敵対的学習を用いた骨格ベースの手話認識

中村友里也、荊雷（会津大学大学院／コンピュータ理工学研究科）

1．背景

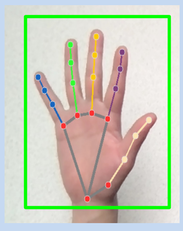
○手話認識の必要性

- 聴覚障害のある方は約４億６６００万人
- 手話を利用しない人にとって手話は難しい

○手話認識の種類

視覚ベース

- カメラを利用
- 非効率的
- 実用的



デバイスベース

- センサーグローブなどの計測装置を利用
- 効率的
- 実用性に欠ける

○手話認識の問題点

- データ収集、アノテーションにコストがかかる

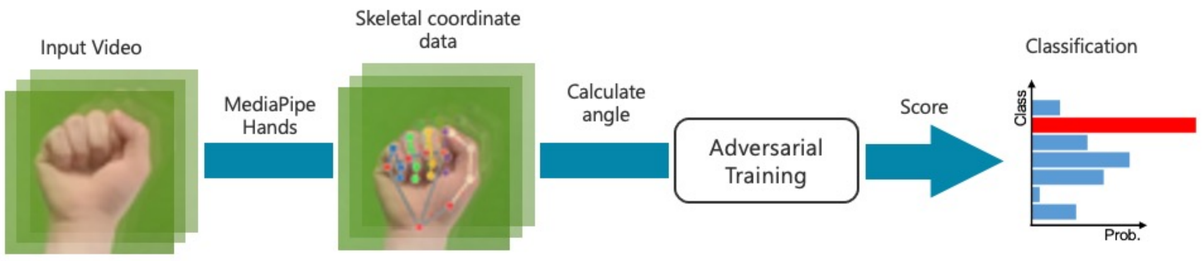
→データ不足

機械学習において精度低下、過学習の原因

目的

小規模な手話データセットを用いた視覚ベースの手話認識の精度を改善し、過学習を抑制したい

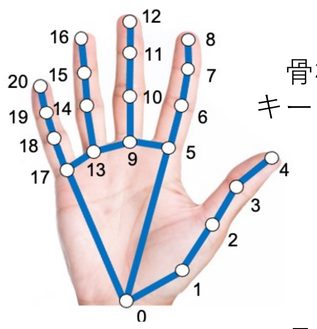
2．提案手法



提案手法のアーキテクチャ

STEP 1　手の姿勢推定

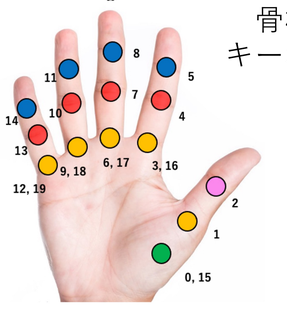
- MediaPipe Handsを用いる
- ３次元骨格座標を出力
- ２１個のキーポイント



骨格座標キーポイント

STEP 2　関節角度の抽出

- 内積を用いる
- ０～１４：垂直角度
- １５～１９：水平角度
- ２０個のキーポイント



骨格角度キーポイント

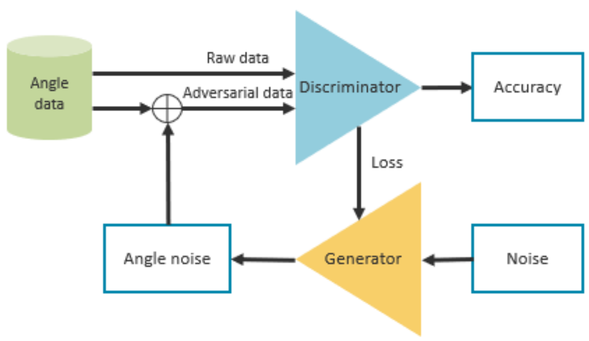
STEP 3　敵対的学習

○生成器(Generator)

- 識別器の損失を増加させる可能性のある敵対的データを出力するように学習

○識別器(Discriminator)

- より良い性能を得るために敵対的データから学習



敵対的データの損失　　生データの損失

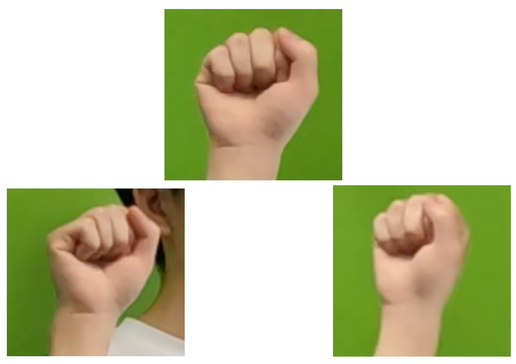
$$\max_G \mathbb{E}_{\substack{x\sim p_{data}(x)\\ z\sim p_z(z)}} \mathcal{L}[D(G(z)+x),y)] - \mathcal{L}[D(x),y)]$$

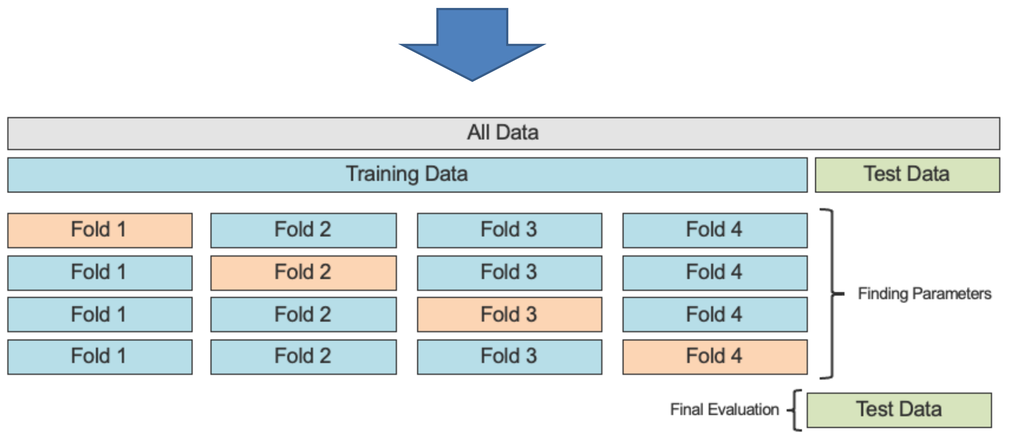
$$\min_D \mathbb{E}_{\substack{x\sim p_{data}(x)\\ z\sim p_z(z)}} \mathcal{L}[D(G(z)+x),y)] + \mathcal{L}[D(x),y)]$$

生成器: G
識別器: D
ノイズ: z
生データ: x
正解ラベル: y
損失関数: $\mathcal{L}$

3．データセット

- 被験者：５人(非母国語手話者)
- カメラ：3台
- 背景：グリーンバック
- 回数：１つの手話に対して５回
- 手話：２５種類の日本語指文字
- 合計：１８７５個の動画(MP4)





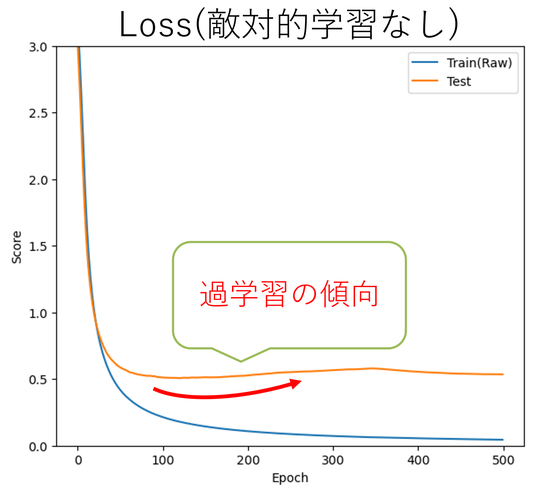
4．評価実験

○実験結果

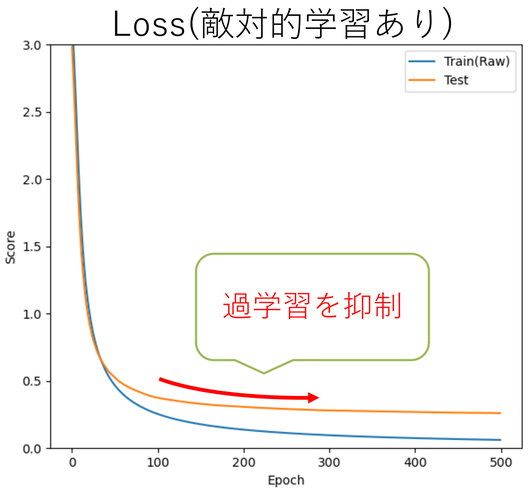
テストデータの評価値

学習方法	accuracy	recall	precision	f-score
敵対的学習なし	89.16	0.864	0.855	0.859
敵対的学習あり	93.75	0.907	0.907	0.907

Loss(敵対的学習なし)



Loss(敵対的学習あり)



○考察

- 敵対的学習ありの結果は敵対的学習なしの結果と比べて優位となった。
- 敵対的学習なしは過学習の傾向があるのに対し、敵対的学習ありでは過学習を抑制していることが確認できる

- 敵対的学習は識別器にとって良い効果を与えていると予想できる

5．おわりに

○まとめ

- 小規模の手話データセットにおける手話認識を敵対的学習を用いて行う手法を提案した
- 過学習を抑制し、精度を改善した

○今後の課題

- 角度データでは手の向きがわからない
- モデルが複雑であり、学習に時間がかかる

→

- 準運動学を用いて角度データを座標データに変換する
- コードの最適化など