Posudek diplomové práce

Matematicko-fyzikální fakulta Univerzity Karlovy

Autor práce	Bs. Yuu Sakaguchi
Název práce	Intelligent Interior Design – Style Compatibility of 3D Furniture Models using Neural Networks
Rok odevzdání	2020
Studijní program	Informatika **Studijní obor** Umělá inteligence

Autor posudku	Mgr. Jakub Střelský **Role** oponent
Pracoviště	Katedra softwaru a výuky informatiky

Text posudku:

This thesis examines the problem of determining style similarity of 3D objects. The first chapter describes the background topics such as neural networks and metric learning. In the second chapter, the author describes prior work relevant to his thesis. The described prior work concerns learning style similarity of 3D objects using multi-view representation. In the third chapter, the author introduces his methods for solving the given problem. The proposed methods are based on point cloud representation of the objects. The fourth chapter provides a description of datasets and evaluation methods which were used to compare prior work with the proposed methods. The final chapter contains summary of the thesis and outlines possible future directions of work.

The topic of the thesis is non-trivial, and has several real applications. It seems that methods based on point cloud representation were not explored before in the setting of style similarity, so the contribution of the thesis is apparent. One of the main strengths of the thesis is that relevant prior work was studied extensively. Comparison of proposed methods with prior work was difficult, because it was impossible to determine the exact datasets which were used to report the results of previous papers. No official implementation of the papers was available either. The author even attempted to contact the authors of previous papers in order to get the necessary information but he did not receive any answer. The author attempted to replicate the results of prior work. Even though the quantitative results of authors replication differ substantially from the original results, the obtained results are still relevant for comparison of proposed methods with the replicated baselines. The author gives an explanation of possible reasons for the differences. The proposed point cloud based method seems competitive with the re-implemented multi-view baselines, which is a positive result. The evaluation was performed on relevant datasets and the designed experiments are fair and reasonable. The results would be more convincing if the replication of baselines was more succesful, but the author put a lot of effort into that and, as stated above, it seems that

information which is available from the prior work is insufficient.

The text is well written and understandable, but some parts (especially the introductory ones) are described vaguely or without proper care. For example, on page 22, the author states that voxel grid requires $\mathcal{O}(n^3)$ space, but it is not clear from the text what the quantity n represents. Also, on page 22, the statement "That increases the complexity cubically with respect to the number of voxels." is not correct. The cubical complexity is with respect to the value of grid resolution in one of the spatial dimensions. Another example is on page 7, where the sentence "The sigmoid functions are only sensitive to changes of input values around 0.5 because any values much larger than 1 or much smaller than 0 are converted close to 1 or 0." is very confusing because the values 0.5, 0, and 1 are most likely referring to the output values of the sigmoid function, instead of the input values. But the second part of the sentence does not make sense unless the author actually meant the input values.

The terminology in the introductory parts should also be used more carefully. For example, on page 10, the equations 1.9 and 1.10 do not really describe the backpropagation algorithm as the author states, but rather a step of a gradient descent algorithm where the gradients are usually obtained by the backpropagation algorithm. The thesis contains only very small number of typographic errors such as "out model" on page 42.

Despite the minor aforementioned flaws of the thesis I consider it successful and I definitively recommend it for the defense.

My potential questions for the defence arise from the sentence: "It should be noted that for building and furniture we fed input point cloud without jitter and rotation for the following reason: since the datasets allow overlapping samples between training and test sets, the extracted features from the overlapping samples would be exactly the same." from page 53. Does this imply that other test datasets were also rotated and jittered? If so, was it also the case for the multi-view baselines? Are the objects from the datasets presented in the thesis manually aligned into canonical orientation, or the objects appear in an arbitrary orientation (before the step of applying random rotation)?

Práci doporučuji k obhajobě.

Práci nenavrhuji na zvláštní ocenění.

V Praze dne 27. 1. 2020

Podpis: