THESIS REVIEWER'S REPORT

I. IDENTIFICATION DATA

Thesis title:	**Dance genre recognition from a video of a dancing pair**
Author's name:	**Štěpán Křivánek**
Type of thesis :	bachelor
Faculty/Institute:	Faculty of Electrical Engineering (FEE)
Department:	Department of Cybernetics
Thesis reviewer:	Mgr. Matěj Hoffmann, Ph.D.
Reviewer's department:	Department of Cybernetics

II. EVALUATION OF INDIVIDUAL CRITERIA

Assignment **ordinarily challenging**

How demanding was the assigned project?

The work builds on previous work of Petr Kouba (MSc. thesis 2021) and much of the infrastructure – datasets in particular – was already in place.

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 is not clear to what extent the point "2. Collect a suitable set of data for both training and testing" was fulfilled. It seems that mostly existing datasets or those used by Kouba (2021) were used. "the very same datasets as used in [1]"

"4. …The method should aim at recognizing the dance from the broadest set of viewpoints. " This point may not have been specifically addressed.

Methodology **partially applicable**

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

The student used a pipeline consisting of several modules based on deep neural networks. Compared to his predecessor (Kouba 2021), he replaced some modules by alternatives (MMPose/HRNet instead of OpenPose for human keypoint extraction and PoseC3D instead of MS-G3D for action recognition). Their choice is not sufficiently explained or motivated, although there are some references to benchmarks.

The schematics in Fig. 4.3 is unclear.

As the names of the action recognition methods suggest (PoseC3D, MS-G3D), they seem to require 3D skeletons on input. Both OpenPose and HRNet normally output 2D keypoints. How are the 3D keypoints obtained?

It is not clear which of the modules in the pipeline were trained by the student.

As admitted by the student, the downside of the method he applied is that the dancing pair is identified based on its size and movement w.r.t. the camera, which is assumed to be static.

Technical level **E - sufficient.**

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?

The student followed up on the work of Kouba (2001), but failed to even exactly replicate his results.

"For a clear comparison we rerun even the pipeline of Petr Kouba again, using the settings that achieves the best results in his work [1]. It achieves slightly worse results than presented in his work. But still decent enough so we do not consider it as an error." (pg. 23)

To my knowledge, "HRNet" is a type of neural network architecture. What was the (pretrained?) model that was used exactly?

There is a lot of evidence in the thesis that the experiments were performed in a rush. Some credit should be given to the student for being honest, but phrases like the examples below are not acceptable for a final thesis and reveal the low quality of the work.

"unfortunately during the experimenting we discovered a bug, which resulted into running all the experiments on a single GPU nd since we already spent a lot of time on these experiments we proceeded to finish the experiments" (pg. 19)

"we realized we did not split the Dance Tutorials Dataset videos by 300 frames, which drastically lowered the results obtained. However, it should not change much about the previous observations and therefore we stick with those results. Since we also already wasted a lot of time and computational power, we still provide some more results" (pg. 23)

Formal and language level, scope of thesis **E - sufficient.**

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?

The work is following up on Kouba (2021), but should still be more self-contained. Without knowing the thesis of Kouba, the reader has difficulty understanding the work – this is apparent already in the abstract.
The Chapters have unconventional names like "Objects of interest" or "Errors".
Table 3.1 results shows some "numbers" – the caption should explain what they are.
The text is full of mistakes and informal language.
Reoccuring errors: "tho" instead of though, "quiet" instead of quite, "depict in" instead of depicted in…
Example of vague and informal formulation: "There also is a few datasets, which we did not use directly during our training or testing, but they were used for pre-training various parts of the model and therefore they deserve to be mentioned as well. Most of them will probably be well known to everyone, who has ever gotten in touch with the action recognition before." (pg. 5)
"which may not seem too long ago, but as we already saw it kind of is in this field"

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?

The thesis mostly refers to github repositories and only few articles are cited. The thesis lacks a proper problem definition and theoretical grounding for the software modules used.

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

Overall, hasty, preliminary results are reported in this thesis.

The grade that I award for the thesis is **E - sufficient.**

Questions for the defense:

1) Can you explain Fig. 4.2 and which modules were trained in your work?
2) Do the action recognition methods (PoseC3D, MS-G3D) need 3D skeleton on input? Where is that obtained?

Date: **1.9.2022** Signature: