DIARY

Thesis at CVG lab



**28.03.2024**



First meeting with Zuria. See **paper notebook**.

**03.04.2024**

Email from Zuria:

Hello Gaia,

I hope you had a nice Easter break.

Apologies for not sending you the email before, but I just totally forgot.

Here is an overview of everything we discussed last week:

Keep in mind that most of the scene graph project have a different end goal than your project is going to have, so they may or may not fit entirely, but it is to give you an overview and understanding of what people are doing in this field:

ConceptGraphs: scene graphs with open vocabulary. Paper: <https://concept-graphs.github.io/assets/pdf/2023-ConceptGraphs.pdf> Code: <https://github.com/concept-graphs/concept-graphs>

* Hydra: scene graph generation/ Paper: <https://www.roboticsproceedings.org/rss18/p050.pdf> Code: <https://github.com/MIT-SPARK/Hydra>
* SGAligner: Scene graphs to align 3D point clouds. Paper: <https://arxiv.org/pdf/2304.14880.pdf> Code: <https://github.com/sayands/sgaligner>
* 3D-Scene graph: Paper: <https://openaccess.thecvf.com/content_ICCV_2019/papers/Armeni_3D_Scene_Graph_A_Structure_for_Unified_Semantics_3D_Space_ICCV_2019_paper.pdf> Code: <https://3dscenegraph.stanford.edu/>
* Kimera: <https://arxiv.org/pdf/2101.06894v3.pdf>
* 3D Dynamic Scene Graphs: <https://arxiv.org/pdf/2002.06289.pdf>
* Dynamic Scene Graph Generation via Anticipatory Pre-training: <https://openaccess.thecvf.com/content/CVPR2022/papers/Li_Dynamic_Scene_Graph_Generation_via_Anticipatory_Pre-Training_CVPR_2022_paper.pdf>

For the semantic scene understanding:

* Mask3D: it does semantic segmentation on Point Clouds.<https://jonasschult.github.io/Mask3D/>

Regarding the detection of dynamic objects in the scene:

* <https://arxiv.org/abs/2307.07635> ⇾ a very good and recent method to track how something moves in a video
* <https://github.com/shikharbahl/vrb> ⇾ they use a very simple mechanism to track what a hand is doing in videos. This can be useful to check if it works for your case as well, to identify which objects are currently manipulated
* <https://arxiv.org/abs/2210.03105> this is a 3D instance segmentation method that we are using a lot

The dataset for your first scene graph generation we chatted about is Hypersim:

* Hypersim: <https://github.com/apple/ml-hypersim>
* For the data: If you have a list of corresponding images, depth maps, and poses, here is a piece of code to integrate them with open3d into a 3D map: <https://github.com/cvg/LabelMaker/blob/main/scripts/utils_3d.py>

And I would say that is enough content for now. Keep in mind that I don't expect you to go over everything. This is really to give you a general overview of what is out there so that you can also check their codes and how they did it.

To find more papers I recommend searching on Google Scholar in the cited section of papers that you think could be related, there you can see what papers were published afterwards and maybe improve on the before achieved results. Another option is Connected Papers: <https://www.connectedpapers.com/>.

Could you provide me with 2-3 possible dates to meet every week so that I can see how I can fit our meetings in my calendar?

If you have any further questions, please let me know 😊

Best,

Zuria

**14.04.2024**

Setting things up (i.e., understand where to put what, organisation).

Starting downloading papers suggested by Zuria.

**15.04.2024**