徳島県立工業技術センター
令和４年度　経常研究

小規模データセットにおける音の分類モデルの高精度化に関する研究

Tokushima Prefectural Industrial Technology Center

工業技術センター　機械技術担当　平岡　忠志

1.研究目的

高精度な音の分類モデルを開発するためには大規模データセットが必要であるが，大規模データセットの作成は困難である．本研究では，大規模データセットで学習されたモデルを小規模データセットのモデル学習に利用する転移学習について，大規模データセットの違いや小規模データセットのサンプルサイズが精度に及ぼす影響を調べた.

2.研究内容

小規模データセットにESC-50，大規模データセットにImageNetとAudioSetを利用した．ESC-50はドアのノック等の環境音が50クラス2,000個ある．ImageNetは一般物体認識用の画像が21,841クラス14,197,122枚ある．このうち1,000クラス1,331,167枚を利用した．AudioSetは楽器等の音が632クラス2,084,320個ある.
大規模データセットの違いが精度に及ぼす影響を調べるため，モデルの初期パラメータについて，ランダムに設定した場合（Scratch），ImageNetで学習されたモデルを設定した場合（ImageNet　Pretrain），AudioSetで学習されたモデルを設定した場合（AudioSet　Pretrain）の3つで比較した．小規模データセットのサンプルサイズが精度に及ぼす影響を調べるため，各クラスの学習用サンプルサイズを1,2,4,8,...と倍々に変化させた.

3.研究成果

小規模データセットのESC-50で学習されたAudio Spectrogram Transformaer（AST）の精度検証結果を表1に示した．ScrachよりもImageNet Pretrain, ImageNet Pretrain よりもAudioSet Pretrainの方が各クラスの学習用サンプルサイズのいずれにおいても高精度だった.

表１．精度検証結果（ESC-50）

各クラスの学習用サンプルサイズ	精度[%]		
	Scratch	ImageNet Pretrain	AudioSet Pretrain
1	6.75	28.25	46.75
2	7.00	36.25	63.50
4	8.25	58.50	78.00
8	17.00	60.75	88.75
16	28.50	79.50	91.00
32	42.75	87.75	93.50

お問い合わせ　工業技術センター　機械技術担当　平岡　忠志　TEL: 088-635-7905