

木材破壊音を分類する CNN モデルに入力するデータの検出方法

東京理科大学 学生会員 ○生内 良汰
東京理科大学 正会員 佐伯 昌之

1. 目的

近年、深層学習などの機械学習理論が発展し、機械学習を用いた音の分析が行われている。当研究室では、畳み込みニューラルネットワーク(以下、CNN)を用いて、地震時に構造部材の木材が破壊される音を検知することで、構造物の被災度を迅速に推定することを目指している。著者らのうち一人は、先行研究において CNN に 3 秒のスペクトログラムを入力し、木材破壊音の有無を分類するモデルを作成した¹⁾。ただし、この CNN は定型の画像を入力値とするため、入力する音データを音源の wav ファイルから抽出して、スペクトログラムに変換して入力する必要がある。また、長時間の音源から全範囲を抽出して CNN に入力すると、計算量が増加し迅速性が失われる可能性がある。そこで、木材破壊音の可能性がある音を検出し、CNN に入力するスペクトログラムを減らす手法を検討した。また、3 秒のスペクトログラムには複数の木材破壊音が描画されることがあるため、230 ms のスペクトログラムを使用して CNN で分類を行った。図-1 にスペクトログラムから木材破壊音の可能性がある音を検出した例を示す。

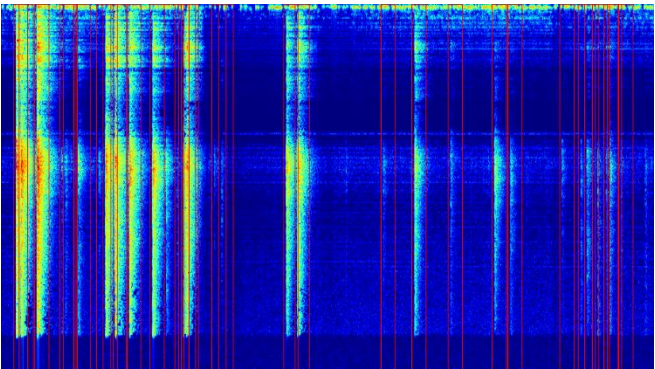


図-1 スペクトログラムに存在する木材破壊音の可能性がある音を検出し赤矩形で囲ったもの
縦軸が周波数、横軸が時間、色が音の強さを表す

2. 本研究で使

本研究において使用したデータは、小野ら²⁾によって
キーワード 木材破壊音, CNN, Spectral Flux, データ検出
連絡先 〒278-8510 千葉県野田市山崎 2641 東京理科大学野田キャンパス佐伯研究室 TEL 047-124-1501

行われた 66 回の木材破壊実験を 6 つの異なるデバイスで録音した 394 個の音圧波形データである。さらに、実環境に近づけた場合の木材破壊音の検出率を評価するため、木材破壊音と環境音を合成したデータ(以下、混合音)を 12214 個生成した。環境音には ESC-50³⁾という日常的な音が録音されたデータセットから 31 種類の音を使用した。使用した音の名称を表-1 に示す。混合音は、木材破壊音と環境音の最大振幅の比率が 2:1 となるように合成し、木材破壊音が支配的となるようにした。

表-1 環境音に使用した ESC-50 の音源名称

Animals	Natural soundscapes & water sounds	Human, non-speech sounds	Interior/domestic sounds	Exterior/urban noises
Rooster	Rain	Crying baby	Keyboard typing	Helicopter
Cat	Crickets	Sneezing	Can opening	Chainsaw
Hen	Chirping birds	Laughing	Washing machine	Siren
Insects(flying)	Water drops	Brushing teeth	Vacuum cleaner	Car horn
	Wind	Snoring	Glass breaking	Engine
	Pouring water			Train
	Toilet flush			Church bells
	Thunderstorm			Airplane
				Hand saw

3. カスケード分類器を用いた検出手法

カスケード分類器は入力値として画像を用い、画像内に検出窓をかけて、物体を検出する機械学習モデルである。また、強識別機を複数連結させた構造を持つことで、検出対象でない画像を初期の段階で排除するので、高速で物体を検出できる。学習に用いた教師画像は木材破壊実験のスペクトログラムから、木材破壊音や類似音 6790 枚を正解画像として、木材破壊音以外の音 8000 枚を非正解画像として切り出したものである。また、学習に用いた特徴量は Haar-like 特徴量と Local Binary Pattern (以下、LBP)特徴量である。

検出結果を表-2 に示す。混合音のスペクトログラムは学習させていないため、木材破壊音に比べて検出率が低下した。また、画像の局所明暗差を用いる Haar-Like 特徴量のほうが、画像の局所輝度差を用いる LBP 特徴量よりも木材破壊音の検出に長けていると示された。

表-2 カスケード分類器の検出結果

特徴量	データ種類	検出率[%]	適合率[%]	検出数
Haar-like 特徴量	木材破壊音	31.51	52.60	17.60
	混合音	26.29	41.60	20.15
LBP 特徴量	木材破壊音	22.26	50.34	12.09
	混合音	17.35	44.15	11.04

4. Spectral Flux, 音圧レベルを用いた検出手法

次に Spectral Flux と音圧レベルの時系列データから閾値を超えるデータを検出する手法を試みた。Spectral Flux とは、音圧波形データを短時間フーリエ変換して得られる振幅スペクトルの変化量を表すものである。時間方向に連続する二つのフレームにおいて、フレーム間の振幅スペクトルの変化量が正であるときの値で表される。そのため音が発生した瞬間を表現することに優れている。

音圧レベルとは、音圧と人間が聞こえる最も小さい音の音圧（基準音圧、 20×10^{-6} Pa）の比を常用対数で表現した量である。今回は Root Mean Square(二乗平均平方根)を使用して、音圧波形データから音圧を計算し、音圧レベルを求めた。

検出結果を表-3 に示す。閾値は入力データの平均値を設定した。Spectral Flux を用いた手法の木材破壊音 と混合音の検出率は 99 %以上で、他の手法を大きく上回った。Spectral Flux は、木材破壊音が広い周波数成分を持つことと、破壊時に振幅スペクトルが急激に上昇する特徴を捉えるのに適切な特徴量であった。また、音圧レベルは、じわじわと大きくなる音や、騒音の検出に優れているが木材破壊音の検出には適していない特徴量だと示された。最適な閾値を調べた結果、Spectral Flux の閾値を 250 とすることで検出率 90 %以上を維持して、検出数を削減することができた。

表-3 Spectra Flux, 音圧レベルの検出結果

特徴量	データ種類	検出率[%]	適合率[%]	検出数
Spectral Flux	木材破壊音	99.89	19.18	189.3
	混合音	99.53	15.39	226.2
音圧 レベル	木材破壊音	78.57	25.16	121.2
	混合音	69.94	20.38	115.2
Spectral Flux (閾値=250)	木材破壊音	99.19	29.14	118.7
	混合音	95.73	22.95	151.2

5. 検知時間の比較

Haar-like 特徴量を用いたカスケード分類器が検出したデータを CNN に入力する方法と、Spectral Flux を用いて検出したデータを CNN に入力する方法、検出せずにスペクトログラムの全範囲を CNN に入力する方法の検知時間を比較した。検知時間とは音源の wav ファイルから CNN に入力するデータを作成し、CNN で分類が完了するまでの時間である。ここで、検知対象としたデータは 3 分のデータである。検知時間を表-4 に示す。その結果、検出の工程を追加することで検知時間が短縮され、CNN に入力するデータ数を削減することが処理の高速化につながると示された。

表-4 各手法の検知時間

手法	カスケード 分類器	Spectral Flux	音圧レベル	全部入力
処理時間[s]	11.41	11.68	12.83	20.23

6. まとめ

本研究では、木材破壊音を分類する CNN モデルに入力するためのデータの検出方法を検討した。その結果、Spectral Flux を用いた検出手法は木材破壊音と混合音から 99 %以上を検出できることを示した。ただ、評価した混合音の合成比率が 2:1 のみであったので、合成比率を変化させた混合音に対する検出を行い、検出器の詳細な評価をする必要がある。また、CNN に入力するスペクトログラムのデータ長は 230ms としたが、データ長を変化させた場合の検知時間や、精度の変化を調べて最適なデータ長を決定する必要がある。

参考文献

1) 竹森司, 佐伯昌之 : 木材破壊音検出を目的とした NN と CNN の比較検討, 土木学会論文集 A2 (応用力学), Vol.77, No.2(応用力学論文集 Vol.24), 2021

2) 小野真由子, 竹森司, 佐伯昌之 : CNN による木材破壊音検出のためのデータ拡張手法の検討, 土木学会 第 25 回応用力学シンポジウム, 2E01-05-03, 2022

3) K. J. Piczak : ESC: Dataset for Environmental sound Classification, Proceeding of the 23rd Annual ACM Conference on Multimedia, Brisbane, Australia, 2015