

# Detekce akustického prostředí z audio nahrávek

Filip Grepl

Vedoucí: Ing. Pavel Matějka, Ph.D.

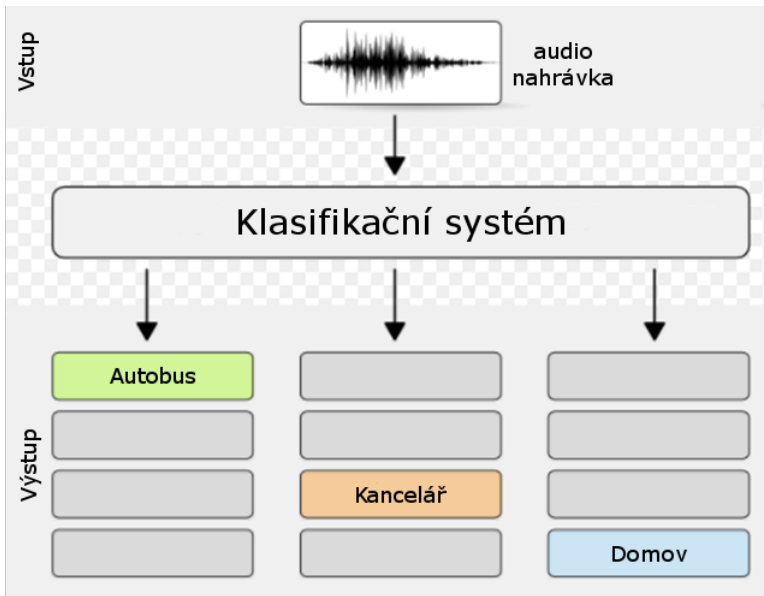
Vysoké učení technické v Brně, Fakulta informačních technologií

Božetěchova 1/2. 602 00 Brno - Královo Pole

[xgrepl05@stud.fit.vutbr.cz](mailto:xgrepl05@stud.fit.vutbr.cz)



Červen 12, 2018



- Celkem 15 různých míst
- Z každého místa 312 nahrávek, dohromady 4680 nahrávek
- Každá nahrávka je dlouhá 10 s

Tabulka klasifikačních míst

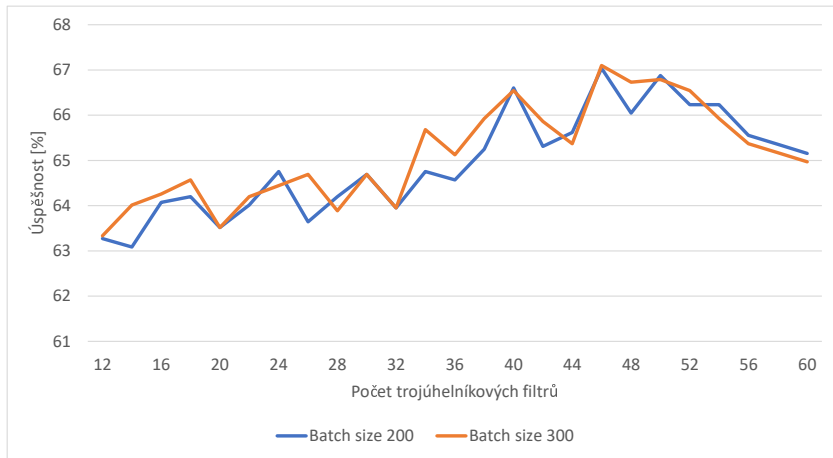
Pláž	Lesní cesta	Kancelář
Autobus	Obchod s potravinami	Park
Kavárna/Restaurace	Domov	Obytná oblast
Auto	Knihovna	Vlak
Centrum města	Stanice metra	Tramvaj

Reprezentace vlastností audio nahrávek:

- Mel-filter bank (MFB)
  - blok Mel-filter bank
- Mel-frequency cepstral coefficients (MFCC)

Klasifikátor realizován pomocí vícevrstvé hustě propojené neuronové sítě

Vstupní vrstva
Dense(ReLu, 200)
Dropout(0.2)
Dense(ReLu, 200)
Dropout(0.2)
Dense(Softmax, 15)



- Původní vzorkovací frekvence 44.1 kHz
- Průměrná hodnota obou kanálů
- blok MFB koeficienty s kontextem 10 rámců
- 46 trojúhelníkových filtrů
- žádná normalizace

Parametr	Hodnota
Délka rámce	20 ms
Překrývání	12 ms
Learning rate	$1 * 10^{-5}$
Počet iterací	100
Batch size	300

Třída	Úspěšnost (%)			
	Datová sada DCASE 2016		Datová sada DCASE 2017	
	Základní systém 2016	Nejlepší systém	Základní systém 2017	Nejlepší systém
autobus	88.5	96.2	38.9	43.5
knihovna	26.9	53.8	30.6	62.0
kavárna/restaurace	69.2	61.5	43.5	59.3
stanice metra	100.0	76.9	93.5	100.0
auto	96.2	100.0	64.8	71.3
kancelář	96.2	100.0	73.1	63.0
centrum města	80.8	84.6	79.6	92.6
obytná oblast	88.5	65.4	77.8	85.2
lesní cesta	65.4	100.0	85.2	85.2
vlak	30.8	46.2	72.2	66.7
obchod s potravinami	88.5	84.6	49.1	61.1
tramvaj	92.3	100.0	57.4	70.4
domov	92.3	92.3	76.9	78.7
park	53.8	92.3	32.4	38.9
pláž	84.6	88.5	40.7	34.3
<b>Celková úspěšnost:</b>	<b>76.9</b>	<b>82.8</b>	<b>61.0</b>	<b>67.5</b>

- Vyzkoušet jiné typy neuronových sítí, zejména konvoluční neuronové sítě
- Fúze systémů

Děkuji za pozornost