



**Fakulta elektrotechniky
a informatiky**



Fakulta elektrotechniky a informatiky

Katedra kybernetiky a umelej inteligencie

Predmet : **Neurónové siete**
kurz 2022 / 2023

Zadanie číslo 1:

Využitie neurónových sietí v spracovaní zvuku

klúčové slová: Nvidia Broadcast, Spracovanie zvuku

Spracovali:

Mykyta Kot, Volodymyr Kolesnikov, Kirill Klabukov



Obsah

1	Úvod	2
2	Typy neurónových sietí	3
3	Príklady použitia neurónových sietí pri spracovaní zvuku	3
3.1	Zlepšenie kvality zvuku	3
3.2	Rozpoznávanie reči	4
3.3	Hudobný priemysel	5
3.4	Systémy na detekciu zvukových anomálií	6
3.5	Hlboké neurónové siete (DNN)	6
4	Viac o NVIDIA Broadcast	7
5	Súčasný vývoj	8
6	Zhodnotenie a záver	9
7	Použitá literatúra	9

1 Úvod

Dobrý deň! Dnes by som chcel hovoriť o tom, ako sa neurónové siete stali v súčasnosti veľmi obľúbeným nástrojom na spracovanie zvukových údajov. Je to spôsobené tým, že množstvo zvukového a video obsahu na internete sa výrazne zvýšilo.

Použitie neurónových sietí na spracovanie zvuku môže zlepšiť kvalitu zvuku, znížiť šum a skreslenie, rozpoznávať reč, vytvárať hlasových asistentov, vopred odhaliť problémy pri výrobe a mnoho ďalšieho. Nástroj sa dá použiť aj na tvorbu novej hudby a analýzu zvukových údajov v rôznych oblastiach.

2 Typy neurónových sietí

Najprv si povedzme, aké typy neurónových sietí možno použiť na spracovanie zvukových údajov. Existuje niekoľko typov neurónových sietí, ktoré možno na tento účel použiť.

- Prvým typom sú rekurentné neurónové siete (RNN). Tieto siete majú schopnosť zapamätať si predchádzajúce vstupné údaje a použiť ich na analýzu nasledujúcich údajov. Vďaka tomu sú obzvlášť užitočné na analýzu časových radov, napríklad zvukových záznamov.
- Druhým typom sú konvolučné neurónové siete (CNN). Tieto siete sa používajú na analýzu obrazu, ale môžu byť aj aplikované na spracovanie zvuku. Môžu sa napríklad použiť na detekciu šumu alebo iných akustických charakteristík v zvukových nahrávkach.
- Tretím typom sú hlboké neurónové siete (DNN). Tieto siete sa používajú na komplexné úlohy, ako je rozpoznávanie reči a klasifikácia zvuku.
- Štvrtým typom sú generatívne neurónové siete (GAN). Tieto siete možno použiť na vytváranie novej hudby a iných zvukových efektov.

Konkrétne typy neurónových sietí však môžu závisieť od konkrétnej úlohy spracovania zvuku, ktorá sa má vykonať. Napríklad na úlohy automatického prepisu reči sa môže použiť kombinácia konvolučných a rekurentných neurónových sietí a na úlohy odstraňovania šumu sa môžu použiť konvolučné neurónové siete využívajúce algoritmy hlbokého učenia.

3 Príklady použitia neurónových sietí pri spracovaní zvuku

Jedným z najčastejších príkladov je zlepšenie kvality zvuku. Konvergentné neurónové siete (CNN) možno použiť na odstránenie šumu a iných rušivých vplyvov zo zvukových nahrávok, aby boli jasnejšie a ľahšie sa počúvali. To môže byť užitočné napríklad pri nahrávaní rozhovorov cez telefón alebo v konferenčných miestnostiach.

3.1 Zlepšenie kvality zvuku

Adobe Podcast a NVIDIA Broadcast sú dve výkonné inovácie, ktoré využívajú neurónové siete na spracovanie zvuku. Sú to vám umožní vytvárať vysokokvalitné podcasty a streamy s minimálnym úsilím a nákladmi.

NVIDIA Broadcast je softvér na spracovanie zvuku a videa, ktorý dokáže zlepšiť kvalitu videokonferencií, streamovania a iného video obsahu. Ide o balík niekoľkých nástrojov, ktoré nás však zaujali:

- Odstránenie šumu odstráni šum na pozadí zo zvukových záznamov, napríklad z telefonických hovorov alebo videokonferencií. Na spracovanie zvukového signálu a odstránenie šumu používa neurónovú sieť.
- Zvukové efekty - umožňuje pridávať do zvukového záznamu rôzne zvukové efekty, napríklad dozvuk alebo blednutie.
- NVIDIA Broadcast používa na spracovanie zvuku a videa technológie hlbokého učenia (DNN), ako sú napríklad neurónové siete. Umožňuje vytvárať vysokokvalitný obsah v reálnom čase bez potreby drahého hardvéru alebo softvéru.

Adobe Podcast je ďalší nástroj vyvinutý spoločnosťou Adobe na tvorbu a úpravu podcastov. Je to integrovaný balík, ktorý obsahuje rôzne funkcie, ako je nahrávanie, editácia a úprava zvukových súborov. Adobe Podcast používa na spracovanie zvuku algoritmy neurónových sietí, ktoré môžu výrazne zlepšiť kvalitu a odstrániť nepotrebný šum. Taktiež ponúka širokú škálu nástrojov na vytváranie hudobných efektov a prispôbovanie zvukovej stopy obsahu videa.

Adobe Podcast používa na spracovanie zvuku niekoľko rôznych typov neurónových sietí. Na zlepšenie kvality zvuku a odstránenie šumu možno použiť najmä rôzne typy konvolučných neurónových sietí (CNN) a na spracovanie hlasu a reči možno použiť rekurentné neurónové siete (Recurrent). Neurónové siete (RNN) a ich varianty, ako napríklad Long Short-Term Memory (LSTM).

Celkovo predstavujú aplikácie Adobe Podcast a NVIDIA Broadcast významný krok vpred vo vývoji technológií spracovania zvuku pomocou neurónových sietí. Výrazne zjednodušujú a urýchľujú proces tvorby kvalitného zvukového obsahu, čo je v dnešnej digitálnej ekonomike mimoriadne dôležité. Vďaka použitiu algoritmov neurónových sietí umožňujú tieto nástroje vysokú vernosť a výkon pri spracovaní zvuku. Okrem toho sú príkladom toho, ako sa umelá inteligencia a strojové učenie dajú využiť na zlepšenie kvality zvuku a vytváranie kreatívnych efektov.

3.2 Rozpoznávanie reči

Ďalším príkladom využitia neurónových sietí je rozpoznávanie reči. Hlboké neurónové siete (DNN) možno vycvičiť na rozpoznávanie reči a jej prevod na text. To môže byť užitočné napríklad na vytvorenie systému rozpoznávania reči pre ľudí so zdravotným postihnutím alebo na prepis zvukových nahrávok do textu.

DeepSpeech: je neurónová sieť, ktorá sa používa na rozpoznávanie reči. Je založená na rekurentných neurónových sieťach a dokáže pracovať s rôznymi jazykmi a prízvukmi.

Aj v oblasti rečových technológií možno neurónové siete použiť na vytvorenie hlasových asistentov, automatických preklad reči, rozpoznávanie emócií a mnoho ďalšieho.

- Vytvorenie hlasových asistentov možno realizovať pomocou hlbokých neurónových sietí, ktoré sa dokážu naučiť rozpoznávať a porozumieť prirodzenému jazyku a odpovedať na otázky používateľov. Takéto systémy sa používajú v populárnych hlasových asistentoch, ako sú napr. Siri od Apple, Google Assistant od Google, Alexa od Amazonu a ďalšie.
- Automatický preklad reči do iných jazykov možno realizovať aj pomocou neurónových sietí. Takéto systémy sa používajú v online prekladačoch a softvérových aplikáciách na preklad reči v reálnom čase. Neurónové siete sa trénujú na veľkom súbore paralelné texty a zvukové nahrávky v rôznych jazykoch, ich úlohou je preložiť reč z jedného jazyka do druhého pri zachovaní jej významu a štýlu.
- Rozpoznávanie emócií sa dá realizovať aj pomocou neurónových sietí, ktoré sa dajú vycvičiť na rozpoznávanie zvukových vzorov spojených s určitými emocionálnymi stavmi, ako je radosť, hnev, smútok atď. Takéto systémy možno použiť na analýzu emocionálneho zafarbenia hlasov v texte, zvukových a obrazových záznamoch, čo môže byť užitočné v marketingu, vzdelávaní, medicíne a iných oblastiach.

3.3 Hudobný priemysel

Tretím príkladom je hudobný priemysel. Rekurentné neurónové siete (RNN) možno použiť na vytváranie novej hudby a spracovanie existujúcich zvukových nahrávok. Neurónové siete možno napríklad vycvičiť na rozpoznávanie hudobných žánrov a predpovedanie ďalších tónov v hudobnej skladbe na základe analýzy rytmu a melódie. To môže pomáhať skladateľom a producentom vytvárať zaujímavejšiu a originálnejšiu hudbu.

- Neurónová separácia zdrojov: Ide o neurónovú sieť, ktorá sa používa na oddelenie zdrojov zvuku, napríklad hlasu a hudby, od zmiešaných zvukových signálov. Je založená na algoritmoch hlbokého učenia, ako sú konvolučné neurónové siete a rekurentné neurónové siete.

V oblasti kompresie zvuku sa neurónové siete môžu použiť na vytvoriť účinnejšie kompresné algoritmy. Môžu sa naučiť predpovedať, ktoré zvukové časti možno bezpečne odstrániť bez výraznej straty kvality zvuku.

- Jedným z hlavných cieľov kompresie zvuku je odstrániť zo zvukového súboru nepotrebné informácie, ktoré nie sú vnímateľné ľudským uchom alebo ktoré sa neskôr obnovia. Neurónové siete možno vycvičiť na predpovedanie, ktoré zvukové údaje možno bezpečne odstrániť, aby sa zachovala kvalita zvuku. To sa dosiahne tréňovaním neurónovej siete na veľkom súbore zvukových súborov, kde sa naučí vzorce zvuku a určí, ktoré aspekty zvuku sú pre ľudské vnímanie najdôležitejšie a ktoré možno bezpečne odstrániť. To môže byť užitočné najmä pri streamovaní zvuku alebo streamovaní zvukových súborov cez internet, kde účinnejšia kompresia môže znížiť náklady na prenos a čas sťahovania.

3.4 Systémy na detekciu zvukových anomálií

Ďalším zaujímavým príkladom použitia neurónových sietí pri spracovaní zvuku sú systémy na detekciu zvukových anomálií. Takéto systémy možno vycvičiť na rozpoznávanie zvukových signálov, ktoré indikujú potenciálne problémy v strojoch alebo zariadeniach. Napríklad zvuky trenia, brúsenia, nárazov alebo vibrácií môžu naznačovať, že zariadenie je v chybnom stave a treba ho opraviť alebo vymeniť.

- Hlboké neurónové siete možno použiť na analýzu tónu a emócií v textoch. To môže pomôcť spoločnostiam pochopiť náladu svojich zákazníkov a zistiť, aké zmeny možno zaviesť na zlepšenie služieb. DNN sa napokon dajú použiť na vytvorenie syntetickej reči, ktorá znie rovnako ako ľudská reč. To by mohlo byť užitočné na vytváranie hlasových asistentov, audiokníh, aplikácií na čítanie textu a ďalších. Neurónové siete možno naučiť rozpoznávať takéto zvukové signály a vydávať príslušné varovania operátorom alebo automaticky spúšťať postupy údržby a opravy zariadení. To môže pomôcť predchádzať nepredvídaným prestojom na výrobnéj linke, znížiť riziká pre pracovníkov a ušetriť peniaze za opravy a výmeny.
- Jedným z najčastejších príkladov použitia systémov na detekciu zvukových anomálií je monitorovanie stavu automobilov a iných vozidiel. Neurónové siete môžu byť vyškolené na rozpoznávanie zvukov spojených s poruchami v motora, prevodovky alebo iných systémov vozidla, ktoré môžu pomôcť prevádzkovateľom a servisom odhaliť a odstrániť problémy skôr, ako sa stanú vážnejšími a povedú k nehode.
- Okrem toho môžu byť systémy na detekciu zvukových anomálií sú užitočné v energetike na zisťovanie porúch zariadení, ako sú turbíny a generátory. Môžu sa používať aj v iných priemyselných odvetviach, kde zvuk zohráva dôležitú úlohu pri diagnostike stavu zariadenia.

3.5 Hlboké neurónové siete (DNN)

Hlboké neurónové siete (DNN) možno použiť pri spracovaní prirodzeného jazyka na vytvorenie dialógových systémov a rozhraní. Okrem toho možno DNN rozšíriť o rôzne techniky, ako je napríklad posilňovanie učenia a genetické algoritmy, čo môže viesť k ešte presnejším výsledkom a širšiemu rozsahu aplikácie.

- Hlboké neurónové siete možno použiť na analýzu tónu a emócií v textoch. To môže pomôcť spoločnostiam pochopiť náladu svojich zákazníkov a zistiť, aké zmeny možno zaviesť na zlepšenie služieb.
- DNN sa napokon dajú použiť na vytvorenie syntetickej reči, ktorá znie rovnako ako ľudská reč. To by mohlo byť užitočné na vytváranie hlasových asistentov, audiokníh, aplikácií na čítanie textu a ďalších.

4 Viac o NVIDIA Broadcast

Nvidia Broadcast používa neurónovú sieť, ktorá pozostáva z viacerých vrstiev a má konvolučnú architektúru. Táto neurónová sieť je navrhnutá na spracovanie zvukových signálov v reálnom čase a dokáže účinne filtrovať šum, odstraňovať hluk a zlepšovať kvalitu zvuku.

Topológiu neurónovej siete Nvidia Broadcast možno opísať takto:

- Vstupná vrstva: Vstup neurónovej siete prijíma zvukový signál v Formát PCM (pulzná kódová modulácia) so vzorkovacou frekvenciou 48 kHz a 16-bitovou bitovou rýchlosťou.
- Konvolučné vrstvy: Zvukový signál potom prechádza niekoľkými konvolučnými vrstvami, z ktorých každá používa niekoľko konvolučných jadier rôznych veľkostí. Každé konvolučné jadro vykoná konvolučnú operáciu na časti vstupného signálu a vytvorí nový atribút na výstupe vrstvy.
- Vrstvy združovania: Po každej konvolučnej vrstve nasleduje vrstva združovania, ktorá znižuje priestorové rozlíšenie výstupných prvkov vykonaním operácie združovania na podmnožine prvkov. Tým sa znižuje počet parametrov modelu a urýchľuje sa jeho tréning a aplikácia.
- Normalizačná vrstva: Na efektívnejšie spracovanie signálu v ďalšej vrstve využíva neurónová sieť dávkovú normalizačnú vrstvu, ktorá štandardizuje hodnoty funkcií v každej vrstve a znižuje pravdepodobnosť nadmerného prispôbenia modelu.
- Plne prepojené vrstvy: Nakoniec sa výstupy posledných konvolúcií a normalizačných vrstiev odovzdajú niekoľkým plne prepojeným vrstvám, ktoré kombinujú informácie zo všetkých funkcií a vytvárajú konečný výstup siete.
- Výstupná vrstva: Posledná vrstva neurónovej siete používa aktivačnú funkciu Softmax na prevod výstupu siete na pravdepodobnosti tried, ktoré zodpovedajú rôznym zvukom efekty, ako je redukcia šumu, odstránenie šumu, vylepšenie hlasu atď.

5 Súčasný vývoj

Mnohé univerzity a spoločnosti v súčasnosti vyvíjajú neurónové siete na spracovanie zvuku.

- Napríklad na univerzite Carnegie Mellon (CMU) Vyvíjajú sa systémy rozpoznávania reči založené na hlbokých neurónových sieťach. Tieto systémy sa používajú v aplikácie na hlasové ovládanie zariadení, ako sú smartfóny alebo inteligentné domácnosti.
- Neurónové siete sa používajú aj v zvukové systémy na detekciu anomálií. Niektoré spoločnosti, ako napr. ako Falconry, vyvíjajú systémy monitorovanie stavu zariadení pomocou neurónových sietí na detekciu zvukov spojených s poruchami.
- Ďalším príkladom je spoločnosť Audeze, ktorá využíva neurónové siete na zlepšenie kvalitu zvuku v slúchadlách.

Toto je len niekoľko príkladov použitia neurónových sietí pri spracovaní zvuku a ich zoznam by mohol byť ešte oveľa dlhší.

6 Zhodnotenie a záver

Na záver možno konštatovať, že spracovanie zvuku pomocou neurónových sietí je veľmi sľubnou oblasťou a má mnoho aplikácií v praxi. Rekurentné, konvolučné, hlboké a generatívne neurónové siete sa môžu použiť na riešenie rôznych problémov pri spracovaní zvuku. Dúfame, že naša prezentácia bola užitočná a vám umožnil hlbšie pochopiť, ako sa dajú neurónové siete použiť na spracovanie zvukových údajov. Ďakujeme!

7 Použitá literatúra

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/>