AudioNet:
Transfer Learning with a Simplified Version of SoundNet

Amy Bryce

Universität Hamburg
Department Informatik
Knowledge Technology, WTM

Bio-Inspired Artificial Intelligence
WiSe 17/18
15.02.2018

Why Build AudioNet?

{Introduce the problem; summarize the abstract}

{Introduce SoundNet and how AudioNet is derived from it.}

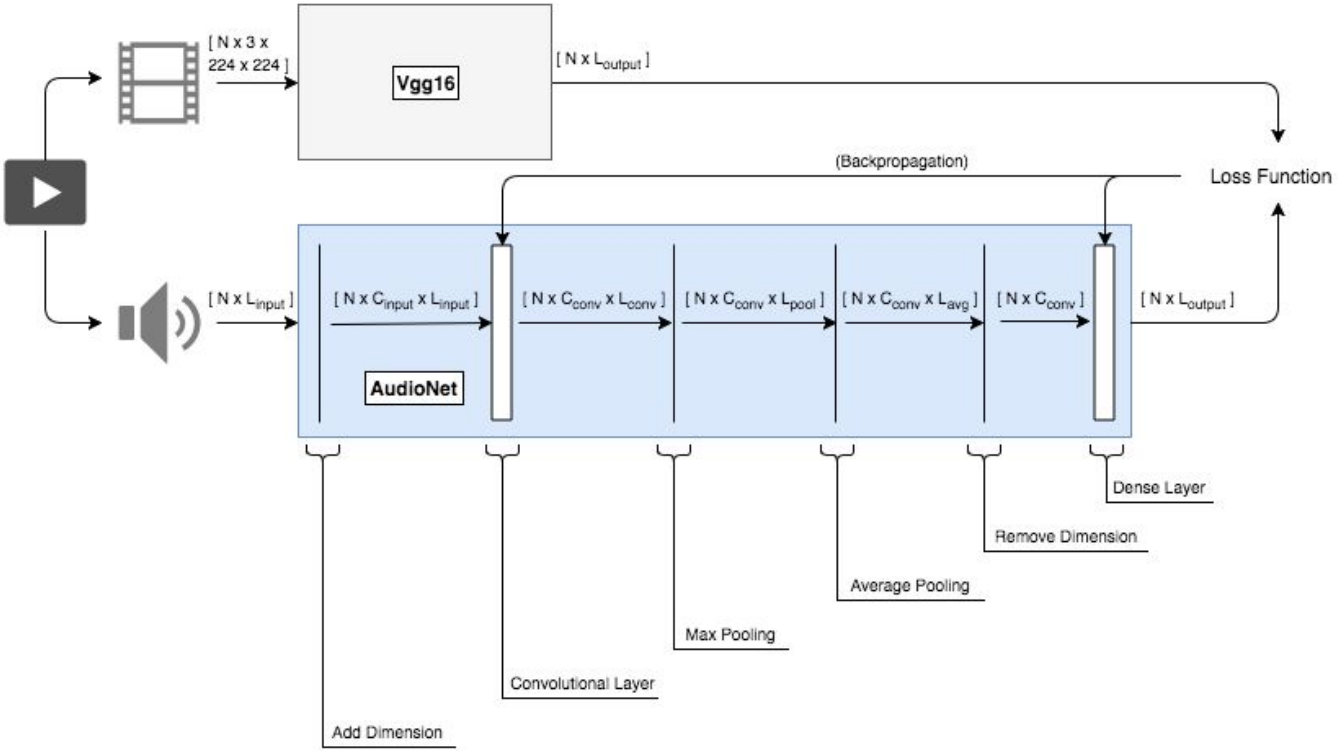
Transfer Learning

Define transfer learning.

Overview

- AudioNet Architecture
- AudioNet Implementation
- Videos Used to Train AudioNet
- AudioNet Evaluation
- Conclusion

AudioNet Architecture

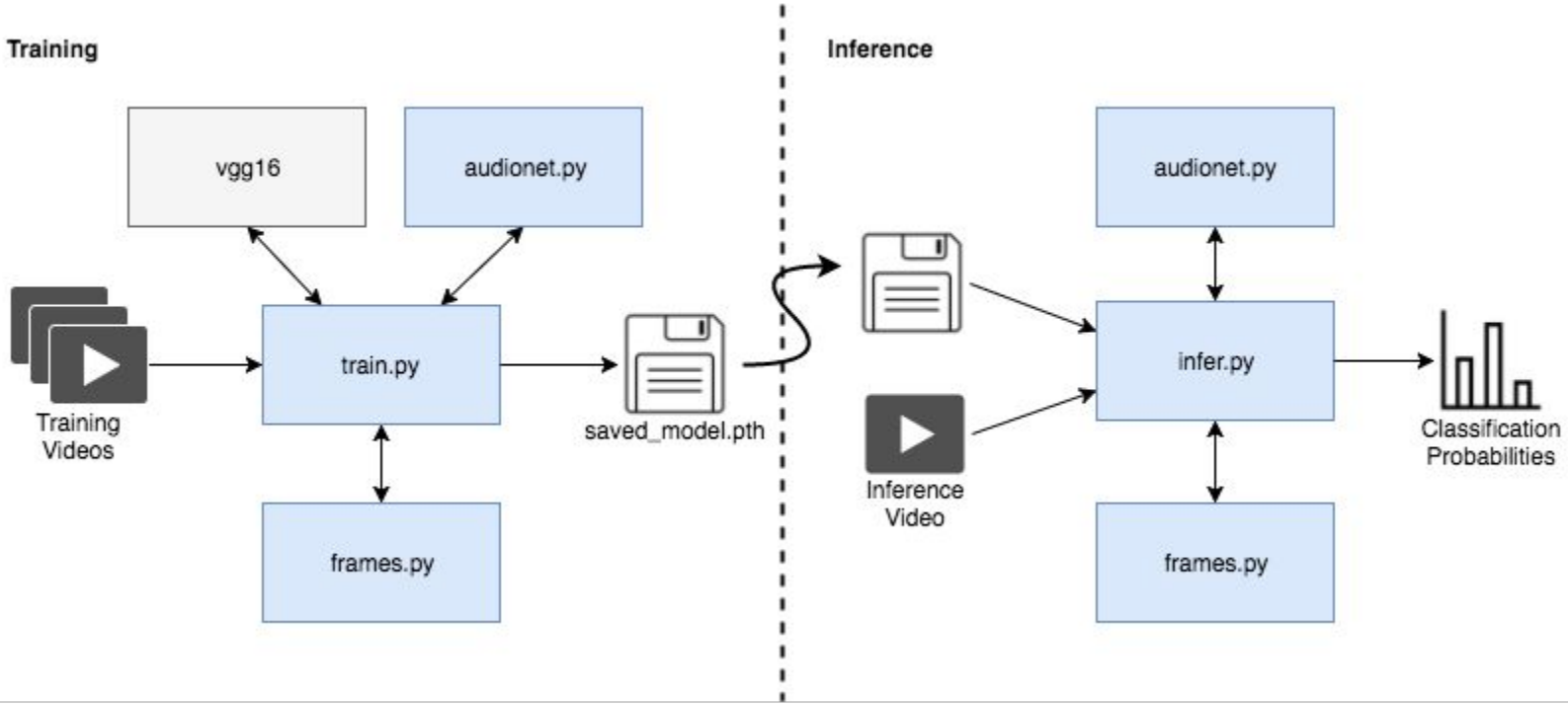


AudioNet Implementation

Layer	Input Dimension	Kernel Size	Padding	Stride	Output Dimension
Convolutional	$N \times 1 \times L_{input}$	64	32	2	$N \times 16 \times L_{conv}$
Max Pooling	$N \times 16 \times L_{conv}$	8	4	1	$N \times 16 \times L_{pool}$
Average Pooling	$N \times 16 \times L_{pool}$	L_{pool}	–	–	$N \times 16 \times 1$
Dense Linear	$N \times 16$	1000	–	–	$N \times 1000$

{Define layers here}

AudioNet Implementation



Videos Used to Train AudioNet

Video Contents	Youtube URL	Start Time	Length
Rooster (Cock)	https://www.youtube.com/watch?v=67GZuUxV27w&t=30	00 : 30	10s
Sewing Machine	https://www.youtube.com/watch?v=9PmzQI8ZYpg&t=30	00 : 30	10s
Fire Truck	https://www.youtube.com/watch?v=_A30xsFBMXA&t=40	00 : 40	10s
Harmonica	https://www.youtube.com/watch?v=BUGx2e70gFE&t=30	00 : 30	10s
Polaroid Camera	https://www.youtube.com/watch?v=eHIlPlNWISg&t=90	01 : 30	10s
Race Car	https://www.youtube.com/watch?v=eV5JX81GzqA&t=150	2 : 30	10s
Electric Guitar	https://www.youtube.com/watch?v=-0AyRsvFGgc&t=30	00 : 30	10s
Tree Frog	https://www.youtube.com/watch?v=rctt0dhCHxs&t=16	00 : 16	9s
Keyboard	https://www.youtube.com/watch?v=rTh92nlG9io&t=30	00 : 30	10s
Magpie	https://www.youtube.com/watch?v=-XilaFMUwng&t=50	00 : 50	10s

Videos Used to Train AudioNet

{These are the videos I trained AudioNet on.}

{Linked video: 100 seconds}

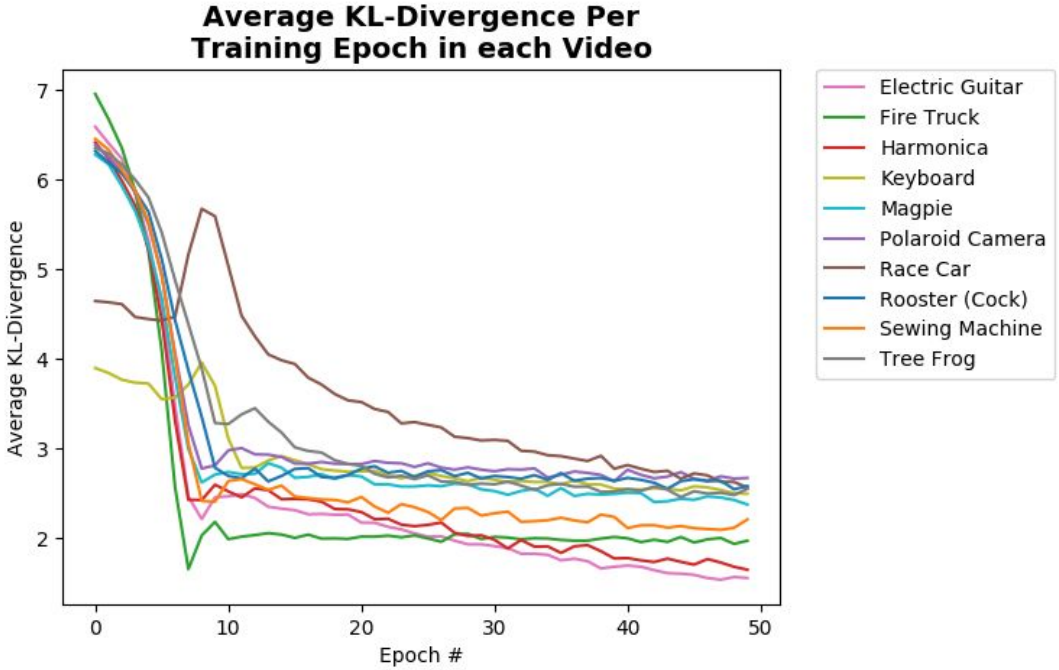
AudioNet Evaluation

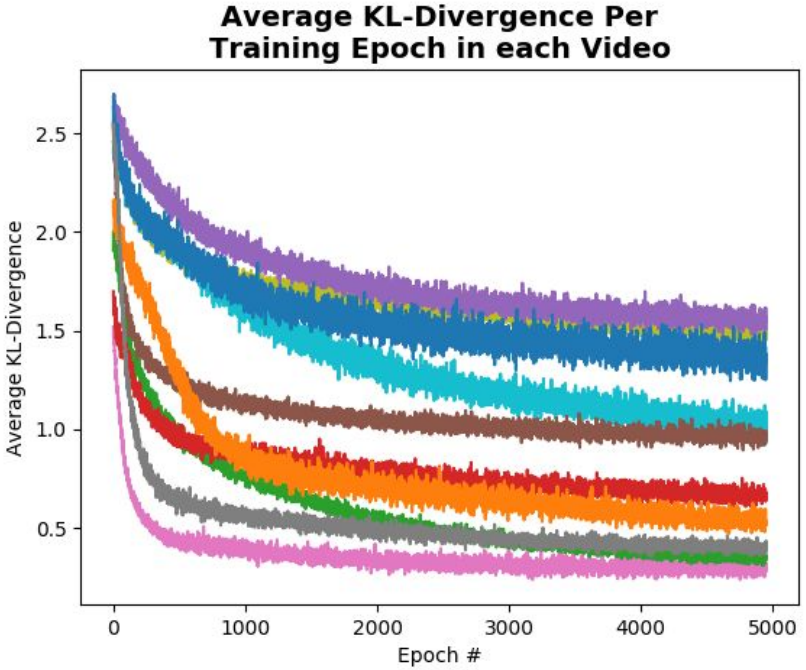
Parameter	Value
Epochs	5000
Batch Size	512
Video Sample Period	$40ms$
Audio Sampling Rate	$16kHz$
Learning Rate	$1e-4$

I used four NVIDIA Tesla M60 GPUs with 8GB of RAM each to train AudioNet with these parameters.

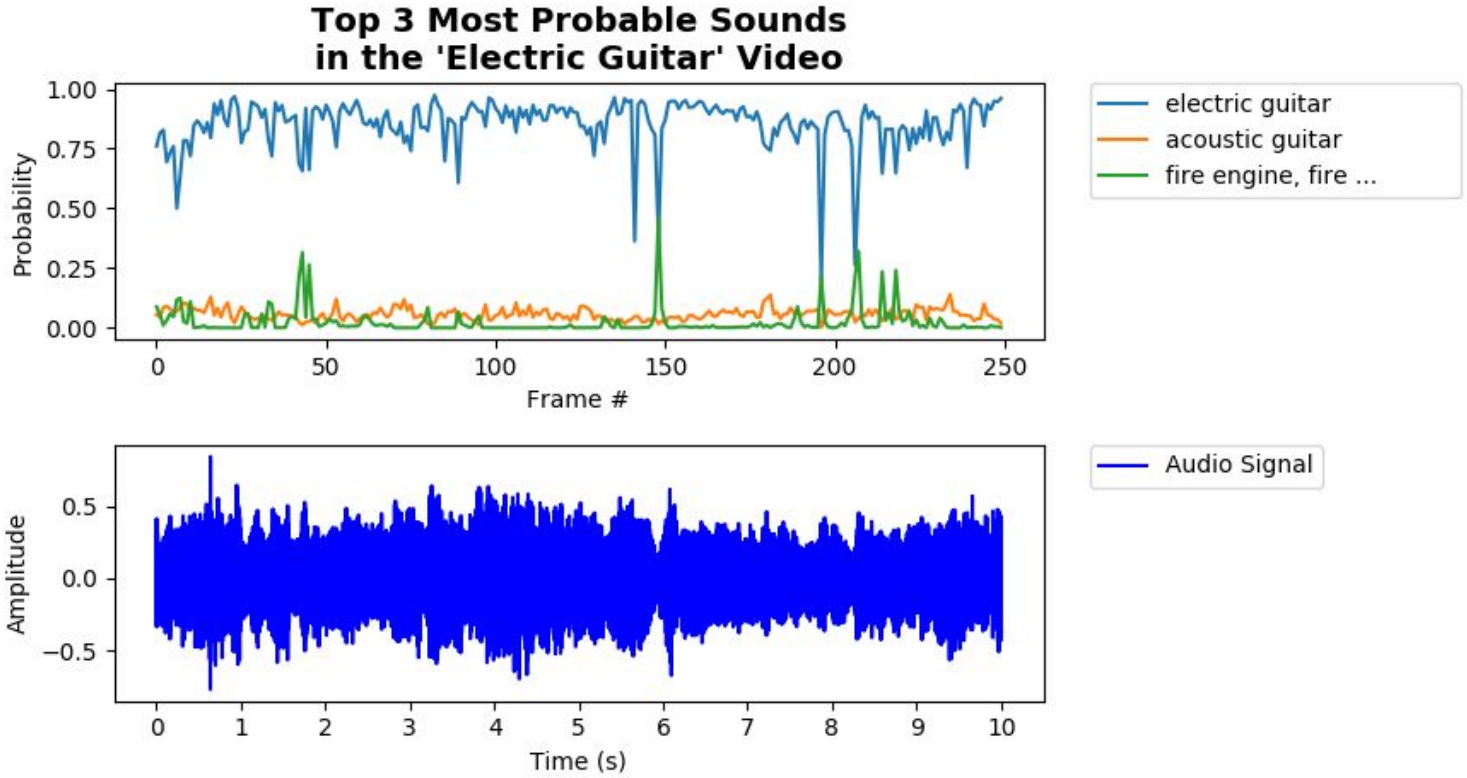
- Epoch definition
- Batch Size definition
- Video Sample Period definition
- Audio Sampling Rate definition
- Learning Rate definition

AudioNet Evaluation

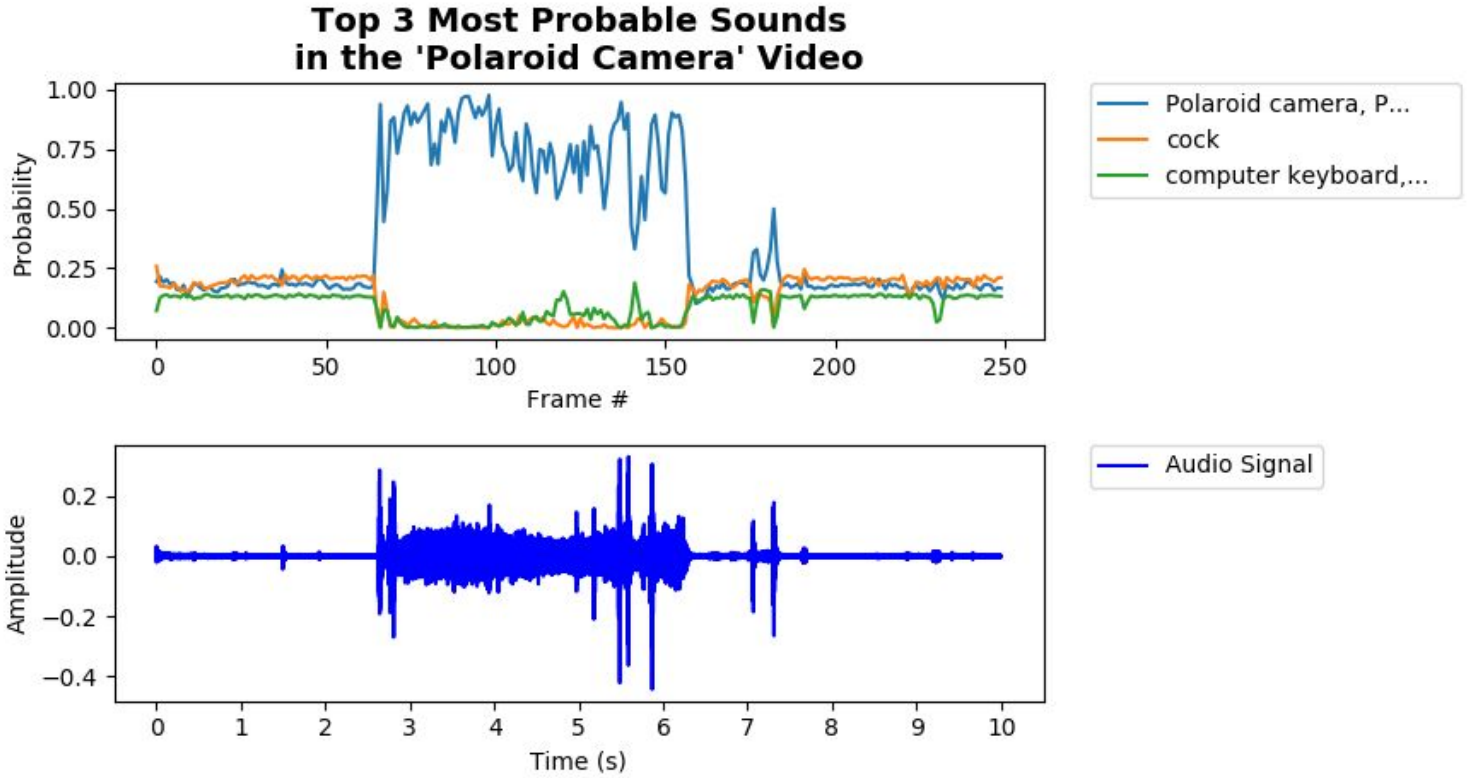




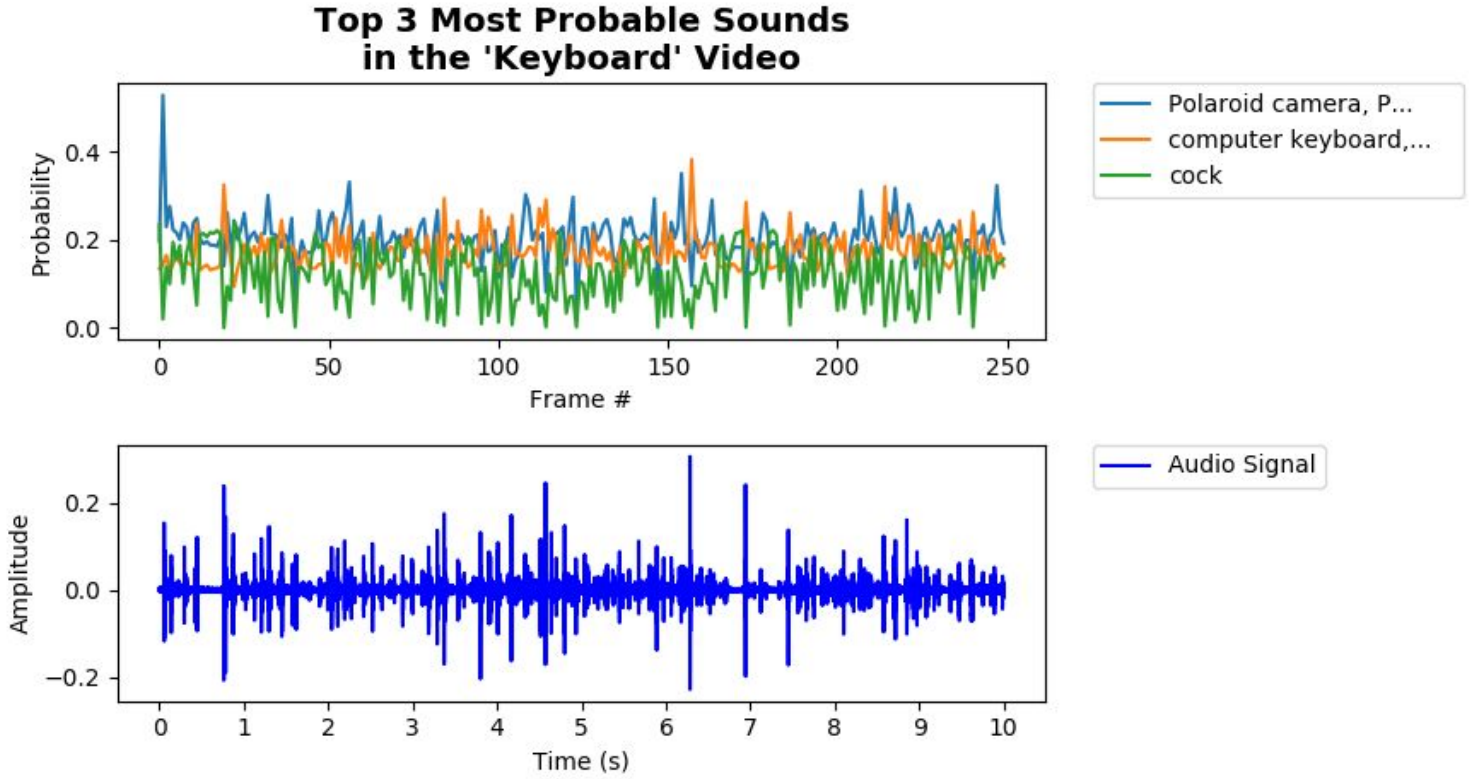
AudioNet Evaluation



AudioNet Evaluation



AudioNet Evaluation



Conclusion

{Final thoughts and the current state of SoundNet}

Thank you for listening!

Questions?