THESIS REVIEWER'S REPORT

I. IDENTIFICATION DATA

Thesis title:	**Semi-Supervised Learning for Spatio-Temporal Segmentation of Satellite Images.**
Author's name:	**Antonín Hruška**
Type of thesis:	master
Faculty/Institute:	Faculty of Electrical Engineering (FEE)
Department:	Department of Cybernetics
Thesis reviewer:	Dmitrij Schlesinger
Reviewer's department:	Institute of Software and Multimedia Technology, Dresden University of Technology.

II. EVALUATION OF INDIVIDUAL CRITERIA

Assignment — **challenging**

How demanding was the assigned project?

One of the thesis goals was to apply a novel method (symmetric learning). Hence, the proposed models (network architectures, learning schemes, etc.) should be designed and implemented from scratch as there is no similar implementations one can start with or use as baselines.

Fulfilment of assignment — **fulfilled with minor objections**

How well does the thesis fulfil the assigned task? Have the primary goals been achieved? Which assigned tasks have been incompletely covered, and which parts of the thesis are overextended? Justify your answer.

It seems (according to the title) that originally it was planned to apply the method to satellite images. However, in the work everything is done for a standard benchmark dataset only. Indeed, this is acknowledged in the thesis and motivated by the need to perform extensive evaluation, i.e., a large dataset with labeled data is necessary. Nevertheless, a proof-of-concept small-scale experiment for satellite images would be desirable. Besides, some additional experiments would be useful to better illustrate the proposed approach (see below).

Methodology — **correct**

Comment on the correctness of the approach and/or the solution methods.

The thesis is based on already known theoretical concepts. Concerning this, everything seems correct.

Technical level — **C - good.**

Is the thesis technically sound? How well did the student employ expertise in the field of his/her field of study? Does the student explain clearly what he/she has done?

There are some aspects which need a more detailed explanation, some additional experiments would be highly appreciated (see below).

Formal and language level, scope of thesis — **B - very good.**

Are formalisms and notations used properly? Is the thesis organized in a logical way? Is the thesis sufficiently extensive? Is the thesis well-presented? Is the language clear and understandable? Is the English satisfactory?

There are just some minor typos. Otherwise, the work is good structured and clearly written. Formalisms and notations are consequently introduced and properly used.

Selection of sources, citation correctness — **C - good.**

Does the thesis make adequate reference to earlier work on the topic? Was the selection of sources adequate? Is the student's original work clearly distinguished from earlier work in the field? Do the bibliographic citations meet the standards?

Some parts of section 2 (State of the art of SSL) are somewhat redundant. For example, it is questionable whether the connection between VAEs and Calculus of Variations is really necessary for the thesis. Moreover, such an extensive discussion about standard VAEs is also not really needed since ELBO maximization is not further used.

Additional commentary and evaluation (optional)

Comment on the overall quality of the thesis, its novelty and its impact on the field, its strengths and weaknesses, the utility of the solution that is presented, the theoretical/formal level, the student's skillfulness, etc.

See below.

III. OVERALL EVALUATION, QUESTIONS FOR THE PRESENTATION AND DEFENSE OF THE THESIS, SUGGESTED GRADE

Summarize your opinion on the thesis and explain your final grading. Pose questions that should be answered during the presentation and defense of the student's work.

There is a couple of specific questions:

Unfortunately, the final architecture used for symmetric learning is not explained to the necessary level of detail. Basically, it is only said that both encoder and decoder have a U-Net like architecture. Note however that they share parameters, i.e., the encoder uses decoder parts in order to e.g., generate from $q_{\theta,\phi}(z|x)$. It is not entirely clear how it works. Are the stochastic variables z_i attached to all resolution levels? Are they attached to both U-Net branches (of decreasing and increasing spatial resolutions)? Is there a skip-connection also between blocks of the original resolution? How many resolution levels there are? So, a more detailed explanation as well as a figure illustrating the architecture and generating / learning process would be highly appreciated.

An important question is the dimensionality of the latent z_i-s, especially for l (which is a part of z_0). Note that the segmentation alone does not include any coloring information, like segment colors (or colors of objects / instances), textures, shadows etc. Hence, in order to generate realistically looking images, such information should be encoded by the latent variables, in particular by l. If its dimension is low, it is obviously not capable to represent this information adequately (btw. if so, perhaps it can explain why the reconstructed images are bad). If, however the dimension of l is high, one has a gigantic input tensor for the decoder (due to replication of l along spatial dimensions), which obviously causes certain technical problems.

Eq. (3.3) is not entirely correct. In the second addend it should be $q_{\theta,\phi}(z|x)$ (i.e., including l and s) instead of $q_{\theta,\phi}(z_{>0}|x)$. It is not crucial if it is just a typo. If, however it corresponds to the implementation, it is a serious bug, because then it would mean that $q_{\phi}(l|x)$ is not learned at all.

Concerning experiments. It is somewhat surprising that there are no "baseline" experiments for symmetric learning like it was done for Mixmatch. Moreover, such baseline experiments could be designed in different ways. For example, one could just learn the segmentation model $q_{\phi}(s|x)$ on fully supervised training data of different sizes. Next, one can apply the "full" symmetric learning according to eqs. (3.2), (3.3) but again on fully supervised training data of different sizes only. Hopefully, comparing these two experiments one can observe some improvement, because additional terms in eqs. (3.2), (3.3) should serve as a regularizer for the segmentation model, and hence improve generalization capabilities. Finally, comparing the final experiments (which are present in the work) with fully supervised symmetric learning, one can draw conclusions about the applicability of the symmetric learning for SSL.

Besides, it would be also interesting to compare symmetric learning with ELBO maximization for the proposed architecture, chosen dataset and considered learning scenarios (supervised, SSL, etc.).

Despite the above criticism it should be noted that the thesis deals with a novel direction in the area of deep generative modeling. The proposed approach, although not perfect yet, is promising and may be a good starting point for further developments.

The grade that I award for the thesis is **B - very good.**

Date: **3.6.2023**

Signature: