Benchmark for Kitchen20, a Daily Life Dataset for Audio-based Human Action Recognition

Marc Moreaux, Michael Garcia-Ortiz, Isabelle Ferrané, Frédéric Lerasle

Benchmark for Kitchen20, a daily life dataset for audio-based human action recognition

Marc Moreaux[123]
AI-Lab[1]
SoftBank Robotics Europe
Paris, France
mr.moreaux@gmail.com

Michael Garcia Ortiz
AI-Lab
SoftBank Robotics Europe
Paris, France
mgarcia@softbank-robotics.com

Isabelle Ferrané
IRIT[2]
Univ. de Toulouse, CNRS,
Toulouse, France
ferrane@irit.fr

Frederic Lerasle
LAAS-CNRS[3]
Univ. de Toulouse
Toulouse, France
lerasle@laas.fr

***Abstract*—This paper introduces a new raw-audio, environmental, kitchen-related, non-vocal dataset to fill what we consider as a gap in the context of audio datasets. Our so-called Kitchen20 dataset is compared to other datasets such as ESC-50 and shown that both datasets can be merged together in what we call ESC-70.**

A human quantitave appreciation of the audio samples contained in Kitchen20 is provided as well as several machine learning benchmarks on both Kitchen20 and ESC-70.

I. Introduction

Companion robots could serve for monitoring and assisting isolated persons, particularly elderly people. Robots which act and perceive in a human environment must rely on embedded sensors of different modalities in order to perceive and recognize human activities, and in order to provide help if necessary. Vision is often used as the principal modality in performing such recognition tasks [12, 17] but audio can contribute and help in these challenging situations by providing additional information. Moreover, a robot does not always have a clear line of sight, and could thus rely on audio to help understand what the human is currently doing. Both the robotic and the audio communities can benefit from a systematic method to assess the performances of their smart devices using an audio dataset that is quantitatively augmented with challenging sounds coming from human interactions with objects of their daily environment. This paper aims at presenting a new dataset that is going in that direction. This new dataset shall be open-source, extensively tested with several state of the art machine learning techniques, and be easily available to the public for further testing. We acknowledge that similar datasets exist but they are not fully attending our requirements.

Thereafter, section II presents our motivation for building Kitchen20, section III provides an extensive description of the dataset, section IV describes the benchmark implementations, section V shows how the results obtained with this dataset compare with state of the art datasets, and finally, section VI concludes on the work.

This work is supported by a collaboration between SoftBank Robotics Europe, LAAS, IRIT, and ANRT

	#class	#samples	sample length	origin	label
Kitchen20	20	800	5s	FS/Ind.	strong
ESC-50	50	2000	5s	FS	strong
DCase 2	41	9500*	300ms/3s	FS	strong
DCase 4	10	2244*	10s	AS	weak

TABLE I
Comparison of selected raw-audio datasets. In this table FS stands for FreeSound, Ind. for individuals, and AS for AudioSet. *Unbalanced dataset.

II. Motivations and justification for Kitchen20

There are three main reasons for assembling this dataset. Firstly, given this context in which smart systems should understand human actions in their houses based on audio streams, we figured that kitchen-related activities would be interesting as it is a closed environment with characteristic noises. Furthermore, skilled robots could rely on this recognition to provide relevant help to humans, making it a very interesting domain of study. Secondly, we wanted a raw-audio dataset to gain flexibility on the algorithms that would be applied to it. In fact many deep learning models have emerged [7, 9, 14, 16] in the later years using raw-audio instead of hand crafted features. Also illustrating this interest, new conferences based on raw-audio inputs have emerged such as DCase [1] that have seen its attendance double within 2 years, reaching 150 attendees in 2018. Thirdly, to the best of our knowledge, the literature does not provide raw-audio datasets containing that many classes related to kitchen sounds. For these reasons, a kitchen related audio dataset with raw-audio samples is created.

To have a better grasp on the novelty of Kitchen20 we compare it with the literature and found that Kitchen20 could be compared to (1) a subset of AudioSet like DCASE 2018 Task 4 (DCase 4): *Large-scale weakly labeled semisupervised sound event detection in domestic environments* [13]; (2) a subset of FreeSound like DCASE 2018 Task 2 (DCase 2): *General-purpose audio tagging of Freesound content with AudioSet labels* [5]; and (3) another subset of FreeSound like

[1] http://dcase.community/

ESC-50 [10]. A summary of the differences in terms of (a) number of classes, (b) number of samples, (c) sample lengths, (d) audio origin, and (e) label strenght between these datasets is given in Table I.

As specified in its documentation, DCase 4, consists of many 10s audio samples extracted from Youtube videos that are weakly labeled. In fact, by listening to the audio clips, one may notice that more actions than the labeled ones may be happening in a sample, for example, people may be speaking while they are extracting elements from the microwave oven. In addition, this dataset only has four classes closely related to kitchen actions which are *Dishes*, *Frying*, *Blender*, and *Running water*. In DCase 2, the samples are extracted from FreeSound, and, in the verified part of the dataset, each sample is strongly (accurately) labeled. Yet the samples of DCase 2 have varying audio-clip timespans ranging from 30ms to 30s and, as DCase 4, there are very few classes specific to kitchen environments with only *Microwave oven*, and *Drawer open or close*.

Finally, ESC-50, is an environmental dataset providing 50 classes of environmental sounds, strongly labeled, comprised of raw-audio samples, and carrying consistent audio clips of 5s. But, the dataset does not include kitchen labels. A subset of ESC-50, composed of 10 classes, has been proposed by the same authors to easily test new algorithms and is called ESC-10.

Seeing that a dataset like ESC-50 matched our expectations in terms of samples per class, audio-types, and audio-lengths, it was decided to build a new dataset called Kitchen20 that would resemble ESC-50 in its file structure and would have our kitchen oriented classes. In contrast to DCase 4, Kitchen20 is more specific in the sense that, by listening entirely to a sample, an annotator can be confident in the label of the clip he is listening to. In contrast to DCase 2, all the samples in Kitchen20 last 5s. In the end, Kitchen20 have 14 classes specific to the kitchen environement. The remaining 6 classes are related to kitchen but could appear in other environements. Because it is built on top of ESC-50, Kitchen20 and ESC-50 can be merged together in a new dataset that we refer to here as ESC-70[2].

III. Kitchen20 Description

Kitchen20 contains 800 sound samples equally split into 20 different classes. The 20 classes are themselves split into (a) 10 classes related to kitchen appliances and (b) 10 classes related to human manipulations. All 20 classes are listed in Table II, together with a detailed description of their corresponding actions.

In Kitchen20, every class is represented by 40 raw-audio samples. Each of these samples are 5s long. Each sample is attributed to one of 5 folds that are later used, in the learning process, to perform cross validation. All the samples were recorded at least at 44.1KHz and downrated to 44.1KHz when needed. The audio samples were either extracted from FreeSound or recorded in various real kitchen environments when samples found on FreeSound were scarce. Overall, 662 samples originated from FreeSound and 138 samples were recorded in 9 kitchens by 8 people using various modern mobile phones. Different non-overlapping samples extracted from the same original audiotrack are clustered together in the same fold when possible. The compatibility of Kitchen20 with ESC-50 accounts for all the similarities in audio-lengths, audio-rates, audio sources, and dataset folding.

Appliances	
Dishwasher	A dishwasher running
Microwave	A microwave being set up and running
Blender	A blender blending and stopping
Fridge	Opening and closing a fridge
Juicer	Using an electric juicer
Stove-fan	Switching and running a stove-fan
Frying-pan	Cooking oily food in a pan
Stove-burner	Switching on and off gaz on a stove-burner
Boiling-water	Water boiling in a kettle or a pan
Water-flowing	Water flowing from a tap
Human Manipulations	
Clean-dishes	Cleaning dishes and cutlery with water
Chopping	Chopping vegetables
Cupboard	Opening and closing cupboards
Drawer	Opening and closing drawers
Cutlery	Manipulating cutlery with noise
Plates	Manipulating plates with noise
Sweep	Sweeping a floor
Book	Opening, closing and going through a book
Peel	Peeling vegetables
Eat	Chewing chips and other foods

TABLE II
THE 20 AUDIO CLASSES PRESENT IN THE KITCHEN 20 DATASET.

Because Kitchen20 was gathered from FreeSound and individual kitchens, there is no overlapping samples between Kitchen20 and public audio-visual datasets such as YouTube-8M [1], or the Epic-Kitchens Dataset [4]. As a result, Kitchen20 may be used as an auxiliary task while learning any of these audio-visual datasets.

Kitchen20 is available on GitHub[3] together with a Pytorch accessor to the dataset. This accessor creates files that are retro compatible with the original implementation of both EnvNet and EnvNetV2[4]. These models are presented in section IV. Overall, the python module can retrieve Kitchen20, ESC-10, ESC-50, and ESC-70 datasets with helper classes. An example of the module is shown in Listing 1.

Together with the dataset, a human based study was conducted to understand the perception difficulties found in Kitchen20. The dataset has been evaluated by a panel of 16 participants consisting of 5 women and 11 men aged from 22 to 45. All of the participants were new to the dataset, and were asked to classify 75 to 80 different sounds each. During the experiments, the participants were allowed to listen to each sample as many times as they wanted prior to classifying them. Before moving to the next sample, the participants were shown

[2]Named as an extension of ESC-50 with the permission of the author of [10]

[3]https://github.com/marc-moreaux/kitchen20

[4]https://github.com/mil-tokyo/bc_learning_sound

```python
import torch
from torch import nn
from torch.utils.data import DataLoader
from kitchen20 import esc
import kitchen20.utils as U

# Get a training set at 16KHz,
# with on-the-fly data augmentation
k20_train = esc.Kitchen20(
    folds=[1, 2, 3, 4],
    audio_rate=16000,
    transforms=[
        U.padding(inputLength//2),
        U.random_scale(1.25),
        U.random_crop(24000),  # 1.5s
        U.normalize(float(2 ** 16 / 2))
    ])

# Get a validation set at 16KHz
k20_val = esc.Kitchen20(
    folds=[5],
    audio_rate=16000,
    transforms=[
        U.random_crop(24000),  # 1.5s
        U.normalize(float(2 ** 16 / 2))
    ])

# Pytorch DataLoaders
train_loader = DataLoader(k20_train,
    batch_size=32, shuffle=True)
valid_loader = DataLoader(k20_val,
    batch_size=32)

```

Listing 1. Creating train and validation sets with the Kichen20 library

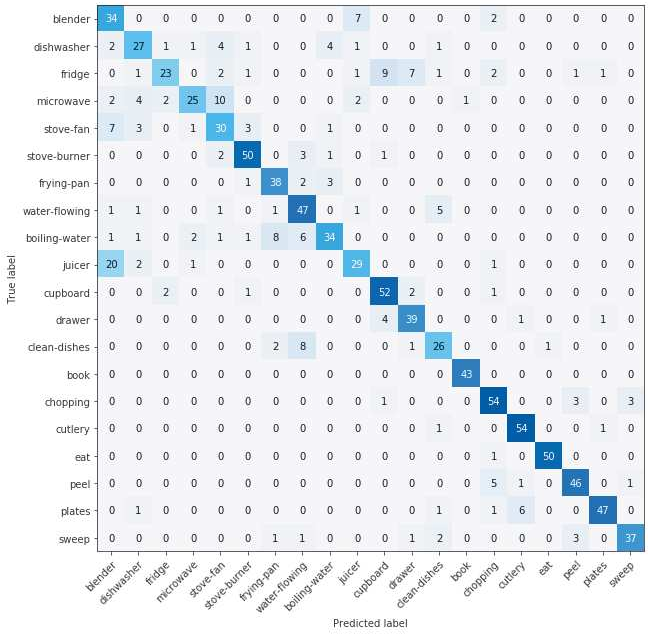

Fig. 1. Human confusion matrix obtained by a panel of 16 participants on Kitchen20.

the ground-truth of the sample they had just classified but they did not have the possibility to change their annotation.

The results of this experiment by our panel shows an accuracy of 78.9% on 996 randomly sampled sounds. The confusion matrix obtained from this experiment is displayed on Figure 1. From this matrix we observe that the classes *Eating*, *Book* and *Cutlery* are the easiest to figure out whereas *Fridge*, *Microwave* and *Juicer* are the hardest. We explain the confusion by (1) the fact that our panel was not used to using the type of juicers present in the dataset, (2) the fact that the fridge slightly sounds like a cupboard when both are closing, and (3) a microwave sounds like a stove-fan when it is running. From these observations, one might conclude that this dataset is not trivial.

IV. Benchmark classifier implementation

This section presents two types of classification baselines evaluated on both Kitchen20 alone and jointly with ESC-50.

The first type of baseline relies on classical machine learning approaches. The methods used are *K-Nearest-Neighbor* (KNN), *Random-Forest* (RF) and *Support Vector Machine* (SVM) [2, 3] using the implementations from Scikit-Learn[5]. All three methods use hand-engineered features that are based on the *zero-crossing rate* (ZCR) [6] and the *Mel-Frequency Cepstrum Coefficients* (MFCC) [11] of the audio at 44.1KHz. To be more specific, the features are (a) the mean and variance of the zero-crossing rate of each consecutive 11.6 ms time window of a given sound together with (b) the mean and variance of the first to thirteenth MFCCs of the same time window. The MFCCs are extracted by using the default parameters of the Librosa library available on Python[6].

The hyperparameters of each method are finetuned using a Bayesian Optimization [8] maximizing the accuracy of each method over the 5 folds. The set of parameters concerned by the optimization are a subset of those available on Scikit-Learn. The KNN is optimized over its number of neighbors. The RF is optimized over the number of features to consider when looking for the best split (Max-features), the number of trees estimators in the forest, the maximum depth of the trees (Depth-tree), and the criterion measuring the quality of a split. Finally, the SVM is optimized over the penalty parameter C, the kernel coefficient γ and the kernel type. Each Bayesian Optimization is performed using the Python module Hyperopt [7] called with the Tree Parzen Estimator (TPE).

The second type of baseline are two deep learning models: *EnvNet* and *EnvNetV2*. These networks were presented with ESC-10 and ESC-50 [10] and form a good baseline for raw-audio classification problems. EnvNet is a 7-layer-deep convolutional-neural-network; it takes as an input a 1.5s raw-audiofile at 16KHz; in its core, the network uses a mechanism to change its time-based convolutions to frequency-based convolutions; and, the network ends with a probability distribution over each classes in the dataset it is trained on. EnvNetV2 is an upscaled version of EnvNet. The core mechanism of the network and the probability distribution

[5]https://scikit-learn.org/

[6]https://librosa.github.io/librosa/index.html

[7]https://github.com/hyperopt/hyperopt

at the end of the network are the same, but EnvNetV2 is a 13-layer-deep convolutional-neural-network with 1.5s raw-audiofiles at 44.1KHz.

The training procedure of this network relies on feeding *Strongly Augmented* data [16] and learning with a *Between Classes* strategy [15]. Strongly Augmented data implies having input clips that are (1) normalized, (2) padded with zeros, and (3) randomly stretched or contracted. Between Classes learning implies feeding the network with two superposed sounds randomly picked from two classes and predicting the gain ratio between these two inputs. Regarding the hyperparmeters of these models, the number of epochs, the learning rate (LR), and the learning schedule are set to the same default values as the ESC-10 learning scheme present in the original code[8] [10] because the default values of the ESC-50 learning led to an exploding gradient. All these hyperparameters are reported on Table III in Section V.

V. Classification Results

This section presents the results of the experiments starting with Kitchen20 and followed by ESC-10, ESC-50 alone or merged with Kitchen20. In either cases, the hyper-parameters optimization results are presented first. They are followed by the accuracies of the models and general remarks.

After 200 steps of Bayesian Optimization on each of the classical machine learning algorithms, applied on Kitchen20, it is found that the best KNN is reached when the algorithm uses a single neighbor; the best RF is reached with 765 estimators, 8 features max, a maximum depth tree of 16 and the entropy criterion; and the best SVM is reached with a linear kernel, a C value of 0.023 and γ equals 4.202.

KNN	Neighbors: 1
RF	Estimators: 765 - Max-features: 8 - Depth-tree: 16 Criterion: Entropy
SVM	Kernel: linear - C: .023 - γ: 4.202
EnvNet	nEpochs: 1200 - LR: 0.01 - Schedule: .5, .7, .9
EnvNetV2	nEpochs: 2000 - LR: 0.01 - Schedule: .3, .6, .9

TABLE III
Hyperparameters used for training given algorithms on Kitchen20.

Given these hyperparameters, the KNN achieves 35.3% accuracy, RF achieves 49.5%, SVM achieves 44.5%, EnvNet achieves 71.8%, and EnvNetV2 achieves 79.1%. All the results of these classifiers are summarized, per folds, on Figure 2, and throughout the folds on Table IV. For further understanding, the confusion matrix obtained with the best baseline system, namely EnvNetV2, is shown on Figure 3.

Similarly to what has been observed in image classification, deep neural networks based on raw-audio outperform classical approaches based on extracted features. Also, a network properly trained with more parameters such as EnvNetV2 is more accurate than its lower dimension counterpart, namely EnvNet.

[8]https://github.com/mil-tokyo/bc_learning_sound/blob/master/opts.py

	KNN	RF	SVM	EnvNet	EnvNetV2
Kitchen20	35.3	49.5	44.5	71.8	79.1
ESC-70	23.5	37.6	33.0	71.3	78.1

TABLE IV
Mean accuracies achieved on cross folded learning of Kitchen20, and ESC-70 by five distinct machine-learning classifiers.

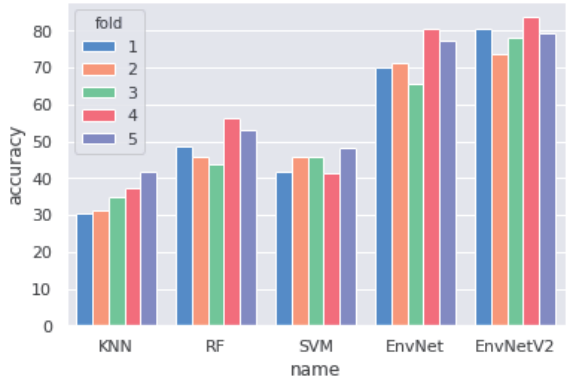

Fig. 2. Accuracies achieved on 5 fold testing of Kitchen20 by K-Nearest-Neighbor, Random-Forest, SVM, EnvNet, and EnvNetV2 approaches.

Looking at the confusion matrix displayed in Figure 3, we see that the classes *Eating*, *Book* and *Cutlery* are the easiest to figure out whereas *Fridge*, *Microwave* and *Juicer* are the hardest.

Jointly looking at the human and the neural confusion matrices, we notice that many confusions made by humans are reflected on the neural learning. For instance, in both cases, *microwaves* are mistaken with *stove-fans*, *blenders* with *juicers*, *clean-dishes* with *water-flowing*, and *dishwashers* with both *stove-fans* and *boiling-water*.

Examining the sole EnvNetV2 confusion matrix, one can also notice the low true-positive-rate of the *microwave* and the *stove-fan*. When listening to the corresponding audio-clips, one may notice that the audio from both classes only differs from one another by a triggering sound in the clip, like a door closing for the *microwave* or a button switched for the *stove-fan*. Finally, *Clean-dishes* is another interesting class as its predictions get mixed-up with *water-flowing*, *plates* and *cutlery* which truly represent what *clean-dishes* is: a mix of plates and cutlery manipulated under water-flowing.

ESC-70:

Now looking at the combination of Kitchen20 with ESC-50, after 200 steps of Bayesian Optimization on each of the classical machine learning algorithms, it is found that the best KNN is reached when the algorithm uses 10 neighbors; the best RF is reached with 980 estimators, 3 features max, a maximum depth tree of 17 and the Gini criterion; and the best SVM is reached with a linear kernel, a C value of 0.023 and γ equals 1.766. All of these values are summarized in Table V.

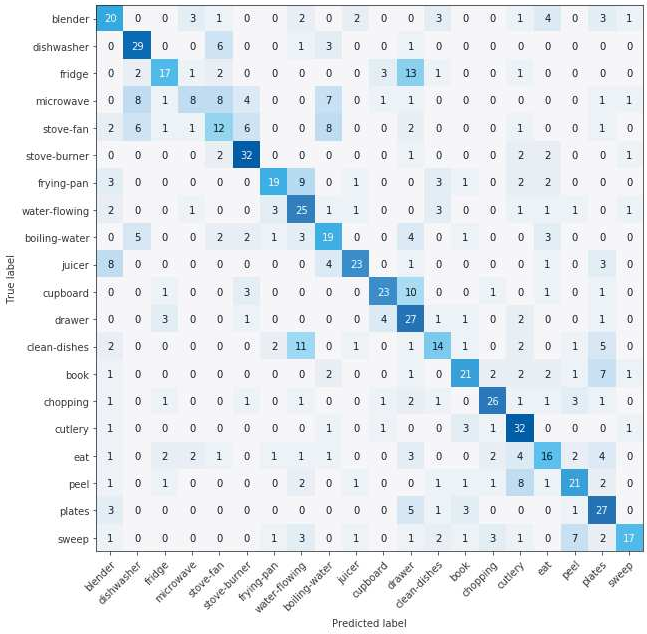

Fig. 3. EnvNetV2's confusion matrix obtained from the validation sets of the 5 folds of Kitchen20.

KNN	Neighbors: 10
RF	Estimators: 980 - Max-features: 3 - Depth-tree: 17 Criterion: Gini
SVM	Kernel: Linear - C: .023 - γ: 1.776
EnvNet	nEpochs: 1200 - LR: 0.01 - Schedule: .5, .7, .9
EnvNetV2	nEpochs: 2000 - LR: 0.01 - Schedule: .3, .6, .9

TABLE V
HYPERPARAMETERS USED FOR TRAINING GIVEN ALGORITHMS ON ESC-70.

Given the hyperparameters shown in Table V, the KNN achieves an accuracy of 23.5%, the RF achieves 37.6%, the SVM achieves 33.0%, EnvNet achieves 71.3%, and EnvNetV2 achieve 78.1%. These accuracies are reported on Table IV. A summary of the accuracies achieved by these machine learning methods applied on different datasets, namely ESC-10, ESC-50, ESC-70, and Kitchen20, is shown in Figure 4.

The dominance of the algorithms observed on Kitchen20 globally stays the same throughout ESC-10, ESC-50, and ESC-70. For every dataset presented, the neural networks are performing better than their traditional machine learning counterparts in terms of accuracy. All baseline approaches have an accuracy increase from ESC-50 to Kitchen20, but both neural networks that show a relative accuracy drop. We hypothesize that this last drop comes from the fact that the hyperparameters used for training EnvNet and EnvNetV2 are the default values found in the original code and are not specific to Kitchen20.

VI. Conclusion

In this paper, a new dataset called Kitchen20 was introduced. Its objective was to leverage the use of the audio modality in a context of helping elderly people in their houses with a robot. Our motivations for building this dataset was (1) the fact that both appliances sounds and sounds based on human interactions in a kitchen are characteristic and (2) the fact that companion robots could be helpful in such environment. Taking into account these motivations, we figured the literature showed a lack of strongly annotated, raw-audio, open-source datasets related to human actions in a kitchen and decided to build one called Kitchen20. Kitchen20 can convienienltly be merged with an existing dataset from the literature called ESC-50 such that a more challenging dataset issued from this merging would be available to the community.

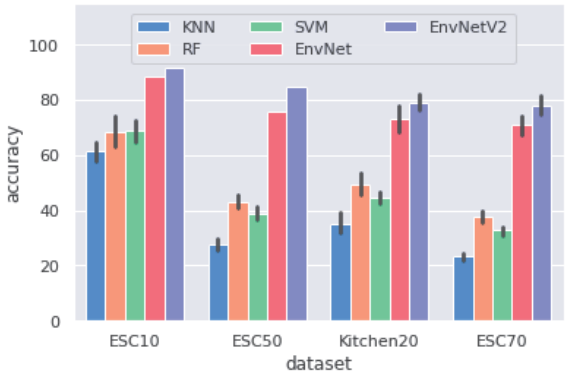

Fig. 4. Mean and variance of the accuracies achieved on cross folded learning of ESC-10, ESC-50, Kitchen20, and ESC-70 by five distinct machine-learning classifiers.

To understand the consistency and what one could expect of Kitchen20, the dataset was evaluated by humans. Going further, many machine learning benchmarks are tested and compared both on Kitchen20 alone and paired with ESC-50. Together they reflect the intra-class similarities that make these datasets learnable.

In later work, we want to analyse to which extend a robot could rely on the audio modality together with the vision modality to understand a human activity in a kitchen environment. To some extend, the Epic-Kitchen Dataset may be used to test such purpose.

VII. Acknowledgment

We thank the authors of ESC-50 for both approving the merging of Kitchen20 with their dataset and for approving the reuse of their naming convention leading to ESC-70.

This work has been carried out under the funding of an industrial doctorates fellowship from National Association for Research and Technology, France

References

[1] Sami Abu-El-Haija, Nisarg Kothari, Joonseok Lee, Paul Natsev, George Toderici, Balakrishnan Varadarajan, and Sudheendra Vijayanarasimhan. YouTube-8M: A Large-Scale Video Classification Benchmark. 9 2016.

[2] Christopher M Bishop. *Pattern recognition and machine learning*. Springer, 2006.

[3] Leo Breiman. Random Forests. *Machine Learning*, 45(1):5–32, 2001.

[4] Dima Damen, Hazel Doughty, Giovanni Maria Farinella, Sanja Fidler, Antonino Furnari, Evangelos Kazakos, Davide Moltisanti, Jonathan Munro, Toby Perrett, Will Price, and Michael Wray. Scaling Egocentric Vision: The EPIC-KITCHENS Dataset, 2018.

[5] Eduardo Fonseca, Manoj Plakal, Frederic Font, Daniel P. W. Ellis, Xavier Favory, Jordi Pons, and Xavier Serra. General-purpose Tagging of Freesound Audio with AudioSet Labels: Task Description, Dataset, and Baseline. *Detection and Classification of Acoustic Scenes and Events*, 2018.

[6] Fabien Gouyon, François Pachet, and Olivier Delerue. On the use of zero-crossing rate for an application of classification of percussive sounds. In *Proceedings of the COST G-6 conference on Digital Audio Effects (DAFX-00)*, Verona, Italy, 2000.

[7] Taejun Kim, Jongpil Lee, and Juhan Nam. Sample-level cnn architectures for music auto-tagging using raw waveforms. *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 366–370, 2018.

[8] J. Močkus. On bayesian methods for seeking the extremum. pages 400–404. Springer, Berlin, Heidelberg, 1975.

[9] Aaron van den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew Senior, and Koray Kavukcuoglu. WaveNet: A Generative Model for Raw Audio. *SSW*, 9 2016.

[10] Karol J. Piczak. ESC: Dataset for environmental sound classification. *Proceedings of the 23rd ACM international conference on MultimediaProceedings of the 23rd ACM international conference on Multimedia*, pages 1015–1018, 2015.

[11] Louis CW Pols and others. *Spectral analysis and identification of Dutch vowels in monosyllabic words*. AmsterdamAcademische Pers, 1977.

[12] Alexander Richard, Hilde Kuehne, Ahsan Iqbal, and Juergen Gall. NeuralNetwork-Viterbi: A Framework for Weakly Supervised Video Learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7386–7395, 2018.

[13] Romain Serizel, Nicolas Turpault, Hamid Eghbal-Zadeh, and Ankit Parag Shah. Large-Scale Weakly Labeled Semi-Supervised Sound Event Detection in Domestic Environments. *Workshop on Detection and Classification of Acoustic Scenes and Events*, 2018.

[14] Yuji Tokozume and Tatsuya Harada. Learning environmental sounds with end-to-end convolutional neural network. *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2017.

[15] Yuji Tokozume, Yoshitaka Ushiku, and Tatsuya Harada. Between-class Learning for Image Classification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5486–5494, 11 2018.

[16] Yuji Tokozume, Yoshitaka Ushiku, and Tatsuya Harada. Learning from Between-class Examples for Deep Sound Recognition. *International Conference on Learning Representations*, 2018.

[17] Chenxia Wu, Jiemi Zhang, Ozan Sener, Bart Selman, Silvio Savarese, and Ashutosh Saxena. Watch-n-Patch: Unsupervised Learning of Actions and Relations. In *IEEE transactions on pattern analysis and machine intelligence*, pages 467–481. IEEE, 2018.