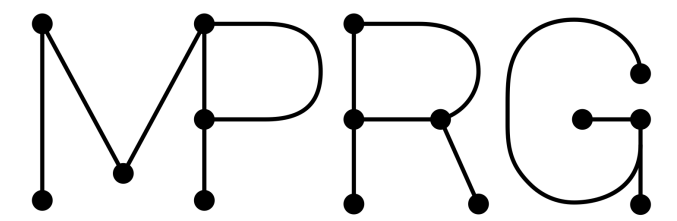
- 研究テーマ
- 実験条件について
- 条件batch_size=128, n_pair=32の場合
- 実験状況

- 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



- 実験条件を論文と揃える目的
 - 論文内の表記を参照しバッチサイズを4096に設定
 - メモリ不足エラーが出現
 - バッチサイズ2048, 1024, 512, 256でもエラーが出現
 - 解決が難航
 - 論文内の表記は直接的なバッチサイズではない値
 - Batch_size × n_pairの値
 - Batch_size : 動画の数
 - N_pair : 動画1本あたりのclip数
 - それぞれ初期値はbatch_size=128, n_pair=32
 - 2つの積は4096

条件batch_size=128, n_pair=32の場合



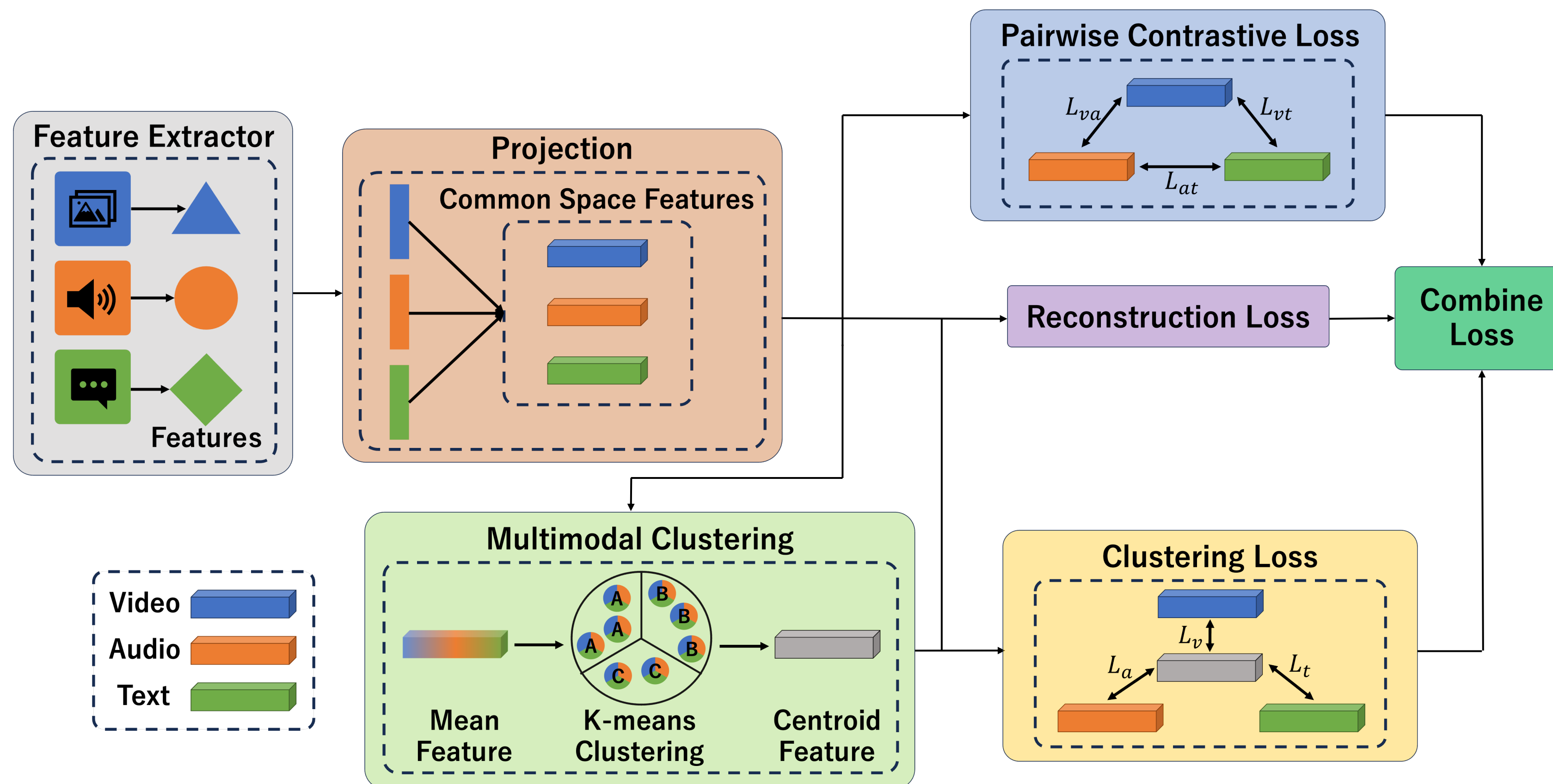
- A 100 × 4で0番の使用量が13~15GB
 - 著者の環境のV 100が16GBモデルだとしても動作可能
 - 情報が合致
- 学習時間
 - 著者環境で2日程度
 - CPU : 80スレッド
 - Num_worker : 74
 - 自分の環境では1週間程度
 - CPU : 32スレッド
 - Num_worker : 4
 - 大きくしすぎるとRAMが不足
 - CPU性能およびRAM搭載量の差が原因の可能性

- デフォルトの条件のまま実行
 - Num_workerの値のみ変更
- Core dumpで停止
 - 原因を調査中

- 実験 : デフォルトに限りなく近い状態で実行中
- 今後の予定 :
 - 中中間の資料作成
 - 他の環境での実験

Multimodal Clustering Network (MCN) [B. Chen+, ICCV'21]

- ラベル付けされていないナレーション付きビデオから学習
 - テキストからビデオの検索, 時系列行動検出が可能
- テキスト, オーディオ, ビデオの3つのモーダルを使用



- MCNが参考になっているAVLnet
 - コードの流用部分が存在
 - 同様にHowTo100Mデータセットを使用
- AVLnet著者の環境
 - GPU : V100×4 MCN著者と同一
 - CPU : 80スレッド
 - RAM : 不明
- MCN著者も同等の環境の可能性大

- 評価タスク：テキストからビデオの検索 on YouCook2
 - R@k
 - R：Recall（再現率）
 - K：各クエリの上位k個の予測のうちの正解数の総正解数に対する割合
- 平均を取る処理の追加で精度向上
- バッチサイズを論文の条件と揃えることが必要

	R@1	R@5	R@10
論文	18.1	35.5	45.2
前々回の実験	1.70	7.00	11.6
前回の実験	5.29	13.2	18.4