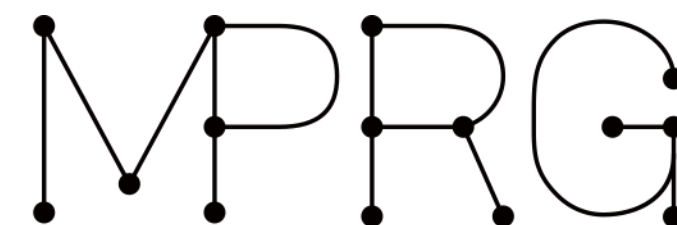


第17回ディスカッション

メモリエラー

ER20038 小林亮太

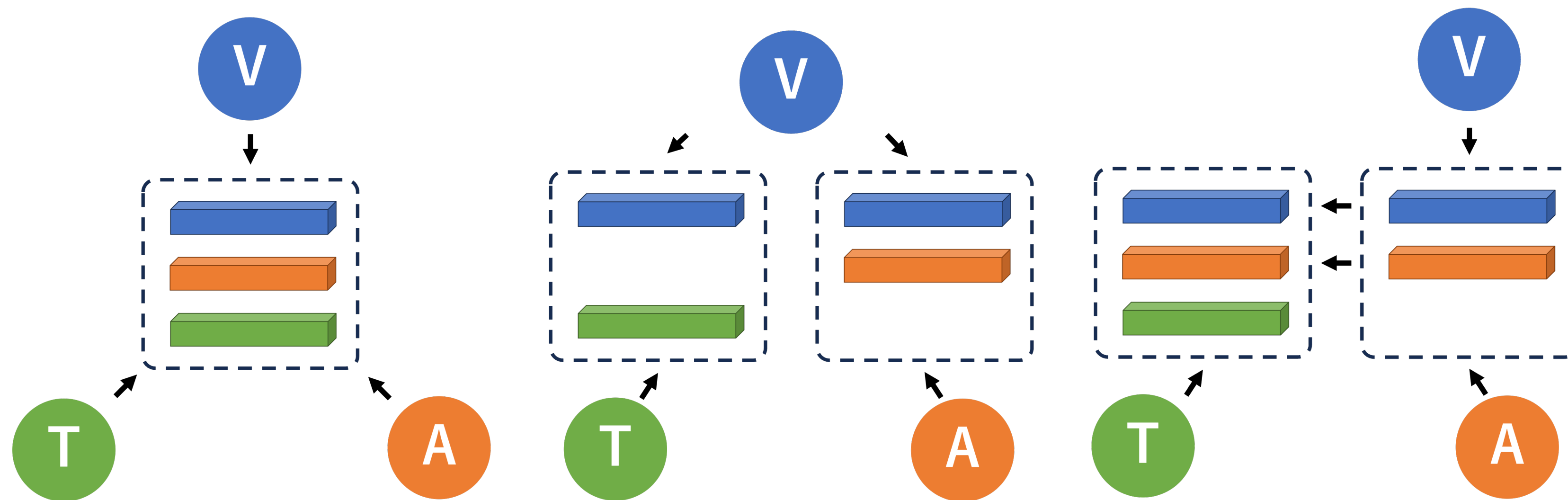
担当：鈴木雅★， 福井， 張



MACHINE PERCEPTION AND ROBOTICS GROUP

- 研究テーマ
- 実験状況
- DDP

- 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



- 評価タスク：テキストからビデオの検索
 - 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

| | R@1 | R@5 | R@10 |
|-------|------|------|------|
| 論文 | 10.5 | 25.2 | 33.8 |
| 前回の実験 | 1.70 | 7.00 | 11.6 |
| 今回の実験 | 5.29 | 13.2 | 18.4 |

- バッチサイズによるメモリエラー
 - GPUメモリの不足 : 現在対処中
 - 論文 : V100 × 4
 - 自分 : A100 × 4
 - ほとんどの場合, 0番が原因
 - DDPの導入で解決する可能性

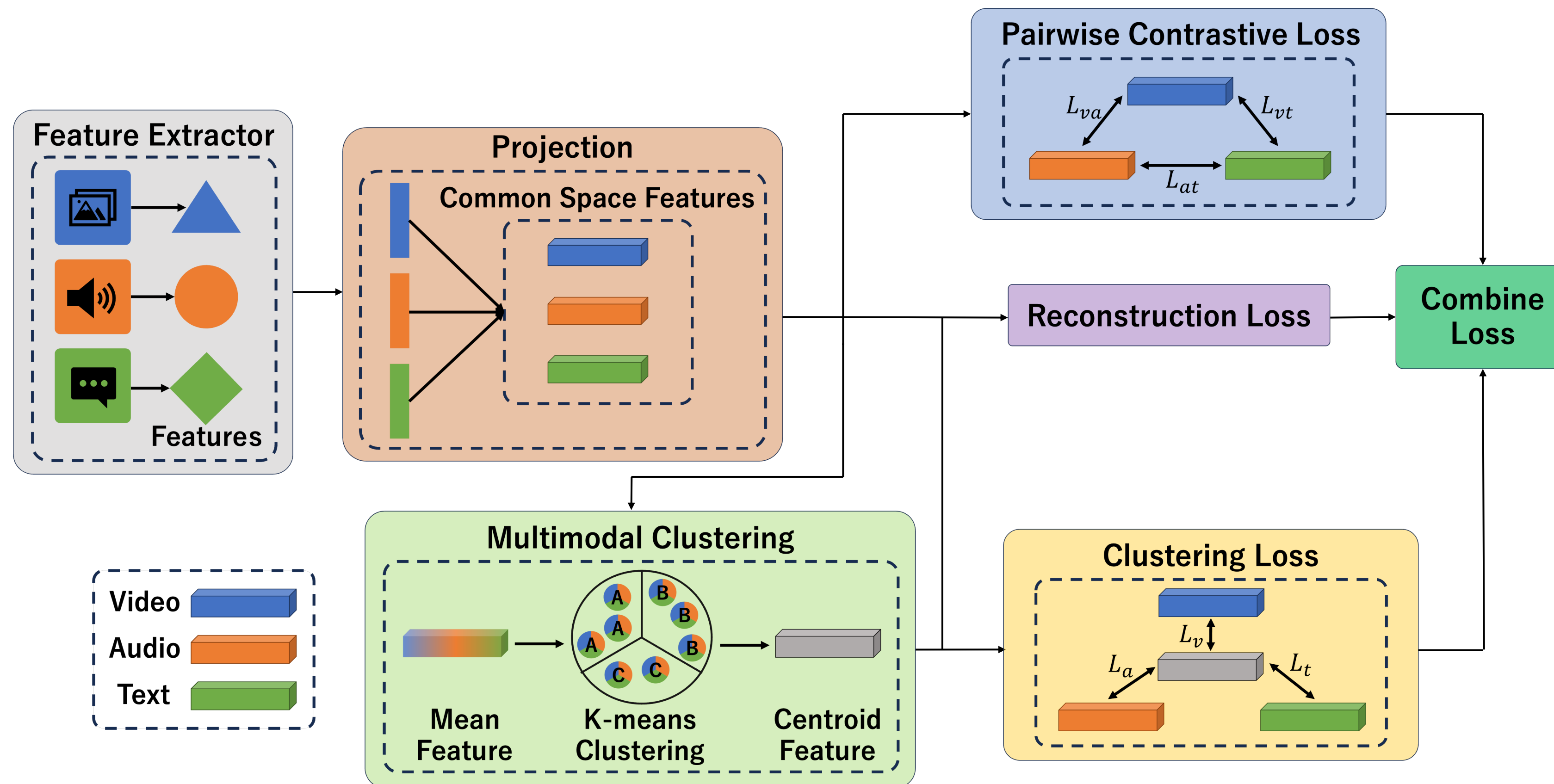
```
Traceback (most recent call last):
  File "train_tri_kmeans.py", line 529, in <module>
    batch_loss, queue_v, use_the_queue, centroid, loss_cont, loss_val, loss_recon = TrainOneBatch(net, optimizer, sample_batch, loss_op, queue_v, use_the_queue, scheduler, save_epoch, i_batch, centroid, apex)
  File "train_tri_kmeans.py", line 361, in TrainOneBatch
    loss_at = loss_fun(sim_audio_text)
  File "/usr/local/lib/python3.7/dist-packages/torch/nn/modules/module.py", line 1194, in _call_impl
    return forward_call(*input, **kwargs)
  File "/home/kobayashi/c/Multimodal-Clustering-Network/loss.py", line 20, in forward
    C2I_loss = F.nll_loss(F.log_softmax(S.t(), dim=1), target)
  File "/usr/local/lib/python3.7/dist-packages/apex/amp/wrap.py", line 28, in wrapper
    return orig_fn(*new_args, **kwargs)
  File "/usr/local/lib/python3.7/dist-packages/torch/nn/functional.py", line 1930, in log_softmax
    ret = input.log_softmax(dim)
torch.cuda.OutOfMemoryError: CUDA out of memory. Tried to allocate 1024.00 MiB (GPU 0; 39.59 GiB total capacity; 37.60 GiB already allocated; 532.62 MiB free; 38.46 GiB reserved in total by PyTorch) If reserved memory is >> allocated memory try setting max_split_size_mb to avoid fragmentation. See documentation for Memory Management and PYTORCH_CUDA_ALLOC_CONF
```

- DDPの導入が複雑
 - Hugging faceのAcceleratorで代用可能？
- Acceleratorの導入
 - 一応導入は完了
 - 現在，実行中

- 実験 : Acceleratorを実装中
- 今後の予定 :
 - プログラムの作成

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



- アーキテクチャ : MCN
- Feature Extractor :
 - ビデオ : ResNet152
 - オーディオ : DaveNet
 - テキスト : Word2vec
- データセット : HowTo100M
 - ビデオ解像度 : 454×256
 - ビデオフレームレート : 30FPS
 - オーディオサンプリングレート : 16kHz
- バッチサイズ : 128
- エポック数 : 30
- 学習率 : 0.0001
- 特徴量次元数 : 4096

- DPの方がGPU:0を起点としたGPU間通信の回数が多い
 - DPの方が処理が遅くなりやすい

| | DP | DDP |
|----------------------------|---------------------------------------|------------------------------|
| Dataloader | GPU:0のみ所有 | 各GPUで所有 (この時点でDataset分割済) |
| Batchの分割 (mini batch作成) | GPU:0で実行 | なし (すでに分割済) |
| Mini batchのGPU間共有 | GPU:0から他のGPUに送信 | なし (すでに分割済) |
| Loss計算 | 他のGPUからGPU:0にoutputs を送信してGPU:0で実行 | 各GPUで実行 |

