Intelligent Music Genre Classification using CNN

Dr Reeja S R
Vellore Institute of Technology, Amaravati, Andhra Pradesh, India

Prof. Ishfaq Yaseen
Prince Sattam bin Abdul Aziz University, Alkharaj, Saudi Arabia

Abstract

In past years the classification of music genre is done various methods. The methodology used for the classification is deep learning and K neural networks. In this paper, we classify the music genre using convolution neural networks. In Music Information Retrieval (MIR), research will always be going on retrieving information from music and on building classifiers with better accuracy. In this paper, we are building a deep learning neural network models to perform a multi-class classification task of labelling music genres using Feed-forward and convolution Neural Network and compare them to see which offers a better accuracy. In our experimental results shows 87.4% accuracy by using GTZAN data set.

Keywords: Music Genres, Convolution Neural Network, Information Retrievals

I. INTRODUCTION

There are different applications of retrieval of information from music. They are segmentation, recognize speech automatically and music genre classification. In music genre classification we will build a program which takes a piece of sample and classifies the sample into its specified genre based on the amplitude, frequency, pitch and sounds of instrument being played in the sample into different genres.

With mainstream music industry being shifted towards online in recent times, where music is being streamed and sold online. Many music streaming platforms like Spotify, iTunes, depend on this classifiers more. The music streaming companies have more uses with classifiers more than ever. The classifiers help companies to design better recommendation systems, which provides audience with music they would like to listen, which will let users to spend more time on those platforms and to generate better playlists. Generating playlist is also an important feature of streaming companies. Playlist contains all the songs of users liking or based on the songs which users have listened previously. These playlist generation is important as users can listen to the songs of there liking, without having to search for each individual songs and these playlist generation also uses classifiers. To help companies better organize their database.

In this paper we classify music into 10 different genres: jazz, reggae, rock, blues, hiphop, country, metal, classical, disco, and pop. The extraction of features from audio samples which we can use to build neural networks from. We will build two different neural networks, First one is feed-forward neural network. We will first model this neural network because CNN is an image based and computationally expensive. We compare this with the proposed model and CNN. We will build a Convolutional Neural Network (CNN) with a specified architecture and compare the accuracy. Both models should generate better accuracy.

II. LITERATURE SURVEY

All Music genre classification is not a new research area in Music Information Retrieval (MIR). Many people to this day working on this area trying to bring new algorithms and improve the models which have been built already. The unwanted noise in audio can be removed by various noise removal techniques[8],[9],[10],[11] and [12].

One such model is by Tao Fang [1], he built multi-class genre classifier for 2,3,4,10-classes. The model tuned well for 2-class, 3-class, and 4-class classification but it didn’t done well for 10 genre classification.

Another research by a group of scholars Aaron van den Oord, Sander Dieleman, Benjamin Schrauwen [2] from Ghent University. Though their work isn’t on classification of music genres. Their work showed interesting results. They used Convolutional neural networks (CNNs) in their Music Recommendation System and the results are great. Various neural network are used[13],14] and [17]. For instance, their model predicted Miley Cyrus, Girls Just Wanna Have Fun & My Chemical Romance, Teenagers on input of Jonas Brothers, Hold on.

So, the basis of this project is to use both of their work. We try to build a model using CNN for 10 genre classification using CNN.

III. Classify the Music Genre

A. *Data Set used*

We are going to be using The GTZAN Genre collection [3] as the audio data to train and test the models we are going to create. There are 100 audio samples for each genre, each sample is of 30s duration of .wav format, neatly labelled. So there isn't any additional work involved in processing and cleaning audio samples. During pre-processing audio noise is removed using LPF [7],[15] and[16].

B. *Extracting Features from Audio Samples*

We are going to extract all the features from an audio signal shown in fig3.1 using a python library called libROSA, which helps us in extracting different features of audio signal like amplitude, frequency, mel spectogram, Mel-frequency cepstral coefficients (MFCC), spectral centroid etc. The audio signal frequencies are shown in figure.

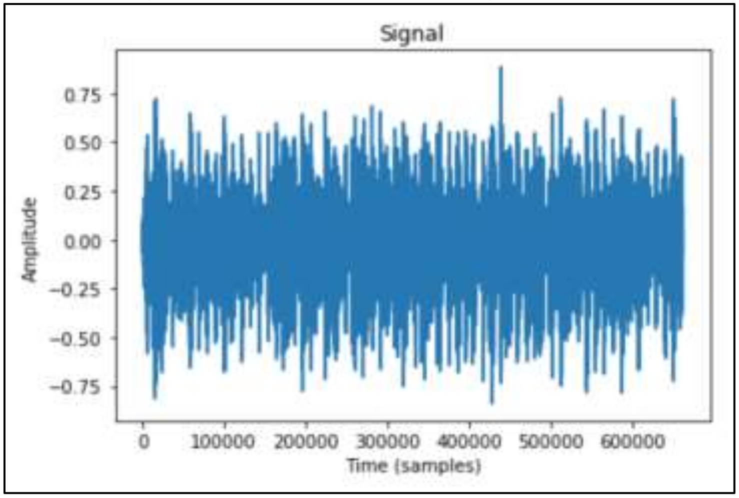


Fig. 3.1: Sample audio signal

We can extract other features from audio samples and use them to model neural networks. But most Music Information Retrieval (MIR) projects like speech recognition, music recommendation systems, and automatic music transcription use Mel Spectogram feature shown in fig.3.2.

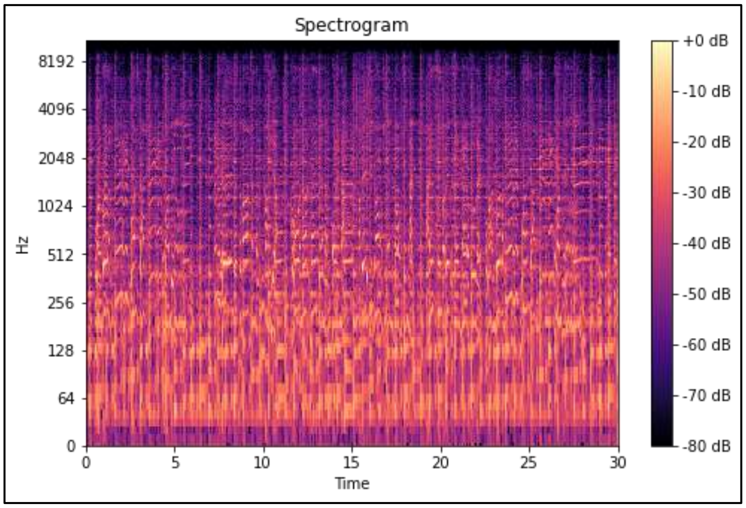


Fig. 3.2: Spectrum Vs Time

mel spectogram is different from normal audio spectrogram. In this spectrum the human being hearing frequency recognize at equal distance. As frequency increases in hertz interval, when mel scale frequency (or simply mels) increases. The name mel derives from melody and indicates that the scale is based on the comparison between pitches. The mel spectrogram remaps the values in hertz to the mel scale. Mel spectrogram data is also suited for use in audio classification applications.A mel spectrogram differs from a linearly scaled audio spectrogram in two ways. A mel spectrogram logarithmically renders frequencies above a certain threshold (the corner frequency). In this mel spectrogram shown in Fig.3.3, the space between those ranges is approximately the same. A mel spectrogram computes its output by multiplying frequency-domain values by a filter bank.The sample builds the filter bank from a series of overlapping triangular windows at a series of evenly spaced mels. The number of elements in a single frame in a mel spectrogram is equal to the number of filters in the filter bank.

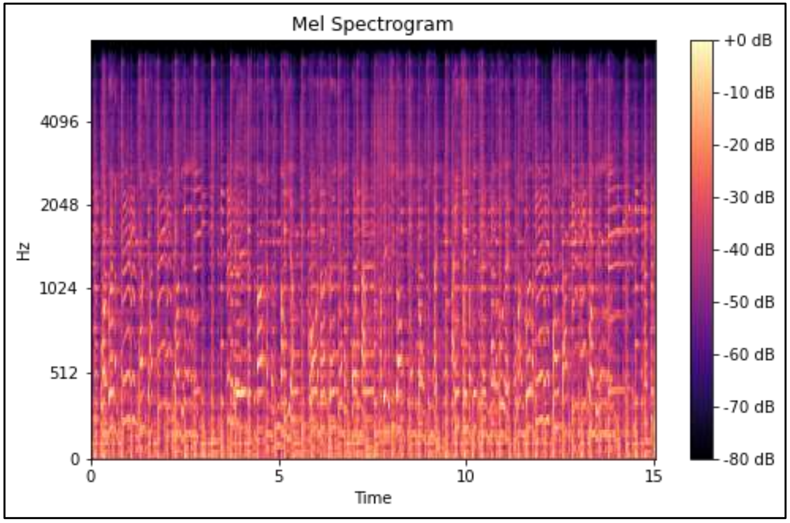


Fig. 3.3: mel spectrogram

IV. EXPERIMENTAL OUTCOME

As CNN is image based neural network and computationally more costly. I thought about using a different neural network first. To see, how it performs and does it perform with same accuracy. In CNN architecture, we use 128 x 660 neurons as an input layer by giving 16 different features with Feature Size: 3 x 3. The 2 x 4 frame size is used in next layer. Again Next layer is equivalent to first layer. In this layer 32 different features are given with Feature Size 3 x 3. The frame size is same but the neurons are reduced to 64. At last in output 10 neurons of 10 different genres. Mel Spectogram's of 10 Genres is shown in Fig. 4.1.

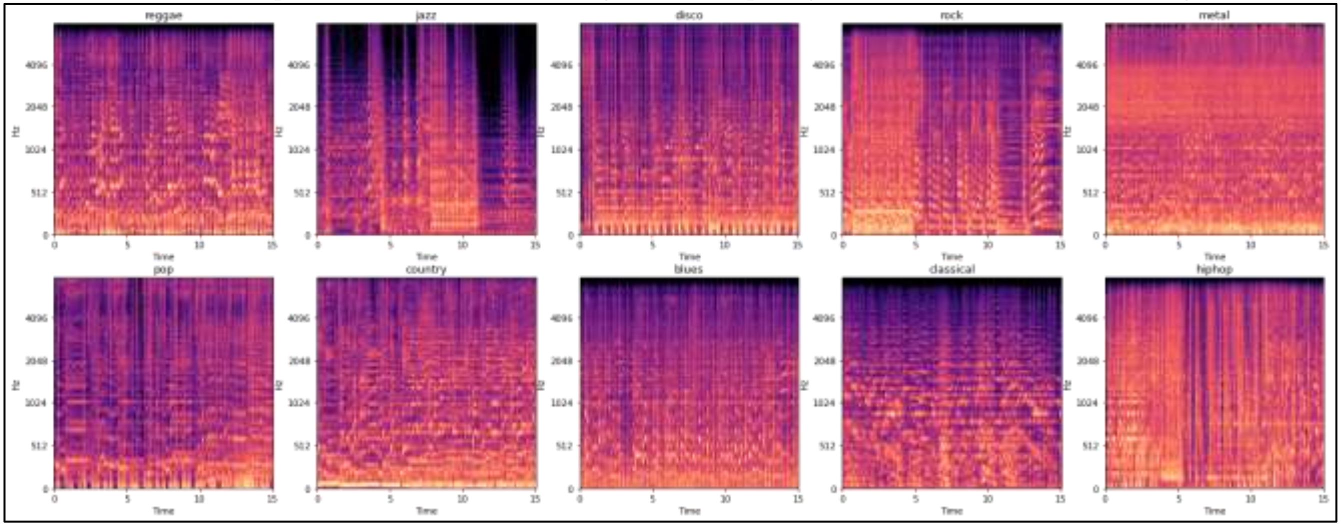


Fig. 4.1: Mel Spectogram's of 10 Genres

A. *Confusion Matrix*

To know what was happening with model, we tried to plot a confusion matrix to visualize the model's predictions against actual values.

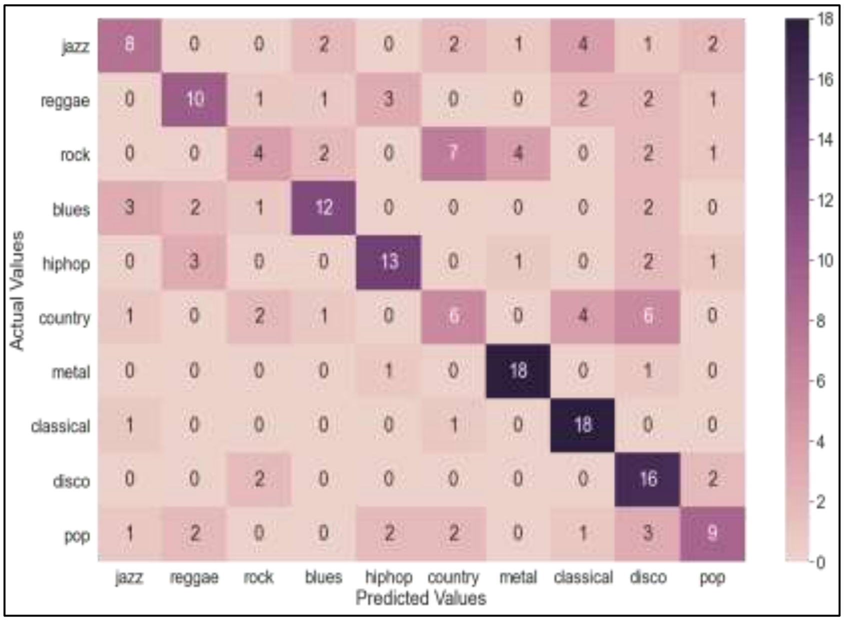


Fig. 4.2: Blues samples have been most mismatched with jazz

The Results show that, the blues samples have been most mismatched with jazz shown in Fig.4.2. This is actually true because blues and jazz are similar and even humans get confused sometimes differentiating between these two.

It also shows that, the jazz samples have been most mismatched with classic. It is also can be noted because most of the classical and jazz music have sounds of common music instruments like Trumpet, Piano, Trombone, Drums etc. And even some people get confused between these two genres and some people even assume jazz is the same as classical music.

Metal and Disco are the most correctly matched samples, it is great because the music can be differentiated because they have different sounds.

Also we tried CNN for binary classification of genres and module has tuned well for binary classification. When tried with metal and classical it matched every sample correctly.

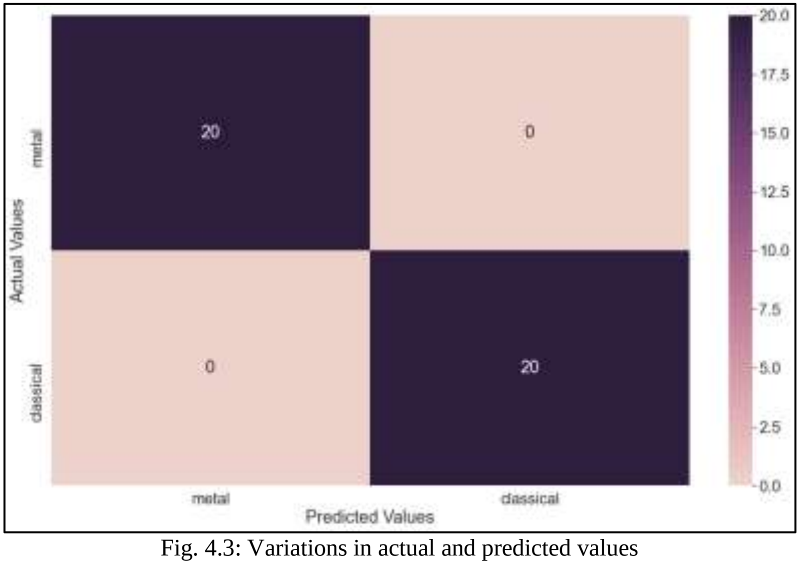


Fig. 4.3: Variations in actual and predicted values

Though the accuracy and predicted values are great the data set is smaller with each genre having 100 samples out of which 80 are used for training shown in Fig.4.3. The model can be prepared better if we have more samples, and also it would be great if we could get real-world music examples to work with. But it is hard to get real-world music as we have to pay some fee for using those samples.

Also musical genres are loosely classified. So, it is also important to take this into consideration because some may sound different but can actually be different genre.

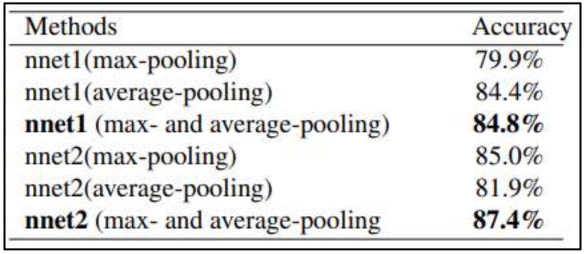
The result shows the comparison table of DNN, feed forward KCNN and convolution Neural Network (CNN), Convolution network had given better accuracy than the other method. The results of various pooling is shown in table. Max and average pooling gives 87.4% of accuracy shown in table 4.1. The training and testing loss and accuracy epoch is shown in fig.4.4 and 4.5

Table – 4.1
Accuracy

Method	*Accuracy*
DNN	*83*
Feed-Forward Neural Network	*82.2*
KCNN	*83.9*
Convolutional Neural Network	*87.4*

Result for various pooling is shown in table 4.2.

Table – 4.2
Pooling result



Methods	Accuracy
nnet1(max-pooling)	79.9%
nnet1(average-pooling)	84.4%
nnet1 (max- and average-pooling)	**84.8%**
nnet2(max-pooling)	85.0%
nnet2(average-pooling)	81.9%
nnet2 (max- and average-pooling	**87.4%**

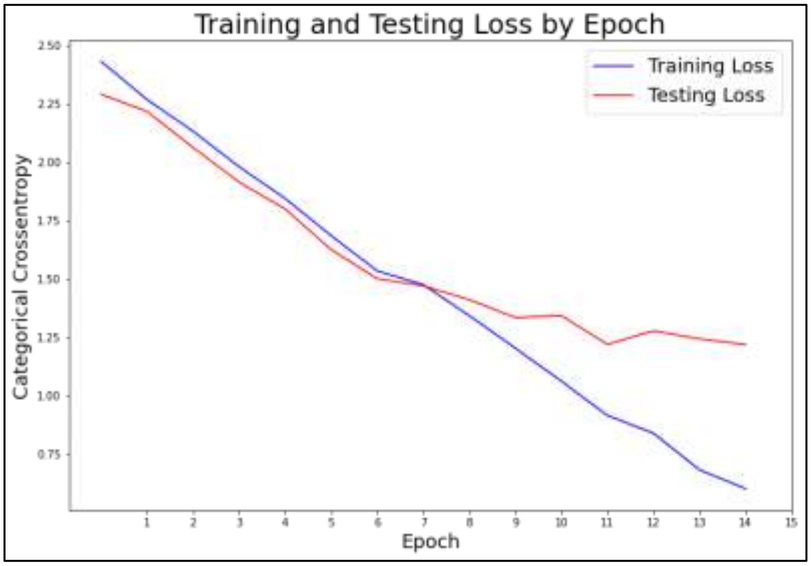


Fig. 4.4: Training and Testing Loss by Epoch in CNN

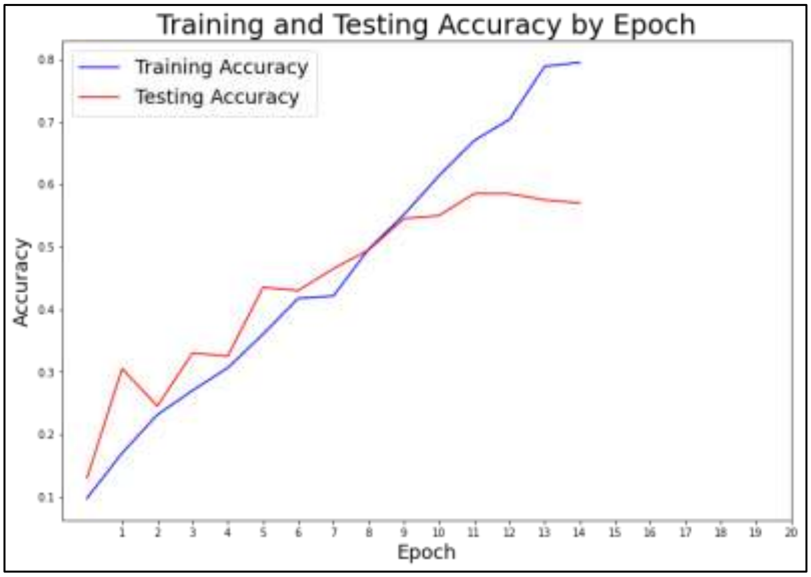


Fig. 4.5: Training and Testing Accuracy by Epoch in CNN

V. CONCLUSION AND FUTURE WORK

Both Feed-forward Neural Network (FFNN) and Convolutional Neural Network (CNN) are giving better accuracy. While later performed better between those two. When looked at Confusion Matrix, we can see that even machine is struggling to differentiate between genres like blues and jazz, and also between jazz and classical which are even mostly confused by humans. Also, in future we can work more on this music genre classification and try to improve accuracy of the models with large data set. We can also go deep and sub-categorize genres and find more data related to those genres, which will help to prepare even better models. We can also try and use different features of audio signals and use a different approach to build models and get better accuracy. The best thing that can be useful in taking this musical genre classification ahead is using real-time data. But real-time data can be costly even for chunks of data. So, getting real-world data can boost the performance of the algorithm.

REFERENCES

[1] Feng T. Deep learning for music genre classification. private document. 2014.Aaron van den Oord, Sander Dieleman, Benjamin Schrauwen, “Deep content-based music recommendation”,

[2] Nakai, Tomoya, Naoko Koide-Majima, and Shinji Nishimoto. "Music genre neuroimaging dataset." Data in Brief 40 (2022): 107675.What is Spectogram,https://pnsn.org/spectrograms/what-is-a-spectrogram

[3] Goldstein, Jerome A., and Mel Levy. "Linear algebra and quantum chemistry." The American mathematical monthly 98.8 (1991): 710-718.

[4] Couellan, Nicolas. "Probabilistic robustness estimates for feed-forward neural networks." Neural Networks 142 (2021): 138-147.

[5] Castellon, Rodrigo, Chris Donahue, and Percy Liang. "Codified audio language modeling learns useful representations for music information retrieval." arXiv preprint arXiv:2107.05677 (2021).

[6] Grattarola, Daniele, and Cesare Alippi. "Graph neural networks in tensorflow and keras with spektral [application notes]." IEEE Computational Intelligence Magazine 16.1 (2021): 99-106.

[7] Srinidhi S Shetty, Reeja S R, “Audio Noise Removal – The State of the Art,” International Journal of Computational Engineering Research, Vol 04, Issue 12, December 2014, ISSN (e): 2250 - 3005, pages 34 -37

[8] S. R. Reeja and N. P. Kavya, "Real time video denoising," 2012 IEEE International Conference on Engineering Education: Innovative Practices and Future Trends (AICERA), 2012, pp. 1-5, doi: 10.1109/AICERA.2012.6306745.

[9] Reeja, S. R., and N. P. Kavya. (2012), "Noise Reduction in Video Sequences-The State of Art and the Technique for Motion Detection." International Journal of Computer Applications 58.8 (2012).

[10] Reeja, S. R., and N. P. Kavya.(2012) "Noise Reduction in Video Sequences-The State of Art and the Technique for Motion Detection." International Journal of Computer Applications 58.8 (2012).

[11] Norman Dias, Reeja S R, “A quantitative report on the present strategies of Graphical authentication,” International Journal of Computer Sciences and Engineering, Vol.06, Issue.10, pp.64-73, 2018.

[12] Kavya, Reeja SR, and N. P. Dr. "An Approach for Noise Removal from a Sequence of Video." International Journal of Scientific & Engineering Research 5.4 (2014): 1266-1270.

[13] S.R. Reeja, Rino Cherian, Kiran Waghmare, Jothimani,Chapter 7 - EEG signal-based human emotion detection using an artificial neural network, Editor(s): Hemanth D. Jude, Handbook of Decision Support Systems for Neurological Disorders, Academic Press, 2021, Pages 107-124, ISBN 9780128222713, https://doi.org/10.1016/B978-0-12-822271-3.00007-4.

[14] R*, Dr.Reeja.S. et al. 2020. Leaf Disease Identification using Convolution Neural Network (CNN). International Journal of Innovative Technology and Exploring Engineering. Blue Eyes Intelligence Engineering and Sciences Engineering and Sciences Publication - BEIESP.

[15] Reeja S R, N P Kavya, Aparna Thampy, Joyel Jose, Real Video Noise Removal in Time, International Journal of Applied Engineering Research 10(86):361-370, March 2022.

[16] Reeja S R, Dr. N P Kavya, Video Denoising:Binary Pattern with Distribution and RTF of PCFFAlgorithm, WJES Volume 1, Issue 4, June-July2013, ISSN: 2320-7213 (PDF) Real Video Noise Removal in Time. Available from: https://www.researchgate.net/publication/358957069_Real_Video_Noise_Removal_in_Time [accessed Apr 13 2022].

[17] Reeja S R, Dr. N P Kavya, Video Denoising:BPD and RTF of PCFF Algorithm, AEMDS-2013,published by Elsevier special issue.