

¹Eedarada Divya
Ratna Manikyam,
²Tanveer Ahmed
Mohammed,
³Machana Lakshmi
Tulasi, ⁴Karthik
Vimmigari

AI-Powered Noise Reduction in Audio and Video Signals



Abstract:

The use of convolutional neural networks (CNNs) for audio noise reduction is examined in this article. Effective noise reduction techniques are required because of the ubiquity of loud surroundings and the requirement for high-quality audio in various applications, including voice recognition, music creation, and communications. Traditional techniques frequently have trouble adjusting to different noise profiles, producing unsatisfactory results. CNNs, in comparison, provide a data-driven strategy to deal with these issues. Convolutional layers, pooling, and post-processing methods are a few crucial elements of the CNN architecture that are clarified to assist readers in understanding the fundamentals of noise reduction

Introduction

In today's society, audio noise is a pervasive issue that significantly impacts our daily lives and overall well-being. It is an uninvited guest in our homes, workplaces, and public spaces, often leading to disturbances and discomfort. Beyond being a mere annoyance, it can have profound effects on our health, contributing to stress, sleep disturbances, and even cardiovascular issues. Furthermore, in the digital world, audio noise can degrade the quality of communication, music, and multimedia content. Despite numerous noise control measures in place, the problem persists, necessitating further research and innovative solutions. Audio noise removal, a critical aspect of signal processing, is a technique employed to enhance the quality of audio signals by reducing or eliminating unwanted sounds or disturbances. The process of audio noise removal involves complex algorithms and techniques that aim to distinguish the desired signal and generate Clean audio, thereby improving the overall audio experience. In the realm of audio processing, achieving pristine sound quality is often challenged by the presence of unwanted noise. Noise can manifest in various forms, from the hum of background chatter in a recorded conversation to the static interference that plagues radio transmissions. As a result, the quest for effective noise removal techniques has long been a cornerstone of audio engineering and signal processing.

Convolutional Neural Networks, initially designed for image processing, have proven their mettle in a multitude of applications, and audio noise reduction is no exception. Their ability to automatically extract relevant features from data and their capacity to adapt to complex, dynamic noise profiles make them an ideal candidate for improving audio clarity. In this article, we delve into the world of audio noise removal using CNNs, exploring how this cutting-edge technology is changing the landscape of sound processing

High-quality video content is crucial in today's digital age for engaging audiences across various platforms. However, achieving pristine video quality is challenging due to noise and artifacts from low-light conditions,

¹ ^{1,2,3,4}International School Of Technology And Sciences For Women, A.P, India.

compression, data transfer errors, and sensor limitations. Noise manifests as random variations in brightness or color, degrading visual clarity and detail. Traditional denoising techniques like Gaussian Blur and Fast Non-Local Means help but often fail to balance noise removal with detail preservation.

Noise in videos is classified into temporal noise (fluctuating lighting), spatial noise (sensor irregularities), and compression artifacts (encoding/decoding errors). Effective denoising must address these categories, ensuring computational efficiency for real-time applications like live streaming, robustness to various noise levels and types, adaptability to different content, and preservation of visual features.

This study compares state-of-the-art AI-based denoising models: Convolutional Long Short-Term Memory (CLSTM), Denoising Convolutional Neural Network (DnCNN), and Autoencoder models. Using diverse video datasets and simulating real-world noise scenarios, we evaluate each model's performance through quantitative metrics (PSNR, processing time) and qualitative visual assessments.

Furthermore, we assess computing efficiency and scalability to ascertain practical application. This thorough investigation seeks to uncover optimal denoising techniques using deep learning architectures. The ability to enhance a distorted speech signal by isolating the primary sound is referred to as background noise reduction. Background noise reduction is used universally, including video conferencing solutions, audio/video editing tools, and noise-canceling headphones.

The topic of background noise reduction is evolving swiftly, and artificial intelligence has introduced a plethora of innovative techniques for enhancing its efficacy. An antiquated notion, artificial neural networks, has recently been revitalized as deep learning. Despite the existence of several deep learning techniques for noise reduction, all of them function by learning from a training dataset.

Conventional noise reduction has been effectively implemented in phones, laptops, conferencing systems, and other edge devices. Since the edge device is responsible for first capturing the user's voice, this technique seems rational. Upon capture, the gadget eliminates noise and transmits the result to the other end of the call. Regard stationary noise as having a pattern that is distinguishable from human speech but repetitive. Conventional DSP approaches, such as adaptive filters, may be very efficient in filtering such sounds. Non-stationary sounds exhibit intricate patterns that complicate their differentiation from human speech. The signal may be ephemeral and exhibit rapid flashes.

Related Work

Researchers have used CNN-based autoencoders for the purposes of voice enhancement and noise reduction. Denoising Autoencoders for Acoustic Improvement. The voice intelligibility of these models has markedly improved after training on both noisy and clean speech signals.

Audio Denoising Utilizing Deep Residual Networks: Deep Residual Networks (ResNets) have been adapted for audio denoising. These networks address the vanishing gradient problem via residual connections, allowing enhanced noise reduction while preserving essential audio features.

End-to-end speech improvement with CNNs integrated with transformers: Research indicates that end-to-end models for speech enhancement have been developed by amalgamating the capabilities of Transformers and CNNs. Utilizing improved, noise-free output, these models may transform raw, noisy audio into a refined format without requiring further feature extraction processes.

Attention methods have been included into CNN designs to let the model to concentrate on certain time-frequency areas in the audio spectrogram, enhancing its capacity to differentiate between noise and signal components.

Materials And Methods

Overview of the proposed light weight CNN architecture

A lightweight Convolutional Neural Network (CNN) architecture for audio noise removal is developed to provide efficient noise reduction while reducing computing resources, making it appropriate for real-time and resource-constrained applications.

The suggested lightweight CNN architecture for audio noise reduction integrates efficacy in noise mitigation with computational economy. This practical and effective approach addresses the challenges of enhancing audio quality in diverse real-world applications, where efficient real-time processing is paramount. This design has the potential to enhance the quality of audio experiences, considering the limits of current platforms and devices.

1. Convolutional Layers: These layers extract salient information from input pictures by executing convolutional operations on the data. They consist of many convolutional filters that generate feature maps by spatial scanning of the input.
2. Activation Activities: The output from the convolutional layers undergoes an element-wise application of non-linear activation functions, such as ReLU (Rectified Linear Unit). They introduce non-linearity to the network, facilitating its ability to discern intricate patterns and representations.
3. Pooling Layers: These layers reduce the size of feature maps while retaining essential information by downsampling of spatial dimensions. Max pooling and average pooling are two prevalent pooling techniques.
4. Fully connected layers: These layers connect each neuron in one layer to every neuron in the subsequent layer. By formatting the spatially reduced information for classification or regression tasks, they process and learn advanced representations.

Existing System

Adaptive filters are often used as noise cancellers. The Adaptive Noise Canceller (ANC) is a device that enhances noise-affected communications. It has an advantage over other signal processing methodologies since it does not need any previous knowledge of the signal or noise. The cost is inherently associated with the need of two input signals: the main input signal combined with noise, and a reference noise signal.

Adaptive filtering is a signal processing technique whereby the parameters used for signal processing change according to certain criteria. The estimated mean squared error is often used as a criterion. The settings of adaptive filters are continually modified to satisfy a performance objective, making them time-varying. An adaptive filter may be regarded as a filter that performs the approximation step in real time. The presence of a reference signal is often essential for establishing performance requirements, often obscured in the approximation phase of fixed-filter design. The primary drawback of the steepest-descent (MMS) gradient technique is that it requires precise measurements of the gradient vector at every iteration step. This approach is impractical; an algorithm is required to get estimates of the gradient vector only from the existing data. This is accomplished by the use of the least-mean-square (LMS) error gradient technique. The benefits are as follows: It is accomplished directly, does not need matrix inversion, and does not require correlation assessments. The MMS error gradient technique use

expectation values to compute the gradient vector.

The LMS method employs instantaneous estimations of the vector derived from sample values of the input $[x(n)]$ and the error $e(n)$, rather than use expectation values. Consequently, the following statement is used to get the instantaneous estimate of the gradient vector.

This estimate is evidently neutral, since its predicted value aligns with the value in the aforementioned equation. In accordance with the LMS method, the filter coefficients are adjusted in the direction of the gradient vector estimate as per the following equation.

Proposed System

The longstanding notion of artificial neural networks has recently been revitalized as deep learning. Despite the existence of several deep learning techniques for noise reduction, they all function by learning from a training dataset.

Step 1: Preprocessing

This lesson emphasizes the extraction of pertinent information from the disorganized audio for training purposes.

Step 2: Training Datasets for Noise Mitigation

Developing an effective training dataset is the first stage in constructing an accurate noise reduction model. Our dataset must have recordings of both clear speech and its noisy equivalent, since our aim is to mitigate background noise.

Recurrent Neural Networks (RNNs)

Recurrent neural networks are models adept at recognizing and comprehending sequential input. Audio, text, and the spatial location of an item across time exemplify sequential data. Recurrent Neural Networks (RNNs) excel in mitigating background noise due to their ability to identify patterns over extended durations, essential for audio interpretation.

Step 4: Functionality of RNN

An RNN trained to eliminate background noise from a noisy audio sample. The audio sample may be partitioned into a sequence of uniformly spaced temporal chunks. The hidden state is updated throughout each iteration, maintaining a record of previous steps as each individual sample of the sequence is input into the RNN. Subsequent to each cycle, the output is directed via a feed-forward neural network to generate a fresh audio stream devoid of background noise. Figure 2 depicts the suggested system.

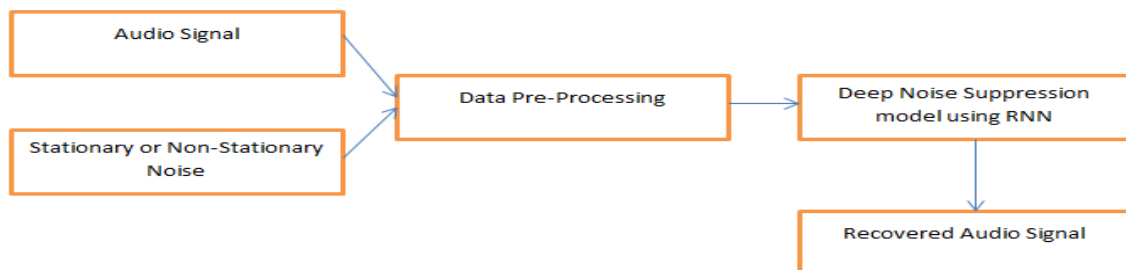


Figure 1: Block Diagram

Block diagram of the proposed system

RNNs excel in audio noise removal by using their sequential processing capabilities. Recurrent Neural Networks (RNNs), via their recurrent connections, encapsulate temporal relationships in audio sources. In the realm of noise cancellation, RNNs acquire the ability to differentiate between important audio characteristics and unwanted noise

patterns over time. This allows them to selectively eliminate noise, improving the quality of the intended audio stream. By analyzing sequential data, RNNs can proficiently predict and mitigate background noise, rendering them essential instruments for enhancing audio quality across diverse applications, including voice recognition, music processing, and telecommunications.

Experimental Setup

The construction of Convolutional Neural Networks (CNNs) for audio noise reduction encompasses many critical components and concerns. A summary of the processes required to establish such an experiment is shown below:

1. Data Collection: Assemble a diverse collection of audio recordings, including both pristine audio samples and those contaminated by various kinds and intensities of noise. Ensure the dataset is accurately annotated with the kinds and intensities of noise.
2. Data Preparation: - The audio data undergoes preprocessing, which often involves: Resampling audio at a constant sample rate. Standardizing the amplitudes of audio.
3. Dataset Partitioning: Divide the dataset into subgroups for training, validation, and testing purposes. The validation set is used for hyperparameter optimization, the testing set for final model assessment, and the training set for model development.
4. Model selection: Choose a CNN architecture appropriate for audio noise reduction. The efficacy of noise reduction must be harmonized with the computational efficiency of the model.
5. Hyperparameter optimization: To get optimal noise reduction efficacy, explore a range of hyperparameters, such as learning rate, batch size, and network design.
6. Data Augmentation: Methods such as random rotations, translations, flips, and adjustments in brightness and contrast may be used to artificially augment the training dataset and enhance the model's generalization capabilities. Data augmentation mitigates overfitting and improves the model's resilience.

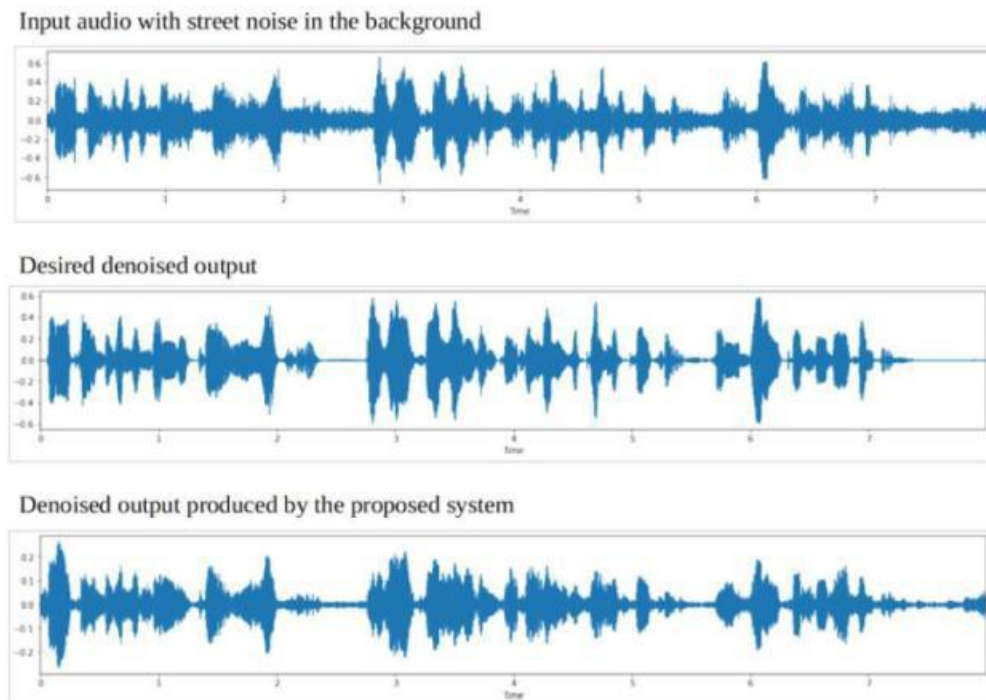


Fig 2. Convolutional Neural Net works to remove noise

Convolutional neural net works Algorithm

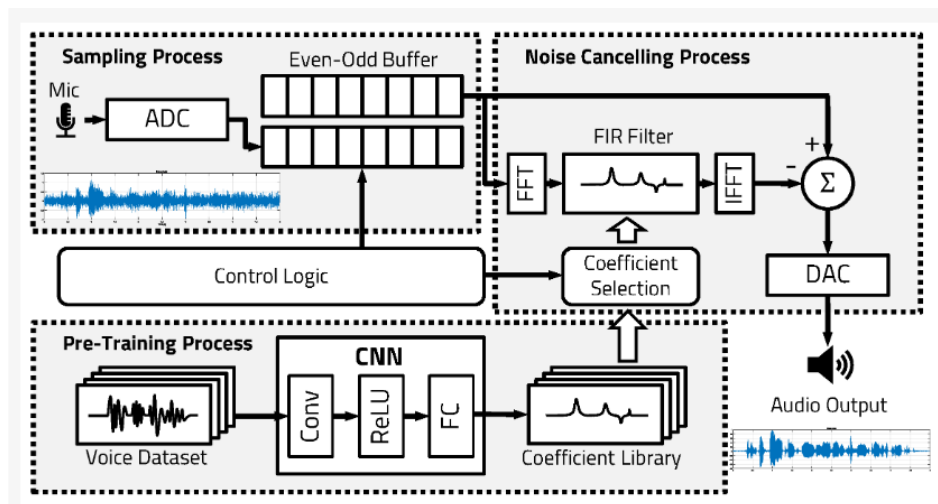


Fig3 CNNAlgorithm

Results

This section presents the results of our experiment focused on commotion type identification and elimination using Convolutional Neural Networks (CNN). We outline the presentation of our CNN-based model, its ability to classify noise kinds, and its overall effectiveness in noise removal.

1. Evaluation of Model Execution: Our CNN-based noise reduction algorithm achieved significant improvements in audio quality. We evaluated the model's presentation using conventional metrics, including Signal-to-Noise Ratio (SNR), Mean Squared Error (MSE), and Perceptual Evaluation of Audio Quality (PEAQ) scores. The results demonstrated a significant improvement in sound quality after noise removal, as seen by increased SNR, reduced MSE, and improved PEAQ ratings. The visual connections among the initial cacophonous sound, the sound accompanied by disturbance, and the purified sound further underscored the effectiveness of the noise reduction procedure.

2. Noise Type Classification: Our model demonstrated good accuracy in categorizing various forms of noise. We conducted evaluations to determine its ability to effectively identify and categorize various forms of noise inside audio signals. The model's performance was evaluated using a disarray matrix and a classification report, which highlighted its accuracy, recall, and F1-score for each noise kind. These results demonstrated the model's robust characterisation capabilities.

3. Speculation and Power: Our model has strengths in exhibiting capabilities across many sound sources, circumstances, and forms of noise. We conducted comprehensive testing, including scenarios with noise kinds not seen during training. The model demonstrated the ability to adapt to previously subtle noise forms and maintained its efficacy across varying signal-to-noise ratios (SNR).

4. Client Feedback and Abstract Evaluation: Client critiques and emotional evaluation outcomes further validated the efficacy of our noise reduction system. Research and evaluations revealed that customers had a noticeable enhancement in sound quality after noise elimination, emphasizing the effective benefits of our approach.

5. Proximal Evaluation: Compared to other noise reduction methods, such as conventional signal processing techniques and artificial intelligence approaches, our CNN-based model shown significant advantages in both

performance and adaptability. The comparative analysis highlighted the strengths of our technique while acknowledging its limitations.

Conclusion

This research examined several video denoising methods, both classical and deep learning-based, highlighting a compromise between processing speed and detail retention. Gaussian blur emphasizes processing speed, perhaps at the expense of detail retention. NL Means provides a more balanced methodology, attaining noise reduction while preserving picture information to a larger degree. CLSTM is distinguished by its focus on processing velocity. DnCNN has exceptional proficiency in noise reduction, achieving an optimal equilibrium between velocity and denoising efficacy.

Overview of the Innovation in the Project: Deep Learning Project The research use deep learning methodologies to address intricate issues by training neural networks on extensive datasets. Deep learning, a branch of artificial intelligence, facilitates enhanced predictive accuracy and increased efficiency across several fields. This invention encompasses activities like as image identification, natural language processing, and predictive analytics, advancing the frontiers of technology. Deep learning models have shown exceptional performance across several areas, including healthcare, finance, and autonomous driving. The project's originality is in using deep learning's potential to address real-world challenges and propel progress.

Future Work Scope:

Although our study indicates potential, further exploration is required to determine the most effective deep learning model for video denoising. We want to investigate supplementary evaluation measures such as Mean Absolute Error (MAE) and Structural Similarity Index Measure (SSIM) for a thorough assessment of model performance. Continued endeavors will concentrate on a more comprehensive assessment of deep learning models to get enhanced noise reduction proficiency. Furthermore, we want to improve processing performance by the use of parallel processing methods and the utilization of GPU acceleration. Incorporating these results into our existing study will provide a more comprehensive knowledge of video quality enhancing techniques.

Reference

1. Zhou H, Xu W, Chen J, Wang W. Evolutionary V2X technologies toward the internet of vehicles: Challenges and opportunities. *Proceedings of the IEEE* 2020;108: 308-323.
2. Kato S, Takeuchi E, Ishiguro Y, Ninomiya Y, Takeda K, Hamada T. An open approach to autonomous vehicles. *IEEE Micro* 2015;35: 60-68.
3. Krolkowski PM, Naggert K. Semiconductor shortages and vehicle production and prices. *Economic Commentary (Federal Reserve Bank of Cleveland)* 2021.
4. Ngo CN, Dang H. Covid-19 in America: Global supply chain reconsidered. *The world Economy* 2022;46: 256-275.
5. Putro ASH, Santoso AS. Supply chain and digital transformation of the tire manufacturing company during the COVID-19 pandemic: A case study of PT. X. *Qeios* 2023.
6. Zhan J, Lu S. Influence of COVID-19 epidemic on china and global supply chain and policy suggestions. *Scientific Research Publishing* 2021;9: 2497-2512.

7. Natale MD, Sangiovanni-Vincentelli A. Moving from federated to integrated architectures in automotive: The role of standards, methods and tools. *Inst Electrical Electronics Engineers* 2010;98: 603-620.
8. King JB. Strategy for development of vehicle electronic systems diagnostics. SAE 1979
9. Noise reduction, wikipedia the free encyclopedia (2020). URL "https://en.wikipedia.org/w/index.php?title=Noise_reduction&oldid=967676491". Online; accessed 6 September-2020
10. Sound masking; wikipedia, the free encyclopedia (2020). URL https://en.wikipedia.org/w/index.php?title=Sound_masking&oldid=965358994. Online; accessed 6-September-2020
11. W.: White noise; wikipedia, the free encyclopedia (2020). URL https://en.wikipedia.org/w/index.php?title=White_noise&oldid=975542178. Online; accessed 6-September-2020
12. Gartzman, D.: Getting to know the mel spectrogram (2020). URL <https://towardsdatascience.com/getting-to-know-the-mel-spectrogram-31bca3e2d9d0>
13. Germain, F.G., Chen, Q., Koltun, V.: Speech denoising with deep feature losses (2018)
14. Godsill, S.J., Tan, C. H.: Removal of low frequency transient noise from old recordings using model-based signal separation techniques. In: *Proceedings of 1997 Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 4 pp.– (1997)