

Potential Graduate Independent Project Ideas

CSCI E-25, Computer Vision

Overview

This document contains some suggestions for potential graduate independent projects. These ideas are intended to help you formulate the details of your project proposal. Since having an appropriate data set is vital to a successful project, the focus here is on available data.

The project ideas and datasets outlined here do not constitute complete project plans or proposals. Nor are these ideas intended to limit your choice of project.

It is strongly recommended that you find a project of particular interest to you. By doing so, you will find your project much more interesting and engaging. Therefore, your project can use any appropriate dataset of your choosing.

Keep in mind the following general guidelines for selecting a successful project:

1. You must select an end-to-end CV specific project using methods and algorithms within the scope of the course.
2. Selection of a suitable dataset is absolutely critical. The dataset you select must have sufficient examples (samples) to perform your project. However, be aware that many CV datasets are very large and require massive storage and computing resources to work with. Video data is inherently larger and harder to work with than still-frame data. In summary, verify your dataset is suitable before you spend too much time on your proposal.
3. If you are unsure about what you want to do or how to do it, speak with the instructor. You can do so during the Monday section meetings or you can arrange another time. Contact: stephen.elston@gmail.com.
4. Your project must be your own work. Simply copying a project from a blog post is not acceptable. You are encouraged to investigate and learn from other sources. If you use posts or papers as background to formulate your project make sure you cite your sources to be consistent with the Harvard DCE [academic integrity policy](#) and the [sources responsibilities policy](#). As long as you have a distinctive aspect to your project, it will be acceptable.
5. A great project is a wonderful addition to your data science portfolio. You should consider creating a blog post and/or GitHub repository to tell with world about your work.

What is the purpose of the proposal?

The process of creating a proposal helps you organize your project. Take the time to explore your data, and perhaps test your ideas with a prototype. Your instructor will review your proposal and provide feedback on the suitability of your project. The ultimate goal is to help ensure your project goes well and you do not end up stuck against difficult problems with the clock running down!

Potential Datasets

1. One of the commonly used benchmark CV data sets can be the basis for a project. Different benchmark datasets have differing intended purposes and can be chosen for projects as such.

These projects can range from You can find curated lists of these benchmark dataset several places, including [here](#).

2. Kaggle hosts a great many datasets, including some reasonably small ones suitable for a project for this course. You can find the list [here](#).
3. The Visual Geometry Group of Oxford University has compiled an extensive [repository of datasets](#).
4. The application of CV to industrial inspection has a very long history. Industrial inspection problems often require reasonable size datasets. You can find a curated list of datasets [here](#).
5. The company V7Labs has created an [open repository of over 500 searchable computer vision project datasets](#). The [V7labs blog](#) covers a wide range of CV solutions and can be a source of project ideas.
6. 3-D CV methods have many interesting applications. You can find a curated list of some benchmark datasets [here](#).
7. The [IDD-lite](#) and [IDD-mini](#) datasets allow you to experiment with [advanced semantic segmentation](#) without the need for massive computing resources.
8. There are many possibilities for applying CV techniques to remote sensing images. You can find an extensive and well-organized [list of datasets in this GitHub repo](#).
9. A large number of CV dataset can be found in the [Cancer Imaging Archive](#). But be warned, many of these data sets are very large and difficult to work with.

Models and Pretrained Models

While deep neural networks are the standard for many CV applications now days, it is not required that you use such methods for your project. However, if you do want to use deep neural networks, I suggest you carefully consider one of the many pre-trained (transfer learning) models available. These models are available of several platforms including a [rich set of possibilities in Keras](#).

Environments

You may find that you require computing resources beyond your local environment to perform the work for your project. Some possibilities for paid environments include [Google Collaboratory](#) (Pro or Pro Plus recommended), [Amazon EC2](#), or [Azure Notebooks within Azure Machine Learning Studio](#).

A Few Ideas

1. A student in the past used a time sequence of remote sensing images to determine the shrinkage in areal extent of mountain glaciers.
2. CV based industrial inspection is a widely used technology. Items inspected can range from machined parts to fruits and vegetables.
3. Blog posts can be a rich (if sometimes overwhelming) source of ideas. For example, one could take a car classification project like [this one](#) and expand or change it in some significant way. For example combine detection and classification of cars in images.
4. Generative models can be useful in extending training data sets. You can find a post discussing this and other applications of generative models [here](#).

What happens if my project does not work out?

Not every good idea works out in reality. You may find that the goals you are hoping your model learns just do not work out. As long as you have tried a reasonable number of model improvement, your project is successful. In the real-world what is important to have tried and understand why things don't work. In summary, your project need not be a smashing success to get a good grade, as long as you have worked systematically on the problem and understand why the methods tried do not work out.

Grading Criteria

As you prepare your proposal and work on you project and report, keep the following grading criteria in mind.

You are required to create and submit a proposal for your intended project. The goal of your proposal is to ensure that your project is appropriate, feasible, and the required data is available. During this process you are expected to perform a preliminary investigation of your data and consider appropriate algorithms and methods. Your instructor is available to help you with formulating your project.

Criteria	Weight
An executive summary with a clear statement of the problem to be addressed, the value of the project for the intended audience and the results sought.	20%
A clear statement of the objectives supported by the evidence and inferences drawn from an initial exploration of relevant literature and data? Is the problem statement and value proposition of the project understandable to a non-technical audience?	15%
Are the data selected for the project appropriate for the problem being addressed?	20%
Are the proposed algorithms and modeling methods appropriate to the problem being addressed?	15%
Are the anticipated data preparation steps clearly laid out?	15%
Does the project plan properly cover communications to non-technical people?	15%
Total	100%

The following criteria are to be used to evaluate and grade you CSCI E-25 independent project presentation and report. The weights are percent of the total score for the project, **which is 25% of your overall grade.**

Criteria	Weight
Introduction with a clear statement of the problem to be addressed and summary of the results? Is this introduction understandable to a non-technical domain expert?	10%
Are there clear conclusions supported by the evidence and inferences drawn from the analysis and models presented, even if the result is negative? Are the results understandable to a non-technical audience?	15%

Given the data set and the stated problem, are the steps used to evaluate and interpret the results appropriate? A few random histograms and scatter plots are not really good enough. This evaluation process will guide you improving algorithms and methods.	15%
Are the modeling methods selected appropriate to the question being addressed? You are expected to consider and compare alternative algorithms and hyperparameter choices for your models.	10%
Has the data been properly prepared in terms of cleaning, filtering, normalization, etc.?	10%
Have appropriate steps been taken to improve the models and analysis methods? Examples of model improvement methods include, testing alternative models, testing multiple hyperparameter values for models, and identifying and correcting data problems. A project applying one or two basic algorithms is not acceptable.	15%
Are the conclusions and inferences drawn from the model clear, concise and addressing the stated problem? Can a non-technical domain expert understand these results?	15%
Is all code used written in a reasonably professional style, including sufficient comments?	10%
Total	100%