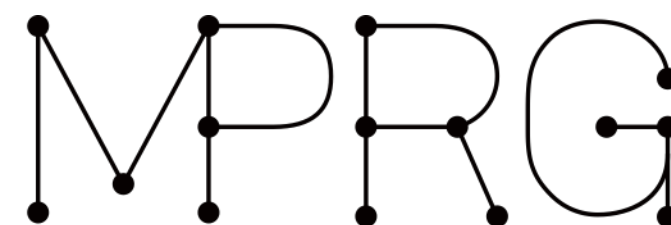


第24回ディスカッション

追加の実験

ER20038 小林亮太

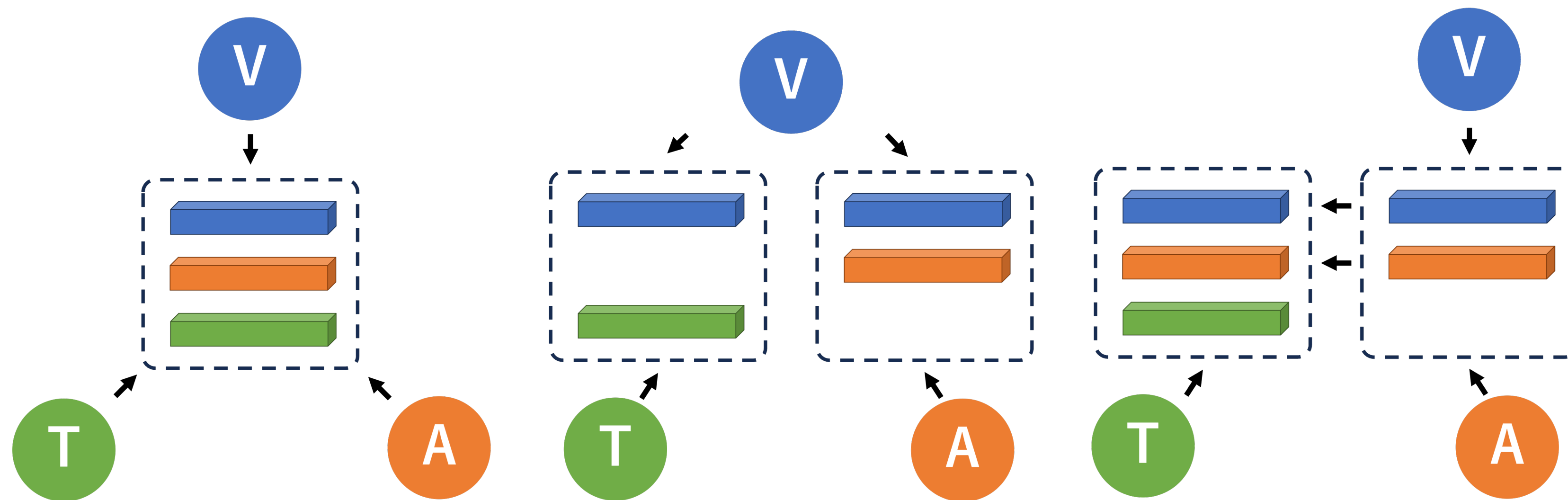
担当：鈴木雅★， 福井， 張



MACHINE PERCEPTION AND ROBOTICS GROUP

- 研究テーマ
- 実験概要
- 実験条件

- 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



- 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で学習した後からAを追加
 - AT_V : AとTで学習した後から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

- 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

- サーバのメンテナンスに伴い中断

- 実験 : サーバのメンテにより中断中
- 今後の予定 :
 - 追加の実験