

Prof. Dr.-Ing. Markus Mayer • THD • Postfach 13 20 • 94453 Deggendorf

Prof. Dr. Patrick Glauner
Dieter-Görlitz-Platz 1
94469 Deggendorf

Telefon-Durchwahl
Tel.: +49 991 3615-
8324

E-Mail
markus.mayer@th-deg.de

Ort, Datum
Landshut,
August 23, 2023

Opinion on the Master Thesis “Evaluation of self-supervised learning for vehicle-type detection in autonomous driving” by Ahmed Manaf

Dear colleagues!

In the following, I present my opinion on the Master Thesis of Ahmed Manaf, with a grade suggestion at the end. The opinion, with positive and negative factors that influence the result, is divided into three sections: Form, motivation and embedding in the scientific field, and method and discussion:

Form:

- Positive: The language used adheres to scientific standards, with minor exceptions, e.g. punctuation errors and broken sentences (pg. 16, 18, 23, 36, 38), mixing of upper/lower cases (pg. 31).
- Positive: The structure of the thesis is well thought.
- Negative: Captions of figures are not self-explanatory and need the full text for to reach an understanding. In addition, the figures of the scientific review seem to be redrawn from the original sources (which is good), but again not transformed into a common system (horizontal vs. vertical network structures, sharp vs. rounded edges....)

Motivation and embedding in the scientific field:

- Positive: The purpose and motivation for the work are well reasoned.
- Positive: The literature overview seems to have captured a multitude of related works, while still being compact.
- Negative: The summaries of the referenced papers and methods have no interconnection to each other and read like abstracts, containing all the buzzwords, but lacking an embedding structure. This can be pointed down to three examples: First, important wordings and algorithmic differentiations like contrastive/non-contrastive, pretext-task, momentum-based are not explained on their own, giving the reader an impression of the SSL

Technische Hochschule
Deggendorf
Dieter-Görlitz-Platz 1
94469 Deggendorf

Tel.: +49 991 3615-
8324
www.th-deg.de

field firsthand, before diving into details. Second: While figures of the NN structures of the presented methods are given, there are no explanations, beside the buzzwords, how the methods actually differ from each other (e.g. precise key ideas or maths). Compare “*The basic idea of SimCLR is to learn representations that are invariant to different augmentations of the same image. This is achieved by using a contrastive loss function that maximizes the similarity between positive pairs and minimizes the similarity between negative pairs*” (pg. 20) and “*The basic idea of MoCo is to learn a momentum-based contrastive loss function that maximizes the similarity between positive pairs and minimizes the similarity between negative pairs.*” (pg. 22). Third: Drawbacks of various methods (eg. Template/Object based detection) just repeat each other (word by word), without summarizing them into one general statement.

In total, this leaves the impression of a lack of in-depth understanding of the mathematical relations.

Method and discussion:

- Positive: Well thought structure of the evaluation, with reasoning given for the design decisions.
- Positive: The given figures and tables allow the reader to reassess the conclusions for him/herself, but still remain compact. The balance between details and a compact view is very well met.
- Negative: The dataset descriptions are too vague, i.e. there is no summary and clear denomination at the end of the data section which datasets contain which images (and how many for each class), which were used for initial training, and which for transfer learning etc. This can be only read from the context, with clear references only seldom written down.
- Negative: There is no clear reasoning given, why out of the many methods presented in the literature overview SimCLR and SimSiam were chosen for the evaluation.
- Error: A comparison is made in the evaluation based on the loss measure used in the training process (section 4.2.1). However, this loss measure is 1) different for the network structures 2) based on vectors of different length (128 vs. 2056). It is clear that a direct comparison is therefore meaningless. The student is not aware of this implication, which is clear by a formulation such as “despite its lower training loss...”.
- Error: ROC curves are shown and Au-ROC is computed. But the ROC curves contain only a single evaluation point, which makes them (and the Au-ROC measure) no more informative than giving recall/precision measures alone (and render them therefore useless in their presented form).

Overall, I was impressed with the very good structure of the work, and the very well designed evaluation. But given the negative factors mentioned above (especially the lacking interconnections in the literature overview, and the unprecise description and usages of the datasets) and the methodical errors in the evaluation metrics, my recommendation for a grade is 2.0 - good.

Best regards,

Prof. Dr.-Ing. Markus Mayer