Project Title

DESIGN DOCUMENT

Team number: sdmay21-32
Advisers: Chandan Kumar, Ali Jannesari
Team Members: Michael Boyle, Dylan Hodge, Dylan Smith, Jeff Kinard, Mark Endeshaw, Sam Hassebroek
Team Email: sdmay21-32@iastate.edu
Team Website: https://sdmay21-32.sd.ece.iastate.edu

Executive summary

Development Standards & Practices Used

List all standard circuit, hardware, software practices used in this project. List all the Engineering standards that apply to this project that were considered.

- Agile Development Practices
- Test Driven Development

Summary of Requirements

- There shall be an algorithm to automatically annotate all objects from any image.
- There shall be an extension of the algorithm to annotate objects from video.
- There shall be a further extension to include masking.
- There shall be a backup algorithm to annotate objects missed by the main algorithm.
- There shall be a method of validating the results for each algorithm.
- There shall be plenty of technical documentation and well-commented code.
- There shall be a focus on enhancing annotation accuracy.
- There shall be a focus on reducing the processing time in large scale video data by utilizing the algorithm.
- Our team shall meet all deadlines set by our instructor and our project contact.

Applicable Courses from Iowa State University Curriculum

- COM S 474/472

New Skills/Knowledge acquired that was not taught in courses

List all new skills/knowledge that your team acquired which was not part of your Iowa State curriculum in order to complete this project.

- Machine Learning methods
- Python coding
- Cocosynth
- Manual masking/annotation

Table of Contents

List of figures/tables/symbols/definitions

1 Introduction

1.1 Acknowledgement

We don’t have any client or organization that contributes or will contribute significant assistance on this project.

1.2 Problem and Project Statement

The problem that we are hoping to solve is automatic labeling of images and grouping of the objects inside of those images. Currently there exists many object detection software that run on trained data sets to detect and label images containing certain predefined objects. We hope that our solution will be able to detect and group objects that are similar to each other with no pre training or prompting from the user about specific objects. In our solution it would be completely anonymous and identify all objects in the image.

1.3 Operational Environment

Our project is completely software based and this will not need to be constrained by a physical environment. It will be completely defined by the computers and operating systems that it can run on. We hope to develop something that is independent of its operating system and will be able to run as an independent python program.

1.4 Requirements

- There shall be an algorithm to automatically annotate all objects from any image.
- There shall be an extension of the algorithm to annotate objects from video.
- There shall be a further extension to include masking.
- There shall be a backup algorithm to annotate objects missed by the main algorithm.
- There shall be a method of validating the results for each algorithm.
- There shall be plenty of technical documentation and well-commented code.
- There shall be a focus on enhancing annotation accuracy.
- There shall be a focus on reducing the processing time in large scale video data by utilizing the algorithm.
- Our team shall meet all deadlines set by our instructor and our project contact.

1.5 Intended Users and Uses

For now, we do not have any specific client/company/organization.

1.6 Assumptions and Limitations

The project is about a fully automated and flexible multi-label classifier image annotation. Automatic image annotation is a difficult task and has some limitations. The first reason is that the algorithm might have pixel predictions or semantic gap problems, which is hard to extract accurate images from low level feature images. The other limitation is that it is difficult to know which regions of the image correspond to keywords in the trained data.

1.7 Expected End Product and Deliverables

The main end project deliverable will be a written document and a python script. The deliverable will be given to the grad student overseeing the project at the end of the class. The documentation will be written not only on how to use the software but also how to continue developing the software after we have left and someone with no knowledge of the project.

2 Project Plan

2.1 Task Decomposition

1. Setting up GitLab with CI/CD pipeline, shell script to download required software and installation of required software including GIMP, python/python3, Anaconda Navigator, Cocosynth Library, Shapley and MATLAB.
2. Manual masking and annotation, each team member annotates approximately 18 images.
3. Backup algorithm to annotate objects that are not annotated by the main algorithm
   3.1. Phase 1 and phase 2
      I. preparing ground truth data
      II. Image processing of dataset
   3.2. Phase 3
      I. Create ML/DL model
      II. Improve accuracy
      III. Validation data
   3.3. Phase 4
      I. Documentation
      II. Presentation

2.2 Risks And Risk Management/Mitigation

Risk description	Probability	Risk mitigation plan	Alternative tool, technology
Low visual recognition between ground dataset and test dataset	0.3	Increase the size of the training set or ground truth data	Using Machine learning/Deep learning to improve the quality of visual recognition
Main algorithm fails to annotate an image	0.4	Develop a backup algorithm	use multiple algorithms
Algorithm fails to detect object in video	0.5	Extend the main algorithm	Apply multiple object track finding algorithms

Table 1. Risk and Risk mitigation

2.3 Project Proposed Milestones, Metrics, and Evaluation Criteria

Setting and Installation of the required software

This project will be designed to run and compile in all major operating systems. All the required software and libraries will set up to run in any operating system so that the end users can compile the project easily.

Manual annotation

Each member of the group did a manual masking to have a better understanding of the algorithm and to use the annotated images for the next phase of the project. The manual masking accuracy will be measured by applying the concept of Neural Networks such as Convolutional Neural Networks (CNN) and Mask R-CNN.

Main and backup algorithms for automated annotation

For this project we will have two algorithms, the main algorithm and the backup algorithm. The main algorithm should be able to automatically annotate/label the objects inside an image or video. The automatic image annotation uses few trained data to annotate images in the dataset with shorter time and with greater efficiency. In this algorithm the metadata is assigned to a digital image using appropriate keywords, it explores the correlation between features of the visual image and semantic meaning to draw a function using machine learning. The backup algorithm is to annotate images that are missed by the main algorithm.

This task has three phases, the first and the second phases are preparing the ground truth data and image processing of the dataset. This includes the following major subtasks,

- Model design: model design is the feature detection to describe local features and composition of the objects.
- Trained datasets: the set of trained images and their features to work with the model and to generate the features that match the given image.
- Test set: the set of images for testing against the trained datasets, this will help us to predict the accuracy of the model to get the correct matches.
- Classifier: classifies the given dataset into different classes for search and match optimization.
- Training and testing: use a set of images to check against the ground truth.

The third task is creating machine learning /deep learning models and validation of data. In this phase we will apply the concept of Neural networks including Convolutional Neural Networks (CNN), Feed Forward Networks, back Propagation and Gradient Descent. Each phase of the project will be evaluated using the images that are manually masked and annotated.

2.4 Project Timeline/Schedule

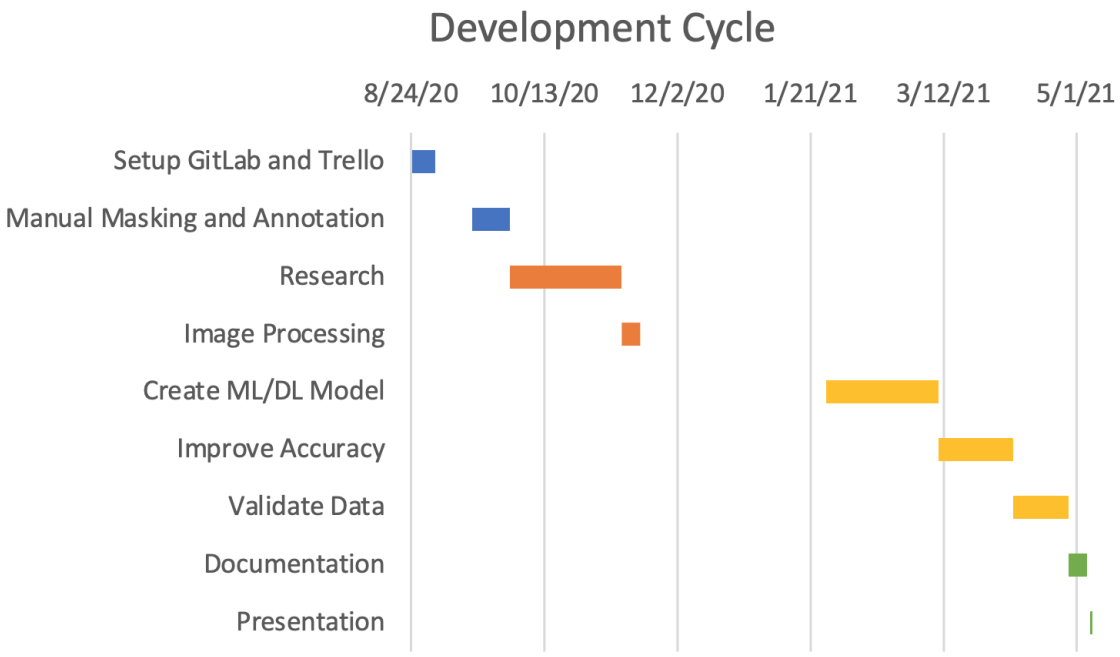

Figure 1 project timeline

2.5 Project Tracking Procedures

We are tracking the project progress using Trello to organize and to prioritize project tasks. Each series of tasks are listed in Trello board, this helps us to break projects down into smaller tasks or create even more detailed to-do lists. The other task managing tool is GitLab which is more than a collaboration and a change tracking tool. GitLab enables development teams to work in asynchronous environments, tracking the progress of the project and working together on a

different branch. The other feature of git is GitLab issues, this helps us to post questions on relevant issue threads, post updates and notify any team member to get them involved in the project. We are also using GroupMe for daily communications and conversation that needs to happen quickly.

2.6 Personnel Effort Requirements

Task	Estimated Person-Hours
Manual masking and annotation (~100 images)	~100 person-hours
Prepare Ground Truth Data and Image Processing (Preparing for Automatic Annotation)	~50 person-hours
Automatic Annotation: - Create ML/DL Model - Improve Accuracy - Validation Data	~120 person-hours total
Documentation and Presentation	~80 person-hours

Table 2. personal effort requirements

Under ideal circumstances, each member of the team will take on work that requires similar amounts of effort to accomplish. Given our estimations of our project task requirements, each member would be required to give roughly 60 hours worth of work to help complete the project on time, for a total of about 350 person-hours to complete the project.

2.7 Other Resource Requirements

As far as other resources go, all we require is a VM with these capabilities:

- Ubuntu1804 OS
- 1 Core CPU
- 4GB Memory
- 50 GB Disk Memory

This will be provided by Iowa State University, and it will serve as a host for our GitLab runner.

2.8 Financial Requirements

As our project is entirely web-based, and requires no paid licenses or materials to conduct, including financial requirements is not necessary.