



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA INFORMAČNÍCH TECHNOLOGIÍ

FACULTY OF INFORMATION TECHNOLOGY

ÚSTAV POČÍTAČOVÉ GRAFIKY A MULTIMÉDIÍ

DEPARTMENT OF COMPUTER GRAPHICS AND MULTIMEDIA

SEPARACE MLUVČÍCH V ČASOVÉ DOMÉNĚ

TIME DOMAIN AUDIO SEPARATION

BAKALÁŘSKÁ PRÁCE

BACHELOR'S THESIS

AUTOR PRÁCE

AUTHOR

JIŘÍ PEŠKA

VEDOUcí PRÁCE

SUPERVISOR

ing. KATEŘINA ŽMOLÍKOVÁ,

BRNO 2020

Zadání bakalářské práce



Student: **Peška Jiří**

Program: Informační technologie

Název: **Separace mluvčích v časové doméně pomocí neuronové sítě**
Time-Domain Neural Network Based Speaker Separation

Kategorie: Zpracování řeči a přirozeného jazyka

Zadání:

1. Seznamte se s problémem separace mluvčích pomocí neuronových sítí.
2. Seznamte se s metodou TasNet pro jednokanálovou separaci signálu v časové doméně.
3. Implementujte danou metodu s využitím vhodného toolkitu (např. PyTorch, Keras).
4. Otestujte systém na vhodném datasetu. Zaměřte se na vyhodnocení vlivu velikosti sítě na přesnost.
5. Navrhněte a diskutujte možné zlepšení použité metody.

Literatura:

- Luo, Yi, and Nima Mesgarani. "Conv-TasNet: Surpassing Ideal Time-Frequency Magnitude Masking for Speech Separation." *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 27.8 (2019): 1256-1266.
- dle doporučení vedoucího

Podrobné závazné pokyny pro vypracování práce viz <https://www.fit.vut.cz/study/theses/>

Vedoucí práce: **Žmolíková Kateřina, Ing.**

Vedoucí ústavu: Černocký Jan, doc. Dr. Ing.

Datum zadání: 1. listopadu 2019

Datum odevzdání: 14. května 2020

Datum schválení: 5. listopadu 2019

Abstrakt

Práce se zabývá využitím konvolučních neuronových sítí pro automatickou separaci mluvčích v akustickém prostředí. Cílem je implementovat neuronovou síť podle architektury TasNet za použití frameworku pytorch, natrénovat síť s různými hodnotami hyperparametrů a porovnat kvalitu separací vzhledem k velikosti sítě.

Architektura oproti dosavadním metodám, které převáděly vstupní směs do časově-frekvenční reprezentace, používá konvoluční autoenkodér, který vstupní směs převádí do nezáporné reprezentace, která je optimalizovaná pro extrakci jednotlivých mluvčích. Samotné separace je docíleno aplikací masek, které jsou odhadnuty v separačním modulu. Modul tvoří opakující se posloupnost konvolučních bloků se zvyšující se dilatací, která napomáhá k modelování časových závislostí ve zpracovávané směsi.

K vyhodnocení přesnosti bylo použita metrika SDR (Sound to Distortion Ratio), která určuje poměr zastoupení šumu a zvuku v nahrávce. Natrénováním několika modelů s různými hodnotami hyperparametrů bylo možno zpozorovat závislost mezi velikostí sítě a hodnotou SDR. Zatímco menší síť dosahovala, po X epochách trénování, přesnosti XY, větší síť dosahovala až XX.

[[Jeste neco? 4. cast?]]

[[Doplnit SDR presnost do odstavce vyse.]]

Abstract

[[Prelozit do anglictiny CZ abstrakt]]

Klíčová slova

neuronové sítě, zpracování řeči, konvoluční neuronová síť, autoenkodér, separace mluvčích, strojové učení, tasnet

Keywords

neural networks, speech processing, convolutional neural networks, autoencoder, speech separation, machine learning, tasnet

Citace

PEŠKA, Jiří. *Separace mluvčích v časové doméně*. Brno, 2020. Bakalářská práce. Vysoké učení technické v Brně, Fakulta informačních technologií. Vedoucí práce ing. Kateřina Žmolíková,

Separace mluvčích v časové doméně

Prohlášení

Prohlašuji, že jsem tuto bakalářskou práci vypracoval samostatně pod vedením ing. Kateřiny Žmolíkové. Další informace mi poskytli... Uvedl jsem všechny literární prameny, publikace a další zdroje, ze kterých jsem čerpal.

.....

Jiří Peška
1. dubna 2020

Poděkování

V této sekci je možno uvést poděkování vedoucímu práce a těm, kteří poskytli odbornou pomoc (externí zadavatel, konzultant apod.).

Obsah

1	Úvod	2
2	Separace mluvních	4
3	Neuronové sítě	5
3.1	Umělý neuron	5
3.2	Aktivační funkce	7
3.3	Proces učení neuronových sítí	9
3.3.1	Objektivní funkce	9
3.3.2	Inicializace parametrů	10
3.3.3	Optimalizační algoritmy	10
3.3.4	Backpropagation	10
3.3.5	Overfitting a generalizace	11
3.4	Konvoluční neuronové sítě	11
3.4.1	Konvoluce	11
4	TasNet - Time-Domain Audio Separation Network	12
4.1	Konvoluční auto-ekodér	12
4.2	Separační modul	13
4.2.1	Konvoluční bloky	14
5	Implementace a trénování sítě	15
5.1	Implementace modelu	15
5.2	Dataset	15
5.3	Trénování	15
5.3.1	Význam validační množiny v trénování	16
5.4	Vyhodnocovací metriky	16
5.4.1	Signal to noise ration	17
6	Experimenty a vyhodnocení	18
6.1	Možná rozšíření a navrhnutá vylepšení	18
7	Závěr	19
	Literatura	20

Kapitola 1

Úvod

[[reference na kapitoly ..ref..]]

1) 5 radku obecný úvod Zpracování řeči hraje v dnešní době důležitou roli v mnoha rozličných oborech. Mezi jedny z hlavních úkolů bezesporu patří separace zdrojů v nějakém zaznamenaném signálu, který může být složen ze signálů N mluvčích, ale i nechtěnému hluku okolí. Vyřešení problému je předpoklad k dalším úkonům jako identifikace konkrétního mluvčího nebo třeba přepis mluveného slova na text. Se stále se zrychlujícím vývojem počítačů a s jejich zvyšujícím se výkonem se do popředí dostávají metody zpracování řeči založené na neuronových sítích, které v mnoha ohledech předčily doposud používané algoritmy.

2) proč je práce v dané oblasti důležitá, význam pro svět a pro nás Separace mluvčích v časové doméně dosahuje mimořádných výsledků v porovnání s dosavadními metodami LSTM založenými na převodu signálu z časové domény do frekvenční domény. Taková reprezentace signálu není optimální pro udržení časových závislostí, které jsou při zpracování řeči podstatné. V referenční studii je vstupní signál převeden do nezáporné reprezentace, která je optimální pro extrakci jednotlivých mluvčích. Silnou stránkou systému je hluboká architektura sítě, která dokáže zachovat dlouhodobé závislosti v signálu.

3) o minulosti, stav dnes a vyhlídka do budoucna

4) proč se tematice chci věnovat, proč je podle mě perspektivní = motivace Téma v oblasti neuronových sítí jsem si vybral, jelikož ty zažívají obrovský rozmach a pomalu se stávají součástí téměř všech odvětví. **[[nejak to jeste dopsat tohle.]]**

5) co je cílem mé práce - vlastními slovy Má práce si klade za cíl implementovat neuronovou síť podle architektury TasNet pro separaci mluvčích, která byla navržena a popsána ve studii^{??}. Následně tuto neuronovou síť natrénovat s různými kombinacemi hodnot hyperparametrů ovlivňujících velikost sítě a její vlastnosti a nakonec porovnat přesnost mezi jednotlivými sítěmi a mezi výsledky studie. Přesnost separace je vypočítána za pomoci míry $si-snr$, udávající poměr mezi chtěným signálem a hlukem na pozadí. **[[reference na sisnr na wiki nebo v nejaké studii??]]**

6) struktura práce - popsat kapitoly **[[prepsat podle novejsich kapitol co jsem navrhoval - doplnit popsání datasetu]]** V první části práce jsou popsány základní prvky neuronových sítí, struktura umělého neuronu, jeho vstupy a výstupy, váhy a role aktivací funkce. V návaznosti na to je popsán proces učení neuronové sítě. Ten se skládá z několika souvisejících částí, které zahrnují výpočet výstupu neuronové sítě metodou feed forward, který transformuje vstupní hodnoty a počítá na základě nich výstup, který propaguje do dalších vrstev neuronové sítě. Dále je rozebrán výpočet chyby, která vzniká během procesu učení, metodou gradient descent a nakonec úprava vah neuronů metodou backpropagation,

která se počítá na základě rozdílu mezi vstupními hodnotami a očekávanými výstupními hodnotami.

Předtím, než jsou představeny konvoluční sítě, tak je vysvětlena samotná konvoluce. Dále je vysvětlen princip konvolučních neuronových sítí, které se používají nejčastěji pro zpracování obrazu kvůli vlastnostem, které umožňují extrahovat příznaky s různou úrovní složitosti od základních útvarů jako úsečka, barva a podobně až po například část obličeje – ucho, nos, či úplně celý obličej. Tohoto lze využít i při zpracování zvuku, kde jsou tyto extrahované příznaky jednorozměrné.

Se znalostí principu konvolučních sítí je představen konvoluční auto-inkodér, který převádí vstupní nahrávku směsi mluvčích na reprezentaci optimalizovanou pro separaci jednotlivých mluvčích.

Druhá část je věnována architektuře TasNet. V této kapitole je popsána podoba separačního modulu, jeho stavební bloky a princip. Postupně je znovu zmíněn konvoluční auto-inkodér, u nějž je vysvětlen jeho úkol v separačním modulu a následně konvoluční blok, který se sám sestává z konvolučních vrstev, normalizací a aktivačních funkcí. Tyto bloky jsou skládány za sebe se zvyšující se časovou dilatací a tvoří jádro separačního modulu.

Třetí část se zabývá implementací neuronové sítě. Je popsán a odůvodněn zvolený framework pytorch, využití prostředky a metody. **[[nejak to jeste dopsat tohle.]]**

Na závěr jsou popsány experimenty s modelem - rychlost učení, vliv hyper-parametrů na učení sítě, na výsledky a přesnost výstupu v závislosti na zvolených parametrech, optimalizacích a počtu konvolučních bloků a pod. Výstup sítě v podobě separovaných mluvčích je porovnán s referenční studií. **[[Zminit zde taky sisnr hodnoceni a pod.]]**

[[Teorie: Co bylo potreba nastudovat;; uvod do problematiky;; pisu to pro nekoho, kdo chce vychazet z me bakalarky.]]

Kapitola 2

Separace mluvčích

- obecně o problému separace; prostředí a využití, cocktail party, multispeech

Kapitola 3

Neuronové sítě

body: co vlastne resi . skladaji se ze vstupni vrstvy, N skrytych vrstev, vystupni vrstvy Co resi neuronove site.

V dnešní době zažívají neuronové sítě díky výkonosti počítačů velký rozmach. Jejich využití prostupuje skrze mnohé vědní obory a nově dokáže řešit celou řadu problémů, ve kterých dosahuje výborných výsledků, které zdaleka předčily dosavadní postupy. Mezi nejčastější úlohy, na které se neuronové sítě používají, jsou klasifikační úlohy, rozpoznávání obrazu a řeči či vzorů na videu nebo ve zvuku až po generování textu. Na základně řešeného problému vzniklo mnoho druhů neuronových sítí, z nichž některé zde budou představeny.

Nejzákladnější neuronová síť je vícevrstvá neuronová síť, neboli MLP (Multi Layer Perceptron). Tento typ sítě je skládá ze třech typů vrstev. Vstupní vrstva slouží k předání hodnot do sítě. Tato vrstva nijak nemodifikuje vstupní hodnoty, které jsou do sítě předávány a nezměněné je kopíruje první skryté vrstvě. Každá vrstva se může skládat z 1 až N neuronů, kde $N \in \mathbb{N}$. Poslední skrytá vrstva je napojena na výstupní vrstvu. Výstupní vrstva má obvykle méně neuronů než předešlé vrstvy a hodnoty na výstupu mohou představovat třídy, do kterých má být zařazen vstup. S počtem jednotlivých vrstev souvisí pojem hloubka sítě, která je rovna počtu všech vrstev neuronové sítě od vstupní až po výstupní vrstvu.

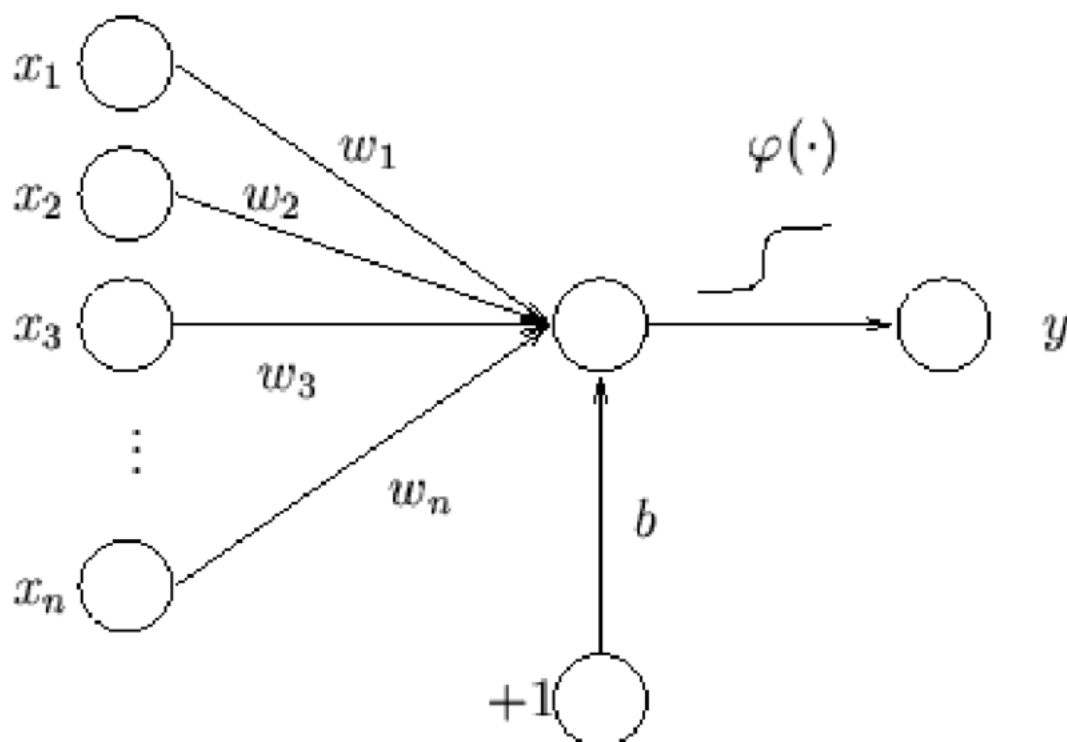
Takto propojené neurony tvoří acyklický graf, který počítá a následně předává hodnoty směrem od vstupní vrstvy skrze skryté vrstvy až k vrstvě výstupní. Nenacházejí se zde žádná zpětná spojení, ve kterých by se výstup vracel zpět do sítě.^[1]

Cílem takové neuronové sítě je aproximovat nějakou funkci f^* . Síti je předána vstupní hodnota x a výstupní hodnota $y^* = f^*(x)$ má být co nejblíže hodnotě $y = f(x)$.

3.1 Umělý neuron

Základní stavební jednotka neuronových sítí je neuron, nebo přesněji perceptron. Tento model je založen na reálných poznatcích o neuronech, které se nacházejí v organismu. Perceptron obsahuje libovolně mnoho vstupních synapsí, přes které se neuronu předá vstupní hodnota, váhy a jeden výstup, jehož hodnota závisí na vstupních hodnotách, vnitřním stavu neuronu (hodnoty vah a biase) a zvolené aktivační funkci. Vstupní hodnoty jsou váhovány, což v praxi znamená, že každá vstupní hodnota je vynásobena s váhou daného vstupu. Váhy v perceptronu představují nějaký vektor vah $w = [w_1, w_2, \dots, w_n]$, se kterým je vynásobený vektor vstupních hodnot $x = [x_1, x_2, \dots, x_n]$.

[[obrazek neuronu a popis]]



Obrázek 3.1: Schéma umělého neuronu – perceptron

Hodnota bias $b \in R$, která je přičtena k sumě násobků vah a vstupních hodnot, modifikuje dobu, kdy se aktivuje perceptron a změni svůj výstup. Je to prahová hodnota, na základě níž je měněn výstup. Matematicky to znamená, že s aktivační funkcí horizontálně pohybuje doleva nebo doprava v závislosti na tom, je-li bias pozitivní nebo negativní. Bias se učí zároveň s ostatními váhami během učícího procesu.



Obrázek 3.2: Vliv hodnoty bias na aktivační funkci

Výstup neuronu se vypočítá jako:

$$y = a\left(\sum_{n=1} w_n x_n + b\right) \quad (3.1)$$

kde a je nějaká aktivační funkce, $x_n \in R$ je vstupní hodnota, $w_n \in R$ je váha, kterou se vstupní hodnota vynásobí a $b \in R$ je hodnota bias, která je přičtena k celkové sumě předtím, než se výsledek předá aktivační funkci.

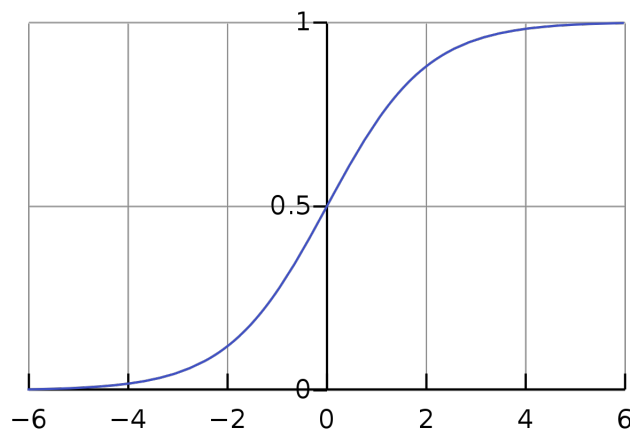
3.2 Aktivační funkce

Aktivační, neboli prahová funkce určuje výstupní hodnotu neuronu. Funkce se vybírá na základě problému, který se má neuronová síť naučit řešit. Správná volba prahové funkce vede k lepší konvergenci učení sítě. Naopak špatná volba může vést ke stále větší odchylce od správného řešení – může divergovat. Povaha problému může vyžadovat specifické vlastnosti aktivační funkce - lineární nebo nelineární – sigmoidní a podobně. Pro správnou volbu aktivační funkce je pro nestandardní problémy potřeba experimentálně zjistit, která bude nejlépe vyhovovat.

Sigmoid

$$f(x) = \frac{1}{1 + \exp(-z)} \quad (3.2)$$

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper. 1



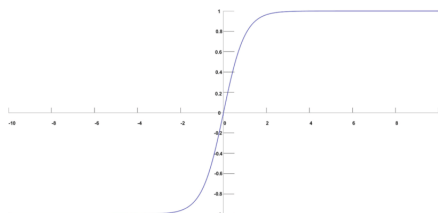
Obrázek 3.3: Graf aktivační funkce sigmoid

Softmax

$$f(x) = \frac{1}{1 + \exp(-z)} \quad (3.3)$$

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla

tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper. 1



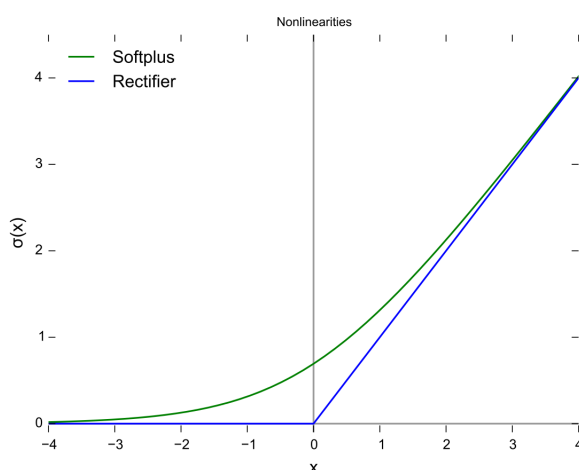
Obrázek 3.4: Graf aktivační funkce softmax

ReLU

Rectified Linear Unit je nejčastěji používaná aktivační funkce. Vyžaduje-li neuronová síť nějakou nelinearitu, je ReLU pro většinu případů ideální. Pro každou zápornou hodnotu x vrací 0 a pro kladnou hodnotu x vrací tutéž hodnotu x .

$$f(x) = \max(0, x) \quad (3.4)$$

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper. 1



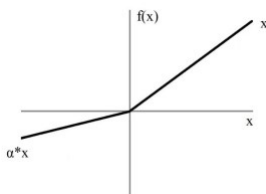
Obrázek 3.5: Graf aktivační funkce ReLU

PReLU

Parametrizovaná ReLU je nelineární aktivační funkce, která se používá v případě, že chceme produkovat na výstup malý nenulový gradient i v případě záporné vstupní hodnoty x . V tom případě je vstupní hodnota vynásobena parametrem α a to představuje výsledek. Parametr α se společně s ostatními váhami učí během učícího procesu.

$$f(x) = \begin{cases} x & \text{if } x \geq 0 \\ \alpha x & \text{if } x < 0 \end{cases} \quad (3.5)$$

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper. 1



Obrázek 3.6: Graf aktivační funkce PReLU

3.3 Proces učení neuronových sítí

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper. 1

3.3.1 Objektivní funkce

= cost funkce -popis, co to je, k čemu to je, proč to je...

MSELoss

- vzoreček

Cross Entrophy

- vzorecek

3.3.2 Inicializace parametrů

[[Mozna, Kniha strana 292]]

3.3.3 Optimalizační algoritmy

[1 deep learning str 301]

Adam

[[Kniha strana 301]]

Adam je jeden z algoritmů s adaptivním učení. Jeho název byl odvozen z fráze "adaptive moments". [1 deep learning str 301]

3.3.4 Backpropagation

- zpetne sireni chyby - adaptacni algoritmus, podil neuronu na chybe, - 3 opakujici se faze uzeni:

[[dodelat zde podkapitoly v lepsim poradí]]

1) feedforward - dopredu 2) zpetne sireni chyby - Backpropagation 3) uprava vah a biasu na zaklade chyby - chain rule

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper. 1

Feed Forward

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper. 1

Gradient descent

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullam-

corper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper. 1

3.3.5 Overfitting a generalizace

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper. 1

3.4 Konvoluční neuronové sítě

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper. 1

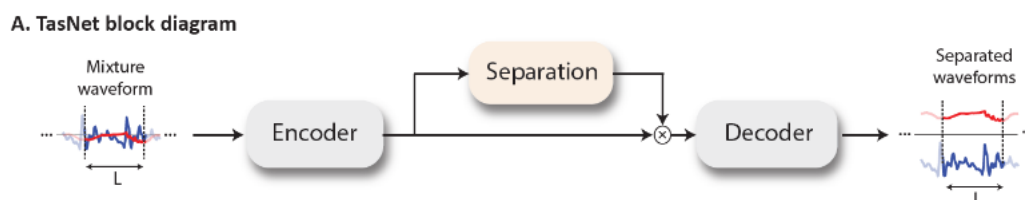
3.4.1 Konvoluce

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper. 1

Kapitola 4

TasNet - Time-Domain Audio Separation Network

[[Architektura full – obrázek, bloky...]]



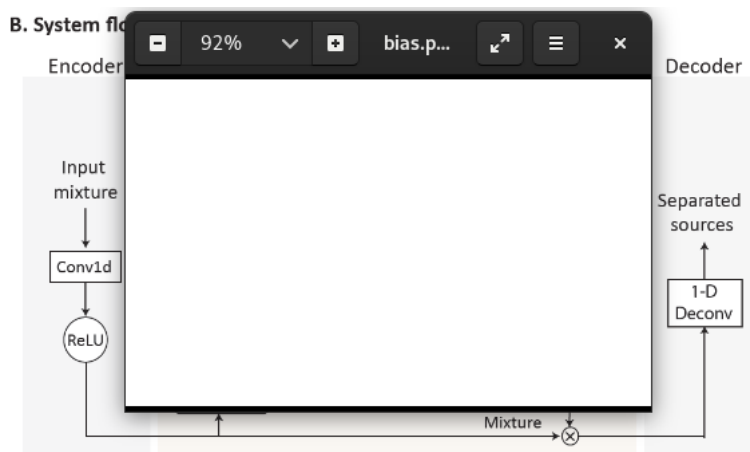
Obrázek 4.1: Zjednodušený model architektury

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper. 1

4.1 Konvoluční auto-enkodér

[[Konvoluční autoenkodér, vstup, výstup...]]

- schema bez separacního modulu - non negative representation of audio

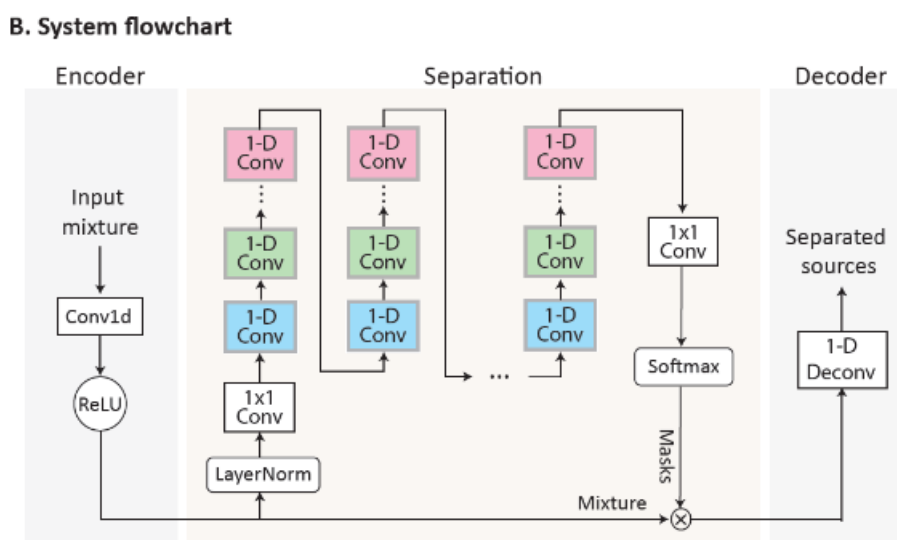


Obrázek 4.2: Schéma konvolučního autoenkodéru

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper. 1

4.2 Separační modul

- odhad masek pro jednotlivé mluvčí - schema se separacním modulem

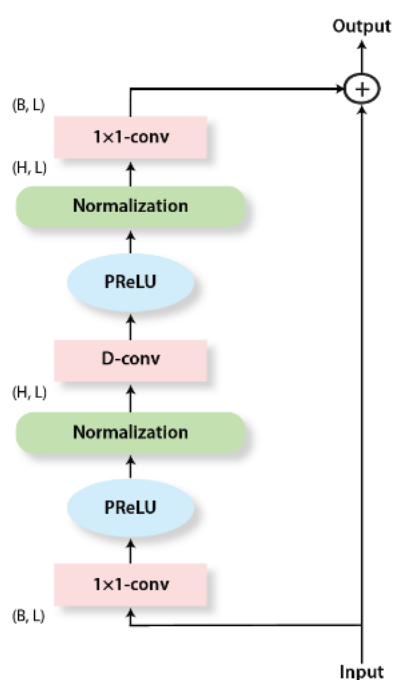


Obrázek 4.3: Schéma architektury TasNet

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper. 1

4.2.1 Konvoluční bloky

- Z čeho se skládá – konvoluční vrstvy, normalizace - diagram konv bloku. - Možná: Dilatace a time perception



Obrázek 4.4: Jeden konvoluční blok

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper. 1

Kapitola 5

Implementace a trénování sítě

Pozn: colab, pytorch, stroj, bash, hyperparams, výkon a čas trénování, seg-len, popis tríd.

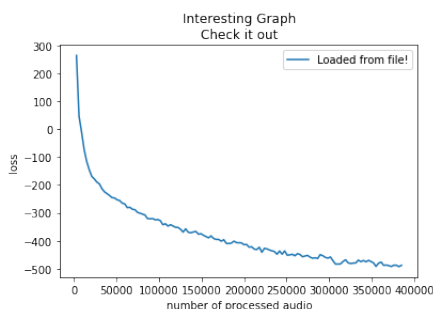
5.1 Implementace modelu

- pytorch, scripty, python3, bash, tridy, moduly, parametry a volby spustění.

5.2 Dataset

[[Ukazat zde vykreslenou vlnu nahravek mix, s1, s2]] **[[popsat co je dataset a k čemu to slouží]]** Trénování a vyhodnocení modelu proběhlo na množině jednokanálových nahrávek směsí dvou mluvčích. Množina byla vygenerována náhodným výběrem různých mluvčích z Wall Street Journal (WSJ0) a vytvořením směsi. Celková délka trénovacích dat je přes 10 hodin a přes 6 hodin validačních dat. Nahrávky jsou převzorkovány na 8kHz a během trénování zarovnány na zero means a jednotkovou varianci[studie str 5 Dataset][49 - ze studie odkaz na script na generování a popis na netu]. **[[doplnit info o zero means a jednotkové varianci]]**

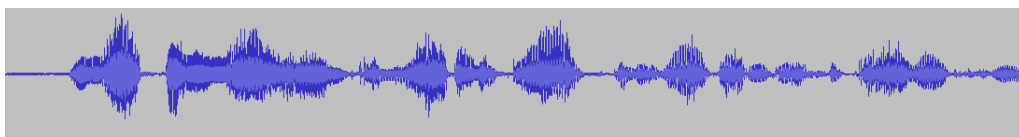
5.3 Trénování



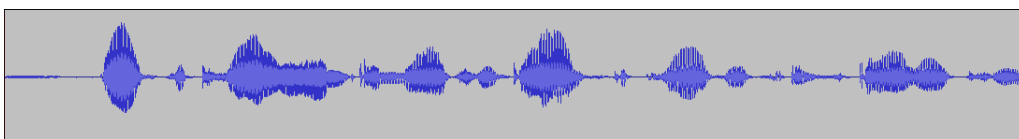
Obrázek 5.1: Příklad grafu loss hodnoty během učení

5.3.1 Význam validační množiny v trénování

Většina algoritmů strojového učení má nějakou sadu hyperparametrů, kterou je upravováno chování algoritmu. Hodnoty hyperparametrů obvykle bývají nastavovány ručně ještě před spuštěním procesu učení a hodnota se v průběhu nemění, protože hodnoty by bylo obtížné optimalizovat. Některá nastavení se nicméně mohou stát hyperparametrem a být upravována během trénování, ale není vhodné je měnit na základě výsledku učení na trénovací sadě, protože by mohlo dojít k přetrénování (overfitting) v důsledku **[[CEHO??]]**. Pro tento případ potřebujeme validační sadu, která je odlišná od trénovací sady. Po každém zpracování trénovací sady následuje validační sada, po jejímž skončení jsou optimalizovány hyperparametry **[[]]**.



Obrázek 5.2: Ukázka nahrávky směsi dvou mluvčích



Obrázek 5.3: První mluvčí ze směsi



Obrázek 5.4: Druhý mluvčí ze směsi

Lze si všimnout, že sečtením signálů separovaných mluvčích na obrázku (ref obr1) a obrázku (ref obr2) dostaneme přesně signál směsi, což lze vyjádřit vztahem

$$x(t) = \sum_{i=1}^C s_i(t) \quad (5.1)$$

, kde $x(t) \in \mathbb{R}^{1 \times T}$ je diskretní signál směsi a $s_i(t) \in \mathbb{R}^{1 \times T}$, kde $i = 1, \dots, C$, je jeden z C zdrojů [ref studie str3 vlevo]. **[[kniha 117-118] kap 5.3 = Hyperparameters and validation set]]** **[[najít jeste nejakej zdroj s popisem a pripadne nejaky zajimavejsi info.]]**

5.4 Vyhodnocovací metriky

- minimalizovat objektivní-hodnotící funkci sisnr .

5.4.1 Signal to noise ration

Source Distortion Ratio – SDR

Artifacts Ratio – SAR

Inference Ratio – SIR

Kapitola 6

Experimenty a vyhodnocení

- trenovani s ruznymi hyperparametry, uspesnost a tabulky s hyper parametry a dosazenymi vysledky a hodnotami sisnr, sdr atd. - model size comparison. - porovnaní s vysledky ze studie - obrazky separovanych mluvčích - signalu. - spektra - grafy trenovani loss a vysledkuu.
- pametova narocnost modelu

6.1 Možná rozšíření a navrhnutá vylepšení

- variabilnější dataset, mikrofony, šum a bordel prostředí - separace více mluvčích - hlučné prostředí - identifikace konkrétního řečníka - realtime separace

Kapitola 7

Závěr

- co jak dopadlo, vysledky a vyhodnoceni velikosti modelu a jaky byl nejlepsi,... Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Literatura

- [1] IAN GOODFELLOW, A. C. *DEEP LEARNING*. MIT Press, 2017. ISBN 9780262035613.