G5

PROJECT DOCUMENTATION

Project Vision

Aerial Imagery Initiative

Index

1. Introduction

The NSW Spatial Services division of the Department of Customer Service (DCS) collects aerial and LiDAR images of the NSW landscape and is responsible for providing reliable spatial information to NSW Emergency Service Organisations (ESOs). In order to better facilitate rapid responses and recovery efforts to emergency flood events, this project aims to develop a machine learning platform that will analyse aerial and LiDAR imagery to generate flood extent maps.

The current standard of practice for DCS during a flooding event is to fly an aeroplane over flood-affected regions of interest (ROI) to capture aerial images. These images are taken as long continuous strips which must be stitched together, colour-balanced, and orthorectified to produce a surface model representation of topographical features. This surface model is distributed to other government departments for identifying flood extent however, this does not provide a discrete boundary with coordinates. Each client performs their own processes for identifying the exact boundary. The aim of this project is to consolidate and standardise this task into a repeatable, accurate, and time-efficient process which will improve the quality of the product offered by DCS and eliminate variance and redundant workloads by DCS's clients.

Programs that are able to generalise relationships between input variables and output variables through inductive inference, whether through supervised, unsupervised, or reinforcement learning algorithms, are categorised as machine learning programs (Guo et al., 2020). Many of the difficulties in contemporary areas of machine learning research deal with defining problems that are simple for people to perform but hard to formally describe because they entail a high level of intuitive and subjective knowledge of the world, such as recognising faces in an image (Goodfellow, 2016). One class of problems that machine learning programs are proving to be effective at solving are classification problems through the use of models such as convolutional neural networks and support-vector machines (Hasan et al., 2019). A machine learning model capable of analysing new images and automatically and accurately classifying specific segments as affected by flood has the potential to greatly assist decision-making capabilities for government authorities managing and monitoring recovery and disaster-relief efforts.

By leveraging machine learning models, this project aims to develop a system capable of accepting new aerial and LiDAR images of the NSW landscape and classifying specific segments as either affected or unaffected by floods. This information will be used to provide accurate flood extent maps to NSW ESOs.

2. Positioning

2.1 Problem Statement

The problem of	providing reliable and expeditious spatial information of flood affected areas
affects	NSW Spatial Services and NSW Emergency Service Organisations
The impact of which is	Implementing disaster-relief and recovery plans based upon current and accurate information
A successful solution would be	A program that automates the process of taking aerial and LiDAR images as input and classifies specific segments as either affected or unaffected by floods, producing a map of peak-flood extent

2.2 Product Position Statement

For	NSW DCS, Spatial Services, ESOs
Who	Need access to reliable spatial information on disaster-affected areas
The	Machine learning model trained in AWS Sagemaker
That	Will be able to take aerial and LiDAR images and classify segments as affected or unaffected by floods, classification will be performed as soon as the images are available
Unlike	Human inspection of images which has a degree of error and lack of time-efficiency
Our Product	Will provide accurate and up-to-date flood extent maps

3. Stakeholder Descriptions

3.1 Stakeholder Summary

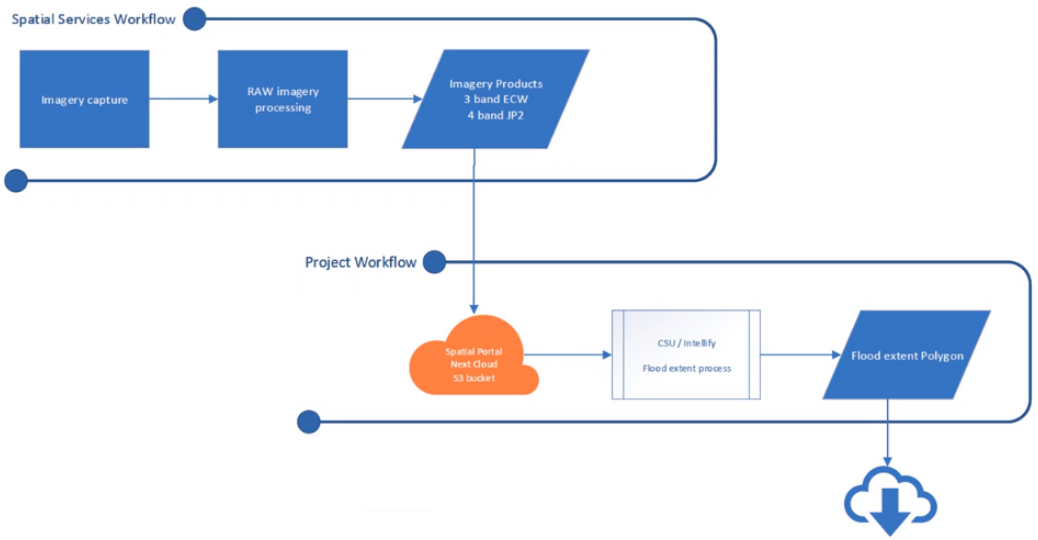
Name	Description	Responsibilities
NSW Department of Customer Service - Spatial Services	Spatial Services is a division of the NSW Department of Customer Service which provides authoritative spatial and land information. Spatial Services has invited Charles Sturt University students for work experience with the Business Technology Services Team until December 2021.	Provide project scope, approach and timelines. Provide roles/responsibilities and governance arrangements. Define high-level requirements of the system. Provide ongoing liaison for project updates. Distribution of final product output by the system.
Intellify	Intellify specialise in the delivery of Artificial Intelligence solutions. Being the second consultant in Australia and New Zealand to be awarded the Amazon Web Services Machine Learning Competency makes Intellify well placed to assist with Machine Learning projects.	Create the AWS environment to use AWS SageMaker and associated AWS services. Upskill Spatial Services & Team 5 in the use of ML/AI. Provide ongoing support to Spatial Services. Outline technical implementation of requirements.
Dr. David Tien	Dr Tien is a Senior Lecturer in Information Technology at Charles Sturt University. He has over 34 years experience in AI related research.	Supervise and Mentor Team 5. Liaise with DCS and Intellify.
Team 5	Team 5 consists of five undergraduate students studying at Charles Sturt University undertaking the ITC 303 and ITC 309 Software Development Project subjects during 2021.	Produce deliverables in line with ITC 303 / 309 requirements and project requirements as requested by DCS.

3.2 User Environment

Certain details of the user environment are yet to be specified, however some general features are established. The DCS users who will utilise the machine learning model to produce the flood extent mapping are familiar with the AWS CLI (command line interface) and ArcGIS (Esri, 2021), a geospatial information system that combines topographic, qualitative, and quantitative data. These users will primarily be situated in the Spatial Services division of DCS. As outlined in the Initial Requirements document, the project is to be developed utilising AWS services to minimise workload and leverage their products to overcome technical difficulties in the machine learning development process. The user environment will therefore consist of the AWS CLI, likely utilising the AWS Lambda (AWS, 2021a) functions as an entry point, AWS S3 Buckets (AWS, 2021b) for storage, and AWS SageMaker (AWS, 2021c) to host the trained machine learning model. The final output will be a surface model object for the ArcGIS software application.

The output of the system will be distributed by DCS to multiple government agencies such as the NSW RFS (Rural Fire Service) and NSW DPI (Department of Primary Industries) and ESOs who will use the flood extent maps for emergency service planning. For example, NSW DPI may use the product to identify stranded herds of cattle and organise food drops. This implies there will be multiple users of the end product with different priorities.

The process of manually processing aerial images into a usable format for ArcGIS applications including image stitching, colour-balancing, and orthorectification takes approximately 48 hours. The system being developed will take these images as its input to generate flood-extent maps as outlined in the figure below.



4. Product Overview

4.1 Needs and Features

The functional requirements and non-functional requirements of the project are outlined in detail in the Initial Requirements Model document. Some of the functional requirements that directly relate to the business needs of the project are briefly outlined below.

Need	Priority	Feature
4.1.1 The system must be able to accept a batch of aerial images and map the extent of flooding within them	Low	AWS Simple Storage Service (S3) for storage and retrieval of images
4.1.2 The system will utilise a machine learning model to classify areas of interest as either flooded or not flooded. **4.1.3** The system will use a process of image segmentation to identify areas of interest within the image for feature extraction.	High High	AWS SageMaker inbuilt machine learning algorithm Use of OpenCV functions for extracting contour lines of flood extent
4.1.4 The system will label the training and validation datasets	Low	AWS SageMaker Ground Truth to be used for dataset preparation
4.1.5 The system will perform feature reduction and data cleaning operations as appropriate before passing the feature vectors into the machine learning model.	Moderate	Use of Jupyter Notebook instance for data reduction and feature extraction within SageMaker Console environment

4.1.6 The output of the system will be a map of the flood extent. The machine learning model will perform classifications at a pixel-by-pixel level. A subsequent process will be necessary to transform this output into a product that can be distributed to clients.	High	To be determined.

4.2 Other Product Requirements

Requirement	Description	Priority
Performance	The output of the system must be accurate. Ideally the model will correctly segment the input images at the pixel level. The resolution of the image will determine the accuracy in terms of metric measurements (e.g. 5m confidence interval).	High
Cost	The development and deployment of the system must not exceed the allocated budget from DCS of $50 000.	Moderate
Security	The storage of Spatial Service data must adhere to the NSW Data & Information Custodianship Policy which outlines data handling and storage requirements for government agencies.	Low
Reliability & Availability	The system will not be in need of continuous use. Usage patterns will depend upon frequency of flood events. However, it is important the system is available and reliable when it is in need as its output will serve as a basis for emergency response planning.	Moderate

References

AWS. (2021a). *AWS Lambda*. https://aws.amazon.com/lambda/

AWS. (2021b). *AWS S3*. https://aws.amazon.com/s3/

AWS. (2021c). *AWS SageMaker*. https://aws.amazon.com/sagemaker/

AWS. (2021d). *AWS SageMaker Ground Truth*. https://aws.amazon.com/sagemaker/groundtruth/

AWS. (2021e). *AWS Command Line Interface*. https://aws.amazon.com/cli/

AWS. (2021f). *AWS CloudTrail*. https://aws.amazon.com/cloudtrail/

Esri. (2021). *ArcGIS Overview*. https://www.esri.com/en-us/arcgis/about-arcgis/overview

Goodfellow, I., Bengio, Y. & Courville, A. (2016). *Deep Learning*. MIT Press

Guo, H., Wu, S., Tian, Y., Zhang, J. & Liu, H. (2021). Application of machine learning methods for the prediction of organic solid waste treatment and recycling process: A review. *Bioresource Technology*. *319*. 124114-124127

Hasan, H., Shafri, H. Z. M. & Habshi, M. (2019). “A Comparison Between Support Vector Machine (SVM) and Convolutional Neural Network (CNN) Models for Hyperspectral Image Classification” In *IOP Conference Series: Earth and Environmental Science* (Vol. 357, No. 1, pp. 012035). IOP Publishing

Jupyter. (2021). *Project Jupyter*. https://jupyter.org/

Title page cover photo.
Milton Coronation Drive flooded. (2011). Retrieved from https://www.abc.net.au/news/2011-06-03/brisbane-floods3a-before-and-after-article/2742794?nw=0