Sound processing with neural networks

Authors: Volodymyr Kolesnikov, Kirill Klabukov, Mykyta Kot

Hello! Today I would like to talk about how neural networks have become a very popular tool in audio data processing these days. This is due to the fact that the amount of audio and video content on the Internet has greatly increased.

Application of neural networks for audio processing allows you to improve sound quality, reduce noise and distortion, recognise speech, create voice assistants, spot problems in production in advance and much more. The tool can also be used to create new music and analyse sound data in various fields.

Plan:

Types of neural networks:

First, let's talk about **what types of neural networks can be used to process audio data.** There are several types of neural networks that can be used for this purpose.

The first type is recurrent neural networks (RNN). These networks have the ability to remember previous input data and use it to analyse subsequent data. This makes them particularly useful for time series analysis, such as audio recordings.

The second type is convolutional neural networks (CNNs). These networks are used for image analysis, but they can also be applied to audio processing. For example, they can be used to detect noise or other acoustic characteristics in audio recordings.

The third type is deep neural networks (DNN). These networks are used for complex tasks such as speech recognition and sound classification.

Finally, **the fourth type is generative neural networks (GAN)**. These networks can be used to create new music and other sound effects.

However, the specific types of neural networks may depend on the specific sound processing task to be performed. For example, a **combination of** convolutional and recurrent **neural networks** may be used for automatic speech transcription tasks, and convolutional neural networks using deep learning algorithms may be used for noise removal tasks.

Examples of the use of neural networks:

One of the most common examples is the improvement of audio quality. Convergent Neural Networks (CNNs) can be used to **remove noise and other interference from audio recordings** to make them clearer and easier to listen to. This can be useful, for example, when recording conversations over the phone or in conference rooms.

Adobe Podcast and NVIDIA Broadcast are two powerful innovations that use neural networks for audio processing. They enable the creation of quality podcasts and streaming shows with minimal effort and cost.

NVIDIA Broadcast is audio and video processing software that improves video conferencing, streaming and other video content. It's a suite of several tools, but we're interested in:

Noise Removal - removes background noise from audio recordings such as phone calls or video conferences. It uses a neural network to process the audio signal and remove noise.

Audio Effects - allows you to add various sound effects to your audio recordings, such as reverb or fade in.

NVIDIA Broadcast uses deep learning (DNN) technologies, such as neural networks, to process audio and video. It makes it possible to create high-quality content in real time, without the need for expensive hardware or software.

Adobe Podcast is another tool developed by Adobe for podcast production and editing. It is an integrated package that includes various features such as recording, editing and editing of audio files. Adobe Podcast uses neural network algorithms for audio processing, which can significantly improve quality and remove unnecessary noise. It also offers a wide range of tools for creating music effects and matching audio track to video content.

Adobe Podcast uses several different types of neural networks for audio processing. In particular, different types of Convolutional Neural Networks (CNNs) can be used to improve audio quality and remove noise, and Recurrent Neural Networks (RNNs) and variants such as Long Short-Term Memory (LSTM) can be used for voice and speech processing.

All in all, Adobe Podcast and NVIDIA Broadcast represent a major step forward in the development of sound processing technologies using neural networks. They greatly simplify and accelerate the process of creating quality audio content, which is particularly important in today's digital economy. Through the use of neural network algorithms, these tools enable high fidelity and productivity in sound processing. They are also an example of how artificial intelligence and machine learning can be used to improve sound quality and create creative effects.

Another example of the use of neural networks is speech recognition. Deep Neural Networks (DNNs) can be trained to recognise speech and convert it into text. This can be useful, for example, to create a speech recognition system for people with disabilities or to transcribe audio recordings into text.

> **DeepSpeech**: is a neural network that is used for speech recognition. It is based on recurrent neural networks and can handle different languages and accents.

Also in the field of speech technology, neural networks can be used to create **voice assistants,** automatic **speech translation** into other languages, **emotion recognition** and more.

> **Building voice assistants** can be implemented using deep neural networks that can learn to recognise and understand natural language and answer user questions. Such systems are used in popular voice assistants such as Siri from Apple, Google Assistant from Google, Alexa from Amazon and others.
>
> **Automatic speech translation into other languages** can also be implemented using neural networks. Such systems are used in online translators and software applications to translate speech in real time. Neural networks are trained on a large set of parallel texts and audio recordings in different languages, and are tasked with translating speech from one language to another while maintaining its meaning and style.
>
> **Emotion recognition** can also be implemented using neural networks that can be trained to recognise patterns in sound associated with certain emotional states such as joy, anger, sadness, etc. Such systems can be used to analyse the emotional colouring of voices in text, audio and video recordings, which can be useful in marketing, education, medicine and other fields.

The third example is the **music industry**. Recurrent neural networks (RNNs) can be used to create new music and process existing audio recordings. For example, neural networks can be trained to **recognise musical genres** and **predict the next**

notes in a musical composition based on rhythm and melody analysis. This can help composers and producers create more interesting and original music.

> **Neural Source Separation**: is a neural network used to separate sound sources, such as voice and music, from mixed sound signals. It is based on deep learning algorithms such as convolutional neural networks and recurrent neural networks.

In the field of audio compression, neural networks can be used to **create more efficient compression algorithms**. They can be trained to predict which sound parts can be safely removed without any noticeable loss of sound quality.

> **One of the main challenges in audio compression** is to remove unnecessary information from the audio file that is not perceived by the human ear or is recovered later. Neural networks can be trained to predict which audio data can be safely removed in order to preserve audio quality. This is done by training a neural network on a large set of audio files, where it learns patterns in sound and determines which aspects of sound are most important to human perception and which can be safely removed.
>
> This can be particularly useful for streaming audio or streaming audio files over the internet, where better compression can reduce transfer costs and download times.

Sound anomaly detection systems are another interesting example of the application of neural networks in sound processing. Such systems can be trained to recognise sound signals that indicate potential problems in machinery or equipment. For example, **sounds of friction, grinding, bumping or vibration** can indicate that equipment is in a faulty state and needs repair or replacement.

Neural networks can be trained to recognise such beeps and issue **appropriate alerts to operators** or automatically trigger equipment maintenance and repair procedures. This can help prevent unforeseen stoppages on the production line, reduce risks to workers, and save money on repairs and replacements.

One of the most common examples of applications for sonic anomaly detection systems is **monitoring the condition of cars** and other vehicles. Neural networks can be trained to recognise sounds associated with **malfunctions in the engine, transmission or other vehicle systems**, which can help operators and vehicle workshops detect and correct problems before they become more serious and lead to an accident.

In addition, sonic anomaly detection systems can be useful in the **power industry** to **detect faults in equipment** such as turbines and generators. They can also be applied in other industries where sound plays an important role in diagnosing the condition of equipment.

Deep Neural Networks (DNN) can be used in natural language processing to **create dialogue systems and interfaces.** In addition, **DNNs can be enhanced** by various techniques such as reinforcement learning and genetic algorithms, which can lead to even more accurate results and a wider range of applications.

Deep neural networks can be used to **analyse tone and emotion** in texts. This can help companies understand the mood of their customers and determine what changes can be introduced to improve service.

Finally, DNNs can be used to **create synthetic speech** that sounds just like human speech. This can be useful for creating voice assistants, audiobooks, text-reading apps, and more.

Read more about NVIDIA Broadcast:

Nvidia Broadcast uses a neural network that consists of multiple layers and has a convolutional architecture. This neural network is designed to process audio signals in real time, and can effectively filter out noise, remove noise, and improve audio quality.

The topology of the Nvidia Broadcast neural network can be described as follows:

Input layer: The input of the neural network is a PCM (pulse-code modulation) audio signal with a sampling rate of 48 kHz and a bit rate of 16 bits.

Convergent layers: The audio signal then passes through several convolutional layers, each using several convolutional kernels of different sizes. Each convolution kernel performs a convolution operation on a fragment of the input signal and creates a new attribute at the layer's output.

Collapse Layer: Each convolution layer is followed by a pooling layer, which reduces the spatial resolution of the output features by performing a pooling operation on a subset of the features. This reduces the number of model parameters and speeds up model training and application.

Normalisation layer: The next step in the neural network is to apply a batch normalization layer, which standardizes the values of the features on each layer and reduces the chance of overtraining the model.

Convex layers: Finally, the outputs of the last convolution and normalization layer are fed into several fully connected layers, which combine the information from all the features and produce a final network output.

Output layer: The last layer of the neural network uses the Softmax activation function to convert the network output into class probabilities that correspond to various sound effects such as noise reduction, noise removal, voice enhancement, etc.

Current developments:

A lot of Universities and companies are currently developing their neural networks in sound processing.

For example, **Carnegie Mellon University (CMU)** is developing speech recognition systems based on

Neural networks are also used in sound anomaly detection systems. Some companies, such as **Falkonry**, are

deep neural networks. These systems are used in applications for voice control of devices such as smartphones or smart homes.

developing equipment condition monitoring systems using neural networks to detect sounds associated with faults.

Another example is **Audeze**, which uses neural networks to improve the sound quality on its headphones.

These are just a few examples of applications of neural networks in sound processing, and the list could still be very much longer.

Conclusion:

In **conclusion**, sound processing using neural networks is a very promising field and has many applications in practice. **Recursive, convolutional, deep and generative neural networks** can all be used to solve different problems in audio processing. We hope that our presentation was useful and gave you a deeper understanding of how neural networks can be used for audio data processing.

Thank you!

Literature:

1. Ramires, A., Yehia, I., & Asaei, A. Deep Learning for Audio Signal Processing. Springer, 2019
2. Graves, A., Liwicki, M., Fernandez, S., Bertolami, R., Bunke, H., & Schmidhuber, J.
   (2006). Connectionist temporal classification: Labelling unsegmented sequence data with
   recurrent neural networks. In Proceedings of the 23rd International Conference on Machine learning (ICML'06) (pp. 369-376). ACM.
3. Reddit's Machine Learning subreddit. (n.d.). Retrieved May 7, 2023, from https://www.reddit.com/r/MachineLearning/
4. International Conference on Machine Learning. (n.d.). Retrieved May 7, 2023, from https://icml.cc/
5. International Speech Communication Association. (n.d.). Interspeech. Retrieved May 7,

2023, from https://www.interspeech2022.org/

6. IEEE Signal Processing Society. (n.d.). Retrieved May 7, 2023, from https://signalprocessingsociety.org/
7. Notion AI. (n.d.). Retrieved May 7, 2023, from https://www.notion.ai/
8. ChatGPT. (2021). OpenAI. https://openai.com