Techniques for Recognising and Classifying Environmental Noise Using Deep Learning

Ludovica Beritelli[1], Maria Grazia Borzì[1], Cristian Randieri[2], Roberta Avanzato[1] and Francesco Beritelli[1]

[1]Department of Electrical, Electronic and Computer Engineering University of Catania, Catania, Italy
[2]Università degli Studi e-Campus, Novedrate (CO), Italy

Abstract

Increasing urbanisation poses new challenges in mitigating noise pollution and preserving quality of life. In this study, we present an innovative approach for the classification of environmental noise, exploiting advanced Deep Learning (DL) techniques. By merging three different public datasets, we created a unified corpus to train and test a convolutional neural network (CNN), with the aim of efficiently recognising and classifying various noise events. The proposed approach overcomes the limitations of conventional methodologies, avoiding the need for data pre-processing that could alter sound characteristics. The experimental results demonstrate a significant improvement in classification accuracy, reaching 96.93% with the test set and 100% by applying a post-processing filter. These results emphasise the potential of DL in the treatment of environmental noise, offering new perspectives for signal processing and telecommunications.

Keywords

Environmental Noise Classification, Convolutional Neural Networks, Signal Processing, Noise Pollution

1. Introduction

The search for sustainable solutions to mitigate the impact of environmental noise has become crucial to preserving the quality of life in our increasingly urban society. In this context, the recognition and classification of environmental noise emerge as key challenges in the field of signal processing and telecommunications, where noise can significantly degrade the quality and intelligibility of transmitted signals [1]. Recently, the advancement of machine learning (ML) and deep learning (DL) techniques has opened new frontiers in the accuracy of noise classification.

Pioneering studies, such as that of Couvreur et al. [2], have demonstrated the effective use of hidden Markov models (HMMs) for the recognition of sound events, offering detailed analysis of sound signals in time and frequency. Despite their effectiveness, these techniques require considerable computational resources, posing challenges in practical implementation [3, 4, 5, 6, 7, 8, 9]. In parallel, the approach by Alsouda et al. [10] presents a machine-learning-based method for urban noise identification using an inexpensive IoT unit and Mel-frequency cepstral coefficient extraction of audio features and supervised classification algorithms (such as support vector machine, k-nearest neighbours, bootstrap aggregation and random forest). This approach achieved noise classification accuracy in the range of 88% to 94%. The integration of HMM, fuzzy logic and neural networks proposed by Beritelli et al. [11] further emphasised the importance of combining different methodologies to improve classification accuracy on large noise databases. Furthermore, a study conducted by Aksoy et al. [12] used advanced deep learning models, including VGG-13BN, ResNet-50 and DenseNet-121, to classify sounds according to their environmental relevance. The results demonstrated high accuracy in classifying sounds, with correctness rates of over 95%, highlighting in particular the VGG-13 BN model that achieved 99.72% accuracy. These results underline the significant potential of the deep learning approach in identifying sounds harmful to the environment. Another contribution is made by Jeon et al. [13], proposing a multi-channel indoor noise database for the development and evaluation of speech processing algorithms. This database includes noise signals generated by physical actions and loudspeakers placed in various locations within an apartment building, allowing for a wide range of noise conditions. A further study, conducted by Ramli et al. [14], proposes a mechanism to reduce background noise in voice communications through the use of a two-sensor adaptive noise canceller. This system demonstrated high convergence rates, significant improvements in the signal-to-noise ratio, and a 65% reduction in computational power compared to traditional methods. The study by Tsai et al. [15] analyses the spatial characteristics of urban noise using noise maps and emphasises the importance of noise maps for a better understanding and management of urban noise.

This study demonstrates how the application of DL techniques can offer effective solutions to the challenges of environmental noise classification, with potential significant benefits for the telecommunication sector and society at large. Our research opens new perspectives for the use of artificial intelligence in urban noise mitigation, promoting a more sustainable environment and a better quality of life.

In section 2 we discuss the importance of developing effective noise classification strategies, which are essential for improving the quality of communication and, consequently, the quality of life in urban areas.

In section 3, we present our innovative approach, which exploits advanced DL techniques for analysing and classifying environmental noise. We will illustrate how, through the use of Convolutional Neural Networks (CNNs), our model works directly with the raw audio data, avoiding the loss of significant information that could result from pre-processing processes.

In section 4, we present the results obtained from our study, demonstrating the effectiveness of the proposed model in classifying environmental noise. The results show a significant improvement in classification accuracy, achieving remarkable performance in the tests performed. We will also discuss the impact of a post-processing filter [16] in

SYSYEM 2023: 9th Scholar's Yearly Symposium of Technology, Engineering and Mathematics, Rome, December 3-6, 2023

beritelli.ludovica@gmail.com (L. Beritelli); borzi.m@studium.unict.it (M. G. Borzì); cristian.randieri@uniecampus.it (C. Randieri); roberta.avanzato@unict.it (R. Avanzato); francesco.beritelli@unict.it (F. Beritelli)

further increasing the accuracy of the model.

2. Environmental Noise

Environmental noise, defined as any unwanted sound generated by the surrounding environment, is a major source of noise pollution. These sounds may come from natural sources such as sea waves or from man-made sources such as vehicle traffic, alarms, voices and electronic devices. Effective management of such noise requires methods that go beyond simply measuring sound pressure levels (dB), including characterising and identifying the type of noise [17]. In the field of telecommunications, environmental noise introduces significant challenges, degrading signal quality and compromising communication efficiency. Research has highlighted the importance of developing advanced noise reduction strategies, through the use of machine learning (ML) and deep learning (DL) techniques, aimed at improving the accuracy of noise classification [18]. The studies in [19, 20] have contributed greatly to the understanding of environmental noise by providing innovative approaches for its analysis and classification. These works emphasise the need for authentic and versatile databases to test and develop signal processing algorithms capable of handling the complexity of real acoustic environments [21]. The accurate identification and classification of environmental noise not only improves the performance of telecommunication systems but also contributes to the health and well-being of individuals by reducing exposure to harmful levels of noise. Therefore, research in this area is crucial to advance the design of more resilient communication systems and to promote a more sustainable sound environment.

3. Method proposed

Advances in Machine Learning (ML) and Deep Learning (DL) techniques have radically transformed the approach to data analysis, allowing us to discover unexpected complex patterns in audio data. In this study, we adopted an innovative methodology that exploits neural networks to directly process audio signals in .wav format. The aim is to evaluate the ability of these networks to accurately classify different sound events without resorting to pre-processing techniques that could compromise data integrity. In the subsection 3.1 we will describe the datasets used, the breakdown of these for training, validation and testing of the neural network and in 3.2 the CNN network used for the classification of ambient noise.

3.1. Dataset

The dataset used in this research was composed by merging three distinct public databases: UrbanSound [18], Demand [17] and Noisex-92 [19]. This fusion created a heterogeneous dataset that includes a wide range of sound classes, specifically excluding the dog bark class from UrbanSound, but incorporating common ambient noise classes from Noisex-92 and Demand. A prepocessing phase is carried out before giving the data as input to the CNN network. Specifically, the recordings were all divided into 2-second sub-sequences and sampled at 22050 Hz. The dataset was randomly divided into two different sets, one used for network training and validation and the other for network testing, ensuring an equal distribution of sound classes between the two. The learning dataset includes classes such as “air_conditioner", “children_playing", and “traffic", with a variable number of sound sequences per class. Similarly, the test dataset maintains a representative proportion of each class, ensuring a valid evaluation of network performance.

3.1.1. Learning dataset

- “air_conditioner": 1271 audio sequences,
- “children_playing": 704 audio sequences,
- “babble": 259 audio sequences,
- “car_horn": 307 audio sequences,
- “drilling": 622 audio sequences,
- “engine_idling": 704 audio sequences,
- “jackhammer": 917 audio sequences,
- “metro": 1800 audio sequences,
- “office": 1800 audio sequences,
- “river": 1800 audio sequences,
- “siren": 956 audio sequences,
- “square": 1800 audio sequences,
- “street_music": 850 audio sequences,
- “traffic": 1800 audio sequences.

3.1.2. Testing dataset

- “air_conditioner": 543 audio sequences,
- “children_playing": 299 audio sequences,
- “babble": 15 audio sequences,
- “car_horn": 129 audio sequences,
- “drilling": 268 audio sequences,
- “engine_idling": 303 audio sequences,
- “jackhammer": 398 audio sequences,
- “metro": 600 audio sequences,
- “office": 600 audio sequences,
- “river": 600 audio sequences,
- “siren": 412 audio sequences,
- “square": 600 audio sequences,
- “street_music": 362 audio sequences,
- “traffic": 600 audio sequences.

3.2. Application of CNNs

Artificial intelligence (AI) represents a vast and evolving field of study that aims to emulate human cognitive capabilities through the development of autonomous hardware and software systems. This ambition to reflect human intelligence in machines has led to the development of technologies capable of autonomous learning, adaptation, reasoning and planning. At the heart of AI are advanced algorithms and computational techniques, which make it possible to replicate typically human behaviours, such as interaction with the environment and decision-making. The applications of artificial intelligence range in different fields, from industrial to domestic, demonstrating its potential to improve both the activities of businesses and public administrations and the everyday lives of people.

Convolutional Neural Networks (CNNs) stand out for their effectiveness in analysing visual and sound data due to their ability to identify complex patterns through the use of convolutional filters.

Our CNN architecture follows a structured model starting with the input layer, proceeding through convolutional and

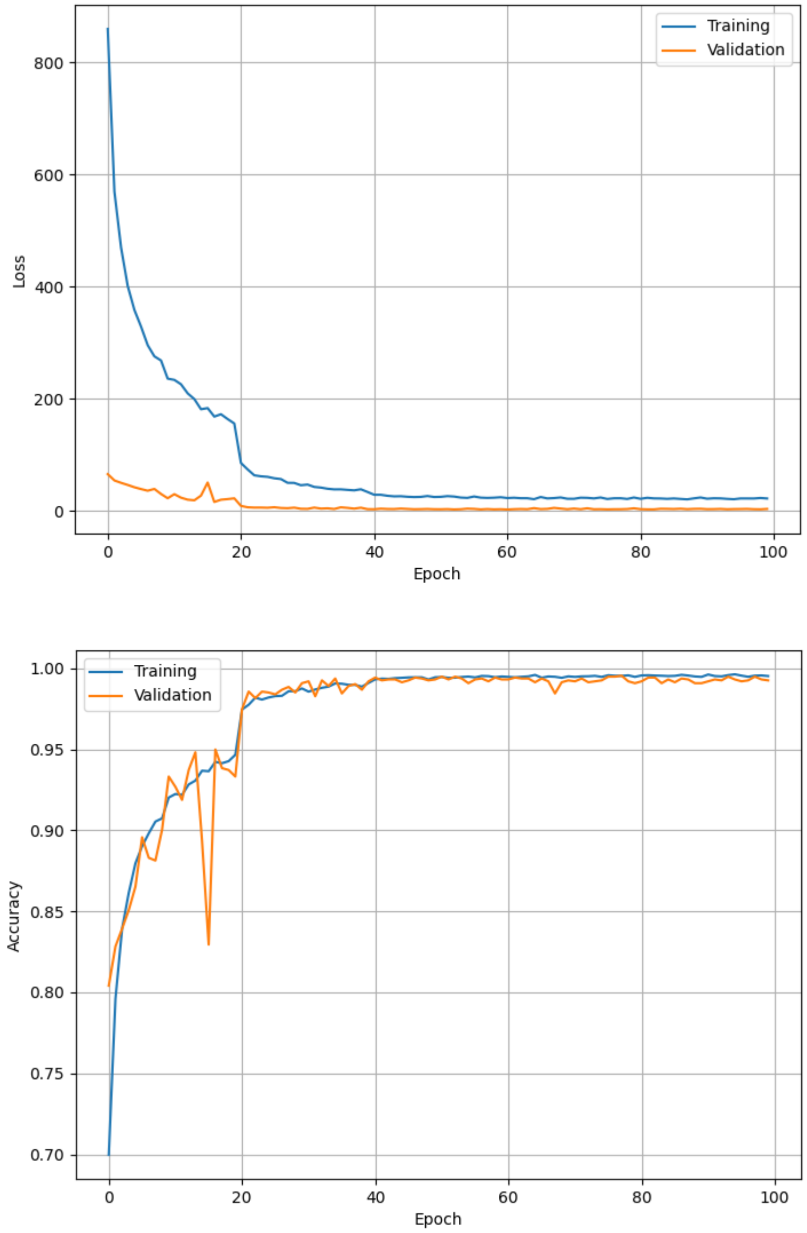

Figure 1: Accuracy and loss trends during training and validation.

activation (ReLU) layers, pooling, and culminating in a fully connected layer for final classification. This design allows the network to process audio features from the simplest to the most complex, facilitating deep and robust data learning. The detailed configuration of convolutional, pooling, and fully connected layers provides a powerful means to extract and interpret sound features, making CNNs particularly suitable for the recognition and classification of complex sound events. Our research aims to demonstrate the effectiveness of this approach in the field of acoustic analysis, contributing significantly to the field of signal processing and audio classification.

The neural network used in this study is based on the architecture of 1D convolutional neural networks and, in particular, on the “M5 (0.5M)" model described in [16]. This network consists of five layers, the first four of which are convolutions (1D Convolution Layer, Batch Normalisation Layer, ReLU Layer and Pooling Layer) and the last layer is the output (Softmax).

The neural network’s input is a vector containing sequences of audio waveforms, each with a duration of $W_{in} = 2$ seconds. The CNN neural network determines the index associated with one of $N_C = 14$ different classes $C_i (i = 1, ..., N)$ using the LogSoftMax function. The network is trained by feeding RAW sequences representing different environmental noises.

4. Experimental Results

The validation process of our approach was carried out through a rigorous experiment involving the direct input

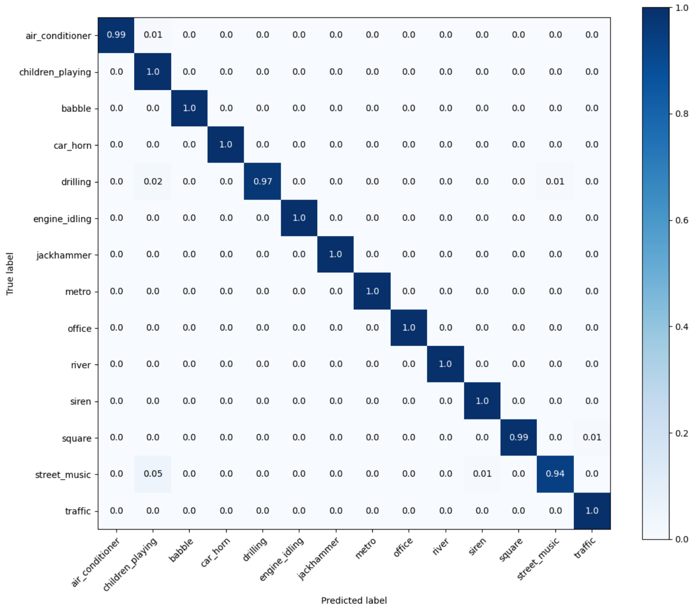

Figure 2: Confusion matrix for the validation dataset.

of raw audio data, in .wav format, into the convolutional neural network (CNN). Below, we present a detailed analysis of the performance obtained during the different phases of training, validation and testing of the network.

4.1. Training and Validation

During the training phase, we observed a progressive improvement in network performance, as illustrated in Fig. 1. This graph shows an increase in accuracy and a decrease in the loss function as the epochs progress, highlighting the effectiveness of the learning process. The dataset was split into a proportion of 70% for training and 30% for validation, as illustrated in Section 3.1.

Fig. 2 presents the confusion matrix obtained from the validation of the model, providing a clear indication of its classification capability across the different sound categories.

4.2. Testing

The effectiveness of the model was further verified through testing on a separate dataset, achieving an impressive accuracy of 96.93%. Fig. 3 illustrates the confusion matrix for this phase, confirming the network's high accuracy in recognising environmental sounds.

4.3. Post-Processing and Time Window Analysis

The introduction of a post-processing filter, called the "recurrence filter" [16], further improved the performance of the model. As demonstrated in Fig. 4, the accuracy of the system increases significantly by extending the analysis time window. In particular, it can be seen that by extending the analysis beyond 28 seconds, the accuracy reaches 100%.

The results underline the effectiveness of our approach based on the use of convolutional neural networks for analysing environmental sound, highlighting the potential of deep learning techniques in overcoming the challenges of accurately recognising complex sound events.

5. Conclusion

This study introduced a new approach for the classification of environmental noise, exploiting the potential of Deep Learning techniques to address one of the most pressing challenges in signal processing and telecommunications. Through the use of a CNN trained on a unified dataset derived from three different public sources, it is shown that high accuracy in the classification of environmental noise events can be achieved without the need for complex preprocessing. The results obtained reveal a marked improvement in classification accuracy, highlighting the effectiveness of our model both in the testing phase and in the application of post-processing techniques. These results not only confirm the value of convolutional neural networks in acoustic analysis, but also open the way for future research to explore the applicability of such methods in broader areas, including urban noise monitoring and the improvement of telecommunication systems. In conclusion, our study contributes significantly to the body of research on signal processing, proposing an effective and efficient model for the classification of ambient noise, with direct implications for environmental sustainability and quality of life in urban areas.

References

[1] F. Beritelli, A. Gallotta, C. Rametta, A dual streaming approach for speech quality enhancement of voip service over 3g networks, in: 2013 18th International Conference on Digital Signal Processing (DSP), IEEE, 2013, pp. 1–5.

[2] C. Couvreur, V. Fontaine, P. Gaunard, C. G. Mubikangiey, Automatic classification of environmental noise

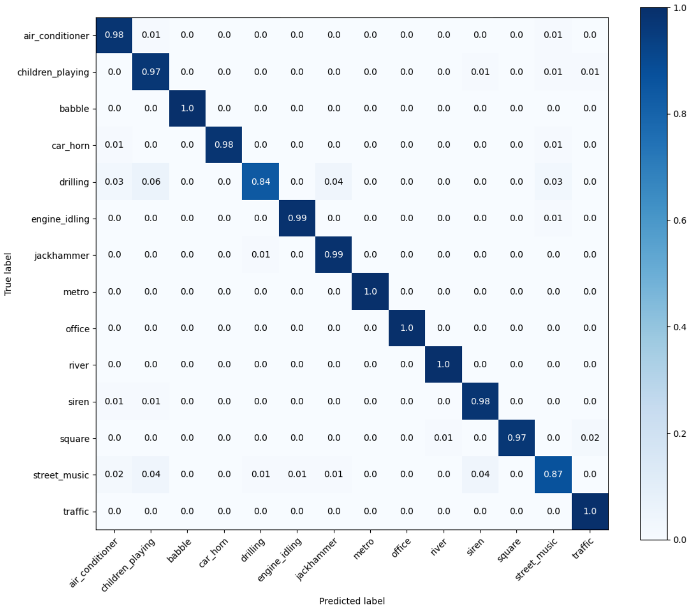

Figure 3: Confusion matrix for the testing dataset.

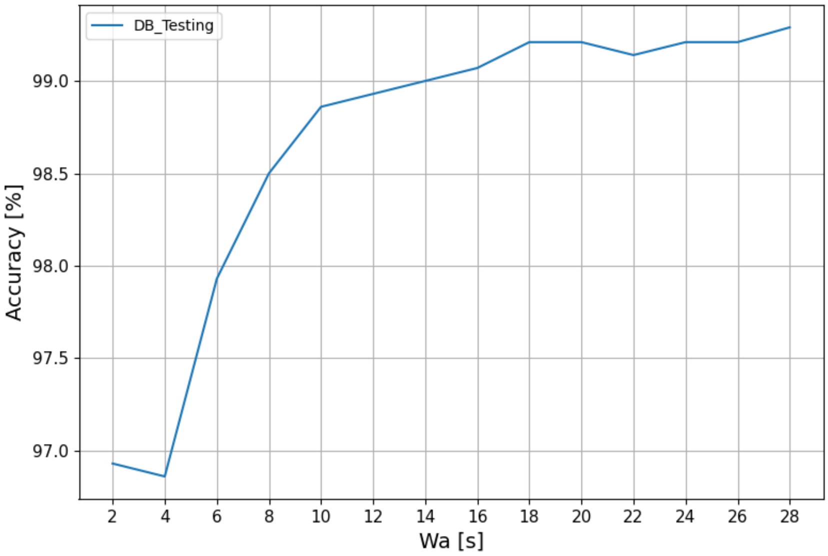

Figure 4: Effect of post-processing filter on accuracy as a function of time window.

events by hidden markov models, Applied Acoustics 54 (1998) 187–206.

[3] G. Capizzi, C. Napoli, L. Paternò, An innovative hybrid neuro-wavelet method for reconstruction of missing data in astronomical photometric surveys, in: Artificial Intelligence and Soft Computing: 11th International Conference, ICAISC 2012, Zakopane, Poland, April 29-May 3, 2012, Proceedings, Part I 11, Springer, 2012, pp. 21–29.

[4] N. Brandizzi, S. Russo, R. Brociek, A. Wajda, First studies to apply the theory of mind theory to green and smart mobility by using gaussian area clustering, volume 3118, 2021, pp. 71 – 76.

[5] F. Bonanno, G. Capizzi, G. Lo Sciuto, A neuro wavelet-based approach for short-term load forecasting in integrated generation systems, in: 2013 International Conference on Clean Electrical Power (ICCEP), IEEE, 2013, pp. 772–776.

[6] V. Ponzi, S. Russo, A. Wajda, R. Brociek, C. Napoli, Analysis pre and post covid-19 pandemic rorschach test data of using em algorithms and gmm models, volume 3360, 2022, pp. 55 – 63.

[7] G. Capizzi, G. L. Sciuto, C. Napoli, M. Woźniak, G. Susi, A spiking neural network-based long-term prediction system for biogas production, Neural Networks 129 (2020) 271–279.

[8] G. De Magistris, M. Romano, J. Starczewski, C. Napoli, A novel dwt-based encoder for human pose estimation, volume 3360, 2022, pp. 33 – 40.

[9] F. Bonanno, G. Capizzi, G. L. Sciuto, C. Napoli, Wavelet recurrent neural network with semi-parametric input data preprocessing for micro-wind power forecasting

in integrated generation systems, 2015, pp. 602 – 609. doi:10.1109/ICCEP.2015.7177554.

[10] Y. Alsouda, S. Pllana, A. Kurti, Iot-based urban noise identification using machine learning: performance of svm, knn, bagging, and random forest, in: Proceedings of the international conference on omni-layer intelligent systems, 2019, pp. 62–67.

[11] F. Beritelli, R. Grasso, A pattern recognition system for environmental sound classification based on mfccs and neural networks, in: 2008 2nd International Conference on Signal Processing and Communication Systems, IEEE, 2008, pp. 1–4.

[12] B. Aksoy, U. Uygar, G. Karadağ, A. R. Kaya, Ö. Melek, Classification of environmental sounds with deep learning, Advances in Artificial Intelligence Research 2 (2022) 20–28.

[13] K. M. Jeon, N. K. Kim, M. J. Jo, H. K. Kim, Design of multi-channel indoor noise database for speech processing in noise, in: 2017 20th Conference of the Oriental Chapter of the International Coordinating Committee on Speech Databases and Speech I/O Systems and Assessment (O-COCOSDA), IEEE, 2017, pp. 1–4.

[14] R. M. Ramli, A. O. A. Noor, S. Abdul Samad, Noise cancellation using selectable adaptive algorithm for speech in variable noise environment, International Journal of Speech Technology 20 (2017) 535–542.

[15] K.-T. Tsai, M.-D. Lin, Y.-H. Chen, Noise mapping in urban environments: A taiwan study, Applied Acoustics 70 (2009) 964–972.

[16] R. Avanzato, F. Beritelli, Heart sound multiclass analysis based on raw data and convolutional neural network, IEEE Sensors Letters 4 (2020) 1–4.

[17] J. Thiemann, N. Ito, E. Vincent, The diverse environments multi-channel acoustic noise database (demand): A database of multichannel environmental noise recordings, in: Proceedings of Meetings on Acoustics, volume 19, AIP Publishing, 2013.

[18] J. Salamon, C. Jacoby, J. P. Bello, A dataset and taxonomy for urban sound research, in: Proceedings of the 22nd ACM international conference on Multimedia, 2014, pp. 1041–1044.

[19] A. Varga, H. J. Steeneken, Assessment for automatic speech recognition: Ii. noisex-92: A database and an experiment to study the effect of additive noise on speech recognition systems, Speech communication 12 (1993) 247–251.

[20] J. Salamon, J. P. Bello, Unsupervised feature learning for urban sound classification, in: 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2015, pp. 171–175.

[21] K. J. Piczak, Esc: Dataset for environmental sound classification, in: Proceedings of the 23rd ACM international conference on Multimedia, 2015, pp. 1015–1018.