G-14

音声ラベルを用いた 3DCNN における効率性と頑健性の検証

A Study of Efficiency and Robustness Validation in 3DCNN using Speech Labels

○王雨桐[1]，　関弘翔[2]，　細野裕行[2]

*Yutong Wang[1], Hiroto Seki[2], Hiroyuki Hosono[2]

Abstract: In this research, we build the human motion classification task by using speech label as the supervised signal and evaluate the availability. We compare the learning conditions for this model using different amounts of data with a categorical label model, leading to the result that the use of speech labels improves the data efficiency. Then, the robustness of the speech label model is examined by the results of adversarial attacks.

1.　まえがき

動画像の分類は，画像の分類ほど確立されていない課題である．人物動作認識で使用されるディープニューラルネットワーク(DNN)には，画像認識で使われる二次元の畳み込みニューラルネットワーク(2D-CNN)を時間軸方向に拡張した三次元の畳み込みニューラルネットワーク(3D-CNN)を使用するモデルなどがある[1]．

教師あり学習における教師信号として音声ラベルを使用することで，学習データ量が少ない場合，より識別性の高い特徴の学習を促すことが報告されている[2]．

本研究の目的は，通常のカテゴリラベルでなく，高次元，高エントロピーの音声ラベルを用いて，頑健性と効率性の高い動画分類器を構築し，有用性を評価することである．

2.　ラベルの処理およびモデルの構築

UCF101 データセット[3]のテキストラベルからテキスト音声合成システムを用いて音声化し，音声をスペクトログラム化という流れで処理を行った．本研究では，この人物動作認識用のデータセットを対象として，動画エンコーダと音声ラベルデコーダの 2 つの部分から構成する動画識別用 3D-CNN を提案し，構築する[4]．

3.　効率性の評価

構築したモデルに対して，カテゴリラベルと音声ラベルを用いて，少量な学習データの場合の認識精度を比較した．不均衡を避けるために，交差検証手法の Stratified K-Fold[5]を利用して層化抽出を行い，元データセットのクラス分布を反映するようにデータ量を削減した．Fig. 1 に音声ラベルとカテゴリラベルの学習状況を示す．結果より，音声ラベルを用いたモデルの検証精度が，最大 10%程度高いことが分かった．

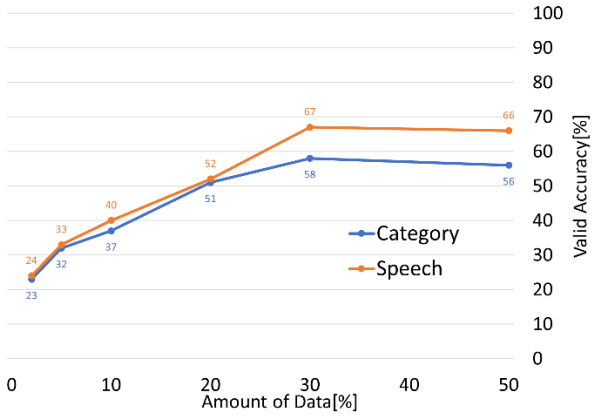

Figure 1. Validation accuracy when the model is trained for 100 epochs with speech labels and category labels

特にデータ量が 20%以上，50%以下の場合，音声ラベルのデータ効率性が比較的高いことを示している．

4.　頑健性の評価

構築したモデルに対して，敵対的攻撃手法である Geo-TRAP[6]を用いて，頑健性を検討する．学習データ量 50%，30%，20%，10%，5%，2%の場合，音声ラベルとカテゴリラベルを用いた二種類のモデルに対して敵対的攻撃を実施する．そして，敵対的攻撃の結果を比較することで異なるモデルの頑健性を検討していく．

5.　まとめ

本研究では音声ラベルを用いた人物動作分類器を提案し，このモデルに対して，異なるデータ量を用いた効率性と頑健性を評価した．カテゴリラベルのモデルと比較し，データ量が一定範囲内では音声ラベルの利用がデータ効率性を向上させる結果になった．今後は，異なる言語の音声ラベルがモデルの学習にどう影響するかの課題を検討する．

参考文献

[1] Hara Kensho, et al. , Proc. of CVPR. pp. 6546-6555. 2018.
[2] Boyuan Chen, et al. , ICLR 2021
[3] Khurram Soomro, et al., CRCV-TR-12-01, 2012
[4] 王雨桐，他，第 66 回日大理工学術講演会. G-1. 2022.
[5] Juan D. Rodriguez, et al., IEEE trans. on PAMI, 32(3), 569-575. 2009.
[6] Shasha Li, et al. NeurIPS2021, 34: 2085-2096. 2021

１：日大理工・院（前）・情報　２：日大理工・教員・情報