Scopus

Documents

Al-Hattab, Y.A., Zaki, H.F., Shafie, A.A.

Rethinking environmental sound classification using convolutional neural networks: optimized parameter tuning of single feature extraction
(2021) *Neural Computing and Applications*, .

DOI: 10.1007/s00521-021-06091-7

Department of Mechatronics Engineering, International Islamic University Malaysia, Gombak, Selangor 53100, Malaysia

Abstract
The classification of environmental sounds is important for emerging applications such as automatic audio surveillance, audio forensics, and robot navigation. Existing techniques combined multiple features and stacked many CNN layers (very deep learning) to reach the desired accuracy. Instead of using many features and going deeper by stacking layers that are resource extensive, this paper proposes a novel technique that uses only a single feature, namely the Mel-Frequency Cepstral Coefficient (MFCC) and just three layers of CNN. We demonstrate that such a simple network can considerably outperform several conventional and deep learning-based algorithms. Through parameters fine-tuning of the data input, we reported a model that is significantly less complex in the architecture yet has recorded a similar accuracy of 95.59% compared to state-of-the-art deep models on UrbanSound8k dataset. © 2021, The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature.

Author Keywords
Convolutional neural networks (CNN); Environmental sound classification; Feature extraction; Mel-frequency cepstral coefficients (MFCC); Urbansound8Kdataset

Index Keywords
Audio acoustics, Deep learning, Learning algorithms, Robots; Audio surveillance, Emerging applications, Environmental sound classifications, Environmental sounds, Learning-based algorithms, Mel-frequency cepstral coefficients, Multiple features, Optimized parameter; Convolutional neural networks

References

- Ali, H., Tran, S.N., Benetos, E., Garcez, A.S.D.A.
  Speaker recognition with hybrid features from a deep belief network
  (2018) *Neural Comput Appl*, 29 (6), pp. 13-19.
- Ghosal, D., Kolekar, M.H.
  Music genre recognition using deep neural networks and transfer learning
  (2018) *Interspeech*, pp. 2087-2091.
- Chachada, S., Kuo, C.C.J.
  Environmental sound recognition: a survey
  (2014) *APSIPA Trans Signal Inf Process*, 3.
- Zhang, Z., Xu, S., Cao, S., Zhang, S.
  Deep convolutional neural network with mixup for environmental sound classification
  (2018) *Chinese Conference on Pattern Recognition and Computer Vision (Prcv)*, pp. 356-367.
  Springer, Cham, pp

- Shkurti, F., Chang, W.D.
  Underwater multi-robot convoying using visual tracking by detection
  (2017) *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 4189-4196.
  IEEE

- Chu, S., Narayanan, S., Kuo, C.C.J.
  Environmental sound recognition with time–frequency audio features
  (2009) *IEEE Trans Audio Speech Lang Process*, 17 (6), pp. 1142-1158.

- Giannoulis, D., Benetos, E., Stowell, D., Rossignol, M., Lagrange, M., Plumbley, M.D.
  Detection and classification of acoustic scenes and events: An IEEE AASP challenge
  (2013) *2013 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 1-4.
  IEEE

- **IEEE international conference on acoustics, speech and signal processing (ICASSP), pp 559–563**
  (2015) *IEEE*,

- (2015) *Deep Learn Nat*, 521 (7553), pp. 436-444.

- Palaz, D., Collobert, R.
  (2015) *Analysis of Cnn-Based Speech Recognition System Using Raw Speech as Input (No. REP_WORK)*,
  Idiap

- Adavanne, S., Virtanen, T.
  (2017) *Sound Event Detection Using Weakly Labeled Dataset with Stacked Convolutional and Recurrent Neural Network*,

- Adavanne, S., Pertilä, P., Virtanen, T.
  Sound event detection using spatial features and convolutional recurrent neural network
  (2017) *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 771-775.
  IEEE

- Zaki, H.F., Shafait, F., Mian, A.
  Modeling 2D appearance evolution for 3D object categorization
  (2016) *2016 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, pp. 1-8.
  IEEE

- Piczak, K.J.
  Environmental sound classification with convolutional neural networks
  (2015) *2015 IEEE 25Th International Workshop on Machine Learning for Signal Processing (MLSP)*, pp. 1-6.
  IEEE

- Meyer, M., Cavigelli, L., Thiele, L.
  (2017) *Efficient Convolutional Neural Network for Audio Event Detection*,
  arXiv preprint axXiv

- Pons, J., Serra, X.
  Randomly weighted cnns for (Music) audio classification
  (2019) *ICASSP 2019–2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 336-340.
  IEEE

- Dai, W., Dai, C., Qu, S., Li, J., Das, S.
  Very deep convolutional neural networks for raw waveforms
  (2017) *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 421-425.
  IEEE

- Salamon, J., Bello, J.P.
  Deep convolutional neural networks and data augmentation for environmental sound classification
  (2017) *IEEE Signal Process Lett*, 24 (3), pp. 279-283.

- Boddapati, V., Petef, A., Rasmusson, J., Lundberg, L.
  Classifying environmental sounds using image recognition networks
  (2017) *Proc Comput Sci*, 112, pp. 2048-2056.

- Abdoli, S., Cardinal, P., Koerich, A.L.
  End-to-end environmental sound classification using a 1D convolutional neural network
  (2019) *Expert Syst Appl*, 136, pp. 252-263.

- Su, Y., Zhang, K., Wang, J., Madani, K.
  Environment sound classification using a two-stream CNN based on decision-level fusion
  (2019) *Sensors*, 19 (7), p. 1733.

- Sharma, J., Granmo, O.C., Goodwin, M.
  (2019) *Environment Sound Classification Using Multiple Feature Channels and Attention Based Deep Convolutional Neural Network*,
  arXiv preprint axXiv

- Virtanen, T., Plumbley, M.D., Ellis, D.
  (2018) *Computational analysis of sound scenes and events*, pp. 3-12.
  (eds), Springer, Heidelberg

- Sahidullah, M., Saha, G.
  Design, analysis and experimental evaluation of block based transformation in MFCC computation for speaker recognition
  (2012) *Speech Commun*, 54 (4), pp. 543-565.

- Shepard, R.N.
  Circularity in judgments of relative pitch
  (1964) *J Acoust Soc Am*, 36 (12), pp. 2346-2353.

- Paulus, J., Müller, M., Klapuri, A.
  State of the art report: Audio-based music structure analysis
  (2010) *Ismir*, pp. 625-636.

- Glorot, X., Bordes, A., Bengio, Y.
Deep sparse rectifier neural networks
(2011) *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics. JMLR Workshop and Conference Proceedings*, pp. 315-323.

- Hinton, G.E., Srivastava, N., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.R.
(2012) *Improving Neural Networks by Preventing Co-Adaptation of Feature Detectors*, arXiv preprint axXiv

- Jung, S.H., Chung, Y.J.
Performance analysis of the convolutional recurrent neural network on acoustic event detection
(2020) *Bull Electr Eng and Info*, 9 (4), pp. 1387-1393.

- Lezhenin, I., Bogach, N., Pyshkin, E.
(2019) *Urban Sound Classification Using Long Short-Term Memory Neural Network*, pp. 57-60.
pp, IEEE

- Salamon, J., Jacoby, C., Bello, J.P.
A dataset and taxonomy for urban sound research
(2014) *Proceedings of the 22Nd ACM International Conference on Multimedia*, pp. 1041-1044.
pp

- Tokozume, Y., Harada, T.
(2017) *Learning Environmental Sounds with End-To-End Convolutional Neural Network.*, pp. 2721-2725.
pp, IEEE

Correspondence Address
Al-Hattab Y.A.; Department of Mechatronics Engineering, Malaysia; email: alhattab.yousef@live.iium.edu.my

Publisher: Springer Science and Business Media Deutschland GmbH

ISSN: 09410643
Language of Original Document: English
Abbreviated Source Title: Neural Comput. Appl.
2-s2.0-85106616922
Document Type: Article
Publication Stage: Article in Press
Source: Scopus

ELSEVIER

Copyright © 2021 Elsevier B.V. All rights reserved. Scopus® is a registered trademark of Elsevier B.V.

RELX Group™