

Master's Thesis

Benchmarking Somatic Variant Callers

Ravi Shankar Chintalapati

Examiner: Prof. Dr. Rolf Backofen

Advisers: Wolfgang Maier, Mehmet Tekman

Albert-Ludwigs-University Freiburg

Faculty of Engineering

Department of Computer Science

Chair for Bioinformatics

April 29th, 2021

Writing period

29.10.2020 – 29.04.2021

Examiner

Prof. Dr. Bugs Bunny

Advisers

Wolfgang Maier, Mehmet Tekman

Declaration

I hereby declare, that I am the sole author and composer of my thesis and that no other sources or learning aids, other than those listed, have been used. Furthermore, I declare that I have acknowledged the work of others by providing detailed references of said work.

I hereby also declare, that my Thesis has not been prepared for another examination or assignment, either wholly or excerpts thereof.

Place, Date

Signature

Abstract

foo bar

Zusammenfassung

German version is only needed for an undergraduate thesis.

Contents

1	Introduction	1
1.1	Template Structure	1
1.2	setup.tex	2
1.3	Advice	3
2	Related Work	5
3	Background	7
4	Approach	11
4.1	Problem Definition	11
4.2	First Part of the Approach	11
4.3	N-th Part of the Approach	11
5	Experiments	13
6	Conclusion	15
7	Acknowledgments	17
	Bibliography	18

List of Figures

1	Structure of a siamese neural network	8
2	Caption that appears in the figlist	14

List of Tables

1	Table caption	13
---	-------------------------	----

List of Algorithms

1	Stochastic Gradient Descent: Neural Network	9
---	---	---

1 Introduction

This is a template for an undergraduate or master's thesis. The first sections are concerned with the template itself. If this is your first thesis, consider reading Section 1.3. Of course, the structure of this thesis is only an example. Discuss with your adviser what structure fits best for your thesis.

1.1 Template Structure

- To compile the document either run the makefile or run your compiler on the file 'thesis_main.tex'. The included makefile requires latexmk which automatically runs bibtex and recompiles your thesis as often as needed. Also it automatically places all output files (aux, bbl, ...) in the folder 'out'. As the pdf also goes in there, the makefile copies the pdf file to the parent folder. There is also a makefile in the chapters folder, to ensure you can also compile from this directory.
- The file 'setup.tex' includes the packages and defines commands. For more details see Section 1.2.
- Each chapter goes into a separate document, the files can be found in the folder chapters.
- The bib folder contains the .bib files, I'd suggest to create multiple bib files for different topics. If you add some or rename the existing ones, don't forget to also change this in thesis_main.tex. You can then cite as usual [1, 2, 3].
- The template is written in a way that eases the switch from scrbook to book class. So if you're not a fan of KOMA you can just replace the documentclass in the main file. The only thing that needs to be changed in setup.tex is the caption styling, see the comments there.

1.2 setup.tex

Edit setup.tex according to your needs. The file contains two sections, one for package includes, and one for defining commands. At the end of the includes and commands there is a section that can safely be removed if you don't need algorithms or tikz. Also don't forget to adapt the pdf hypersetup!!

setup.tex defines:

- some new commands for remembering to do stuff:
 - `\todo{Do this!}`: **(TODO: Do this!)**
 - `\extend{Write more when new results are out!}`:
(EXTEND: Write more when new results are out!)
 - `\draft{Hacky text!}`: **(DRAFT: Hacky text!)**
- some commands for referencing, 'in `\chapref{chap:introduction}`' produces 'in Chapter 1'
 - `\chapref{}`
 - `\secref{sec:XY}`
 - `\eqref{}`
 - `\figref{}`
 - `\tabref{}`
- the colors of the Uni's corporate design, accessible with `{\color{UniX} Colored Text}`
 - UniBlue
 - UniRed
 - UniGrey
- a command for naming matrices `\mat{G}`, **G**, and naming vectors `\vec{a}`, **a**. This overwrites the default behavior of having an arrow over vectors, sticking to the naming conventions normal font for scalars, bold-lowercase for vectors, and bold-uppercase for matrices.
- named equations:
`\begin{align}`

```

d(a,b) &= d(b,a)\\ \eqname{symmetry}
\end{align}

```

$$d(a, b) = d(b, a) \tag{1}$$

symmetry

1.3 Advice

This section gives some advice how to write a thesis ranging from writing style to formatting. To be sure, ask your advisor about his/her preferences.

For a more complete list we recommend to read Donald Knuth's paper on mathematical writing. (At least the first paragraph). http://jmlr.csail.mit.edu/reviewing-papers/knuth_mathematical_writing.pdf

- Don't use passive voice. It's harder to read, more likely to produce errors, and most of the times less precise. Of course there are situations where the passive voice fits but in scientific papers they are rare. Compare the sentence: 'We created the wheel to solve this.' to 'The wheel was created to solve this', you don't know who did it, making it harder to understand what is your contribution and what is not.
- If you use formulas pay close attention to be consistent throughout the thesis!
- Usually in a thesis you don't write 'In [24] the data is..'. You have more space than a paper has, so write 'AuthorXY et al. prepare the data... [24]'. Also pay attention to the placement: The citation is at the end of the sentence before the full stop with a no-break space. ... last word~\cite{XY}.
- Pay attention to comma usage, there is a big difference between English and German. '...the fact that bla...' etc.
- Do not write 'don't ', 'can't' etc. Write 'do not', 'can not'.
- If an equation is at the end of a sentence, add a full stop. If it's not the end, add a comma: $a = b + c$ (1),
- Avoid footnotes if possible.
- Use ‘‘’’ for citing, not "".

2 Related Work

Give a brief overview of the work relevant for your thesis.

3 Background

Explain the math and notation.

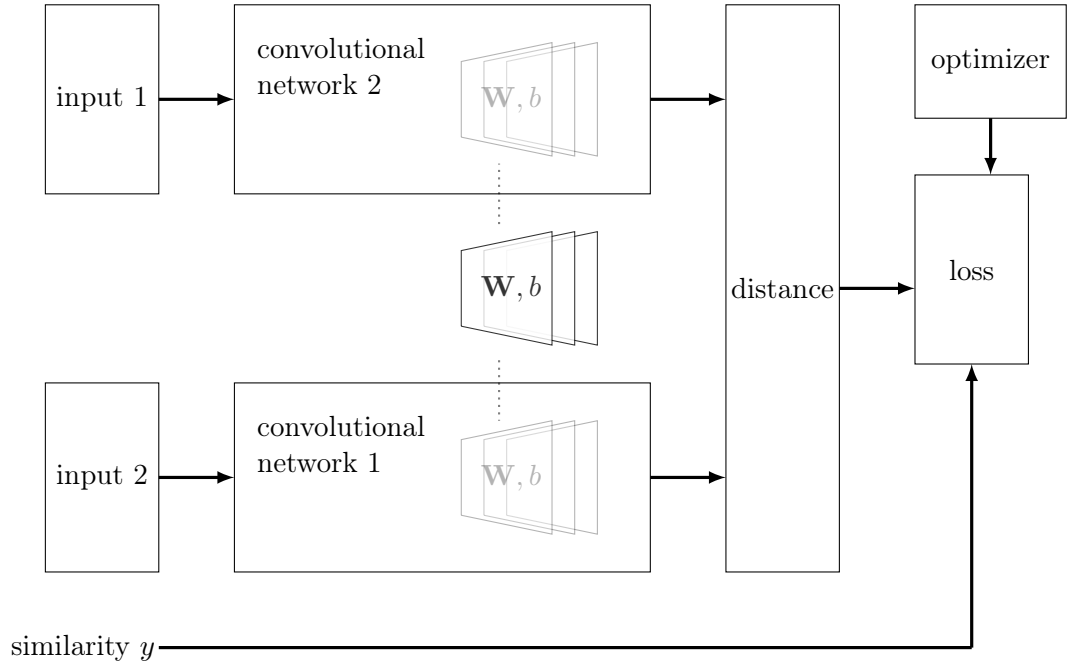


Figure 1: Structure of a siamese neural network. The two convolutional networks share the weights \mathbf{W} and biases b . The distance between the output vectors of the two networks is fed into the the contrastive loss function. The contrastive loss penalizes small or large distances, depending on the similarity label y .

Algorithm 1 Stochastic Gradient Descent: Neural Network

Create a mini batch of m samples $\mathbf{x}_0 \dots \mathbf{x}_{m-1}$

foreach sample \mathbf{x} **do**

$\mathbf{a}^{\mathbf{x},0} \leftarrow \mathbf{x}$ ▷ Set input activation

foreach Layer $l \in \{1 \dots L-1\}$ **do** ▷ Forward pass

$\mathbf{z}^{\mathbf{x},l} \leftarrow \mathbf{W}^l \mathbf{a}^{\mathbf{x},l-1} + \mathbf{b}^l$

$\mathbf{a}^{\mathbf{x},l} \leftarrow \varphi(\mathbf{z}^{\mathbf{x},l})$

end for

$\delta^{\mathbf{x},L} \leftarrow \nabla_{\mathbf{a}} C_{\mathbf{x}} \odot \varphi'(\mathbf{z}^{\mathbf{x},L})$ ▷ Compute error

foreach Layer $l \in L-1, L-2 \dots 2$ **do** ▷ Backpropagate error

$\delta^{\mathbf{x},l} \leftarrow ((\mathbf{W}^{l+1})^T \delta^{\mathbf{x},l+1}) \odot \varphi'(\mathbf{z}^{\mathbf{x},l})$

end for

end for

foreach $l \in L, L-1 \dots 2$ **do** ▷ Gradient descent

$\mathbf{W}^l \leftarrow \mathbf{W}^l - \frac{\eta}{m} \sum_{\mathbf{x}} \delta^{\mathbf{x},l} (\mathbf{a}^{\mathbf{x},l-1})^T$

$\mathbf{b}^l \leftarrow \mathbf{b}^l - \frac{\eta}{m} \sum_{\mathbf{x}} \delta^{\mathbf{x},l}$

end for

4 Approach

The approach usually starts with the problem definition and continues with what you have done. Try to give an intuition first and describe everything with words and then be more formal like ‘Let g be ...’.

4.1 Problem Definition

Start with a very short motivation why this is important. Then, as stated above, describe the problem with words before getting formal.

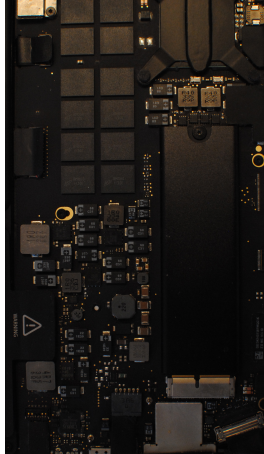
4.2 First Part of the Approach

4.3 N-th Part of the Approach

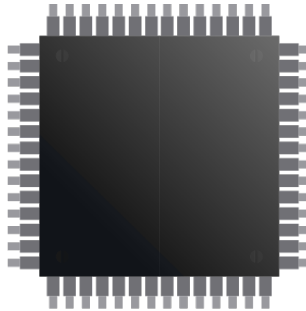
5 Experiments

Type	Accuracy
A	82.47 ± 3.21
B	78.47 ± 2.43
C	84.30 ± 2.35
D	86.81 ± 3.01

Table 1: Table caption. foo bar...



(a) Some cool graphic



(b) Some cool related graphic

Figure 2: Caption that appears under the fig This plot shows...

6 Conclusion

7 Acknowledgments

First and foremost, I would like to thank...

- advisers
- examiner
- person1 for the dataset
- person2 for the great suggestion
- proofreaders

Bibliography

- [1] D. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *arXiv preprint arXiv:1412.6980*, 2014.
- [2] J. Bromley, J. W. Bentz, L. Bottou, I. Guyon, Y. LeCun, C. Moore, E. Säckinger, and R. Shah, “Signature verification using a “siamese” time delay neural network,” *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 7, no. 04, pp. 669–688, 1993.
- [3] M. Muja and D. G. Lowe, “Fast approximate nearest neighbors with automatic algorithm configuration.,” *VISAPP (1)*, vol. 2, no. 331-340, p. 2, 2009.

