Reviewer form[1]

Seminar: „Software Engineering“, Semester WS 23/24
Prof. Dr.-Ing. Samuel Kounev, Lehrstuhl für Informatik II, Software Engineering,
Institut für Informatik, Universität Würzburg

General data

Name of the reviewer:Nicoleta Ciausu

Title of the peer-reviewed article: Making Neural Networks Smaller: An Overview of existing approaches

Name of the author of the peer-reviewed article: Rohit Sebastian Peter

Date: 11. Januar 2024

Summary of the peer-reviewed text

The seminar paper describes methods for making neural networks smaller, as well as

Rating

Please enter a rating in the table according to the following scale. This scale is based on the typical English review classification: 1 = accepted, I would argue for it (strong accept and I would argue for it); 2 = tend to accept (weak accept); 3 = tend to reject (weak reject); 4 = rejected, I would argue for it (strong reject and I would argue for it).

In order to standardize the evaluation criteria, some questions are given in each case that should be considered in the evaluation. In addition, you must refer to our thesis writing guidelines[2] and the guidance on academic papers[3] on our homepage as evaluation criteria. Please justify your evaluation on the next page in the free text fields.

[1]Please fill in all fields marked with „[TODO]“ completely.
[2]https://se.informatik.uni-wuerzburg.de/software-engineering-group/teaching/guidelines-fuer-das-schreiben-von-abschlusarbeiten/
[3]https://se.informatik.uni-wuerzburg.de/software-engineering-group/teaching/wissenschaftliches-schreiben/

Criteria	Rating (number)
Abstract Is the choice of words in the abstract appropriate? Is the level of abstraction appropriate? Does the summary clearly indicate what to expect in the article? Is the summary understandable without having read the article?	2
Formal aspects Does the document have the required length and formatting? Does the document contain spelling or grammatical errors?	2
Outline / Structuring Does the introduction explain the outline? Are the word choices the level of abstraction appropriate for headings? Are terms explained before they are used? Are headings and consistent with section content?	2
Writing style and readability How well does the article use technical terms, equations, pseudocode, figures, and tables? Is the argumentation clear, comprehensible, and understandable? Is the text too prosaic, unscientific, or overly scientific? Does the comprehensibility suffer from a pseudo-scientific style? Are sentences and paragraphs of an appropriate length?	3
Reference use Are references provided when they are necessary? Are the references given relevant? Are verbatim text passages marked as citations?	2
Bibliography Are all entries of the bibliography complete and correct? Would references be mentioned that are missing in the article? Are all critical statements supported by a lliteraturerreference?	3
Level of detail Is the presentation of background=concepts appropriate for the intended readers? Does the paper maintain an appropriate balance between technical details and „high-level“-concepts? (Expected to be an overview article). Are the breadth and level of detail appropriate for an article of this length?	2
Completeness Is the text complete?	1
Conclusion / Bottom line Is the content of the article adequately summarized? Is the content presented reflected (critically) in the concluding part?	1
Overall rating	2

Explanations

In this section, provide detailed reasons for your evaluation above and constructive suggestions for improvement, if any.

Abstract The abstract does a good job of succintly presenting the reasons why researchers would want to make neural networks smaller. It, however, does not go further in describing what the paper goes into, as well as the results it obtains.

The abstract could be improved by:

- Improving inconsistent capitalization and punctuation
- Rephrasing of circular sentences ("the need to make DNNs work with limited resources has become a need")
- extending the abstract to also include details about the paper's content, not just the motivation behind it

Formal aspects The document has the required length and formatting. The bibliography is not separated into its own section.

The document contains a non-negligible amount of spelling, capitalization, punctuation and spacing errors, spanning all the way from the abstract to the conclusion (the very chapter header is mispelled as Conlusion). No grammar errors were noted. I would strongly suggest use of a spelling checker to improve the quality of the writing.

Language and writing style: The language used throughout the paper is good, leaning neither too scientific nor too colloquial.

Changes to the structure of the paper could bring the biggest improvement to the work. As it is written currently, following the author's train of thought is quite difficult to do. The paper jumps directly from the introduction (which provides background for the reasons these techniques exist) to in-depth descriptions of the techniques that are interwoven with examples and performance ratings and improvements. The information itself is good. However, it would benefit greatly from some reorganization.

Some suggestions:

- The paragraphs themselves are long and could benefit from splitting whenever the main idea changes. This would help the reader prepare for a shift in the train of thought and make going through the content easier.
- I think the paper would benefit from a better separation between describing the techniques themselves and the experiments done and the results that were obtained. There are a number of ways this could be done, one example (the way I would do it) is, for each technique, have two sub-headings, "Explanation" and "Experiments". By splitting the existing content as such it would be a lot easier for a reader that is not so familiar with Machine Learning to focus their reading on which part they are more interested in (understanding the techniques themselves vs. the benefits in speed they bring)
- The actual examples and figures chosen are good and relevant for the purpose of the paper.

- A very good summarisation and explanation of the techniques is done in the Conclusion chapter. While the content itself is good, I would argue that the conclusion is not the right place to put it, being more suited for the Introduction (where these four techniques are just listed). As it is right now, the paper jumps from naming the techniques to explaining them in great detail, thus breaking the Pyramid Principle. It would be great if an intermediate, higher-level description of these techniques would exist between the introduction and the main content.

Bibliography/Reference use Citation style is unfortunately unsatisfactory. While citations exist, they are only referenced throughout the paper in the beginning of paragraphs or sporadically across the paper, the most common formulation being "In [citation number], [rest of paragraph]". This makes it unclear whether the text that follows is paraphrasation, summarisation or the author's own interpretation of the cited sources. No figure used throughout the paper has its source cited. The bibliography heading is also missing. Citation style in the bibliography seems to have been done by hand rather than using BibTeX or a similar solution, and is inconsistent in style, sometimes referencing year in parantheses, othertimes not, sometimes offering DOI, othertimes not (even though the cited work has a DOI number). Overall, the work needs significant revamping of its citations, both in style and thoroughness - making it clear through the paper which statements and figures are from the referenced material and which are the author's contribution.

Level of detail: The level of detail is generally appropriate for an overview-style article. More emphasis could be put on how the techniques actually work and less on the experiments done and their results. While a good amount of background concepts are explained, it would be really useful to also write an introduction for the types of neural networks used throughout the paper (CNN, DNNs, and others). As it currently stands, they are not introduced at all even though they are essential background knowledge required for understanding this work.

Strongest aspect(s) of the paper: The paper collects good, valuable information on the approaches for making neural networks smaller. The actual high-level explanations for the techniques used are well done.

Weakest aspect(s) of the paper: The paper has a lot of good information that unfortunately suffers to be delivered to the reader due to the writing style. As an overview-type work, the paper lacks the "term definition" and "introduction" stages that would be essential in guiding the reader to understanding the approaches.

Miscellaneous

Comments that cannot be assigned to any other category.

- Images are small and quite hard to read, especially Fig 4