

P20: KoSign データを拡張するためのData Augmentation手法の検討

中村友里也、荊雷（会津大学大学院）

1. 研究背景・目的

➤ 手話認識の重要性

聴覚に障害のある人とのコミュニケーションを円滑に

➤ 手話認識の現状

多くのデータセットにおいて一つの手話単語に対するデータの不足
→過学習、精度の低下

➤ 研究目的

データ拡張を用いた手話認識手法の提案

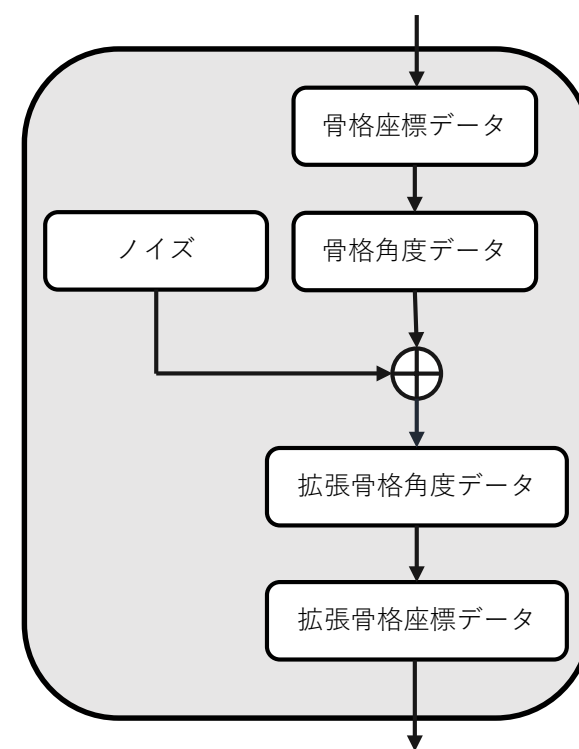
2. 手法

➤ 骨格推定

- MediaPipe[1]を用いて手、顔、腕の骨格座標を推定

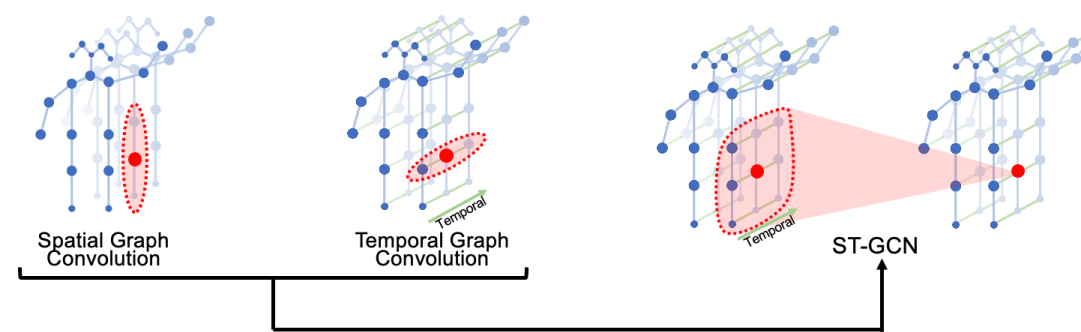
➤ データ拡張

- 骨格座標から指の角度を求める
- ガウス分布に従ったノイズを加える
- 順運動学より角度データを座標データに変換



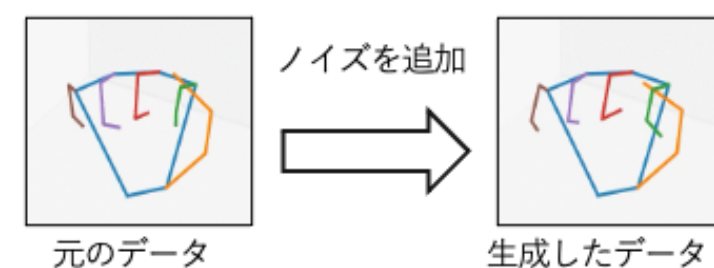
➤ 手話認識

- Spatial Temporal Graph Convolutional Networks (ST-GCN)[2]を用いて、手話を分類する。
- 関節座標をノード、関節の関係性をエッジとしたグラフ構造と見なす
- 骨格データを時間、空間の2つのグラフ構造として捉える



3. データセット

- 男性手話者一名を三方向(左右、正面)から撮影したmp4動画を利用
- 無作為に50個の手話単語を選択し、合計150動画を使用
- 提案したデータ拡張を用いてデータサイズを5倍に拡張



4. 結果

➤ 学習方法

- 訓練用に二つの視点の動画を使い、テスト用に残った視点の動画
- それぞれの視点を一回ずつテスト用とし、訓練を行い平均精度を求める

➤ 損失関数

- 交差エントロピーを使用

$$H(p, q) = - \sum_x p(x) \log(q(x))$$

p : 真の確率分布
 q : 推定した確率分布

➤ 実験結果

	Without DA	DA
Accuracy	63.6 %	71.2 %

Data Augmentation (DA)



5. まとめと今後

- 本研究はデータ拡張を用いた手話認識モデルを検討し、精度の向上を達成した。
- 今後は本研究のデータ拡張で用いているノイズの最適化を検討する。

6. 参考文献

- [1] Lugaresi, C., "MediaPipe: A Framework for Building Perception Pipelines", 2019.
[2] Yan, S., Xiong, Y., and Lin, D., "Spatial Temporal Graph Convolutional Networks for Skeleton-Based Action Recognition", 2018.

7. 謝辞

本研究では、国立情報学研究所のIDRデータセット提供サービスにより工学院大学から提供を受けた「工学院大学 多用途型日本手話言語データベース (KoSign)」を利用した。

