Radboud University Nijmegen

Faculty of Science

Convolutional Neural Networks applied to Keyword Spotting using Transfer Learning

Thesis in Automatic Speech Recognition
(LET-REMA-LCEX10)

Author:

Christoph Schmidl
s4226887
c.schmidl@student.ru.nl

Supervisor:

dr. L.F.M. ten Bosch

May 20, 2019

Contents

1 Introduction

1. Problem
2. Background (literature overview)
3. Research Question, Hypotheses, intro to experiment

1.1 Literature review

- Speech Commands: A Dataset for Limited-Vocabulary Speech Recognition [10]
- Speech Recognition: Keyword Spotting Through Image Recognition. [3]
- Convolutional neural networks for small-footprint keyword spotting [6]
- Small-footprint keyword spotting using deep neural networks [2]
- Convolutional recurrent neural networks for small-footprint keyword spotting [1]
- Honk: A PyTorch reimplementation of convolutional neural networks for keyword spotting [7]
- An experimental analysis of the power consumption of convolutional neural networks for keyword spotting [9]
- Transfer learning for speech recognition on a budget [4]
- Learning and transferring mid-level image representations using convolutional neural networks [5]
- Deep residual learning for small-footprint keyword spotting [8]

2 Method

1. methodology, types of analyses, selection of the method

3 Set-up

1. selection of the speech data, description of the data, tuning/adaptation model parameters
2. types of experiments (generalizations to which unseen conditions, etc.)

4 Experiments

5 Analysis and Results

6 Discussion

7 Conclusion

8 References

9 Appendix

aspect	experimental (max. points)	theoretical (max. points)
Research Question (RQ)	20	20
Literature embedding of the RQ	20	40
Method	20	
Justification experiment(s)	10	
Set-up experiment(s)	30	
Discussion and Conclusion	30	70
Use of figures and tables	10	10
Overall completeness	20	20
Overall clarity, transparency	20	20
Overall coherence (from intro to conclusion)	20	20
Total	*200*	*200*

Figure 1: Weighted grading

- the experiment(s) may be carried out in collaboration with others. In that case: specify in the "author's statement" everybody's contribution
- the thesis itself is written individually and assessed individually
- the ASR performance itself is not relevant for the assessment of the thesis
- the RQ, the literature embedding of the RQ, the description of the method, the justification and set-up of the experiment are relevant for the assessment
- the general university guidelines apply (e.g., with respect to plagiarism)
- there is no minimum number of pages for the thesis

10 Complex stuff

10.1 Domains

Let's start with the following definition:

Definition 10.1. A set $U \subseteq \mathbb{C}$ is a *domain* if:

- U is open in $\mathbb{C}$, and
- U is connected.

10.2 Yumyumyumyum

TO WRITE: an introduction and some examples

Theorem 10.2. *Suppose $n \in \mathbb{Z}$, then the following are equivalent:*

i. $n > 5$.

ii. $5 > 5$. This doesn't seem right...

iii. *For each $n \in n$, we have:*

$$n > n + 1 > n + 1^2 > \cdots > n + 7. \tag{1}$$

where 7 is an arbitrary element of

$$\oint_a^b \operatorname{supersin} \alpha + i \operatorname{supercos} \beta db(a).$$

Remark. Interesting!

Proof. See [**?**]. □

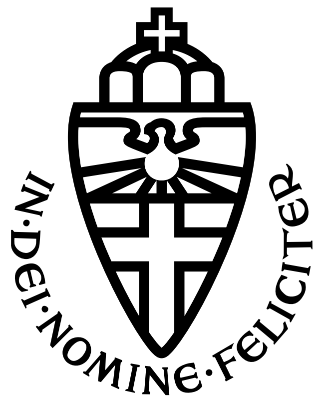

Figure 2: Motivational illustration. Similar to [**?**, **?**].

Corollary 10.2.1. *Suppose $U \subseteq \mathbb{C}$ is a domain (see Definition 10.1), and $f : \overline{U} \to \mathbb{C}$ is continuous on $\overline{U}$ and holomorphic on U. If $z \mapsto |f(z)|$ is constant on ∂U, then f has a zero in U.*

Proof. If not, consider $\frac{1}{f}$. □

The proof of this theorem is illustrated in Figure 2.

References

[1] Sercan O Arik, Markus Kliegl, Rewon Child, Joel Hestness, Andrew Gibiansky, Chris Fougner, Ryan Prenger, and Adam Coates. Convolutional recurrent neural networks for small-footprint keyword spotting. *arXiv preprint arXiv:1703.05390*, 2017.

[2] Guoguo Chen, Carolina Parada, and Georg Heigold. Small-footprint keyword spotting using deep neural networks. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4087–4091. IEEE, 2014.

[3] Sanjay Krishna Gouda, Salil Kanetkar, David Harrison, and Manfred K Warmuth. Speech recognition: Keyword spotting through image recognition. *arXiv preprint arXiv:1803.03759*, 2018.

[4] Julius Kunze, Louis Kirsch, Ilia Kurenkov, Andreas Krug, Jens Johannsmeier, and Sebastian Stober. Transfer learning for speech recognition on a budget. *arXiv preprint arXiv:1706.00290*, 2017.

[5] Maxime Oquab, Leon Bottou, Ivan Laptev, and Josef Sivic. Learning and transferring mid-level image representations using convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1717–1724, 2014.

[6] Tara Sainath and Carolina Parada. Convolutional neural networks for small-footprint keyword spotting. 2015.

[7] Raphael Tang and Jimmy Lin. Honk: A pytorch reimplementation of convolutional neural networks for keyword spotting. *arXiv preprint arXiv:1710.06554*, 2017.

[8] Raphael Tang and Jimmy Lin. Deep residual learning for small-footprint keyword spotting. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 5484–5488. IEEE, 2018.

[9] Raphael Tang, Weijie Wang, Zhucheng Tu, and Jimmy Lin. An experimental analysis of the power consumption of convolutional neural networks for keyword spotting. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 5479–5483. IEEE, 2018.

[10] Pete Warden. Speech commands: A dataset for limited-vocabulary speech recognition. *arXiv preprint arXiv:1804.03209*, 2018.