School of Computing

Year 4 Project Proposal Form

SECTION A

Project Title: Stolen Image Identification

Student Name: Thomas Doyle

Student ID: 15350316 Stream: CASE

Project Supervisor Name: Brian Stone

[Note: It is the student's responsibility to ensure that the Supervisor accepts your project and this is only recognised once the Supervisor assigns herself/himself via the project dashboard. Project proposals without an assigned Supervisor will not be accepted for presentation to the Approval Panel.]

SECTION B

General area covered by the project:

This project will cover the area of computer vision and image classification as well as large scale data collection and web scraping. It will be primarily to compare images and make a prediction to whether that is an original image or it has elements taken from another image. It will be concerned with images across various sites on the internet.

Outline of the proposed project:

This project will attempt to scrape a website finding images and cross checking them against a database of known metadata about images. It will then give an opinion if this is the true origin of the image or if it has seen the image in another location on the internet with a high confidence.

Background:

As a photographer it is a general fear and something that I have often seen across the internet. Images are easy to steal and hard to track down. With this automated search tool it will take out the human element of searching for the image origin, should you want to buy the image.

Achievements

This should be collecting data about images constantly and storing the metadata and computed data about the image. This will then generate a report of stolen images and alert the origin of these images.
It will also allow a user to find the origin of an image if the project has seen it before.

Justification

This will allow photographer to search if anywhere else is using their images.
This will also allow a customer to find the real origin of an image to purchase

Programming language(s)

The programming language I will use is python. I feel it has many libraries for computer vision production ready and it is very fast for me personally to write and develop in. Given the short time frame of the project I don't believe I have the time to learn and become comfortable in a new language.

Programming tools / Tech stack

For my database I am going to choose postgreSQL, apache web server, imageAI library for python computer vision and most likely make use of Amazon Web Services for the high scalability.

Learning Challenges
I will have to learn how to design the entire project to scale with a vast amount of data
I will have to learn computer vision
I will have to learn how to apply machine learning to get the confidence in predicting stolen images correctly

Hardware / software platform
The software will be design to run on linux in the cloud
There is no special hardware for this project

Note: The final revision of your proposal form should be converted to a **PDF** in your GitLab repo from where it will be automatically collected.