

# Assignment, WASP Software Engineering Course Module 2024

Josef Bengtson

June 2024

# 1 Introduction

The focus of my project is to work on the task of novel view synthesis, generating photorealistic images of a scene from new angles. The aim is to do this using a pre-trained generative model. Current methods for this lack geometric consistency due to the absence of an explicit 3D representation, necessitating heavy fine-tuning for improvement. We seek to develop a lightweight approach for ensuring and evaluating consistency without such constraints. Additionally, current methods typically focus on images of single objects where the background has been removed, we aim to extend this to images of scenes where the background is included.

These advancements hold promise for 3D scene understanding, particularly in applications like autonomous driving. Future plans include utilizing these methods to map dynamic environments around vehicles and simulate realistic driving scenarios. Moreover, these methods enable the generation of 3D objects and scenes from images or text, which is valuable for VR and game development.

## 2 Principles from Robert's Lectures

### 2.1 Software Engineering "Principles"

The first idea from Robert's lectures that I will discuss are the Software Engineering "Principles" he presented. I will discuss some of the principles that I feel are relevant for my research. A first principle was the **need for some process (overview, structure and identity)**. I see an importance for this in my research, since my field is rapidly changing, making process important in bringing structure and counteracting confusion. A second principle was that **understanding the problem is (more than half) the problem**. In my research I have experienced that it takes a significant amount of time to figure out what a relevant problem is and which type of solutions could be suitable. Getting deeper understanding for relevant problems can help with going deeper in research, and not just making smaller improvements to existing methods. A third principle was to **test early, test often, test throughout and test automatically**. I have found testing very important in my research and it has sometimes been challenging to figure out what to test and how to test it in a valid manner. I work towards integrating testing in all steps of the research development process and to find ways to effectively scale up the testing as the method I work on gets

more and more developed.

## 2.2 Validation and Verification

I found the ideas of validation and verification to be very interesting. In my area of research I find that the primary method of evaluation is verification based on simple metrics, such as PSNR (Signal to noise ratio). There exists more advanced metrics, such as SSIM and LPIPS, that in a more advanced fashion strive to capture the desired properties of image similarity. But there are still many cases where these metrics are not able to in an adequate way evaluate performance, e.g. generative tasks where there is no single correct output to compare with. One way of handling this is to use metrics that evaluate if an image looks realistic without comparing it to a specific image. Another way of performing validation is to use human user tests to evaluate if generated images look realistic, but this can be expensive and hard to guarantee that the answers are not biased. Currently my research is on the task of guaranteeing multi-view consistency when generating new views of a scene and trying to find better ways of guaranteeing that the generated views actually are consistent. I find this idea of validation, reflecting about how to fulfill the needs of the end user of the developed method, very helpful. Not just aiming to get higher performance metrics, but thinking about how to in a better way evaluate and measure desired performance.

## 3 Principles from Guest Lectures

### 3.1 Why is Software Engineering critical?

I will now discuss what Per Lenberg said regarding importance of Software Engineering at Saab, and how this can relate to the topic area of my research, i.e. generative models and computer vision. His first reason was that they work with complex systems with many developers, and a structured approach is therefore needed. Large generative models are definitely complex, and many people are often involved, with there often being many different teams extending on existing models. There is therefore a need for a structured approach, which could make continued development more stable and efficient. His second reason was that the systems they work on are safety critical and need to meet regulatory standards. Certain systems utilizing computer vision are definitely safety critical, e.g. autonomous vehicles, while other applications might not be safety critical, e.g. an AR based game/app. For large generative models there is an increasing amount of

regulatory standards that will have to be met in the future. These things combined shows that a Software Engineering approach should be critical when developing and applying large generative models in computer vision, especially if working on a safety critical application.

### **3.2 Using LLMs to automate complicated pipelines**

Parthasarathy Dhasarathy talked about ML for Software Engineering in his guest lecture, and mentioned that AI is good at SW engineering. He said that the targeted use of LLM:s can automate complicated pipeline requiring many engineers, mentioning how they at Volvo Trucks have started to use LLM:s to help design test cases. In a recent research project I was working on we used an LLM as part of processing image captions. We could have written specific processing function for this task, but there were many special cases, so it would have taken significant time to get something to work well. But with the help of using an LLM as an function, with a suitable prompt, we were able to quickly process the captions. I want to in the future keep using LLM:s to help automate suitable steps in my research projects.

## **4 Analysis of Papers**

### **Navigating Privacy and Copyright Challenges Across the Data Lifecycle of Generative AI**

#### **a. Core Ideas**

The topic of this paper is privacy and copyright challenges for Generative AI, which are magnified by GenAI's reliance on expansive datasets. The core idea is to investigate and devise solutions based on the lifecycle perspective, leading to holistic approaches. Their analysis centers on two critical legislative frameworks: General Data Protection Regulation (GDPR) that relates to privacy and the Copyright Law of the United States that relates to copyright. The main contribution of the paper is to map key challenges onto different stages of the data lifecycle. They finish the paper by highlighting and calling for further research into lifecycle-centric approaches to these challenges. These ideas are important to ensure that privacy and copyright are respected throughout the whole process of developing and deploying an AI system.

### **b. Relation to my research**

In my research I utilise large diffusion models, which are GenAI models trained on huge datasets of images. I want to build on and extend these models so that they can be used to generate multiview consistent images of a scene. This paper's reflection upon challenges regarding privacy and copyright issues for these types of models is thus of relevance.

### **c. Integration in larger project**

An example of a larger AI-intensive software project where my research could fit in would be a AR/VR app that generates 3D content and integrates it with the surrounding environment. The ideas in this paper could be helpful in the project to ensure that the app creates, uses and distributes data in such a way that privacy and copyright laws are respected through the whole data lifecycle. My WASP research would fit into the project as a step to ensure that 3D content can be generated in a way that is multiview consistent, and that the appearance of the content can be adjusted if desired.

### **d. Adaptions to research to improve AI engineering**

One way my research could be adapted would be to be aware of potential privacy and copyright issues at different stages of the data lifecycle. One example of this is to in both Data Collection and Model Deployment stages making sure that it is transparent which datasets have been used to train the model. My research can also be adapted to improve multiview consistency without requiring additional data, instead using a geometry based optimization. This type of approach could in the future reduce the need for finetuning on large datasets, alleviating some privacy and copyright concerns.

## **Engineering Challenges for AI-Supported Computer Vision in Small Uncrewed Aerial Systems**

### **a. Core Ideas**

This paper looks at engineering challenges when deploying deep learning based computer vision on limited resource devices and with a reliance on accurate and timely perception to perform critical tasks. They perform a systematic literature review and classify the issues and solutions they found as related to Computer Vision pipeline (CV), system hardware (HW) and/or Software Engineering (SW). One conclusion they made was that there in the

reviewed papers was very little discussion about the Software Engineering aspect. Their core idea is that the complex and safety-critical nature of these types of systems warrants a rigorous Software Engineering process. They propose the need for an systematic effort across all three aspects (CV, HW and SW), with each being tested individually and then carefully integrated with each other. This is important to have in mind when engineering AI systems, to remember the importance of Software Engineering and to test and verify each different aspect of the system.

#### **b. Relation to my research**

My research is on the topic of deep learning based computer vision. I am not currently working with limited resource devices or safety critical tasks, but in the future my research might relate to Autonomous Driving Applications that are safety critical. There is in my research no significant HW restrictions (since I utilize clusters where I have access to A100 GPUs), but the analysis regarding CV and SW aspects is of interest, showing the need for both separate testing and integration.

#### **c. Integration in larger project**

A potential project where my research could fit would be for a surveillance program deployed on a drone used to scan regions after a disaster. The ideas from the paper on testing the CV, HW and SW aspects of the system separately and then carefully integrating them could benefit the project, especially since it would be a safety critical system run on limited HW. My WASP research could fit into the project by taking the images gathered by the drone and quickly generating additional views of the scene allowing for a better 3D understanding of the situation.

#### **d. Adaptions to research to improve AI engineering**

My research could be changed to include clearer tests of the SW and CV components specifically and how to integrate them with each other. Another change would be to test and evaluate the method for input images of varying quality, with issues such as blurring, over-exposure and obscured regions. Another change would be to adapt the code to make it less reliant on a large GPU (e.g. by using gradient check-pointing or other such techniques) and large storage (e.g. by removing unnecessary software packages).