

Software Engineering assignment

Paula Carbó Cubero Spring 2025

1 Introduction

Visual localization and mapping is using images only to accurately find the position of a camera in a previously mapped environment. In this context, sparse features are a representation of the images that captures salient points of interest (with a detection algorithm that selects keypoints) and its respective feature vector representing its visual appearance (using a description algorithm to obtain a descriptor). There are several methods to then, from sparse representations, find correspondences in between points of interest across different viewpoints, and once several correspondences are found, a pose can be estimated robustly. New points can then be triangulated and added to the existing map. There are different types of sparse features, each with different advantages: some of them are lightweight and fast (in the extraction and correspondence estimation steps), making them suitable for simultaneous localization and mapping (SLAM) applications, at the expense of accuracy; others provide the opposite benefits: slow but accurate features.

My research consists in enabling sparse feature interoperability. That is, researching methods to allow different features to be used interchangeably. Because of how keypoints are detected, using two different algorithm might yield non-overlapping points of interest (no keypoint repeatability), making the correspondence estimation step unsolvable. At the same time, different descriptors exist in different feature spaces, so they are not directly comparable. My research aims at solving these issues.

2 Lecture principles

Verification and Validation I think for any research it is extremely important to follow a principled approach to ensure the right method is being developed, and the method is being developed right. In my case I can identify a relevant problem and a possible way of solving it, but it's necessary to validate that the problem and needs are being met as the method evolves, for example, the method might be changed to work with a different type of data for simplicity which might mean the original use case is not applicable anymore. At the same time, my research relies on heavy empirical evidence of the performance of the method to check whether it actually addresses the

original problem, so verification that the implementation (neural net design, training, etc.) matches the theory of the method is extremely important and must be detail when writing research papers.

Metamorphic testing In my case, this concept applied to computer vision is used extensively for training with the name of data augmentation. For example, adding noise or illumination changes to an image should have an invariant metamorphic relation at the output. Similarly, to test for robustness to different cameras, adaptive brightness, etc., same image augmentation can be benchmarked against the expected results.

3 Guest lecture principles

Problem vs. Solution space in requirements engineering: In research, I think it is extremely important to understand what is the problem and which limitations the current methods face to ensure the novelty in the solution makes sense given the problems. In my case, it is relatively easy to find related methods (translation of descriptors – something more related to the solution space) and apply it to perform a feature translation in the heterogeneous setting, neglecting a good design of the problem space. I spent some time doing this and although successful and better than the baselines, it does not address one of the main problems of keypoint repeatability. From the get-go, I could have aimed for a solution tailored to my problem space.

Stakeholder elicitation: I'm doing an industrial PhD with Ericsson, which means there are more stakeholders other than myself and my academic supervisor that have interest in the research I'm making. As I involve more people in my research, the stakeholder list grows, and having a clear idea of what each person/group wants or needs out of my research can be really useful. One good example is the separation of business and system goals. Some people from my company might push the business agenda through apparent system goals, however I have to correctly identify these as business goals and treat them as such so they don't dictate my research outside its business/strategic purpose.

4 Data scientists vs. Software engineers

> Do you agree on the essential differences between data scientists and software engineers put forward in these chapters? Why or why not?

I agree, I think there is a clear focus and set of skills that differentiate both roles enough, and that makes it extremely hard for a single person to be an expert in both. At the same time, there is a high overlap between the two, and a lot of benefits from a data scientist having general knowledge in software engineering and vice versa. As I myself have only knowledge in data science and not in the software engineering direction, I can see how software engineering requires knowledge about good execution and structuring of both small and large pieces of code, and more importantly how to make sure several people can contribute to the same project and it's software ready for a product. I think data scientists are more used to working by themselves or in smaller teams, and more so for academic purposes.

> Do you think these roles will evolve and specialise further or that “both sides” will need to learn many of the skills of “the other side” and that the roles somehow will merge? Explain your reasoning.

I don't think roles will merge, but rather, I think another one or two differentiated roles might appear in the future, as deep learning becomes much more of an everyday tool for the whole population, for example, new needs will in turn require new specialised knowledge. At the same time, it is important that all of these have overlap to a degree, as that facilitates coordination and integration.

5 Paper analysis

5.1 DDPT: Diffusion-Driven Prompt Tuning for Large Language Model Code Generation [1]

This paper talks about optimizing prompt engineering. As Large Language Models (LLMs) become ubiquitous, we need to of course guarantee it's optimal use. Slightly different prompts can alter the quality of the output, and that's where this paper studies prompt diffusion for optimal prompt generation for the task of code generation. This of course is of high relevance given the content of this course, and as I was mentioning right before, new specialised knowledge is starting to be required at the intersection of software engineering and data science. This work is a clear example of how the output of data science (in this case, LLMs) might require extra work for it to be used adequately by software engineers.

The paper relates to my research not because it's close to what I work with, but because I can highly benefit as an end-user. LLMs are still not

reliable to accurately write code for very specialised computer vision algorithms, and on top of that, I do not have the knowledge nor the time to perfect my writing of prompts.

An AI/ML intensive project that can benefit from this is AI solutions development tailored to specific users. As each user has their own requirements, it is hard to reuse code. Instead, an LLM is used that, together with the method presented in this paper, can generate high quality user-specific code to then be developed as a solution. My own research could be used for these user-specific demands, as image processing and localization are essential for autonomous platforms, but using DDPT would make my work much easier.

As to how I could evolve my research to support and further develop the ideas of this paper, I could study how DDPT works in the context of computer vision: benchmark use of standard LLMs to aid in code generation of computer vision applications, and compare that to DDPT. I could further retrain DDPT for this particular context, and also research other methods beyond diffusion for optimal prompt generation. On the other hand, I could also provide valuable examples to train DDPT for specialised computer vision applications, as LLMs still have difficulties generating code for particular use cases. If I were to research in the paper’s direction, this could become a very powerful tool that could in turn help me in my research, as instead of having to spend time developing and debugging computer vision software, I would have the means of generating high quality code from simple prompts, and I could develop my own software library.

5.2 ImageBiTe: A Framework for Evaluating Representational Harms in Text-to-Image Models [2]

This paper presents an approach to detect and assess representational harms in text-to-image (T2I) models, that is, where an ambiguous text produces an image with a clear bias in ethnicity, age, gender, etc. In particular, ImageBiTe highlights the presence of stereotyping, under-representation and ex-nomination, three of the five identified representational harms in the art. The work compiles a series of target ethical concerns, such as age, disability and gender; as well as the dimension for these concerns, such as occupation, personal traits or socio-economic factors. For a given T2I model, ImageBiTe provides the dimensions for each of the representational harms given a prompt and every concern.

Although not from the human-centric perspective, this work is relevant for my research in terms of understanding harmful biases present in T2I models. As I claim my research targets indoor scenes, I am unaware that by that I actually mean "western-looking office spaces", as that is mostly the data that I have available to train my models.

An AI/ML intensive project would be any solution developing object recognition for autonomous driving. As more of these solutions are trained with data generated with these T2I models, where the ground truth label required for supervision is the prompt given to the model, it is extremely important to have methods such as ImageBiTe that make sure the data used for training does not suffer from representational harms, otherwise in the autonomous driving use case, safety could be compromised based on race or age. My own research could also contribute to this larger project, and it would be very important to test that my solutions can provide accurate positioning regardless of location and presence of different cultures.

My research, long-term, could be modified to incorporate location and cultural bias recognition. A model inspired by ImageBiTe could be trained to evaluate whether the models that I research to extract relevant features and accurate positioning suffer from biases in location, that derive mostly from cultural architectural differences, presence of ethnically different people in the environment, and different patterns in how things look in general depending on the country my research is deployed. By using a limited set of images to train my model and my own environment to evaluate its performance, I cannot guarantee my solutions are not biased to western architecture, and would not fail in other environments. Instead of evaluating my models on data from all around the world, which is unfeasible, a model akin to ImageBiTe can directly highlight the biases present in my model.

6 Reflection

I selected the papers only from the Cain 2025 long paper proceedings, based on the titles that appeared more interesting for me personally, based on my current AI/ML knowledge. This is of course in itself extremely biased, as I selected papers that use techniques I am already familiar with, like diffusion, LLMs and T2I. I was not misled by any title, as I think in the papers I selected, the title reflects really well what the paper is about. Since

these papers are accepted to a peer-reviewed conference, I trust to a large extent that its contents and references are accurate, and I did not cross-reference citations myself. As per the originality of my work, I only consulted the papers themselves and general knowledge from the web for concepts I was not so familiar with (for example, I read the Wikipedia article for *Representational Harm*).

References

- [1] Li, Jinyang, Sangwon Hyun, and M. Ali Babar. "DDPT: Diffusion-Driven Prompt Tuning for Large Language Model Code Generation." *2025 IEEE/ACM 4th International Conference on AI Engineering–Software Engineering for AI (CAIN)*. IEEE, 2025.
- [2] Morales, Sergio, Robert Clarisó, and Jordi Cabot. "ImageBiTe: A Framework for Evaluating Representational Harms in Text-to-Image Models." *2025 IEEE/ACM 4th International Conference on AI Engineering–Software Engineering for AI (CAIN)*. IEEE, 2025.