ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology
(IJARTET) Vol. 10, Issue 10, October 2023

Smart Data Processing for Energy Harvesting System Using Ambient Noise with Deep Learning

Miss. Aishwarya K[1], Prof. P.Nisha Priya[2]

Student, Computer Science Engineering, CSI College of Engineering, Ketti, India [1]

Professor, Computer Science Engineering, CSI College of Engineering, Ketti, India [2]

Abstract: Electrical energy makes our lives simpler and fuller. One of the not much researched and developed but abundantly available non-conventional energy sources is noise. The method of conversion from sound energy to electrical energy based on the Piezo Electric principle is one of the efficient ways. To recognize the maximum energy efficiency the noise is classified based on signal processing. The proposed system classifies ambient noise that produces high energy output from the piezoelectric material. The process begins with the noise files, convert them into spectrograms, generate the attention part, feed it to the backbone, VGG16-CNN, and produce predictions about the class to which the noise belongs.

Keywords: Noise classification, Neural networks, Deep learning, Energy Harvesting.

I. INTRODUCTION

Sound is a mechanical wave which is carried over a medium. Commonly, the properties of sound can be understood through its characteristics such as amplitude, frequency, speed, direction and pressure. When this sound becomes undesirable it causes harm to humans. Different alternatives have been used to fully neutralize the noise and which is indeed better for all mankind. There has been ample amount of approaches that were used to sort out this issue. This undesirable form of energy can be converted into useful electricity which could benefit mankind. This could be achieved by using Peizoelectric materials. To get the maximum amount of electricity, the frequency has to be high. This requires classification of the noise from the environment based on its frequency. In order to obtain this setup, classification models can be used. Particularly, deep convolutional neural networks (CNN) are very much suitable to the problem of environmental noise classification. Deep learning is almost similar to Machine learning but the main difference between deep learning and machine learning is, that machine learning models functions efficiently but the model requires some leading from our side. Whereas, the neural networks tries to learn the functioining of the human brain by learning from huge amount of data. This model comproises of three types of layers: input, hidden layers, and output. The signal is received by the input layer, the processing is done by the hidden layer and finally the output layer makes a decision or predicts about the input data. The computation is done by the interconnected nodes. To construct an ML model that can do predictions, specification of what input features the model will analyse in predicting an output is done. The neural networks learn the features right from the data by which they are trained, so there is no need to get the features manually.

The introduction of convolution neural networks has been a game changer for the field of Image Classification. Now this is utilized to get higher level of the features in the image. When the eyes are looking at a picture, our brain takes in large amount of data. Each neuron in our brain functions in its own receptive field and is attached to other neurons in such a manner that they cover the complete visual field. Similar to a neuron that reacts to the changes only in the restricted region of the visual field that is the receptive field, each neuron in a CNN initiates data analysis only in its receptive field. The layers of the CNN are arranged in a manner that they detect basic patterns first and more complex patterns later on.

II. RELATED WORK

This section will present the other works that have been conducted by several people regarding the conversion of ambient noise into electricity by the Deep Learning method. The main objective of this section is to compare the work between various works and find the problem for improvement. The work done is analyzed by different authors in this field and added our work to it. The work done by different authors is as follows:

Mustaqeem, Soonil Kwon, *"CNN-Assisted Enhanced Audio Signal Processing for Speech Emotion Recognition"*, proposed an artificial

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET) Vol. 10, Issue 10, October 2023

intelligence-assisted deep stride convolutional neural network (DSCNN) architecture using the plain nets strategy to understand the salient and discriminative features from the spectrogram of speech signals that are improved in previous steps for better performance.

Aditya Khamparia, Deepak Gupta, Nguyen Gia Nhu, and Ashish Khanna, *"Sound Classification Using Convolutional Neural Network and Tensor Deep Stacking Network"*, proposed an approach for sound classification using spectrogram images of sounds which is effective to develop sound classification and recognition systems.

Pablo Zinemanas, Martín Rocamora, Marius Miron, Frederic Font, and Xavier Serra *"An Interpretable Deep Learning Model for Automatic Sound Classification"*, explain a model for automatic sound classification and its predictions depending on the similarity of the input to a set of learned prototypes in a latent space.

Divya, Swati Panda, Sugato Hajra, Rathinaraja Jeyaraj, Anand Paul, Sang Hyun Park, Hoe Joon Kim, Tae Hwan Oh, *"Smart data processing for energy harvesting systems using artificial intelligence"*, gives an understanding about the combination of NG output with machine learning algorithms for a range of applications.

Karol J. Piczak, *"Environmental sound classification with convolutional neural networks"*. The model performs baseline implementations depending on mel-frequency cepstral coefficients and produces results comparable to other approaches.

Wei Chen, Qiang Sun, Xiaomin Chen, Gangcai Xie, Huiqun Wu, Chen Xu, *"Deep Learning Methods for Heart Sounds Classification: A Systematic Review"*, focuses on improving the accuracy of heart sounds classification, an in-depth systematic review and an analysis of existing deep learning methods were performed, with a significance on the convolutional neural network (CNN).

Aditya Khamparia; Deepak Gupta; Nhu Gia Nguyen; Ashish Khanna; Babita Pandey; Prayag Tiwari, *"Sound Classification Using Convolutional Neural Network and Tensor Deep Stacking Network"*. Here the spectrogram images of environmental sounds are used to train the convolutional neural network (CNN) and the tensor deep stacking network (TDSN).

Alessandro Maccagno, Andrea Mastropietro, Umberto Mazziotta, Michele Scarpiniti, Yong-Cheol Lee & Aurelio Uncini , *"A CNN Approach for Audio Classification in Construction Sites"*. The architecture works on the mel-spectrogram representation of the input audio frames and it shows its effectiveness in environmental sound classification (ESC) achieving a high accuracy.

Justin Salamon, Juan Pablo Bello, *"Deep Convolutional Neural Networks and Data Augmentation for Environmental Sound Classification"*, proposes a deep convolutional neural network architecture and audio data augmentation is used for overcoming the problem of data scarcity.

III. PROPOSED METHODOLOGY

The proposed system focuses on the ambient noise of the environment which can be converted into electricity by piezoelectric transducers. The system recognizes the noise having higher acoustic properties and classifies them to achieve maximum power using Harmonic-Percussive Source Separation (HPSS) and VGG-16 along with Mel Spectrogram. This helps in increasing the efficiency of the energy harvesting system, to identify the places which give higher electricity.

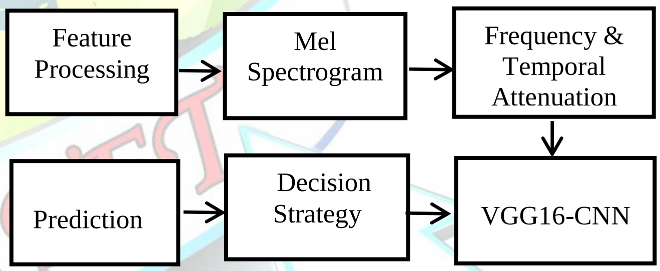

Fig. 1 Block diagram of the propsed system

Librosa is one of the most significant library that is used in audio and music field. Similarly TensorFlow is more suitable library used while working with Deep learning models.

A. Log-Mel Spectrogram

The raw audio format is initially unified and later converted into mono form by average double channel mode. Next, a Hamming window with a size of 46 ms (1024frames, sampling rate 22050 Hz) and overlap of 50\% is used to perform a short-time Fourier transform (STFT) on the data to extract the amplitude spectrogram. Now this spectrogram is being let into the 128-Mel filter bank of bands which is then converted to a logarithmic scale. This will help us produce a Log-Mel spectrogram. The Log-Mel spectrogram is split into 64 frames with 50% overlap, and the zero-

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET) Vol. 10, Issue 10, October 2023

padding method to complete the sub-segments with a length of less than 64 frames. Eventually, the Log-Mel feature of each sub-segment can be shown as a feature vector of size $128 \times 64 \times 1$ (corresponding to frequency × time × channel).

B. Harmonic-Percussive Source Separation

The HPSS algorithm is introduced to the input Log-Mel spectrogram, which can be separated to produce Harmonic Spectrograms and Percussive Spectrograms. The harmonic spectrogram shows the frequency distribution and frequency band activity of the audio data. Now, the convolution kernels with sizes of (1×3) and (5×1) are used to perform convolution operation on both the spectrograms to extract nonlinear features, until the time dimension of harmonic spectrograms and the frequency dimension of the percussive spectrograms are reduced to 1, and then the channel information has to be compressed thereby the (1×1) convolution will be utilized. By this procedure, two one-dimensional matrices AF and AT with sizes (F, 1) and (1, T) is obtained. Eventually, the Softmax function is used to normalize these two matrices to form the frequency weight matrix Fw and the temporal weight matrix Tw.

C. VGG16 Architecture

VGG16 is a convolutional neural network model that is used for image recognition. It consists of only 16 layers that have weights, rather than depending on a large number of hyper-parameters. A VGG network consists of small convolution filters. This model has three fully connected layers and 13 convolutional layers. VGG16 takes input tensor size as 224, 244 with 3 RGB channels. The first two layers consists of 64 channels of a 3*3 filter size and the same padding. Later a max pool layer of stride (2, 2), two layers have convolution layers of 128 filter size and filter size (3, 3) which is followed by a max-pooling layer of stride (2, 2) which is similar to the previous layer. Next there are 2 convolution layers of filter size (3, 3) and 256 filters. Moving ahead, there are 2 sets of 3 convolution layers and a max pool layer. Each of these layers have 512 filters of (3, 3) size along with the same padding. Now, this image is moved on to the stack of two convolution layers. In AlexNet and ZF-Net the filters that are used are of 3*3 and not 11*11 and 11*11 respectively in the convolution and max-pooling. In some layers, it also uses 1*1 pixel which is to manipulate the number of input channels. Padding of 1-pixel (same padding) is required to be done after each convolution layer. This is to avoid the spatial feature of the image.

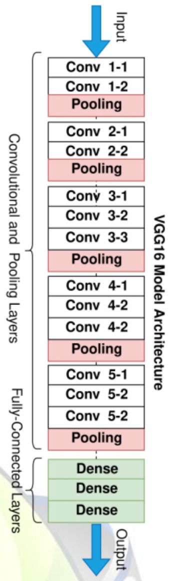

Fig. 2 VGG16 Architecture

The most unique thing about VGG16 is that instead of having a large number of hyper-parameters they focused on having convolution layers of 3x3 filter with stride 1 and usually the same padding and maxpool layer of 2x2 filter of stride 2.

D. Decision Strategy

During the training phase, each sub-segment is put into the network for training, and it predicts the category for each sub-segment. In the final test phase, it is important to predict the complete audio category and use the strategy of probabilistic voting to produce the predicted results of multiple sub-segments for judgment.

$$\mathrm{C} = \arg\max\left(\frac{1}{\mathrm{N}}\sum_{\mathrm{j}=1}^{\mathrm{N}} \mathrm{f}_{\mathrm{ji}}\right), 1 \leq \mathrm{i} \leq \mathrm{K}$$

where N denotes the number of sub-segments divided into each audio sample, K represents the number of categories in the dataset, and f is the prediction result for each segment.
Compile the Model
To compile the model the loss function is defined which is categorical cross-entropy, then accuracy metrics which is the accuracy score, and an optimizer which is Adam.

E. Train the Model

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology
(IJARTET) Vol. 10, Issue 10, October 2023

The model is trained and saved in HD format. The training is done on the model for 25 epochs and a batch size of 32.

F. Check the Test Accuracy

The evaluation of the model is done on test data.

IV. Experimental Results

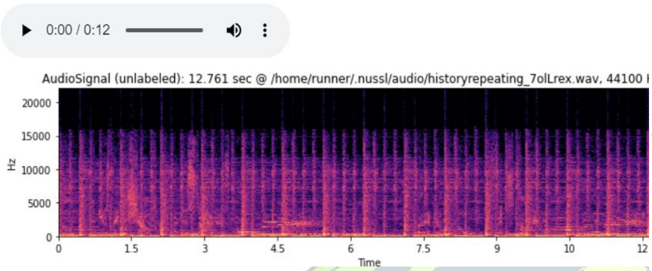

Fig.3 Audio loaded and converted into Spectrogram

The selected audio from the dataset is initially uploaded and converted into the spectrogram as shown in Fig.3

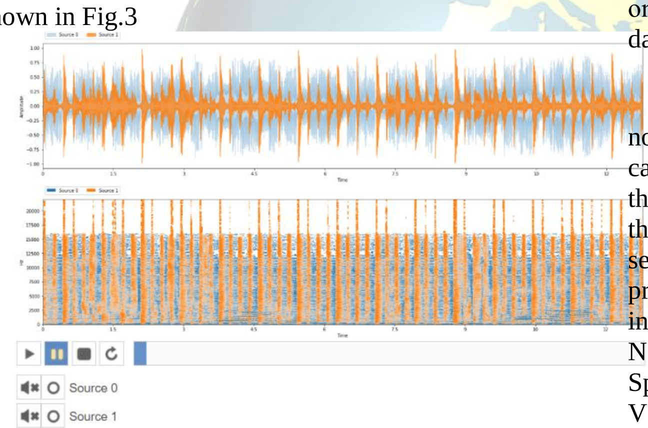

Fig. 4 Splitted Harmonic and Percussion Spectrogram

The HPSS algorithm splits the original audio into two spectrograms as shown in Fig.4 which shows frequency distribution and frequency band activity.

Test Accuracy: 0.85

Fig. 5 The obtained Test accuracy

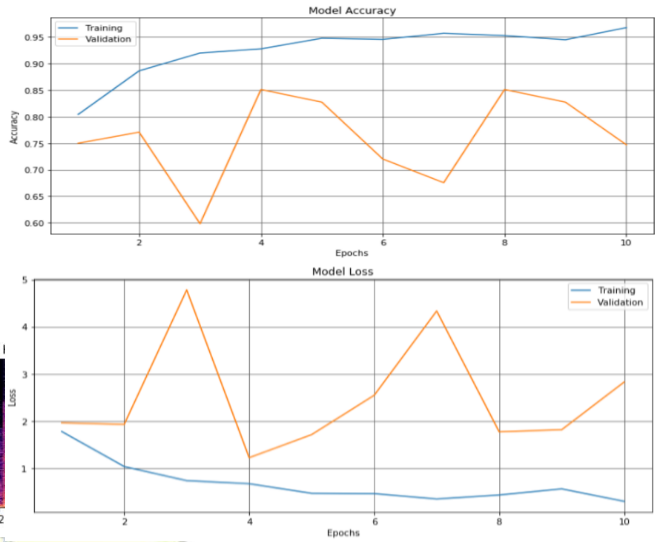

Fig.6 Plotting Accuracy and Loss Overtime

The recorded history is used during the training to get a plot of accuracy metrics and also plot the loss on both the training (“loss”) and test (“val_loss”) data.

V. Conclusion

Work has been done to process the signal of the noise and to classify where the maximum energy can be converted into electricity. The processing of these massive amounts of information is beyond the human eye's capability as the amount of data sets increases daily. Here the ambient noise is processed from the surroundings and converted it into Spectrograms which is acceptable by a Neural Network. This is done with the help of Mel Spectrogram and Temporal Frequency-based VGG16 Architecture.

Applications: The development of the Internet of Things concept, wearable devices, and wireless technologies has led to the need for self-powered systems due to the inaccessibility of batteries for changing. A solution for these self-powered systems is to harvest mechanical energy using piezoelectricity. The proposed model could be used in self-driven IoT devices that are being exposed to the environment. The noise or sound which is produced in the surroundings could be optimized which leads to the functioning of the IoT system without the need of replacing the battery.

References

1. A Khamparia, D. Gupta, N. G. Nguyen, A. Khanna, B. Pandey, and P. Tiwari, "Sound classification using convolutional neural network and tensor deep stacking network", IEEE Access, vol. 7, pp. 7717-7727, 2019.

2. M. Toğaćar, Z. Cömert, and B. Ergen, "Classification of brain MRI using hyper column technique with convolutional neural network and feature selection method", Expert Syst. Appl., vol. 149, Jul. 2020.

3. Y Guo, H. Cao, J. Bai, and Y. Bai, "High efficient deep feature extraction and classification of the spectral-spatial hyperspectral image using cross-domain convolutional neural networks", IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 12, no. 1, pp. 345-356, Jan. 2019.

4. S Srivastava, Y. Wang, A. Tjandra, A. Kumar, C. Liu, K. Singh, and Y. Saraf, "Conformer-based self-supervised learning for nonspeech audio tasks," arXiv preprint arXiv:2110.07313, 2021.

5. Y Verbitskiy, V. Berikov, and V. Vyshegorodtsev, "Eranns: Efficient residual audio neural networks for audio pattern recognition," arXiv preprint arXiv:2106.01621, 2021.

6. L. N. Smith and N. Topin, "Super-convergence: Very fast training of neural networks using large learning rates," in Artificial intelligence and machine learning for multi-domain operations applications, vol. 11006. International Society for Optics and Photonics, 2019, p. 1100612.

7. Kong, Y. Cao, T. Iqbal, Y. Wang, W. Wang, and M. D. Plumbley, "Panns: Large-scale pre-trained audio neural networks for audio pattern recognition," arXiv preprint arXiv:1912.10211, 2019

8. K Koutini, J. Schluter, H. Eghbal-zadeh, and G. Widmer, "Effi- ¨ cient training of audio transformers with patchout," 2021.

9. Ford, H. Tang, F. Grondin, and J. Glass, "A Deep Residual Network for Large-Scale Acoustic Scene Analysis," in Proc. Interspeech 2019, 2019, pp. 2568–2572. [Online]. Available: http://dx.doi.org/10.21437/Interspeech.2019-2731

10. P. Nisha Priya, S. Anila ”Fetal head biometrics measurements using convolutional neural network and mid-point ellipse drawing algorithm” in Springer Multidimensional Systems and Signal Processing (2023), 17 August 2023.

11. S Divya, Swati Panda, Sugato Hajra, Rathinaraja Jeyaraj, Anand Paul, Sang Hyun Park, Hoe Joon Kim, Tae Hwan O, "Smart data processing for energy harvesting systems using artificial intelligence" at Nano Energy, Volume 106, February 2023, 108084.

12. Muhammad Zawish; Nouman Ashraf; Rafay Iqbal Ansari; Steven Davy, "Energy-Aware AI-Driven Framework for Edge-Computing-Based IoT Applications" in IEEE Internet of Things Journal (Volume: 10, Issue: 6, 15 March 2023).

13. Tim Bäßler, Robin Bäßler, Markus Kley, "Augmented mel-spectrogram VGG-16 model for axial and radial load classification at wire-race bearings" Published by Oldenbourg Wissenschaftsverlag June 28, 2022.

14. Tursunov Anvarjon, Mustaqeem, Soonil Kwon, "Deep-Net: A Lightweight CNN-Based Speech Emotion Recognition System Using Deep Frequency Features" Sensors, Volume 20, Issue 18, 10.3390/s20185212, 12 September 2020.

15. Xiaoyang, Wang, Zhe Zhou, Zhihang Yuan, Jingchen Zhu, Guangyu Sun, Yulong Cao, Yao Zhang, Kangrui Sun, "FD-CNN: A Frequency-Domain FPGA Acceleration Scheme for CNN-based Image Processing Applications" ACM Transactions on Embedded Computing Systems, https://doi.org/10.1145/3559105, 13 September 2022.