International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Deep Learning for Next-Gen Audio Enhancement Systems

G Naga Jyothi[1], Paidimalla Naga Raju[2], Raghu Kalyana[3], D.N.V.S.Vijaya Lakshmi[4]

Submitted: 10/09/2024 **Revised:** 15/10/2024 **Accepted:** 25/10/2024

Abstract— This article reviews the latest advancements in deep learning approaches for audio signal processing. Speech, music, and ambient sound processing are examined concurrently to elucidate similarities and contrasts within these domains, emphasizing common methodologies, challenges, significant references, and opportunities for cross-fertilization between fields. This study covers the principal feature representations, including log-mel spectra and raw waveforms, with deep learning models, including convolutional neural networks, variations of the long short-term memory architecture, and specialized audio neural network models. Subsequently, significant domains of deep learning applications are addressed, including audio recognition (automatic speech recognition, music information retrieval, environmental sound detection, localization, and tracking) and synthesis and transformation (source separation, audio enhancement, generative models for speech, sound, and music synthesis). Ultimately, critical concerns and prospective inquiries about the use of deep learning in audio signal processing are delineated.

Index Terms*—. deep learning, connectionist temporal memory, automatic speech recognition, music information retrieval, source separation, audio enhancement, environmental sounds*

INTRODUCTION

In recent years, deep learning methodologies have garnered significant attention as a means of constructing hierarchical representations from unlabeled data. However, to our knowledge, these deep learning methodologies have not been thoroughly examined for auditory data. In this study, we use convolutional deep credence networks on audio data and conduct empirical evaluations across several audio classification tasks. In the context of verbalization data, we demonstrate that the acquired characteristics align with phones/phonemes. Additionally, our feature representations trained on unlabeled audio data demonstrate exceptional performance across many audio classification tasks. We anticipate that this article will stimulate further research on deep learning methodologies applied to a vast array of audio perception challenges. Artificial neural networks have garnered significant interest in three distinct waves: 1) the introduction of the perception algorithm in 1957, 2) the development of the backpropagation algorithm in 1986, and 3) the triumph of deep learning in voice recognition and picture classification in 2012. Deep learning encompasses deep feedforward neural networks, convolutional neural networks (CNNs), and long short-term memory (LSTM) networks. In recent days, deep learning has facilitated practical applications across several domains of signal processing. In this latest wave, deep learning first acquired prominence in image processing, subsequently being extensively used in voice processing, environmental sound processing, music, and several other domains like as genomics and quantum chemistry. Deep learning models surpassed traditional approaches in performance. Data plays a crucial part in deep learning applications. Numerous deep learning methodologies are formulated on the foundation of image processing techniques.

Significant distinctions exist between the realms of picture and audio, particularly for the configuration of the vector. Audio samples in their raw form are one-dimensional time series signals, fundamentally distinct from two-dimensional visuals.

Audio signals may be transformed into two-dimensional time-frequency representations for processing. The two axes, frequency and time, differ from the vertical and horizontal axes. Advanced Deep Learning Methods for Audio Processing

IMETHODS

An overview of audio analysis challenges (II-A), the representation of input data and characteristics often used in audio applications (II-B), the various models utilized for distinct applications (II-C), the datasets

[1,2,3,4]*International School of Technology and Sciences for Women*

implemented for deep learning experiments (II-D), and the techniques for result assessment. The primary challenges or tasks associated with audio signals include Audio Classification, Audio Fingerprinting, Automatic Music Tagging, Audio Segmentation, Audio Source Separation, Onset Detection, Music Transcription, Music Retrieval, Music Recommendation, Beat Tracking, Audio Enhancement, and Voice Activity Detection. One significant difficulty in the domain of audio processing is audio categorization. Derive several characteristics from audio and categorize the audio according to these characteristics. Numerous beneficial applications related to audio categorization, including speaker identification, instrument recognition, and speech recognition, are accessible. Diverse deep learning frameworks, such as TensorFlow and Keras, are used for audio classification. Real-time voice recognition is achievable with deep learning methodologies. The objective of audio fingerprinting is to determine the digital "summary" of the audio. This may be executed to identify the audio from an audio sample. Shazam exemplifies the implementation of audio fingerprinting well.

It acknowledges the musical concept of the first two to five seconds of a composition. Nonetheless, there are instances in which the system malfunctions, particularly in environments with significant ground noise. Music tagging might be considered a more intricate kind of audio categorization. This scenario allows for numerous classes to which each audio may belong, known as a multi-label classification issue. An application of this activity is to generate metadata for the audio to facilitate future searches. Deep learning has addressed this challenge to a certain degree.

Segmentation refers to the division of a designated item into segments according to a specified set of attributes. Segmentation is an essential pre-processing technique for audio data analysis. This occurs because we may divide a prolonged and intense audio input into brief, uniform parts (concise audio sequences) that are used for further processing. The work involves cardiac sound segmentation, specifically identifying sounds unique to the heart.

Audio Source Separation entails the extraction of one or several source signals from a composite of signals. A prevalent use of this technology is the identification of lyrics from audio for simultaneous translation, such as in karaoke. The objective is to monitor the status of each drill inside a collection of audio samples.

Beat tracking may be used to automate laborious procedures necessary for synchronizing events with music. It is beneficial in several applications, including as video editing, audio editing, and human-computer improvisation. Music recommendation systems assist in managing information overload by autonomously suggesting new music to listeners. Content providers such as Ganna and Spotify have created very advanced music recommendation algorithms. These models use the user's historical listening data, among several other attributes, to generate personalized suggestion lists. Music Retrieval, one of the most challenging challenges in audio processing, fundamentally seeks to create a search engine based on audio content. While we can do this by addressing sub-tasks such as audio fingerprinting, this task involves far more than that. For example, we must address many minor issues related to different kinds of music retrieval, and timbre recognition would be advantageous for gender identification. At yet, no alternative solution has been designed to meet the anticipated industry requirements. Music transcription is a formidable audio processing endeavor. It involves annotating audio and producing a "sheet" for further music generation. The physical labor required for transcribing music from recordings will be substantial. The variability is substantial, contingent upon the intricacy of the music, the proficiency of our auditory talents, and the level of detail desired in our transcription.

Onset detection is the first step in analyzing an audio or music sequence. For several activities outlined above, it is essential to execute onset detection, namely identifying the commencement of an audio event. Onset detection was fundamentally the primary objective that researchers aimed to address in audio processing.

A Voice Activity Detector (VAD) is used to identify the presence or absence of speech in audio signals. This article will provide a tutorial on constructing a simple Voice Activity Detector (VAD) that outputs 1 when speech is detected and 0 when it is not. The function of a Voice Activity Detector (VAD) is to accurately ascertain the presence of speech, even amongst background noise. Under optimal circumstances, even a basic energy detector may accurately identify speech; nevertheless, the Voice Activity Detection (VAD) may falter in the presence of noise. Deep learning-based Voice Activity

Detection (VAD) provides almost accurate predictions about the presence of noise in the signal. Speech denoising has been a persistent issue. We want to isolate the unwanted noise from an input noisy signal without compromising the desired signal. Envision an individual conversing during a video conference as music plays in the background. In this context, a voice denoising system is responsible for eliminating background noise to improve the clarity of the spoken signal. This application is crucial for video and audio conferencing, since noise may substantially impair intelligibility.

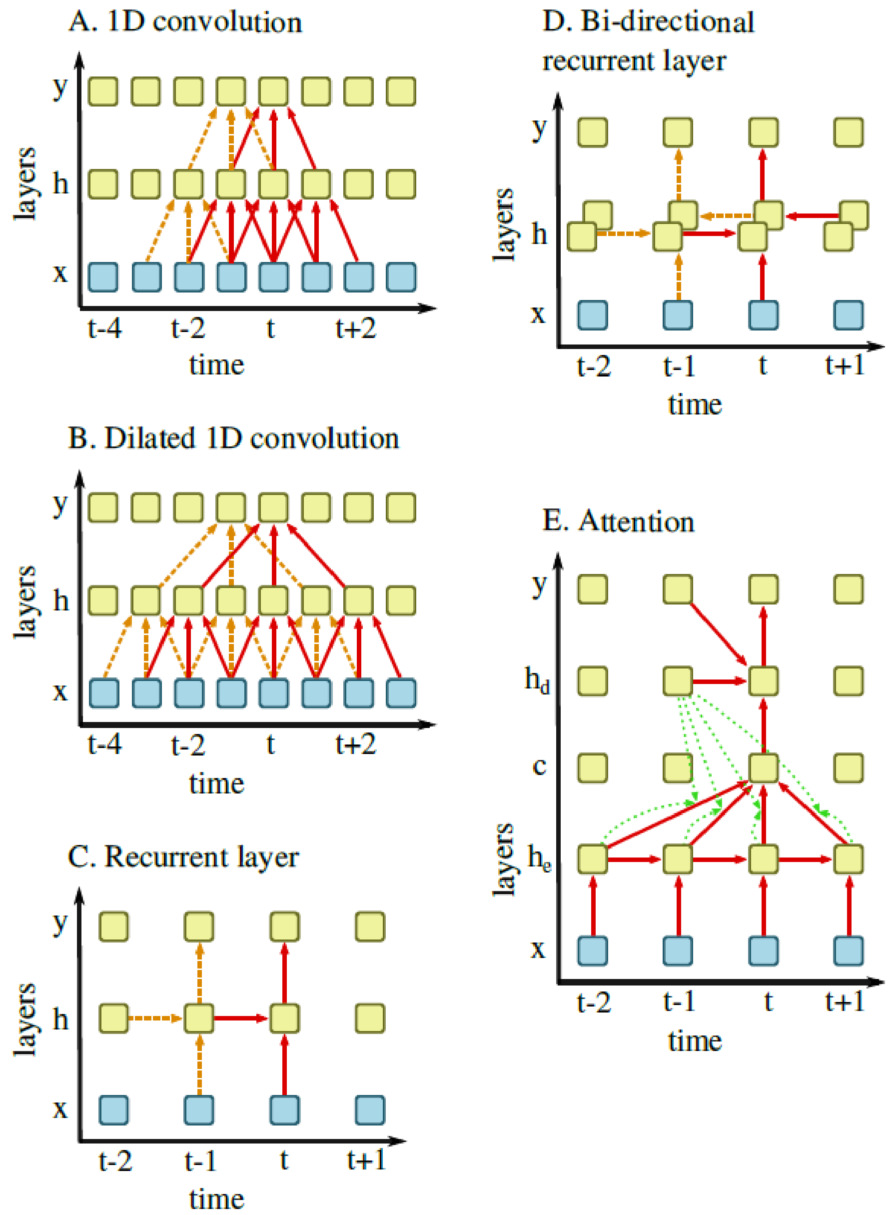

Figure 1. Different ways of processing temporal context.

ARCHITECTURES

As models improve, they increasingly use many architectures simultaneously, rendering efficient categorization unfeasible. Consequently, each subsection will have models applicable to many subsections, however they have been categorized according to the author's rationale for optimal clarity. In contrast to the acoustic characteristics, several distinct designs exist. This section will outline the architectures most often used in audio generation.

Generative adversarial networks

Another notable generative architecture is generative adversarial networks (GAN). It consists of two models that fulfill distinct functions: the generator and the discriminator.

The generator's purpose is to transform a random input vector into a data sample. The random vector is often smaller since the generator emulates the decoder component of the auto-encoder [1].

In contrast to VAE, which enforces a distribution for realistic data generation, GAN employs an additional network known as the discriminator [1]. The system evaluates the generator's output or a dataset sample, aiming to categorize it as authentic or counterfeit. The generator is fined according to the discriminator's capacity to distinguish between authentic and counterfeit. Conversely, if the discriminator fails to differentiate between the generator and the genuine data points, it is likewise punished. The two neural networks compete in a two-player minimax game. As per [123], the optimal result for network training is for the discriminator to possess 50% confidence in determining whether the input is authentic or fraudulent. In reality, we train the generator via the discriminator by minimizing the likelihood that the sample is counterfeit, while the discriminator conversely increases this likelihood for counterfeit data and maintains it for authentic data.

Figure 2b depicts the generator receiving a random vector input as the discriminator tries to differentiate between authentic and counterfeit samples.

This fundamental configuration enables the synthesis of samples that mimic those in the dataset; however, it does not permit conditional generation. The random vector used in the generator fails to align with the semantic characteristics of the input [5]. Numerous datasets contain further information about each sample, such as the item type shown in a picture. Utilizing the supplementary information to condition the generator and produce outputs from a subset of the learnt results would be advantageous. Conditional GAN (cGAN) adds supplementary structure by including extra information into the inputs of both the generator and discriminator. The generator incorporates further information into the random vector, whereas the discriminator integrates it into the data for differentiation purposes. Notable studies using cGAN include MidiNet [124], Michelsanti and Tan [125], Chen et al. [126], Neekhara et al. [127], and V2RA-GAN [128].

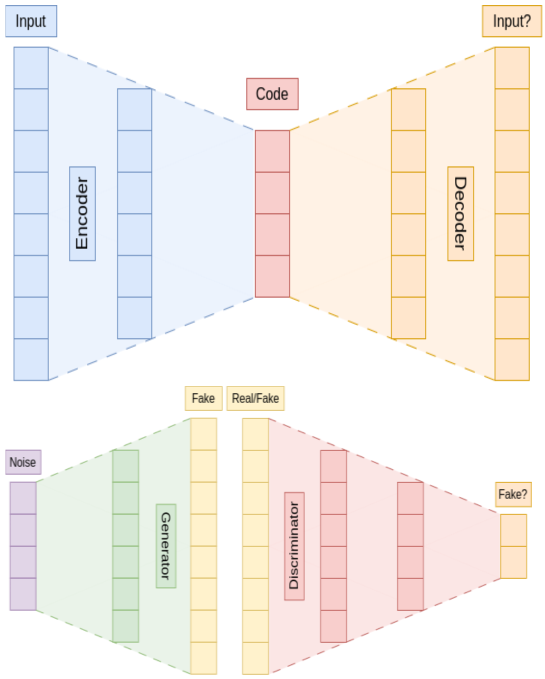

Figure 2. Two deep learning architectures that appear to have little in common until we look closer. The generator mimics the auto-encoder's decoder, whereas the discriminator resembles the encoder.

Prevalent challenges associated with GANs include mode collapse, unstable training, and the absence of a definitive assessment measure [2]. Mode collapse transpires when the generator concentrates only on a limited number of outputs that deceive the discriminator into seeing them as authentic. Although the generator satisfies the discriminator's criteria, it cannot be used to generate more than a limited number of samples. This may occur due to the discriminator's inability to compel the generator to exhibit diversity [123]. The Wasserstein Generative Adversarial Network (WGAN) is a prominent alternative for tackling this issue. WGAN reassigns the discriminator's role from differentiating between authentic and counterfeit data to calculating the Wasserstein distance, sometimes referred to as the Earth Mover's distance. A variation to facilitate WGAN convergence has been suggested; it employs a gradient penalty instead of weight clipping and is referred to as WGAN-GP. MuseGAN [129], WaveGAN [130], TiFGAN [131], and Catch-A-Waveform [132] employed WGAN.

APPLICATIONS

Deep learning-based audio processing may be used in several domains. Initially for the analysis of speech (Sec. III-A1), music (Sec. III-A2), and ambient sounds (Sec. III-A3), followed by the synthesis and transformation of audio source separation (Sec. III-B1), speech augmentation (Sec. III-B2), and audio creation (Sec. III-B3).

A. Analysis

1) Speech

Utilizing vocalization to get information and engage with the surroundings may represent a fundamental and instinctual mode of communication for humans. Speech recognition, which transforms auditory speech into word sequences, may be essential for any speech-based interaction. Initiatives to develop automated speech recognition systems have been on for more than fifty years. Nonetheless, the widespread implementation of such systems in practical applications has transpired only in recent years. For several years, the triphone-state Gaussian mixture model (GMM) / hidden Markov model (HMM) was the preeminent selection for speech modeling. These models provide several benefits, including their mathematical beauty, which results in various principled solutions to practical issues such as speaker or task adaptability. Circa 1990, discriminative training was shown to provide superior performance compared to models trained by maximum likelihood.

Hybrid models based on neural networks were suggested to replace Gaussian Mixture Models (GMMs). In 2012, deep neural networks (DNNs) with several parameters, trained on hundreds of hours of data, significantly reduced the word error rate (WER) across multiple voice recognition tasks. Moreover, in addition to the successful performance of deep feedforward and convolutional networks, LSTMs and GRUs have shown superior efficacy compared to feedforward deep neural networks. Subsequently, a sequence of convolutional, LSTM, and feedforward layers, known as the convolutional long short-term memory deep neural network (CLDNN) model, demonstrated superior performance compared to models using solely LSTM. In CLDNNs, an input frame window is initially processed by two convolutional layers with max pooling layers to reduce frequency variance in the signal, subsequently projected into a lower-dimensional feature space for the LSTM layers to model temporal correlations, and finally refined through additional feedforward layers and an output softmax layer.

2) Music

In contrast to speech, music recordings often include a broader array of sound sources of interest. In many musical genres, their occurrence adheres to prevalent limits of time and frequency, establishing intricate relationships both within and across sources. This presents a favorable array of options for the automated description of music recordings. Tasks include low-level analysis (onset and offset detection, first harmonic estimation), rhythm analysis (beat tracking, meter identification, downbeat tracking, tempo estimation), Fourier analysis (key detection, melody extraction, chord estimation), high-level analysis (instrument detection, instrument separation, transcription, structural segmentation, artist recognition, genre classification, mood classification), and high-level comparison (discovery of repeated themes, cover identification, music similarity estimation, score alignment). Initially, each of these was addressed using hand-crafted algorithms or features in conjunction with shallow classifiers; however, they are currently being attacked using deep learning techniques. This document highlights selected instances that include diverse activities and methodologies. Kindly confer with me for a more comprehensive list. An example of a multi-class labeling challenge is chord recognition. The objective is to assign root notes and chord classes to any given steps in a Western music recording. Conventional hand-crafted techniques rely on folding several octaves of a spectral representation into a 12-semitone chromagram, temporal smoothing, and alignment with established chord templates. Linear-magnitude spectrograms, subjected to contrast normalization and amplified using pitch shifting techniques, are often used. Contemporary systems use temporal modeling and are transformed into a collection of identifiable chords. McFee and Bello recently used a CRNN, which consists of a 2D convolution for learning spectro-temporal characteristics, succeeded by a 1D convolution that integrates information across frequencies, and culminates in a bidirectional GRU. They utilized side targets to include relationships across a specific set of 170 chord classes. Korzeniowski et al. use an unconventional approach by training CNNs using log-frequency spectrograms, which are employed not only for

chord prediction but also for enhancing chromagram representation, beneficial for tasks extending beyond chord estimation.
In the context of sequence categorization, a fundamental job is to assess the global pace of a piece. A logical option is to ground it in beat and downbeat tracking. Downbeat tracking may use tempo estimate as the primary constraint for determining downbeat placements. Nonetheless, despite the frequent avoidance of onset detection in beat tracking, Schreiber and Müller demonstrated that convolutional neural networks (CNNs) may be trained to directly estimate tempo from 12-second spectrogram extracts, yielding superior results and accommodating tempo variations or drift within a recording. Tag prediction, as an overarching sequence classification challenge, seeks to determine which labels from a limited vocabulary a user would assign to a certain music composition. Tags may relate to instrumentation, tempo, genre, and other elements, but they consistently pertain to a whole recording, devoid of time information. The transition from an input sequence to global labels has been addressed using many methods, which are valuable for comparison. Dieleman et al. train a CNN using brief 1D convolutions (i.e., temporal convolution) on 3-second log-mel spectrograms and average predictions over successive extracts to get a global label. In contrast, they train a CNN on unprocessed data, selecting the first layer filter size to correspond with standard spectrogram frames, however get worse results. Choi et al. use a fully convolutional network (FCN) consisting of interspersed 3×3 convolutions.

DISCUSSION AND CONCLUSION

In contrast to speech, music recordings often include a broader array of sound sources of interest. In many musical genres, their occurrence adheres to prevalent limits of time and frequency, establishing intricate relationships both within and across sources. This presents a favorable array of options for the automated description of music recordings. Tasks include low-level analysis (onset and offset detection, first harmonic estimation), rhythm analysis (beat tracking, meter identification, downbeat tracking, tempo estimation), Fourier analysis (key detection, melody extraction, chord estimation), high-level analysis (instrument detection, instrument separation, transcription, structural segmentation, artist recognition, genre classification, mood classification), and high-level comparison (discovery of repeated themes, cover identification, music similarity estimation, score alignment). Initially, each of these was addressed using hand-crafted algorithms or features in conjunction with shallow classifiers; however, they are currently being attacked using deep learning techniques. This document highlights selected instances that include diverse activities and methodologies. Kindly confer with me for a more comprehensive list. An example of a multi-class labeling challenge is chord recognition. The objective is to assign root notes and chord classes to any given steps in a Western music recording. Conventional hand-crafted techniques rely on folding several octaves of a spectral representation into a 12-semitone chromagram, temporal smoothing, and alignment with established chord templates. Linear-magnitude spectrograms, subjected to contrast normalization and amplified using pitch shifting techniques, are often used. Contemporary systems use temporal modeling and are transformed into a collection of identifiable chords. McFee and Bello recently used a CRNN, which consists of a 2D convolution for learning spectro-temporal characteristics, succeeded by a 1D convolution that integrates information across frequencies, and culminates in a bidirectional GRU. They utilized side targets to include relationships across a specific set of 170 chord classes. Korzeniowski et al. use an unconventional approach by training CNNs using log-frequency spectrograms, which are employed not only for chord prediction but also for enhancing chromagram representation, beneficial for tasks extending beyond chord estimation.
In the context of sequence categorization, a fundamental job is to assess the global pace of a piece. A logical option is to ground it in beat and downbeat tracking. Downbeat tracking may use tempo estimate as the primary constraint for determining downbeat placements. Nonetheless, despite the frequent avoidance of onset detection in beat tracking, Schreiber and Müller demonstrated that convolutional neural networks (CNNs) may be trained to directly estimate tempo from 12-second spectrogram extracts, yielding superior results and accommodating tempo variations or drift within a recording. Tag prediction, as an overarching sequence classification challenge, seeks to determine which labels from a limited vocabulary a user would assign to a certain music composition. Tags may relate to instrumentation, tempo, genre,

and other elements, but they consistently pertain to a whole recording, devoid of time information. The transition from an input sequence to global labels has been addressed using many methods, which are valuable for comparison. Dieleman et al. train a CNN using brief 1D convolutions (i.e., temporal convolution) on 3-second log-mel spectrograms and average predictions over successive extracts to get a global label. In contrast, they train a CNN on unprocessed data, selecting the first layer filter size to correspond with standard spectrogram frames, however get worse results. Choi et al. use a fully convolutional network (FCN) consisting of interspersed 3×3 convolutions.

B) Pre-process the signal and Features

While MFCCs are the predominant representation in conventional audio signal processing, log-mel spectrograms are the preeminent feature in deep learning, followed by raw waveforms or complicated spectrograms. Raw waveforms eliminate the need for manually crafted features, hence enhancing the potential to use the advanced modeling capabilities of deep learning models, which learn representations tailored for specific tasks. Nevertheless, this entails increased computational expenses and data requirements, and the advantages may also be difficult to comprehend in reality. In analytical tasks such as Automatic Speech Recognition (ASR), Music Information Retrieval (MIR), or environmental sound recognition, log mel spectrograms offer a more concise representation. Approaches utilizing these features typically require less data and training, achieving classification performance that is, at the current state of the art, comparable to setups employing raw audio. In tasks aimed at synthesizing high-quality audio, such as source separation, audio improvement, text-to-speech, or sound morphing, using (log-mel) magnitude spectrograms presents the issue of reconstructing the phase. In this context, raw waveforms or intricate spectrograms are often favored as the input format. Algorithms for the calculation of various frame-level and clip-level feature extraction methods have been developed. The efficacy of the clip level characteristics for multiclass audio classification is evaluated via the use of several classifiers inside the audio classification chapter. The multidimensional frame-level characteristics, such as MFCC and LPC, will be used to represent the tools in Table and to characterize a speech signal, as elaborated in chapters 5 and 6, respectively. The characteristics such as K-L divergence distance using Line Spectral Frequencies will be used for audio segmentation. A distinctive characteristic, 'harmonicity,' distinguishes the harmonic essence of music from the antithetical nature of speech. A sub-band energy ratio characteristic for voiced/unvoiced categorization is often advised.

C) Models

Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Convolutional Recurrent Neural Networks (CRNNs) are used well across several disciplines, without a distinct preference for any one kind. All three can represent temporal sequences and address sequence categorization, sequence labeling, and sequence transduction problems.

Convolutional Neural Networks (CNNs) possess a fixed receptive field, limiting the temporal context considered for predictions; yet, this characteristic simultaneously facilitates the adjustment of the context's breadth. RNNs may potentially use extensive temporal context for predictions; however, they must first train to do so, which may need model changes (such as LSTM) and limits direct control over context size. Moreover, they need sequential input processing, rendering them less efficient for training and evaluation on contemporary hardware compared to CNNs. CRNNs provide a balance, including the benefits and drawbacks of both CNNs and RNNs.

D) Computational Complexity, Interpretability and Adaptability

Computational Complexity The efficacy of deep neural networks relies on advancements in rapid and extensive computational capabilities. In comparison to conventional methods, cutting-edge deep neural networks often need more computational power and an increased volume of training data. CPUs are not well equipped for training and assessing extensive deep models.

Instead, processors designed for matrix operations are often used, namely general-purpose graphics processing units (GPGPUs) and application-specific integrated circuits such as proprietary tensor processing units (TPUs). The model parameters are acquired using gradient descent applied to the loss associated with pairs of inputs and targets, or only through inputs for unsupervised training. The relationship between the layer parameters and the actual job is challenging to comprehend.

REFERENCES

[1]. Computational Analysis of Sound Scenes and Events; Virtanen, T., Plumbley, M.D., Ellis, D., Eds.; Springer International Publishing: Berlin, Germany, 2018; doi:10.1007/978-3-319-63450-0.

[2]. D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," Nature, vol. 323, no. 6088, p. 533, 1986.

[3]. G. Hinton, L. Deng et al., "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," IEEE Signal Processing Magazine, vol. 29, no. 6, pp. 82–97, 2012.

[4]. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in NIPS, 2012.

[5]. A.-R. Mohamed, G. Dahl, and G. Hinton, "Deep belief networks for phone recognition," in NIPS workshop on deep learning for speech recognition and related applications, vol. 1, no. 9, 2009, pp. 39–47

[6]. Qian, K.; Ren, Z.; Pandit, V.; Yang, Z.; Zhang, Z.; Schuller, B. Wavelets Revisited for the Classification of Acoustic Scenes. In Proceedings of the Detection and Classification of Acoustic Scenes and Events Workshop (DCASE), Munich, Germany, 16–17 November 2017. 12.

[7]. Ren, Z.; Pandit, V.; Qian, K.; Yang, Z.; Zhang, Z.; Schuller, B. Deep Sequential Image Features for Acoustic Scene Classification. In Proceedings of the Detection and Classification of Acoustic Scenes and Events Workshop (DCASE), Munich, Germany, 16–17 November 2017.

[8]. A. Narayanan, A. Misra et al., "Toward domain -invariant speech recognition via large scale training," in SLT, 2018, pp. 441–447. [77] N. Jaitly and G. E. Hinton, "Vocal tract length perturbation (VTLP) improves speech recognition," in ICML Workshop on Deep Learning for Audio, Speech, and Language Processing, vol. 117, 2013.

[9]. K. Ullrich, J. Schlüter, and T. Grill, "Boundary Detection in Music Structure Analysis using Convolutional Neural Networks," in ISMIR, 2014. Vesperini, P. Vecchiotti, E. Principi, S. Squartini, and F. Piazza, "A neural network based algorithm for speaker localization in a multi-room environment," in proc. in IEEE International Workshop on Machine Learning for Signal Processing, 2016.

[11]. J. Chen, J. Benesty, Y. A. Huang, and E. J. Diethorn, "Fundamentals of noise reduction," in Springer Handbook of Speech Processing. Springer, 2008, pp. 843–872.

[12]. A. Pandey and D. Wang, "A New Framework for Supervised Speech Enhancement in the Time Domain," in Interspeech, 2018.

[13]. Y. Wang, A. Narayanan, and D. Wang, "On Training Targets for Supervised Speech Separation," IEEE/ACM Transactions on ASLP, vol. 22, no. 12, pp. 1849–1858, 2014.

[14]. T. N. Sainath, O. Vinyals, A. Senior, and H. Sak, "Convolutional, Long Short-Term Memory, Fully Connected Deep Neural Networks," in ICASSP, 2015.

[15]. B.-H. Juang and L. R. Rabiner, "Automatic speech recognition–a brief history of the technology development," Georgia Institute of Technology. Atlanta Rutgers University and the University of California. Santa Barbara, vol. 1, p. 67, 2005.