Senior Software Engineering Project

Milestone 1

Blake Cecil, 10/9/20

Intro

This paper will explore three different software engineering projects. Included will be a brief intro to each project, the engineering requirements and constraints, and a proposed timeline. The projects all focus on applying deep learning to various real life tasks, from design automation to expediting the ballot collection process. After reading this paper you should have a sense of the vision of each project, as well as some of the more technical details and challenges each project faces.

Contents

1 AI Based Webapp for Microfluidic Design Automation

HP is one of the world leaders in printing technology. In order to stay innovative and competitive vast amounts of time, resources, and money are poured into research and design. Because of how critical the R&D process is, anything that can improve it is a massive boon. This project seeks to employ machine learning to help HP accelerate the process of designing key components of their printing technology.

HP will be the lead technical partner for this project supplying several proprietary tools for image processing, and testing the designs the ML model generates. The end users of this project are the R&D team of HP. The vision is to be able to integrate the ML model into their workflow, expediting the process of generating novel designs. This means that HP is the principal stakeholder in the project, due to the substantial decrease in R&D costs automation offers.

1.1 Engineering Requirements

There are two key parts of the system, the ML model, and the front end application. The ML model will require the following components to be implemented

- ML model that generates new designs.
- Software that prepares data for the ML model.
- Automated system that feeds the ML model new data to learn from.

The front end application is then responsible for providing an interface for the model as detailed in the following requirements

- Mechanism for users to interact with the ML model.
- Ability to turn model outputs into concrete designs.
- Interact with a database.
- Simulate design and report results.
- Deployable on a server.

1.2 Constraints

One of the most integral constraints is the speed of the web application. HP requires a throughput of approximately 100 design iterations per second. This puts a constraint on the ML aspect of the system, which could be a potential bottleneck. We may also be constrained by the amount of data available to train the ML model with, as well as the amount of compute power available to perform said training. Another constraint is the type of designs we can generate based on what the design emulator can reasonably simulate. The application also needs to be easily deployed on a server and strike a balance between exposing enough of the ML model for the user to tweak it, but not so much that it requires extensive knowledge to use.

1.3 Project Outline

1.3.1 Required Skills

This projects skills will require some basic hardware knowledge and extensive software skills. One will need at least a basic understanding of the microfluidic devices the network will eventually generate. This means knowing their basic functionality and how various design parameters influence the end product.

The machine learning aspect will require an understanding of standard deep learning methods. More specifically knowledge of generative machine learning methods, and ML frameworks such as pyTorch will be essential. Experience crafting production level data pipelines would be desirable but not strictly necessary.

On the front end, expertise in creating web applications that scale will be critical to meeting the projects throughput requirements. This will require experience designing a full stack web application with a focus on efficient architecture. Additionally, understanding how to structure and query databases efficiently would be a desirable skill. Finally, experience with designing usable and intuitive UI's will be necessary to meet the R&D team's needs.

1.3.2 Equipment and Resources

For the machine learning core of the application we will need a sufficient data set to train the network. In addition we will require GPU's to speed up training. For the front end we will need access to a server and database for the web application to utilize. It would also be beneficial to have access to a member of the R&D department. This would help with understanding what they are looking for in the UI. They could also provide valuable technical insights that the ML could integrate into the model.

1.3.3 Development Timeline

The project will begin with familiarizing new team members with the projects current code base, tools, and workflow. After that, the front end team will develop a prototype for the web application, and ML team will prototype neural net architectures and the data pipeline. With prototypes agreed upon the next phase will be producing a minimal working product. Then user feedback will be collected and any new requirements that were discovered will be addressed. With this new information teams will return to the product, integrate any new features or missing requirements, and bring the product to scale. The product will then be verified against the constraints and customer requirements at which point the team will either release the system to the R&D department or will iterate on the current system to meet requirements. It would be reasonable to expect to complete the fully scaled version of the system within the 9 month time period.

2 Biodiversity Monitoring with Edge Neural Networks

Collecting vast amounts of data about the world has never been easier. Despite this many industries have yet to take advantage of this fact. One example is the forestry and ecology sector who have yet to integrate modern data collection techniques into the studies they perform. The principle issue is that it is simply not feasible to set up a camera in the forest and watch all the footage that is produced, as novel events are rare in this setting. Syntiant is working to solve this problem by using machine learning to only collect data when a novel event is occurring, such as movement or noise.

Syntiant is the lead technical partner on this project and is providing the team with access to its cutting edge, energy efficient chips, designed to run neural networks. By integrating these chips into cameras and microphones, large scale forestry and ecology projects will enjoy the efficiency of automating the data collection process. As such the core users of this project are forestry organizations across the country, as well as ecological research groups. The primary stakeholders in this project are Syntiant's many private backers (). It is also possible that the scale of this project means state forestry agencies may have an active stake in this project.

2.1 Engineering Requirements

This project will require an extensive understanding of both hardware and software as the ML model will be deployed on specialized hardware made by Syntiant which will then be integrated into cameras or microphones. For the ML model the requirements are roughly

- Create a Input data processing pipeline.
- Design ML model(s) capable of classifying video and microphone data.
- Create a system to evaluate the ML models output and make decisions based on it

The hardware focused engineering requirements are

- Integrate the ML model(s) with Syntiants chip.
- Develop a system that integrates the chip into a camera/microphone.
- Develop a monitoring system for the camera/microphone to feed data into the ML model(s).

2.2 Constraints

There are two key constraints in this project, the hardware and the speed of the ML model. An integral part of the project is showing off Syntiants new energy efficient chip for deep learning. Thus, all of the software must be written to run on this platform, and the chip must be integrated into the camera/microphone that will be used for data collection. The issue of speed of the ML model arises from the mass amounts of data it will need to process. The model will be retrieving images/sound from a continuous feed controlling the overall system so it is imperative that it can act quickly.

2.3 Project Outline

2.3.1 Required Skills

The project will require skills ranging from both hardware and software. From a hardware perspective the ML model is constrained to working on Syntiants chip so it will be essential to understand the chip and its computational limitations. One will also need a good understanding of writing systems control code that allows the ML model to interface with the camera/microphone.

As for the software this project requires a very strong understanding of computer vision techniques as well as deep learning knowledge. The nature of the application means having to process vast amounts of data. As such it will be crucial to know how to efficiently prepare and process huge data sets as well as design and create efficient deep neural networks.

2.3.2 Equipment and Resources

As with any machine learning project two key resources will be computational power in the form of GPU's as well as a sufficiently large data set for training the model. This project will also require access to cameras and microphones, likely several different styles and brands as some may integrate with Syntiants hardware better than others. It would also be valuable to have a representative user to consult with so as to better understand which events the ML model should be classifying as novel.

2.3.3 Development Timeline

The first step in the project will be familiarizing the team with the hardware Syntiant intends the application to run on. Once the team sufficiently understands the hardware constraints, they should break up with one group focused on prototyping the ML model and the other on the integration of the Syntiant chip and model into the cameras and microphones. Once a prototype is agreed upon the teams should begin working on the system. Because of the systems dependencies it is unlikely that the team will be able to work on each aspect in isolation. Instead the team should focus on realizing the ML model, migrating it to Syntiants chip, and then integrating the chip into the camera/microphone. In the best case scenario it would be reasonable to expect to finish the project within the 9 month period. However given that the chip is Syntiants experimental hardware, integration may prove difficult in which case one might expect the project to exceed the project deadline.

3 Doing Good for Oregonians: Digitization of initiative petition requestor data (HTR)

A fundamental principle of democracy is that a government and its laws should be a reflection of its citizens desires. To this end Oregon law guarantees its citizens the right to create legislation if enough citizens vote. Tragically, the process required to collect and record such ballots is outdated, a problem Solving for Progress aims to fix. By using tools from machine learning such as hand written text recognition and image processing, ballot collection and processing can be expedited.

Solving for Progress is the main technical partner in this project and will be procuring the ballot information, and government connections necessary for the project. The citizens of Oregon and its government will be the primary users as well as stakeholders in this project. The citizens rely on their ballots being counted correctly, and the state would benefit from being able to expedite this process.

3.1 Engineering Requirements

The requirements for this project can be split into the data processing and ML aspects. In some projects these are one in the same but this one poses unique challenges. The amount of incoming data will be large (over 100K ballots estimated). Furthermore the ballots are on paper meaning that digitizing the ballots will be a challenge on its own. For the machine learning part of the project the engineering requirements will include

- ML model to recognize digits.
- ML model to recognize names.
- Interface that translates model output into voter info.
- Database for storing info.

The data processing part of the project will require a blend of both physical and software engineering

- Flywheel system that feeds ballots into image processor.
- Image processor that take photos of ballots.
- Software that prepares ballot photos and feeds them to ML model.

3.2 Constraints

The largest constraint of this project will be the security required to handle voting/personal information appropriately. This will likely require encryption of the photo data required as well as the database where voter info is stored. The project will also have to be efficient, although it is the physical, not software aspects that will likely cause the bottleneck. As such the system that is responsible for digitizing ballots will need to find the right blend of speed and accuracy.

3.3 Project Outline

3.3.1 Required Skills

The project will require a blend of hardware and software skills. The ability to design and engineer simple physical systems is a must for the data processing pipeline. One will also need an understanding of computer vision techniques for imaging and processing the ballots. For the ML aspect of the project deep learning skills will be essential, with a special focus on image recognition.

3.3.2 Equipment and Resources

The project will require several potentially expensive pieces of equipment. The camera used for image processing will need to strike a balance between affordability and quality. The ballot flywheel could be expensive depending on whether it is purchased or engineered from individual parts. Finally is the question of the ML model. Unlike the other ML tasks described, this one has been solved many times over. This means that one could actually circumvent the expensive training process and use a pre trained model. However if Solving for Progress desires to build a model from scratch then one would need GPU's to train it on.

3.3.3 Development Timeline

The timeline for this project is heavily dependent on whether or not a pre trained ML model is used because of this two different timelines will be envisioned here.

If a pre trained model is used then the first step of the project will be prototyping the device that feeds in and images ballots, as well as the database that will be used to store ballot info. After this the team should break up with some members working on the flywheel for feeding in ballots, others on the imaging and processing software as well as the database. Next will be combining the components with the ML model to create an end to end system. Once the system is fully integrated the team can begin the testing phase. If a pre trained model is used the project could be easily completed withing the 9 month time frame. There may even be an opportunity to revisit aspects of the project and quality of life features.

If a pre trained model is not used then the prototyping phase will also require prototyping the ML architecture to be used. From here the team would still break up but some members will be dedicated to developing the ML model instead. From here the project would proceed as described as above. The biggest difference here is the timeline. The project would still be feasible to complete in the 9 month time frame, but would likely take all 9 months if the model is made from scratch.