COMP9517: Computer Vision

2024 Term 2

Group Project Specification

Maximum Marks Achievable: 40

The group project is **worth 40% of the total course mark**.

Project work is in Weeks 6-10 with deliverables due in Week 10.

Deadline for submission is Friday 2 August 2024 18:00:00 AET.

Instructions for online submission will be posted closer to the deadline.

Refer to the separate marking criteria for detailed information on marking.

Introduction

The goal of the group tudents to solve a computer vision prob form.

Group members can ks 6-10 during the usual consultation se

The group project is copy ideas or any materials from other re for some of the tasks, these must be giarism and will be penalised according t

Note that high marks ot used before for the project task. We but the more you use existing code (wh pect you to show creativity and build o terature.

Description

In order for autonomous vehicles to navigate safely and accurately in natural environments, it is important that they are able to recognise the different types of scenarios and objects they may encounter along the way. For example, a vehicle may need to proceed more cautiously when travelling through sand or mud, or when there are many trees around, than when driving over gravel or asphalt in a clear area, while water must be avoided at all times. Compared to urban environments, perception in natural environments is more challenging, as these generally contain highly irregular and unstructured elements.

The first step toward comprehensive scene understanding is to perform fine-grained semantic segmentation of the images captured by the vehicle's cameras. That is, to assign a label to each and every pixel in the images, indicating to which class that pixel belongs.

Task

The goal of this group project is to develop and compare different computer vision methods for semantic segmentation of images from natural environments.

Dataset

The dataset to be used in the group project is called WildScenes (see links and references at the end of this document). This is a recently released multimodal dataset consisting of five sequences of 2D images recorded with a normal video camera during traversals through two forests: Venman National Park and Karawatha Forest Park, Brisbane, Australia. The dataset also contains 3D point cloud representations of the same scenes recorded using a lidar scanner, but in this g d use only the 2D images. In total, the c ach and every one of the images has bee

Methods

Many traditional, ma on methods could be used for this task course and other techniques from litera formance. At least two different method

Although we do not e xpect to see some new combination of r use of more state-of-the-art methods th

As there are virtually e detailed criteria, but as a general guid han copy straight from elsewhere, the better. In any case, always do cite your sources.

Training

If your methods require training (that is, if you use supervised rather than unsupervised segmentation approaches), you can use the same procedure for splitting the dataset into training, validation, and test subsets, as the creators of the WildScenes dataset. In their paper (see references below) they describe the procedure in detail and provide code for this in their GitHub repository (see references below). The procedure ensures that the training, validation, and test subsets have a uniform class distribution.

Even if your methods do not require training, they may have hyperparameters that you need to fine-tune to get optimal performance. In that case, too, you must use the training set, not

the test set, because using (partly) the same data for both training/fine-tuning and testing leads to biased results that are not representative of actual performance.

<u>Testing</u>

To assess the performance of each of your methods, compare the segmented images quantitatively with the manually annotated (labelled) images by calculating the intersection over union (IoU), also known as the Jaccard similarity coefficient (JSC), for each class and then taking the mean over all classes in the whole test set. Notice that although the annotations contain more classes, only 15 classes are to be used for evaluation (see further details in the supplementary material of the paper referenced below).

Show these quantitative results in your video presentation and written report (see deliverables below). Also show representative examples of successful segmentations as well as examples where your methods failed (no method generally yields perfect results). Give some explanation why you believe your methods failed in these cases.

Furthermore, discuss [illegible] or worse than the methods already eva[illegible]s reported in the paper referenced bel[illegible] future research to further improve the s[illegible]

<u>Practicalities</u>

The WildScenes datas[illegible]s and annotations are needed for this p[illegible] be challenging in terms of memory usa[illegible]our own laptop or desktop computer fo[illegible]able, you are free to use only a subset [illegible] a correct splitting procedure to ensure [illegible] the performance of your methods to g[illegible]our approach, this will not negatively im[illegible]

Deliverables

The deliverables of the group project are 1) a video presentation, 2) a written report, and 3) the code. The deliverables are to be submitted by only one member of the group, on behalf of the whole group (we do not accept submissions from multiple group members). More detailed information on the deliverables:

<u>Video</u>

Each group must prepare a video presentation of at most 10 minutes showing their work. The presentation must start with an introduction of the problem and then explain the used

methods, show the obtained results, and discuss these results as well as ideas for future improvements. For this part of the presentation, use PowerPoint slides to support the narrative. Following this part, the presentation must include a demonstration of the methods/software in action. Of course, some methods may take a long time to compute, so you may record a live demo and then edit it to stay within time.

The entire presentation must be in the form of a video (720p or 1080p MP4 format) of at most 10 minutes (anything beyond that will be ignored). All group members must present (points may be deducted if this is not the case), but it is up to you to decide who presents which part (introduction, methods, results, discussion, demonstration). In order for us to verify that all group members are indeed presenting, each student presenting their part must be visible in a corner of the presentation (live recording, not a static head shot), and when they start presenting, they must mention their name.

Overlaying a webcam recording can be easily done using either the video recording functionality of PowerPoint itself (see for example this YouTube tutorial) or using other recording software su ny others. It is up to you (depending on use, as long as the final video satisfies th

Also note that video fi vel of compression used). To avoid stora ill be 100 MB per group, which should f your video file is larger, use tools like H

The video presentati sentations). If the markers have any co may contact the group members by er

Report

Each group must also . 10 pages of main text, and any number

The report must be submitted as a PDF file and include:

1. Introduction: Discuss your understanding of the task specification and dataset.
2. Literature Review: Review relevant techniques in literature, along with any necessary background to understand the methods you selected.
3. Methods: Motivate and explain the selection of the methods you implemented, using relevant references and theories where necessary.
4. Experimental Results: Explain the experimental setup you used to test the performance of the developed methods and the results you obtained.
5. Discussion: Provide a discussion of the results and method performance, in particular reasons for any failures of the method (if applicable).
6. Conclusion: Summarise what worked / did not work and recommend future work.

7. References: List the literature references and other resources used in your work. All external sources (including websites) used in the project must be referenced. The references section does not count toward the 10-page limit.

Code

The complete source code of the developed software must be submitted as a ZIP file and, together with the video and report, will be assessed by the markers. Therefore, the submission must include all necessary modules/information to easily run the code. Software that is hard to run or does not produce the demonstrated results will result in deduction of points. The upload limit for the source code (ZIP) plus report (PDF) together will be 100 MB. Note that this upload limit is separate from the video upload limit (also 100 MB).

Plagiarism detection software will be used to screen all submitted materials (reports and source codes). Comparisons will be made not only pairwise between submissions, but also with related assignments in previous years (if applicable) and publicly available materials. See the Course Outline fo

Student Contribution

As a group, you are fr bers, but all group members must contr oding, making the video, and writing th up members only prepare the video an e.

An online survey will ymously evaluate the relative contribut s will be reported only to the LIC and th moderate the final project mark for ind they contributed substantially less thar

---

References

Webpage: WildScenes: A Benchmark for 2D and 3D Semantic Segmentation in Natural Environments. CSIRO 2023. https://csiro-robotics.github.io/WildScenes/

Paper: K. Vidanapathirana et al. (2023). WildScenes: A Benchmark for 2D and 3D Semantic Segmentation in Large-scale Natural Environments. arXiv:2312.15364. https://arxiv.org/abs/2312.15364

Dataset: CSIRO Data Access Portal: WildScenes Dataset. Version 2 (12 June 2024). https://doi.org/10.25919/5hzc-5p73

Released: 28 June 2024