Monitoring of Riparian Infrastructure and Riverine Environment using AI and Air Vehicles

2024, September

Shijun Pan

Graduate School of Environmental and Life Science

(Doctor's Course)

OKAYAMA UNIVERSITY, JAPAN

Monitoring of Riparian Infrastructure and Riverine Environment using AI and Air Vehicles

A Thesis

Submitted to Graduate School of Environmental and Life Science

Okayama University

In partial fulfillment of the requirement for the degree of

Doctor of Engineering

Submitted by

Shijun Pan

Division of Environmental Science

Graduate School of Environmental and Life Science

Okayama University, Japan

2024, September

Title:

Enhancing Riparian Infrastructure and Environmental Monitoring with AI and Air Vehicles

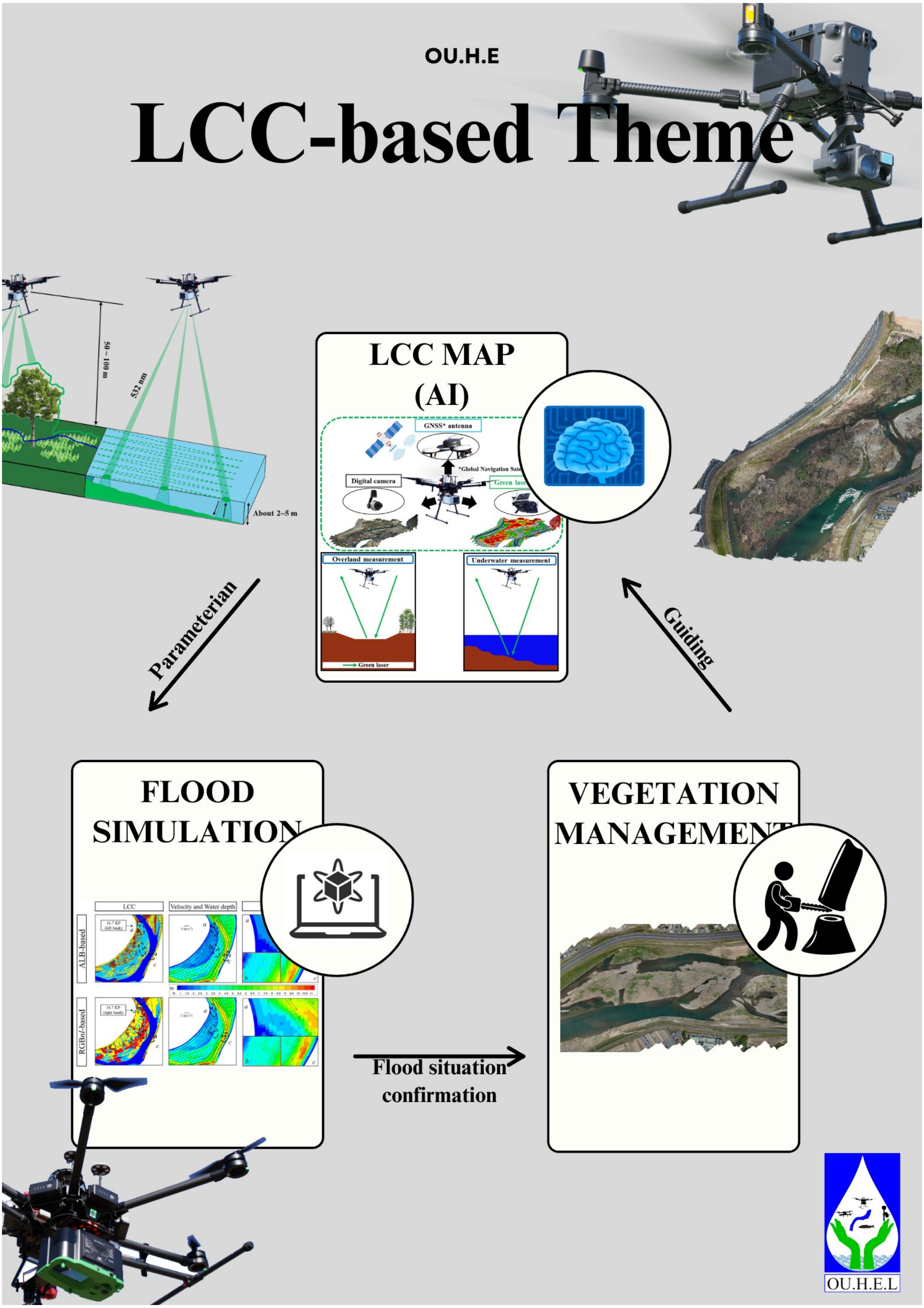
- Chapter 1 General Introduction
- Chapter 2 Application of Digital Camera and AI in Riparian Land Cover Classification (LCC)
- Chapter 3 Application of Deep Learning and Drone Camera in Riparian Area Monitoring (Riverbed Waste Detection)
- Chapter 4 Application of Deep Learning and Drone Camera in Riparian Area Monitoring (Riverbank Topped-paved Crack Detection)
- Chapter 5 Improvements and Suggestions on River Patrolling Methods
- Chapter 6 Conclusion

Land Cover Classification EST. 2020 OU.H.E.L

Riverine Garbage Detection EST. 2022 OU.H.E.L

Road Crack Detection EST. 2022 OU.H.E.L

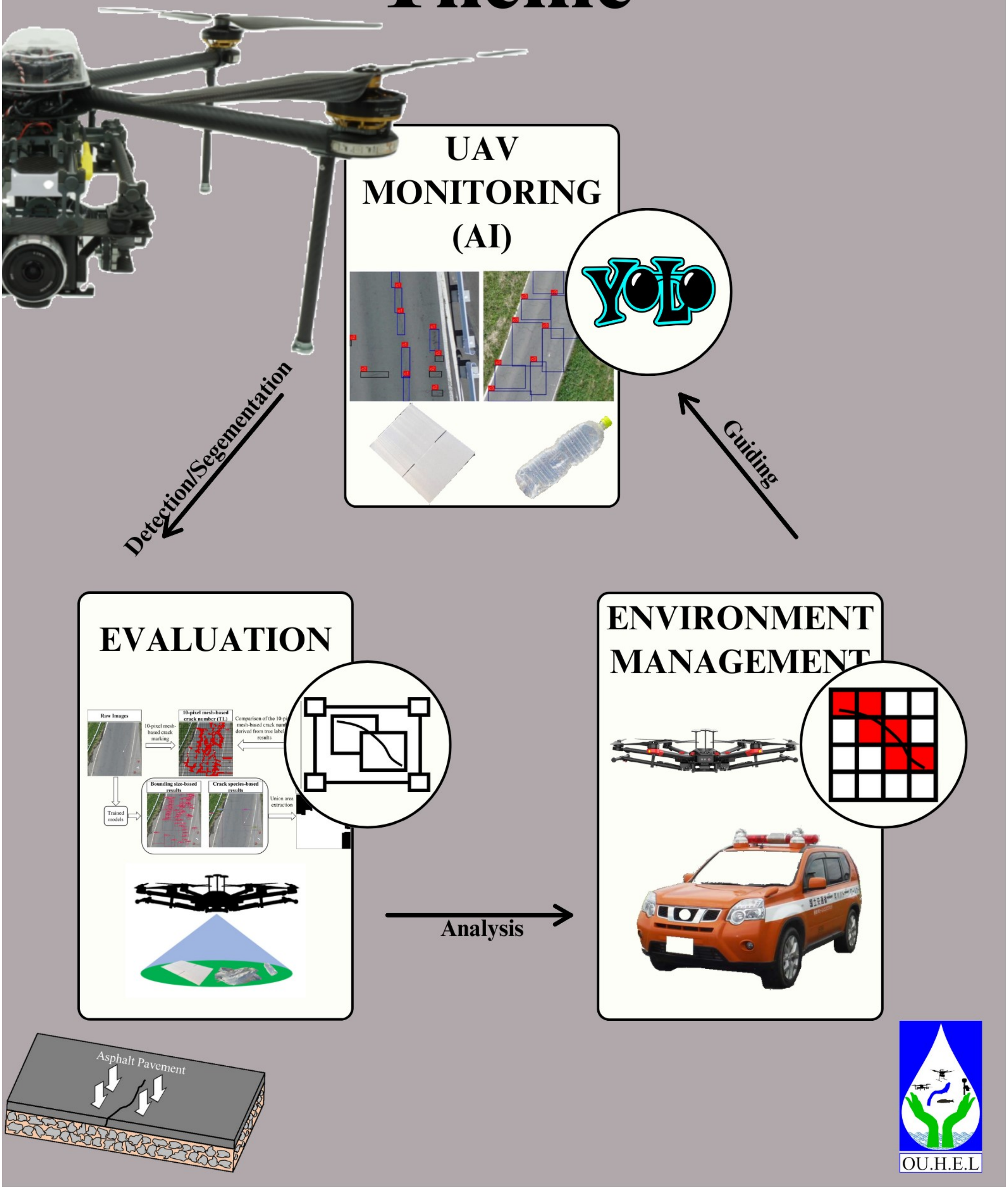
OU.H.E.L

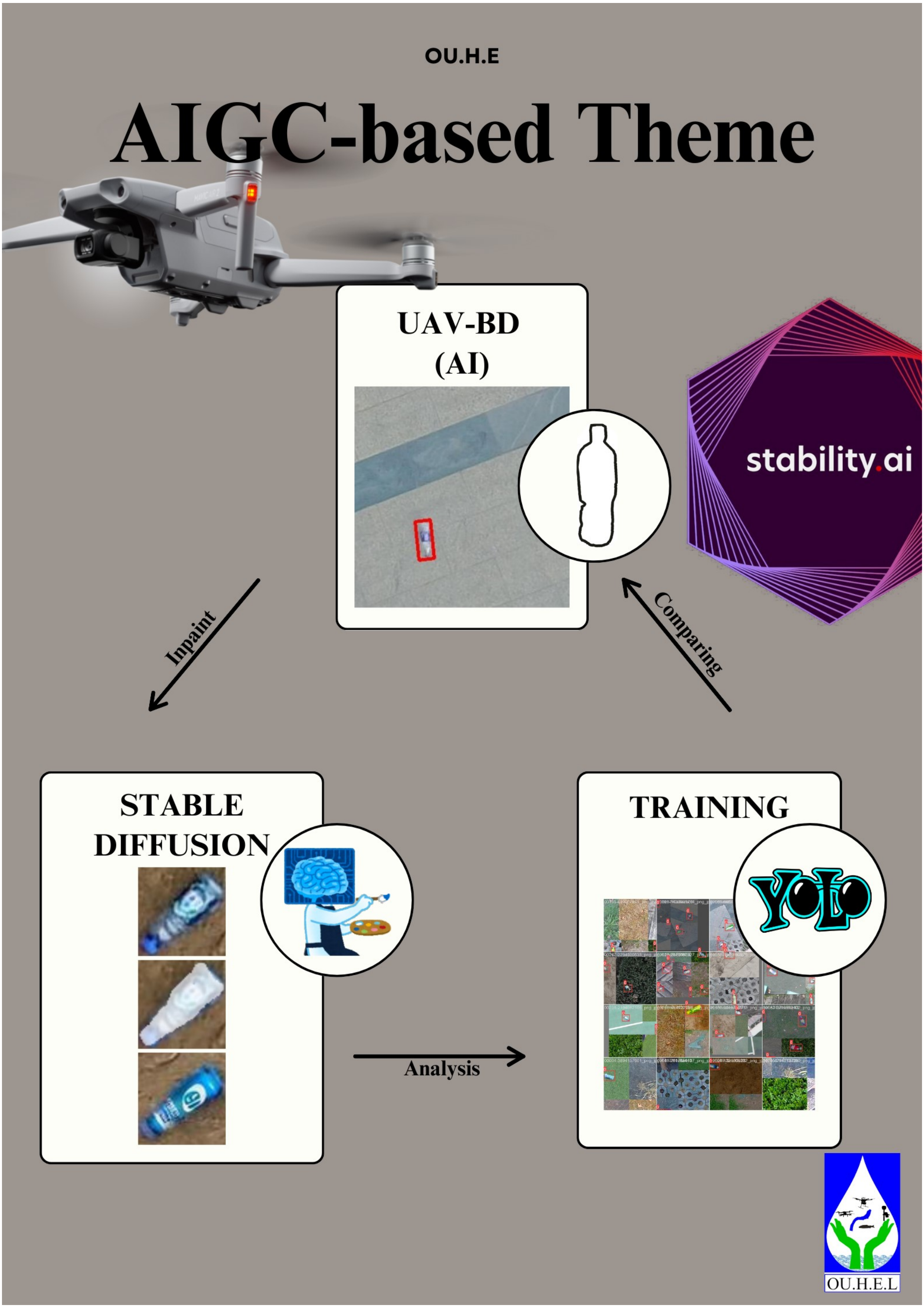


OU.H.E
LCC-based Theme
50 ~ 100 m
532 nm
About 2~5 m
LCC MAP
(AI)
GNSS* antenna
Digital camera
Overland measurement
Underwater measurement
Green laser
Parameterian
Guiding
FLOOD
SIMULATION
LCC
Velocity and Water depth
ALB-based
VEGETATION
MANAGEMENT
Flood situation
confirmation
OU.H.E.L

OU.H.E

UAV-based Monitoring Theme




OU.H.E
AIGC-based Theme
UAV-BD
(AI)
stability.ai
Inpaint
Comparing
STABLE
DIFFUSION
TRAINING
YOLO
Analysis
OU.H.E.L

CONTENTS			PAGES

List of Figures

List of Tables

CHAPTER 1

General Introduction

1.1 Background and Study

In recent years, large-scale floods have occurred one after another. Looking at the past three years, the East Japan Typhoon of 2011, the torrential rains of July 2020, and the heavy rains of August 2021 have caused flooding over a wide area, both internally and externally, and all of them have caused extensive damage **(MLIT, 2020-1)**. In particular, the East Japan Typhoon of 2021 caused record-breaking rainfall over a wide area centering on Shizuoka Prefecture, the Kanto Koshin region, and the Tohoku region, due to the influence of well-developed rain clouds in the typhoon's main body and moist air around the typhoon. As a result, 14 levee breaches occurred on 7 rivers in 6 river systems administered by the national government and 128 levee breaches occurred on 67 rivers in 20 river systems administered by prefectural governments, inundating approximately 35,000 hectares **(MLIT, 2019)**.

Figure 1.1 Damage to the Chikuma River caused by East Japan Typhoon in 2019. (i.e., from Google Earth).

Figure 1.1 shows the damage caused by the Chikuma River. The levee overtopped the river and burst for 70 meters, causing inundation damage. Some of the houses were located in the Hokuriku Shinkansen rail yard. The amount of assets protected by levees has been increasing with urbanization, and this year saw the highest amount of flood damage in history at 2.15 trillion yen **(e-tat, 2019)**.

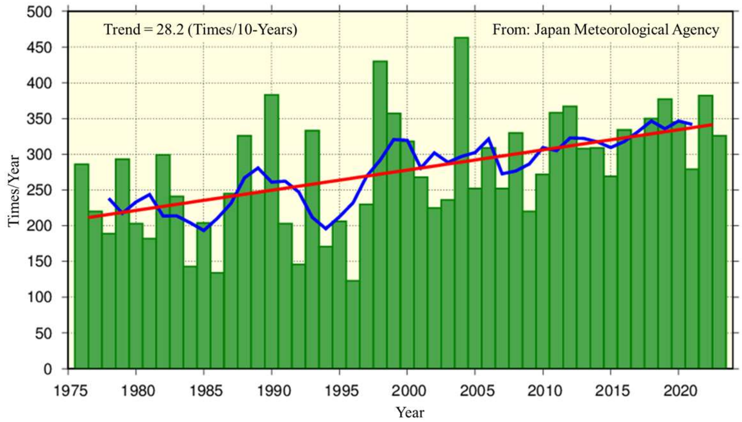


Figure 1.2 Trends in the number of short duration heavy rainfall events exceeding 50 mm in rainfall per hour**.** (i.e., from https://www.data.jma.go.jp/cpdinfo/extreme/extreme_p.html.)

One of the reasons for the increase in flood damage is that the rainfall pattern in Japan is changing. Rainfall in Japan is becoming more localized, concentrated, and intense due to climate change, and the number of short duration heavy rainfall events exceeding 50 mm in duration and heavy rainfall events exceeding several hundred to one thousand mm in total has been increasing. The average number of annual rainfall events of 50 mm or more per hour for the 10-year period from 1976 to 1985 was 226, while the average number for the 10-year period from 2011 to 2020 was 334, indicating an increase of about 1.4 times (**Figure 1.2**). Under the climate change scenario RCP2.6, which assumes that the future temperature increase will be kept below 2°C, it is estimated that rainfall will be 1.1 times higher, river discharge will be 1.2 times higher, and flood frequency will be twice as frequent at the end of the 21st century compared to the end of the 20th century, and there are concerns that water-related disasters will become more frequent and severe **(MOEJ, 2014)**.

Japan has a history of building an agricultural civilization based on rice cultivation around large rivers, and by extension, since the Meiji era, urbanization and industrialization have progressed and population and assets have been located on flood plains. Currently, about 1/2 of the population and 3/4 of the assets are located on flood plains protected by levees. Therefore, river levees are important social infrastructure facilities that protect human lives and assets in the country. Therefore, it is necessary to maintain and manage rivers (river channels and river management facilities) appropriately to ensure that river management facilities such as levees and sluice gates perform their required functions and prevent or mitigate flood damage caused by levee failures and river overflows. On the other hand, many river management facilities were constructed during the period of rapid economic growth, and in recent years, these facilities have been aging and will enter a period of renewal **(MLIT, 2012)**. However, in light of budget cuts due to the future financial situation and the future shortage of staff due to the declining birthrate and aging population, more efficient maintenance and management of river management facilities are required.

The PDCA cycle (**Figure 1.3**) is based on the long-term repeated monitoring of river conditions through river inspections and river patrols, and the analysis and evaluation of findings from these processes, which are then reflected in the river maintenance management plan and its implementation **(MLIT, 2011-1).**

River inspection and river patrol are two of the river maintenance and management services in Japan. The two are clearly distinguished, and river inspections are carried out twice a year for the purpose of understanding the condition of rivers in detail. On the other hand, river patrols, according to the example of river patrol regulations **(MLIT, 2011-2)**, "patrol rivers regularly and systematically as part of river management under normal conditions, detect abnormalities and changes, and generally monitor the river", and play an important role in understanding the ever-changing conditions of rivers. It plays an important role in grasping the ever-changing state of the river. Currently, river patrols are conducted by human eyes on patrol by vehicle, and are basically conducted at least twice a week on large rivers managed by the government. The following are some of the issues with the current method.

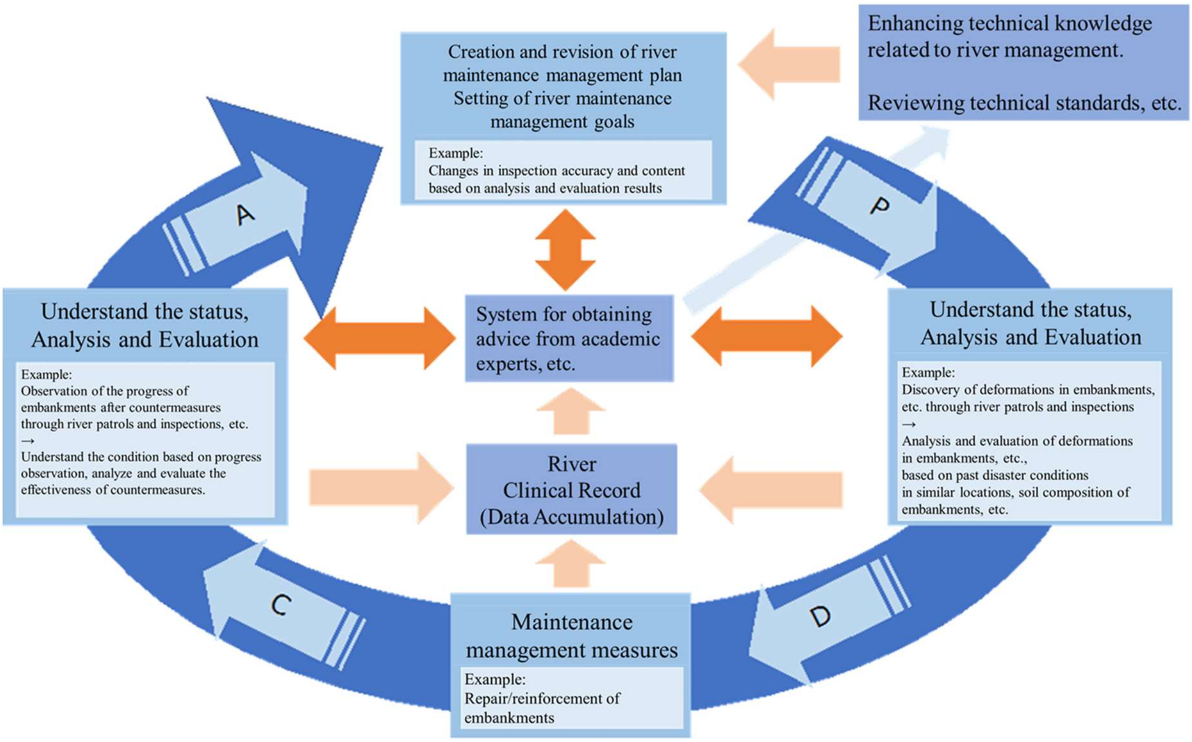


Figure 1.3 Cycle type maintenance management image (i.e., PDCA).

The method and frequency of patrols require a lot of time and effort to patrol long embankments. Lack of consistency in patrols due to qualitative judgments based on the experience of engineers. The risk of overlooking something is considered because only the visible area from the route that can be passed by vehicles, etc. can be checked. In addition to these issues, the recent shortage of engineers and the frequent occurrence of floods make it imperative to improve the efficiency and sophistication of riverine patrol operations.

In the construction industry, the construction DX approach is being used to improve the sophistication and efficiency of operations **(MLIT, 2020-2)**. DX (Digital Transformation) is a term used to distinguish it from conventional digitalization, and its goal is not simply to introduce digital technology, but to establish next-generation systems through the use of digital technology. The aim is not merely to introduce digital technology, but to establish a next-generation system through the use of digital technology. Efforts include i-construction, which introduces digital technology such as BIM/CIM, ICT, and IoT to improve productivity at construction sites, and the use of Artificial Intelligence (i.e., AI), big data management for maintenance management, and drones to improve operational efficiency.

In November 2016, the Ministry of Land, Infrastructure, Transport and Tourism launched the "Innovative River Technology Project" to promote technological development by utilizing superior cutting-edge technologies owned by companies and other organizations in order to solve issues in river administration **(MLIT, 2018)**. So far, the project has developed technologies such as an all-weather drone, a low-cost water level gauge specialized for flood observation (crisis management type water level gauge), and a simple river monitoring camera, which have been implemented at river sites including small and medium-sized rivers. In 2019, as the fifth phase of the project, the advancement of river monitoring was raised as a proposition, and many efforts were made. This study was conducted as part of the fifth phase of the Innovative River Technology Project.

As one of the cutting-edge technologies in the world wide, AI have been practical in the engineering-related research areas that can “pick up” (i.e., detect) or “separate” (i.e., segment) the objects from the scenes **(Zhao *et al.* 2019)**. Considering of the riparian environmental management, the thought-out objects are mainly two types: natural (i.e., tree, bamboo, water area) and artificial ones (i.e., crack and waste pollution). In this research, the author mainly focuses on the solutions of detecting or segmenting the objects from the scenes, and the approach of applying these extracted targets into riparian management.

Based on the previous experience of the riparian management derived from the MLIT in Japan, there are several factors of technologies need to be considered importantly: good cost-performance (i.e., data collection speed per time unit), higher accuracy and easy-to-share. Thus the author developed the following topics to examining the possibility of matching the mentioned factors in the case of applying AI-assisted digital camera:

Topic-1: Application of DeepLabV3+, Ortho-photography and LiDAR in Riparian Land Cover Classification (LCC, mainly natural-related targets)

Topic-2: Application of YOLO and Drone Camera in Riparian Area Monitoring (Crack & Waste Pollution) (mainly artificial-related targets)

The study site in this research is located on the lower Asahi River, a Class I (state-controlled) river with 1810 km2 catchment in Japan, flowing through Okayama prefecture. The average river discharge at the Makiyama hydraulic station, which is at the 20 kilometer post (KP) upstream of the targeted domain, was 57.12 m3 s-1 during 1965–2005 **(MLIT 2007)**. Throughout this research, the KP value means the longitudinal distance (kilometer, km) from the targeted river mouth. Additionally, the riverbed slope is approximately 1:600. The channel width is about 300 m in the targeted reach. More recently, widely diverse vegetation has been observable at the targeted site, which has raised severe concerns about effective flood control and ecosystem management measures.

1.2 Research Objectives and Motivation

Based on the mentioned content, the research objectives are mainly around the natural or artificial objects existing in the riparian environment. And from the management perspective, the author attempt to extract the necessary objects for analyzing, parameterizing or counting. Considering of the motivations for this research around the objectives, there are several tips that the author prefers to list:

Motivation-1: Searching the approach/solution using AI technology to solve the practical riparian management tasks (i.e., derived from the factors shown in **Figure 1.4**).

Motivation-2: Searching the approach/solution on the supplement of digital-camera data for solving the data lack tasks (i.e., derived from the issues shown in **Figure 1.5**).

Motivation-3: Evaluating the possibility of the AI-related application on the riparian management (i.e., mainly derived from the projects shown in **Figure 1.6**).

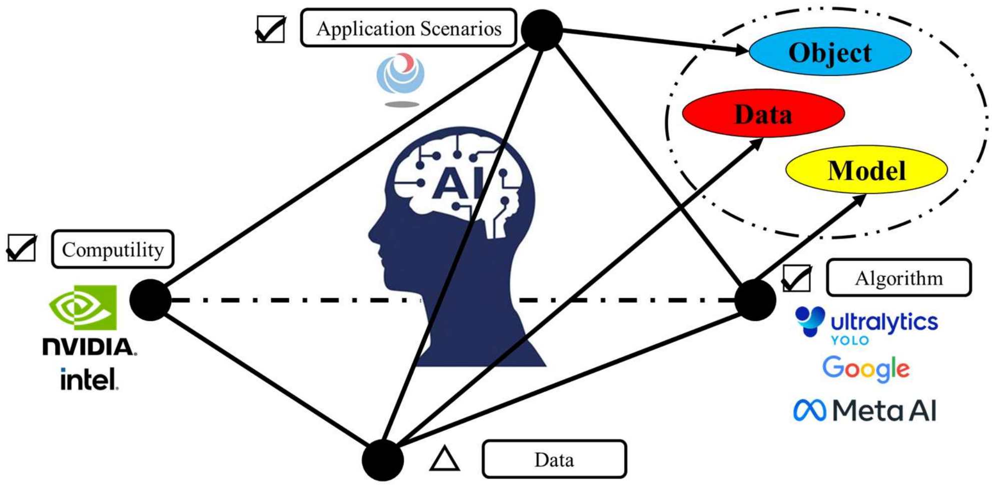


Figure 1.4 Four basic factors for AI: Application Scenarios: Application scenarios are the basic framework for the industrialized implementation of AI, and researchers can match the corresponding resources according to the needs of the scenarios; Data: Data is the source of intelligence for AI. A large amount of high-quality data can be used to train models, such as deep learning models. Currently, data-driven AI can simulate human perception, but it still can't learn knowledge like humans; Algorithm: Algorithm is the fundamental way to realize AI. Algorithms such as deep learning are capable of handling complex tasks such as object recognition, semantic segmentation, and so on; Computing Power: Powerful computing power supports the operation of AI systems. Together, these four elements constitute the AI system, enabling it to play a practical value in different scenarios.

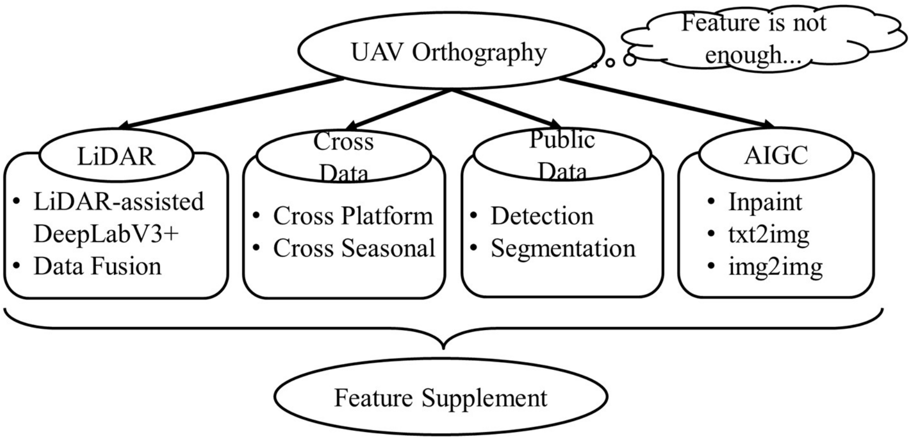


Figure 1.5 Feature supplements of applying different data.

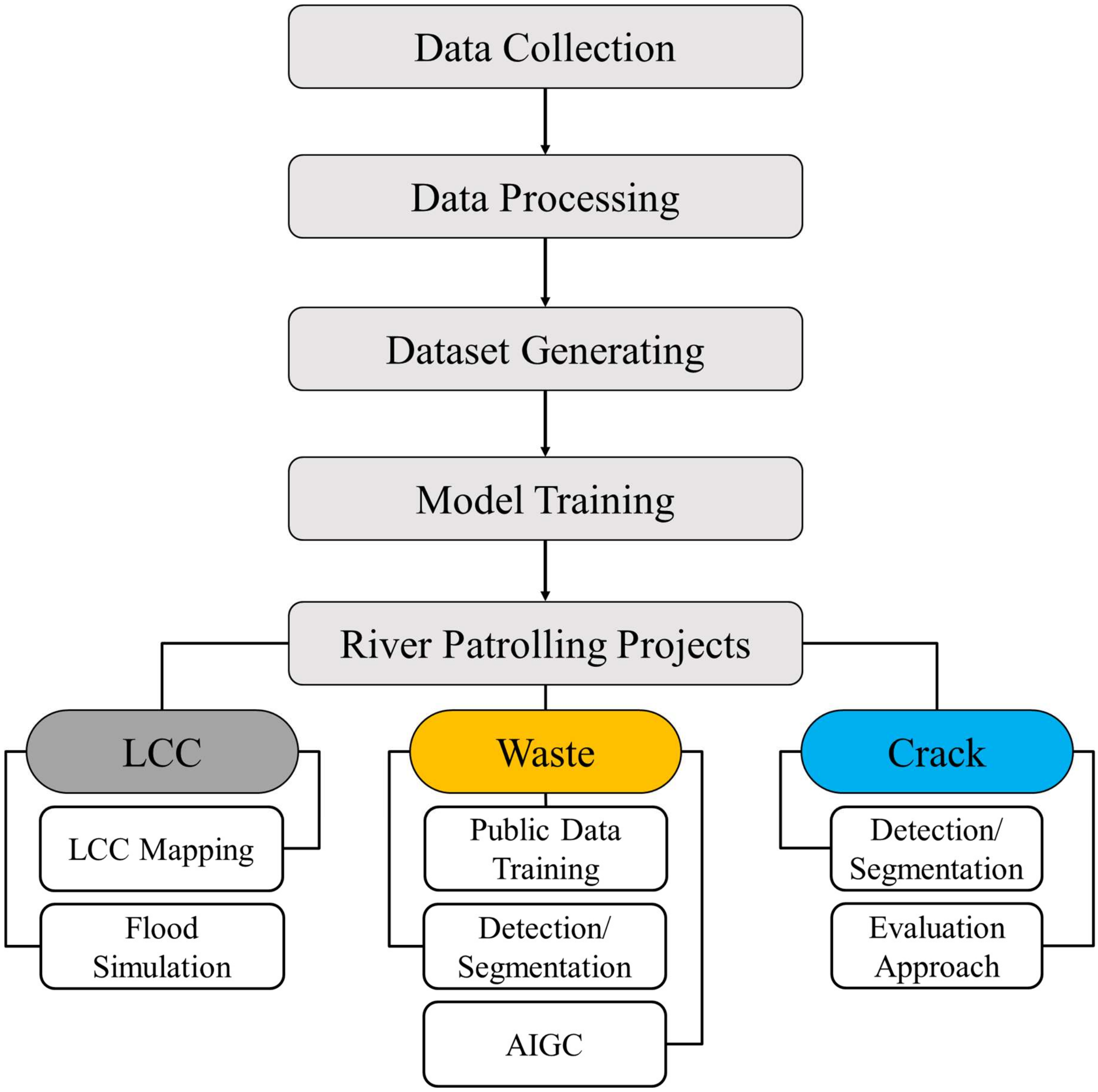


Figure 1.6 Flow chart of river patrolling.

Based on the provided content, the research purposes are shown as follows:

1. Management perspective purpose: This research aims to find several efficient approaches on the multiple riparian management tasks (i.e., infrastructure and natural environment monitoring).
2. AI application perspective purpose: This research seeks to explore and develop AI-based solutions for practical riparian management challenges.
3. Data supplementation purpose: This research attempts to find approaches to supplement digital camera data to address data scarcity issues in riparian studies.

1.3 Datasets Features

In this research, the datasets were mostly collected by the air-platform like airplane or UAV (i.e., unmanned aerial vehicle), and there are common parts based on these responding platforms: If the Orth-photography are in consideration, the files were saved as .tiff (i.e., 3-channel, RGB) format which include high quality color information; The other image-related datasets included the files which used .jpg (i.e., 3-channel, RGB) or .png (i.e., 4-channel, RGBA) format; Except of the .tiff, .jpg and .png, the author also used the LiDAR data (i.e., 1-channel, n or l, which stands for point numbers in per unit and DSM-DTM, respectively) as the supplement for the color information. There were several UAV that have been applied in this research (i.e., Mavic 2 Pro, Phantom 4 Pro and Zenmuse from DJI) derived from different purposes. In this research, even there are several difference derived from parameters setting or hardware, except of the flight height, the author did not take any other parameters setting into considerations.

1.4 Related Models

1.5 DeepLabV3+ Model

DeepLabv3+ **(Chen *et al.* 2019)** is a model which was built upon the previous DeepLab model versions (i.e., v1, v2, v3) derived from the researchers at Google AI. DeepLabv3+ aims to address the semantic segmentation tasks, which need to assign the class label to each pixel in the input image. DeepLabv3+ include several innovations, i.e., Encoder-Decoder Architecture, Atrous Convolution, Atrous Spatial Pyramid Pooling (ASPP), Encoder-Decoder Feature Fusion and Depthwise Separable Convolutions. Until now, DeepLabv3+ was evaluated on various benchmark datasets, i.e., PASCAL VOC, Cityscapes, and ADE20K. DeepLabv3+ achieved the state-of-the-art results when it was publicized. And this model has been already applied in various domains, including but not just autonomous driving, medical imaging, and remote sensing

1.6 YOLO series Model

The introduction of YOLOv5 **(Ultralytics, 2020)**, YOLOv7 **(Wang, 2023)**, and recent YOLOv8 **(Ultralytics, 2023)** represent major developments in the YOLO (You Only Look Once) series of object-based detectors/segmentors in recent years. Compared to the Darknet implementation in the past time, YOLOv5 represented a transition to the PyTorch framework, offering improved ecosystem integration and deploy-ability. The automated anchor box learning was also added for adjusting to different datasets. To increase precision and speed, YOLOv7 suggested architectural changes such as the E-ELAN backbone, compound model scaling, re-parameterization planning, and auxiliary heads. While not intended for CPU-based mobile deployment, YOLOv7 beat earlier YOLO models across all of its variations on the MS COCO dataset, reaching high accuracy and speeds ranging from 5-160 FPS. YOLOv8, the recent version, demonstrated higher throughput with comparable parameter counts when it was published in January 2023. YOLOv8 appears to be focused on restricted edge device deployment with high inference speed based on testing. Future versions should see further architectural changes and performance improvements as this growing series pushes the limits of real-time object identification.

1.6.1 Stable Diffusion Model

Stable Diffusion (i.e. SD) is an artificial intelligence based image generation model **(Stability AI, 2022; AUTOMATIC1111, 2024)** which involves different components i.e. text encoder, image generator and diffusion process. In this diffusion process, SD requires gradually adding noise to the image for training a noise predictor. And this predictor not only predicts but also removes the noise from the noisy image. And the makers of SD use a compressed image "latent space" representation in SD instead of using the full pixel space in order to improve the efficiency of the operation.SD also employs techniques such as classifier-free guidance, which utilizes the text prompts themselves to guide the diffusion process to the desired output. In addition to the above features, SD includes several additional features such as image-to-image generation, inpainting, and depth-to-image generation

1.6.2 Segment Anything Model

The Segment Anything Model or SAM **(Kirillov *et al.* 2023)** from Meta's FAIR lab represents a substantial improvement in computer vision (i.e., segmentation tasks). SAM is able to use promptable segmentation tasks as a foundation model, and it also allows the users to separate objects with high accuracy utilizing user interfaces (i.e., text prompts, bounding boxes, or point clicks). The architecture of SAM includes an image encoder, a prompt encoder, and a mask decoder for producing the exact segmentation masks. And SAM was trained using the dataset called SA-1B, which includes 1 billion masks derived from 11 million pictures. One of SAM's key features is zero-shot, that SAM can segment the input images without any additional training. SAM have the flexibility in solving several practical tasks i.e., mask generation (Everything mode), text-to-mask conversion (prompt mode). SAM is also an innovation in a range of industries, including retail, medical imaging, agriculture, and others. As an open-source model, this model improves the annotation's accuracy and effectiveness in subsequent work.

References (Chapter 1)

AUTOMATIC1111, 2024. https://github.com/AUTOMATIC1111/stable-diffusion-webui

Chen L. C., Papandreou G., Kokkinos I., Murphy K. & Yuille A. L. 2017 DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFS. IEEE Transactions on Pattern Analysis and Machine Intelligence 40 (4), 834–848. https://doi.org/10.1109/TPAMI.2017.2699184.

e-stat: Water Damage Survey, 2019. https://www.e-stat.go.jp/stat-search/files?page=1&toukei=00600590&tstat=000001150926

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A. C., Lo, W. Y., Dollár, P & Girshick, R. Segment anything. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 4015-4026). 2023.https://openaccess.thecvf.com/content/ICCV2023/papers/Kirillov_Segment_Anything_ICCV_2023_paper.pdf

Ministry of Land, Infrastructure, Transport and Tourism (MLIT): River Flow Situation of the Asahi River and Water Quality, 2007. https://www.mlit.go.jp/river/basic_info/jigyo_keikaku/gaiyou/seibi/asahi_index.html.

Ministry of Land, Infrastructure, Transport and Tourism (MLIT): MLIT River Sand Prevention Technical Standards Maintenance Management, 2011. https://www.mlit.go.jp/river/shishin_guideline/gijutsu/gijutsukijunn/ijikanri/

Ministry of Land, Infrastructure, Transport and Tourism (MLIT): Example of River Patrol Regulations,2011. https://www.mlit.go.jp/river/shishin_guideline/kasen/pdf/kasen_junsikiteirei_h230511.pdf

Ministry of Land, Infrastructure, Transport and Tourism (MLIT): Towards Strategic River Maintenance and Management, 2012. https://kenmane.kensetsu-plaza.com/bookpdf/163/fa_03.pdf

Ministry of Land, Infrastructure, Transport and Tourism (MLIT): Innovative River Technology Project, 2018. https://www.mlit.go.jp/river/gijutsu/inovative_project/project1.html

Ministry of Land, Infrastructure, Transport and Tourism (MLIT): Damage caused by Typhoon No. 19 in 2019. https://www.mlit.go.jp/saigai/saigai_191012.html

Ministry of Land, Infrastructure, Transport and Tourism (MLIT): Flood Damage Report 2020, p6. https://www.mlit.go.jp/river/pamphlet_jirei/pdf/suigai2020.pdf

Ministry of Land, Infrastructure, Transport and Tourism (MLIT): Promotion of DX in the MLIT, 2020. https://www.jacic.or.jp/kenkyu/22/data/r02_6_hirose.pdf

Ministry of the Environment, Japan (MOEJ): Summary of the IPCC Fifth Assessment Report, 2014. https://www.env.go.jp/earth/ipcc/5th/pdf/ar5_wg1_overview_presentation.pdf

Stability-AI, 2022. https://github.com/Stability-AI/stablediffusion

Ultralytics, 2020. https://github.com/ultralytics/yolov5

Ultralytics, 2023. https://github.com/ultralytics/

Wang, C. Y., Bochkovskiy, A., & Liao, H. Y. M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for realtime object detectors. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp.7464-7475). 2023. https://doi.org/10.48550/arXiv.2207.02696

Zhao, Z. Q., Zheng, P., Xu, S. T., & Wu, X. Object detection with deep learning: A review. *IEEE transactions on neural networks and learning systems*, *30*(11), 3212-3232. 2019. https://doi.org/10.1109/TNNLS.2018.2876865

My PhD research focused on the following specific goals.

- To develop a novel methodology (i.e., LiDAR-assisted DeepLabV3+ Model) of applying the Airborne LiDAR Bathymetry (i.e., ALB) data on improving the Aerial Photography feature (i.e., the difference between bamboo and tree);
- To understand the distribution of the riparian land cover in the study site with the

assistance of air-platformed Digital Camera and LiDAR using AI technology, which improve from the previous methodology with higher accuracy on land cover classification (i.e., LCC) task;

- To apply the improved LCC mapping on the 2018 Asahi Flood Simulation and optimized the water level result for around 1m from the previous simulation results;
- To apply the methodology derived from LiDAR-assisted DeepLabV3+ Model on the UAV-platformed Digital Camera and LiDAR or Green LiDAR System (i.e., GLS);
- To develop a novel methodology of fusing Aerial Photography and LiDAR, that used high contrast color scale for well-performance on expand the data feature;
- To discuss the possibility of the exchangeability of using the ALB and GLS to each other for the LCC tasks;
- To apply the open source object detection/instance segmentation model on the UAV-based asphalt-paved cracks detection/segmentation tasks;
- To develop a novel evaluation methodology of assessing the remote sensing-based crack detection/segmentation;
- To edit a manual of how to apply the object detection/instance segmentation models ob the crack detection/segmentation tasks;
- To apply the open source object detection model on the UAV-based waste pollution detection tasks;
- To develop a novel methodology of creating the AI-based Generative Content (i.e. AIGC) for increasing the data amount;
- To prove the possibility of applying the AIGC on the object detection task for the Augmentation on waste pollution detection task;
- To prove that with the assistance of the AI-powered UAV technology the river patrolling tasks can be improved the effectively and efficiently.

CHAPTER 2

Application of Digital Camera and AI in Riparian Land Cover Classification (LCC)

2.1 Airborne-derived Orthophotograph aided with Airborne LiDAR on LCC

2.1.1 Introduction

In recent years, climate change has led to frequent extreme and record-breaking flooding events worldwide. For instance, in mid-July 2021, European floods not seen in decades ravaged Germany, Belgium, and the Netherlands, killing hundreds of people and inundating villages and towns. Furthermore, China has been at high risk from disastrous flooding in 2020 **(Wei *et al.*, 2020)**. In fact, a flood struck Henan province of China in mid-July 2021. The severe rainfall constituted an average year's amount, but falling during just three days. It is noteworthy that, because of recent climate change, an extreme rainfall event struck western Japan, our study region, in early July 2018, causing flooding and sediment damage, inundating residential areas, and killing 81 people in Okayama prefecture (**Yoshida *et al.*, 2021**). Researchers today are constantly confronted by new challenges posed by unprecedented river floods in such an ever-changing global hydrological environment (Global Floods 2021). Although riverbed excavation and embankment upgrading can be effective flood mitigation methods to address river flood control issues, the river's flow capacity in the current state must be assessed appropriately before either of these engineering terrain modifications can be implemented efficiently.

The important hydraulic engineering task of assessing flood capacity is based primarily on cross-sectional area changes and flow resistance factors (**Shih & Chen 2021**). In recent studies (e.g., **Dimitriadis *et al.*, 2016**), researchers demonstrated through benchmark simulations that the variability and uncertainty of flood propagation are primarily caused by channel geometry and roughness as compared to other factors such as inflow, longitudinal gradient, floodplain roughness, model structure, etc. On-site bed level surveys and land cover classification (LCC) mapping both play fundamental roles in quantifying such crucial parameters as attributable flow resistance. Especially for the case of vegetated streams, Green (2005) reported that total flow resistance is affected by Manning's roughness coefficients of the following factors: riverbed materials, surface irregularities, shape and size of the channel cross-section, obstructions, vegetation, channel meandering, and so on. In addition, **Nikora *et al.* (2008)** demonstrated that, in addition to stream dimensions, an excellent parameter for estimating hydraulic resistance is the spatial distribution of plant patches. However, regular field surveys of riparian vegetation properties are traditionally required for flood control exercises (**Sun *et al.*, 2010**). In earlier cases, several river management projects were conducted, but strong emphasis was not placed on the spatial distribution of vegetation species and their height, although this task is now regarded as important in balanced river management (**Nepf 2012**). One practical approach for quantifying river channel and floodplain roughness is to use reference values related to flow resistance based on visual confirmation of aerial photographs, considering all factors affecting flow resistance (e.g. **Chow 1959**). However, this time-consuming, unrepeatable, and laborious method of actual measurement has been demonstrated to have limitations for large-scale use. Accordingly, accurate LCC mapping, including information of vegetation attributes, is necessary for balanced river management including measures such as flood risk and ecosystem management.

Over the years, remotely sensed technologies have proven to be effective for application to riparian vegetation attribute surveys, relying on the acquisition of stable digital surface model (DSM) and digital terrain model (DTM) data. For instance, **Mason *et al.* (2003)** used airborne laser scanning (ALS) to derive riparian vegetation heights for floodplain friction parameterization in hydrodynamic modelling. In addition, **Straatsma and Baptist (2008)** evaluated an ALS-based approach to derive hydrodynamically relevant surface features using multispectral data, demonstrating the importance of ALS for mapping vegetation height and for density attribution. Furthermore, Vetter et al. (2011) used dense ALS point cloud data to investigate the vertical vegetation structure for determining hydraulic roughness. More recently, airborne laser bathymetry (ALB) systems have been applied to acquire vegetation height and topo-bathymetric data (**Yoshida *et al.*, 2020**), demonstrating superiority in collecting bed elevation data from submerged areas. Results show that ALB enables data collection on both land and underwater areas simultaneously. Although the ALB system has advantages for collecting underwater data, it also has shortcomings for collecting terrestrial data. For example, in dense vegetation, the near-infrared (NIR) laser used in ALB can penetrate the ground surface only ineffectively, resulting in a lack of laser points on the underlying structure (**Tian *et al.*, 2021**). Because of this phenomenon, it is difficult for an unsupervised method using ALB data to distinguish between detailed species (e.g., woody vegetation and bamboo grove in our targeted vegetation). Furthermore, this method commonly depicts LCC mapping based on each specifically sized grid, such as the 2 m grid used in an earlier study (Yoshida et al. 2020), with no regard for the LCC of the surrounding grid. Such limitation produces a "salt and pepper" effect (**Yu *et al.*, 2006; Blaschke *et al.*, 2000**) in LCC mapping, which strongly affects the accuracy of LCC prediction results. Furthermore, the existing method of manually setting thresholds (**Do *et al.*, 2019; Yoshida *et al.*, 2020**) for various parameters used in LCC (e.g., voxel-based points and vegetation height considered in our current study as discussed hereinafter) has drawbacks because the criteria or proposed values might differ greatly with different remotely sensed datasets. Because of these issues, it is crucially important to establish an appropriate approach to improve the accuracy of LCC predictions using ALB data.

More recently, a few fluvial researchers attempted to classify riparian vegetation with machine learning (ML) methods using only lidar point cloud (e.g. **Fehérváry & Kiss 2020**). For that approach, they used decision trees to identify each cell's land cover. Then they compared the results with a field survey of randomly selected cells. Although they identified LCCs with acceptable accuracy, the ML method used larger two-dimensional (2-D) cells (e.g. 15 m × 15 m) to read the object's features, although each cell accommodated only one label for classification. In such a case, they might have a risk of missing out on other important land cover information. To overcome limitations of the earlier study, we intend to use a smaller square mesh of around 2 m in our current study. Furthermore, **Carbonneau *et al.* (2020)** attempted to assess LCC such as water, dry exposed sediment, green vegetation, senescent vegetation, and roads using red–green–blue (RGB) images from 11 rivers in different countries based on a modified model: "convolutional neural network-supervised classification." Their findings with higher identification accuracy might be beneficial for ecological conservation in fluvial environments. However, they did not test their results for flow-resistance parameterization attributable to riparian vegetation, which we are particularly addressing here for river flood flow simulation. In addition, the ML technique demonstrates its shortcomings in pixel-based image classification (i.e. LCC mapping), where complicated feature extraction is necessary (**Dargan *et al.*, 2019**). Furthermore, manual feature extraction (e.g. data analysis, interpretation) is necessary for ML methods, whereas automatic feature extraction functions have been used widely for deep learning (DL) (a type of ML) models in recent years, particularly for models with encoder–decoder modules. For instance, the DL image processing techniques of DeepLabv3+ (**Chen *et al.*, 2018**) have demonstrated their benefits in overcoming challenges in semantic segmentation, while classic U-Net (**Ronneberger *et al.*, 2015**) can also extract features from images using the familiar encoder-decoder structure as DeepLabv3+. However, the DeepLabv3+ model can extract features more efficiently when assisted by the atrous spatial pyramid pooling module (Chen et al. 2017). Therefore, for the first time, we used the DeepLabv3+ model (RGB) in conjunction with ALB-derived voxel-based laser points and vegetation height information to infer LCC mapping, demonstrating the novelty of our current study.

In early July 2018, our targeted study site, the vegetated lower Asahi River, Okayama prefecture, Japan, details of which are described hereinafter, was struck by extreme flooding with discharge of approximately 4500 m3 s-1. Because of riparian vegetation such as heavy density of woody and bamboo groves, the studied river reached record water levels (**Yoshida *et al.*, 2021**). Therefore, appropriate LCC mapping is necessary to estimate flow-resistance parameters attributable to riparian vegetation in flood modelling. In light of the issues described earlier, this study was conducted to examine a proposed DL-based methodology for LCC mapping in riparian areas considering ALB-derived voxel-based laser points and vegetation height. In the current study, ALB measurements include overland and underwater bed elevation surveys using near-infrared and green lasers. In addition, during the lidar campaigns, we captured aerial photographs and leveraged the RGB information to assess LCC using the current DL approach. Consequently, airborne surveys assist us in determining the flood flow capacity by providing bed elevation data and generating LCC mapping for use as inputs in numerical simulation. However, the new LCC mapping approach presented herein, is particularly expected to perform better than earlier unsupervised methods (clustering) at distinguishing the most dominant riparian vegetation species (i.e. woody vegetation and bamboo grove) in our targeted area. Finally, the proposed LCC reasonably estimated spatially distributed hydrodynamic roughness in the 2018 Asahi River flood modelling. Overall, this study is expected to aid policymakers in developing a balanced scenario for both flood control and ecosystem management tasks while considering riparian LCC.

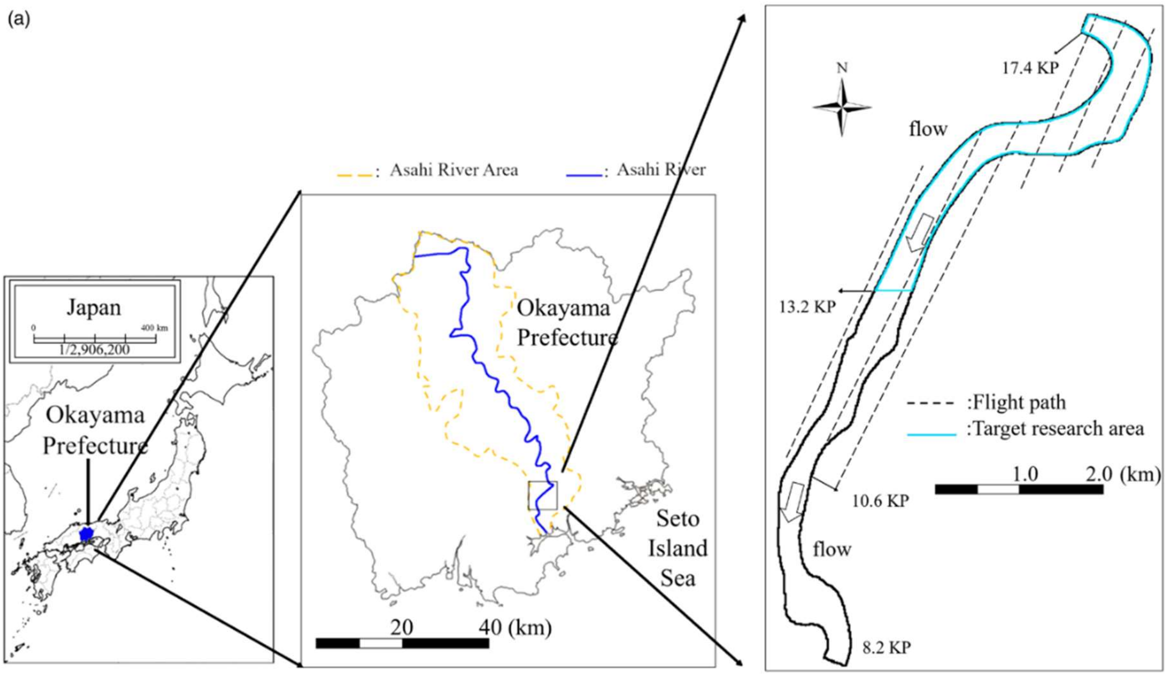
2.1.2 Study site

Figure 2.1a depicts our study site, which is located on the lower Asahi River, a Class I (state-controlled) river in Japan, flowing through Okayama prefecture into the Seto Inland Sea. The catchment area of the targeted river is 1810 km2. The average river discharge at the Makiyama hydraulic station, which is at the 20 kilometer post (KP) upstream of the targeted domain, was 57.12 m3 s-1 during 1965–2005 (**MLIT 2007**). Throughout this study, the KP value denotes the longitudinal distance (kilometer, km) from the targeted river mouth. Furthermore, the riverbed slope is approximately 1:600. The channel width is about 300 m in the targeted reach. The targeted domain was 13.2–17.4 KP, as shown in **Figure 2.1a (right)**, for both the LCC and flood simulation cases. Furthermore, more recently, widely diverse vegetation has been visible at the targeted site, which has raised severe concerns about effective flood control and ecosystem management measures. Irrespective of those concerns, the riparian vegetation for the current LCC study is divisible into three types based on flow resistance: bamboo grove, herbaceous species (grass), and woody species (tree). **Figure 2.1b** represents the dense situation of riparian vegetation in our targeted river, which must be trimmed in a planned manner for flood control tasks, whereas riparian environment management, such as wildlife conservation, must be considered.

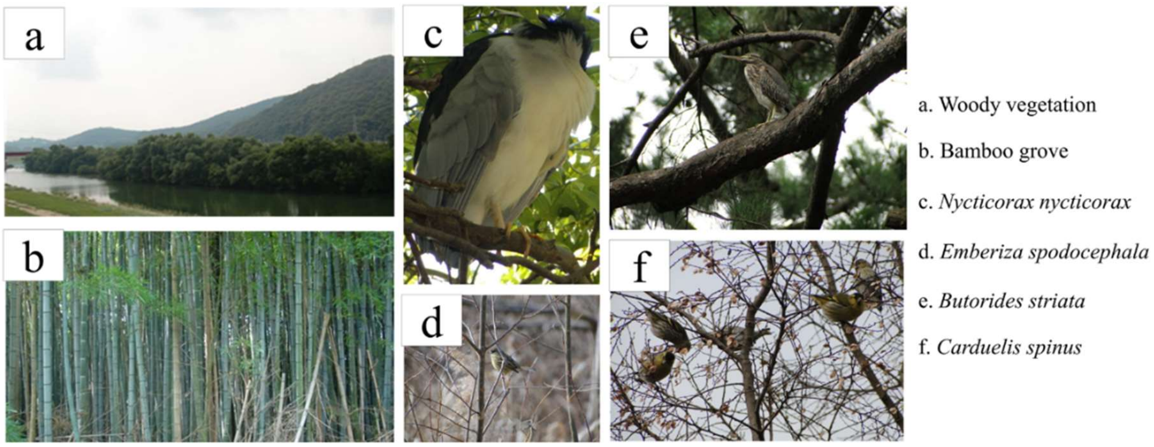
2.1.3 Data collection and processing

2.1.3.1 Data collection

For this study, we conducted ALB (Leica Chiroptera II; Leica Corp.) surveys in March, July, and November 2017 along a 6.2-km reach of the lower Asahi River (10.6–17.4 KP) controlled by the national government. As shown in **Figure 2.1** (right), multiple flight operations were conducted in both leaf-off (March and November 2017) and leaf-on (July 2017) conditions to achieve overlapping coverage of the target area. The current system scanned the river channel for LCC using aircraft-mounted near-infrared and green lasers (**Figure 2.2**).



(a)



(b)

Figure 2.1 Perspective of targeted research area: (a) location of the Asahi River in Japan with the kilometer post (KP) values representing the longitudinal distance (km) from the river mouth and (b) vegetation and birds' species in the targeted research area.

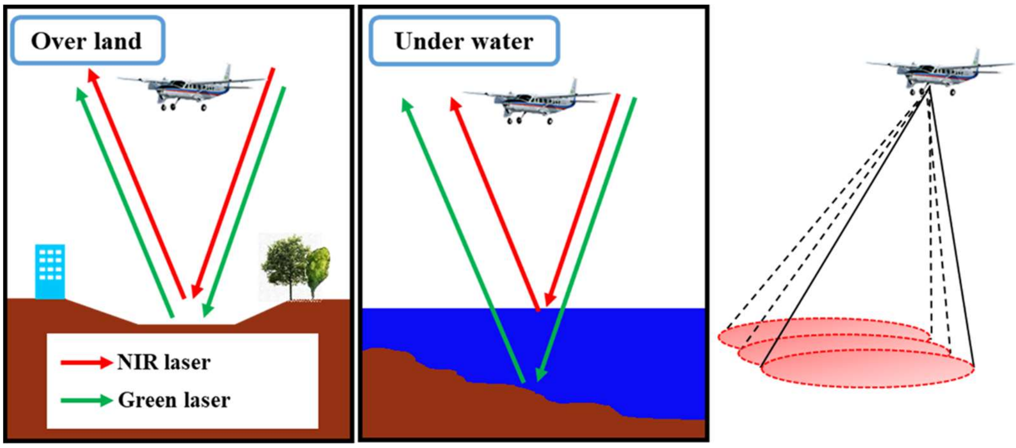


Figure 2.2 Airborne laser bathymetry system using a NIR laser for overland surveys and a green pulsed laser for underwater surveys.

Table 2.1 Specifications of the present ALB system and measurement conditions in the targeted river reach.

Item			Measurement date of ALB & Aerial photograph		
			Mar. 2017	Jul. 2017	Nov. 2017
Equipment specifications	Laser wavelength range (nm)	NIR*	1,064	1,064	1,064
		Green	515	515	515
Measurement specifications	Number of laser beams (s^{-1})	NIR	148,000	148,000	148,000
		Green	35,000	35,000	35,000
	Ground altitude (m)		500	500	500
	Flight speed (km h^{-1})		220	220	111
	Density of measurement points (m^{-2})	NIR	9.0	9.0	18.0
		Green	2.0	2.0	4.0
Photograph specifications	Resolution (cm $pixel^{-1}$)		10	10	10
Water quality	Turbidity** (degree***)		2.9	3.8	3.2

*: Near-InfraRed; **: Ministry of Land, Infrastructure, Transport and Tourism hydrological water quality database (Asahi River, Otoide Weir); ***: One degree of Japan Industrial Standard (JIS K0101) is the same as when 1 mg of standard substance (kaolin or formazine) is contained in 1 L of purified water.

The device commonly uses the green laser to detect underwater (bottom) surfaces because green light can penetrate the water column to some degree. By contrast, the near-infrared laser is used to detect terrain surfaces, including vegetation identification, because it is readily reflected by the air–water interface. The laser beam of this measurement device is specially processed, considering the refraction angle of the green laser at the air–water interface, so that the laser incident at the air–water interface has an elliptical footprint (**Figure 2.2**). Moreover, during each ALB measurement, a digital camera mounted directly beneath the aircraft took aerial photographs of the target river. **Table 2.1** shows specifications of the equipment, measurement parameters, and river water quality at the time of measurements. Because the magnitude of turbidity in a river can strongly affect the amount of light incident into the water column, its value was confirmed before each ALB measurement. The water quality of the three target periods was reasonable for measuring the underwater terrain surface.

2.1.3.2 Data processing

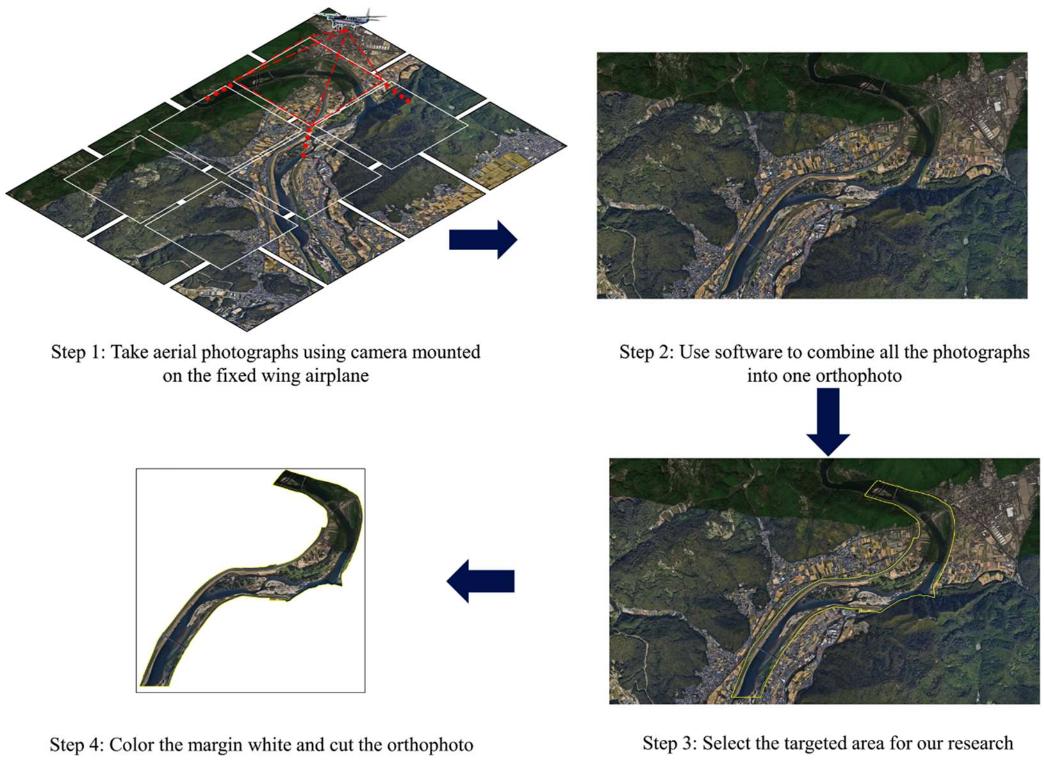


Figure 2.3 Ortho-aerial photograph operation steps.

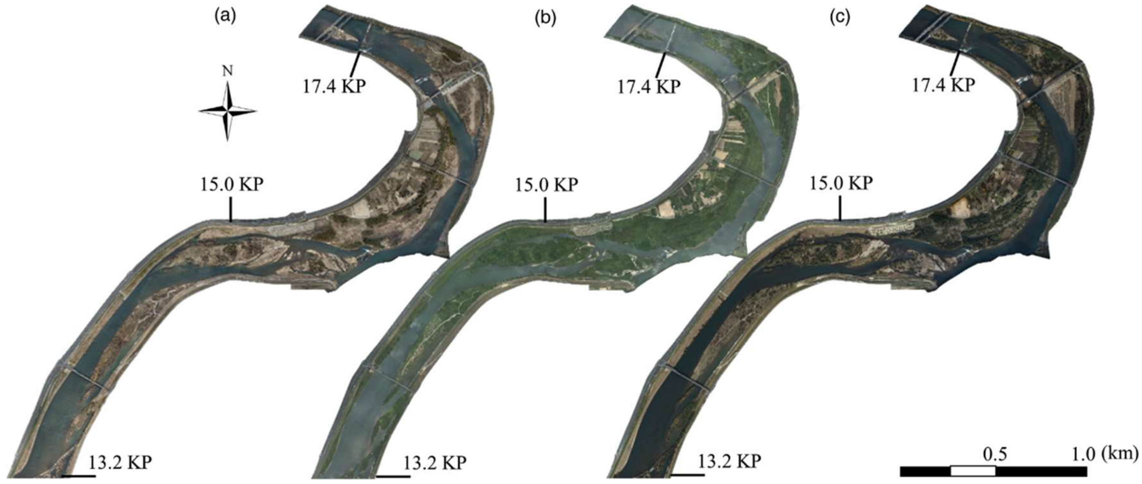


Figure 2.4 Ortho-aerial photographs of the targeted area in (a) March, (b) July, and (c) November 2017.

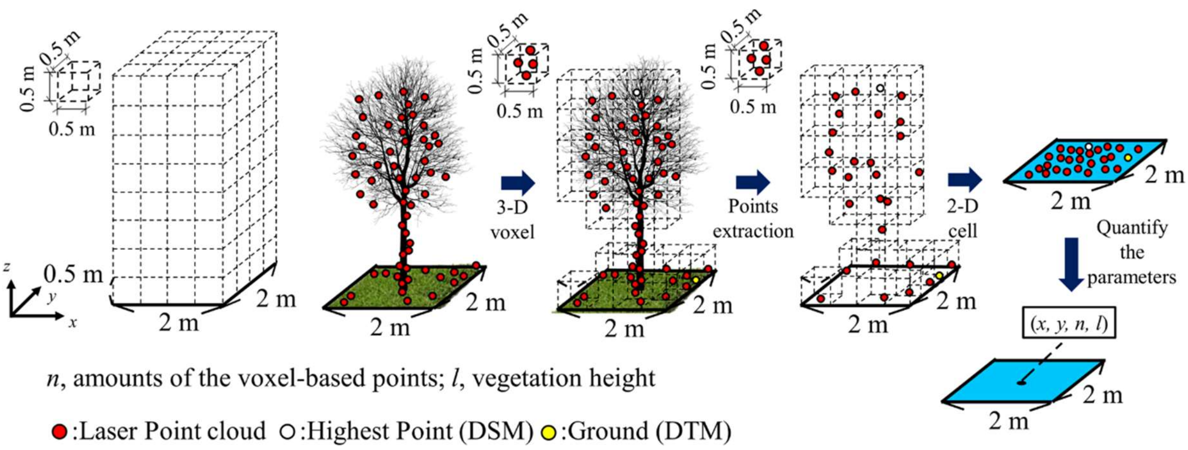


Figure 2.5 Voxel-based ALB data processing.

To remove tilt and relief effects, the aerial photographs were converted to orthophotos, as shown in **Figure 2.3**. Herein, the aerial photographs' overlap and side-lap ratios were respectively greater than 60% and 30%. As shown in **Figure 2.4**, aerial photographs from 13.2 KP to 17.4 KP captured during the targeted three periods were processed sequentially using the four steps depicted in **Figure 2.3**. **Figure 2.5** shows the ALB data processing, beginning with establishment of a Cartesian grid in the target domain comprising three-dimensional (3-D) voxels. Each voxel, which has 0.5 m side length, can only hold one laser point data point using a filter to maintain uniform laser point density. In other words, we kept only the highest one from the ALB measurement rather than all the points for each voxel

(**Yoshida *et al.*, 2017**). For use as a parameter in subsequent 2-D flood simulations, a horizontal 2-D cell that can include all laser points in the processed 3-D voxel was created. The points in each 2-D cell are designated as voxel-based points (n). We identified the ground (riverbed or digital terrain model, DTM) after such processing by filtering the point cloud data near the lower part of the 2-D cell. Later, we calculated the vegetation height (l) by locating the highest point in each 2-D cell (digital surface model, DSM) after subtracting the DTM.

Table 2.2 LCC using ALB data with unsupervised method named as ALB-based LCC method.

LCC	*n*	*l*
Bare ground	0	Under 30 cm
Tree	Over 13 points	Over 30 cm
Grass	Between 5 and 12 points	Over 30 cm
Water	0	0 cm
Bamboo*	-	-

*: Bamboo is not distinguished for trees using the present ALB dataset (Yoshida et al., 2020)

Finally, as a reference for comparison, we attempted to define the LCC using an unsupervised approach (**Yoshida *et al.*, 2020**), as shown in **Table 2.2**, based on the ALB data manipulations described above. In addition, because several bridges cross the Asahi River in the target region, data from the surrounding riverbed were used to approximate the bed height at the pier.

2.1.4 Processing of LCC mapping using DL method

Mapping of the LCC using the DL method is divisible into two parts: data pre-processing and the processing using the modified DeepLabv3+ module. The true label (TL) and datasets were prepared in advance as input data for the following modules during the pre-processing stage. In addition, the modified DeepLabv3+ module part is divisible into two sections: the conventional RGB-based method and the newly proposed RGB*nl*-based method. The RGB-based process uses only ortho-aerial photographs to train the DeepLabv3+ module and predict the LCC. By contrast, the RGB*nl*-based approach achieves some improvement by using ortho-aerial photographs, 2-D voxel-based points n, and vegetation height l to train and predict the DeepLabv3+ module.

2.1.4.1 Pre-processing

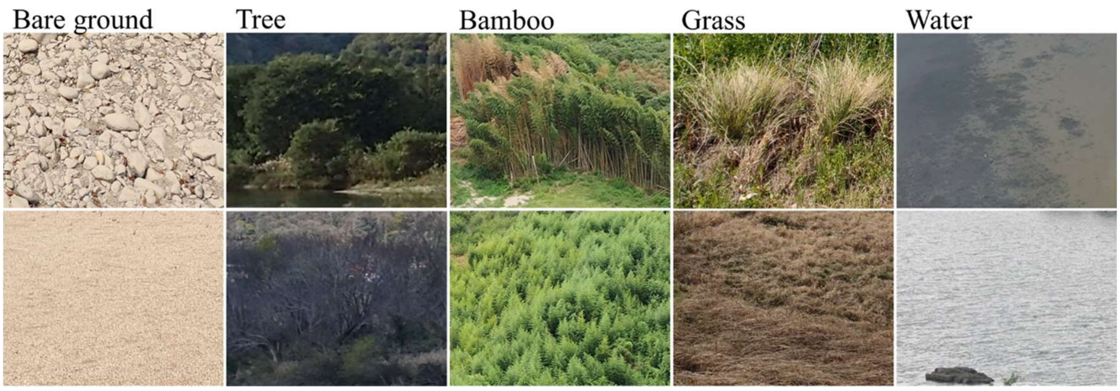


Figure 2.6 Earlier field observation photograph samples of the targeted area for five typical categories of LCC.

Table 2.3 Previous field observation results for the targeted area.

Items (labels)	Objects
A. Bamboo	Moso bamboo, Japanese timber bamboo
B. Tree	Salix eriocarpa, Salix chaenomeloides, Ulmus parvifolia, Juglans mandshurica var, Aphananthe aspera, Quercus variabilis, Rhus javanica, Melia azedarach, Robinia pseudoacacia, Persica, Citrus, Diospyros kaki, etc.
C. Grass	Phragmites australis, Miscanthus sacchariflorus, Miscanthus sinensis, Phragmites japonica, Typha domingensis, Ambrosia trifida, Pleioblastus simonii, Eragrostis curvula, Rosa multiflora, Planted turf, etc.
D. Bare ground	Shoal, Road, Construction site, Agriculture ground
E. Water	Shallow water, Deep water

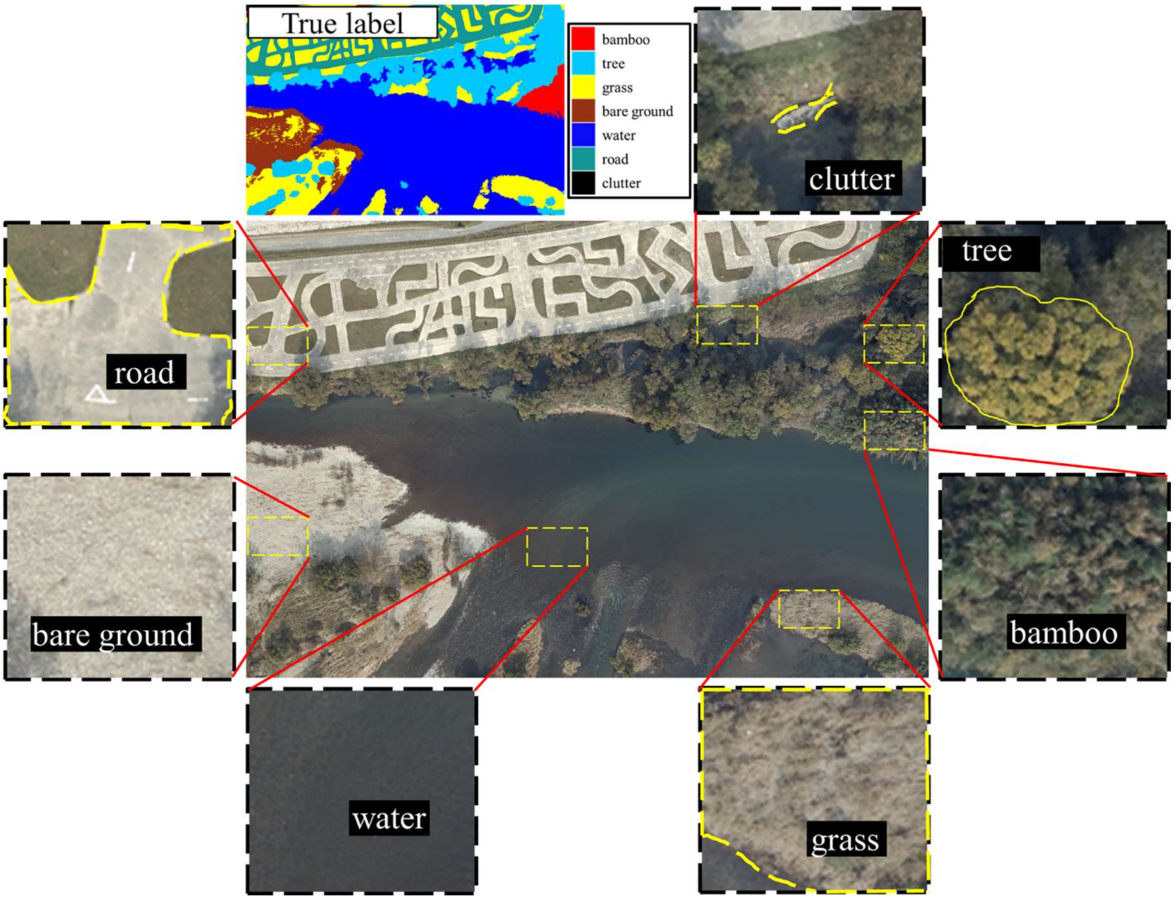


Figure 2.7 Sample of true label mapping in November 2017.

In mapping the LCC pre-processing stage, we considered earlier field observation experience to obtain critical assistance in mapping the TL. As **Figure 2.6** shows, earlier field observation photographs can provide information about the study target LCC characteristics such as texture and color difference. Consequently, based on earlier field observations of our target area presented in **Table 2.3**, we roughly categorized the objects in the study area with five labels: bamboo, tree, grass, bare ground, and water. Furthermore, we sought to produce a rough distinction between natural and artificial areas in true-label mapping. Therefore, aside from the five labels described above, we also considered the other two labels, such as "road" and "clutter" (i.e. all artificial objects except for roads), in the DL method's LCC mapping. **Figure 2.7** presents an example of TL mapping based on orthophotos taken in November 2017.

Furthermore, some high-density areas of bamboo combined with some trees were noted in the targeted region. In orthophotos, trees with leaf-on conditions are difficult to distinguish from bamboo. In such cases, orthophotos from different periods must be compared to differentiate those targeted species. In contrast, "grass" and "trees" are more accessible to differentiation because of shadows and color differences, which vary depending on the height difference between the two. It is noteworthy that, even in a leafless state, where "grass" or "bare ground" can be distinguished clearly under a "tree," the area is still labeled as a "tree" when the TL is drawn because regions that are too small during the deep learning process are not fully learned. For this study, we considered only mudflats and farmland as "bare ground," which can be difficult to define. In addition, orthophotos include sufficient information to identify "water," "roads," and "debris" (anthropogenic landscape components other than roads). Finally, it is noteworthy that our TL is based primarily on orthophotos. Therefore, we were unable to present information that existed but which was not represented in orthophotos (e.g. grass under a tree with leaves). Based on the standards above, we labeled the land cover with seven labels for three periods of orthophotos of the target area, ranging from 13.2 KP to 17.4 KP, as shown in **Figure 2.8**.

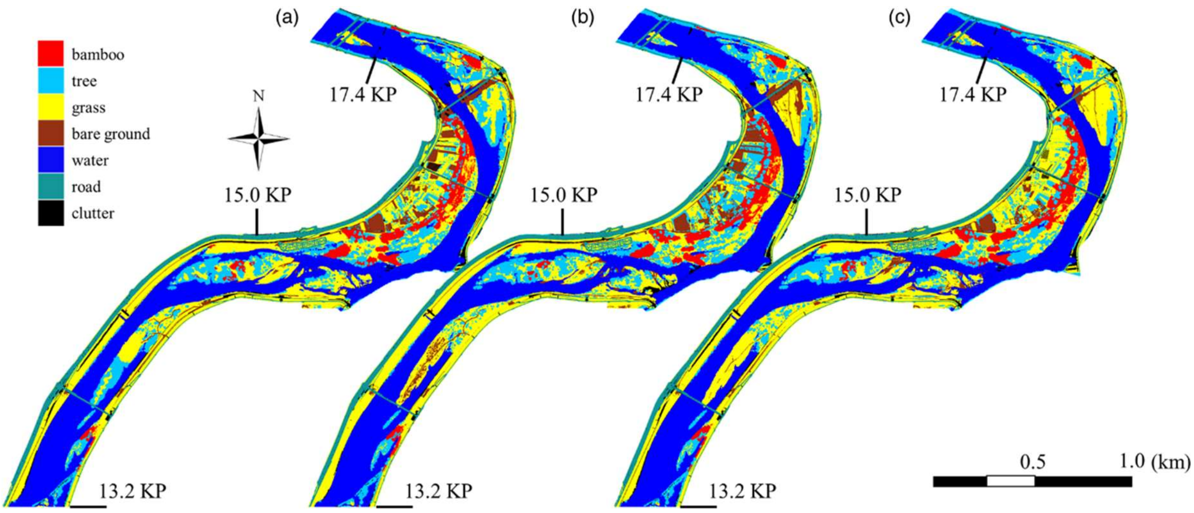


Figure 2.8 True label mapping for the three targeted periods: (a) March, (b) July, and (c) November 2017.

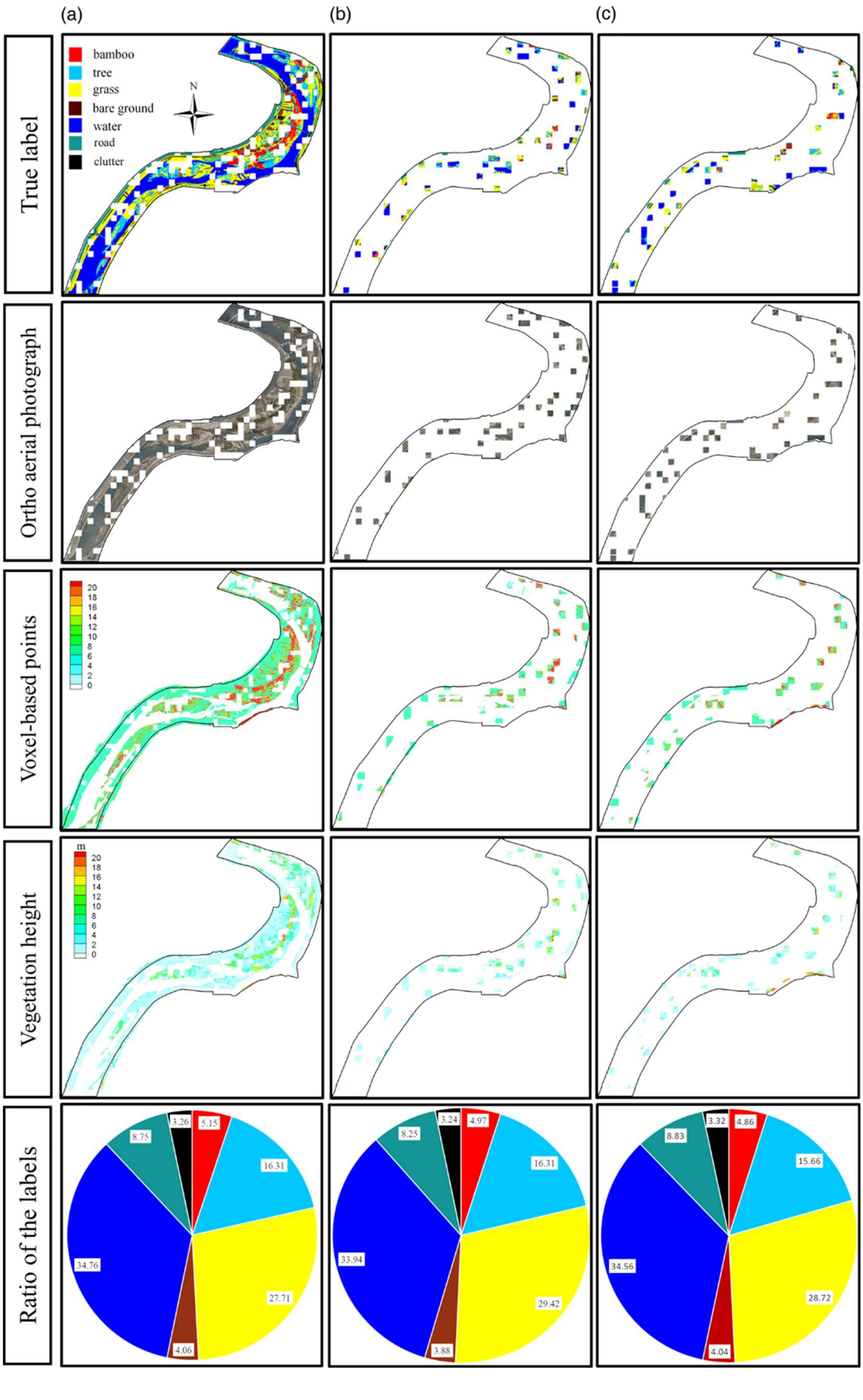


Figure 2.9 Spatial distribution of the dataset in the (a) training, (b) valid, and (c) test areas (March 2017 dataset as example).

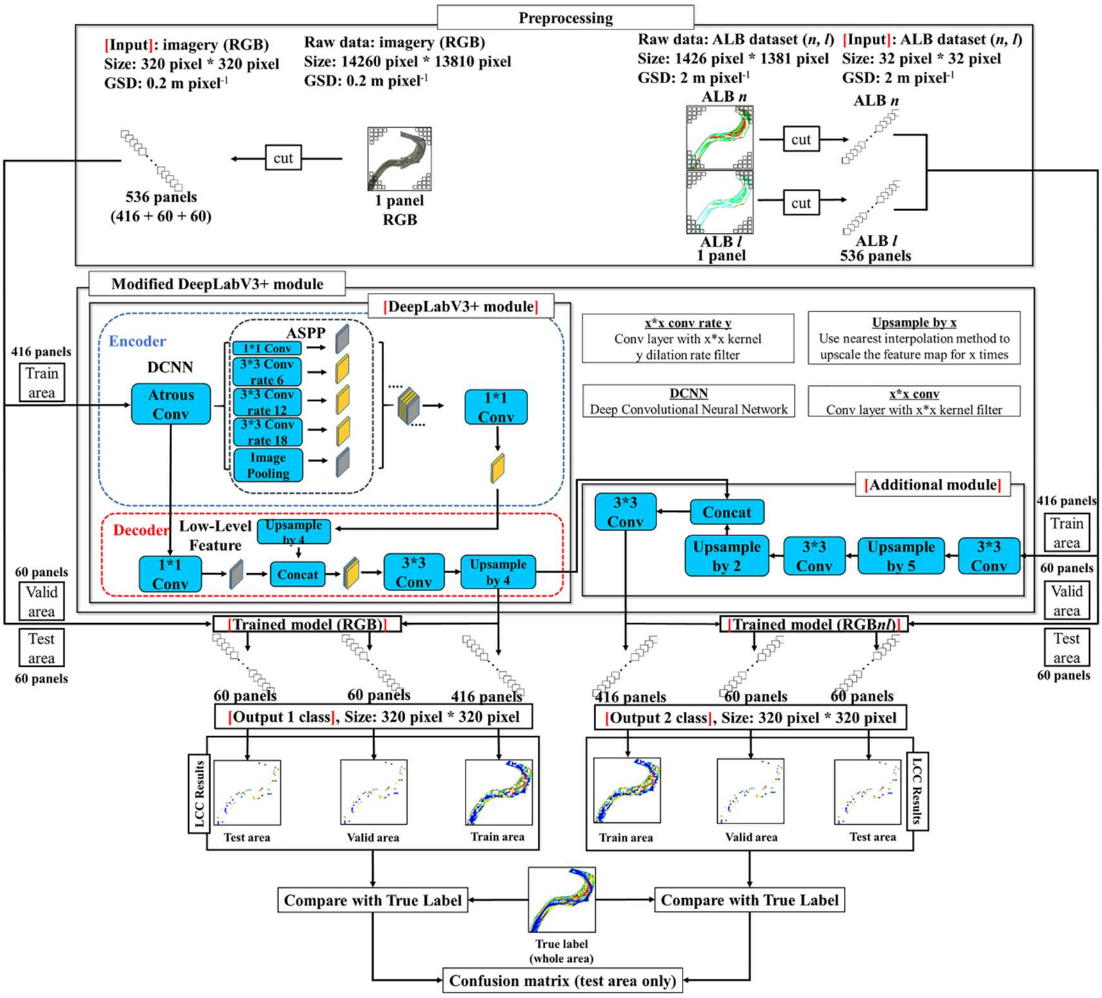


Figure 2.10 Processing of mapping LCC with two DL methods.

As an example, **Figure 2.9** depicts a dataset from March 2017. Datasets (true labels, ortho-aerial photographs, and ALB dataset) from March were divided into three parts: (a) approximately 80% of the dataset for training, (b) approximately 10% of the dataset for validation, and (c) the remaining 10% of the dataset for testing. Datasets for the other two periods (July and November 2017) were assigned similarly. For this study, the modified module chose spatial resolution ratio of 1:10 between the ortho-aerial photographs and the ALB datasets based on DeepLabv3+ model specifications and data resolution. Accordingly, we set the spatial resolution for the ortho-aerial photographs to 0.2 m pixel^{-1} and for the ALB data to 2 m pixel^{-1}. **Figure 2.10** depicts workflows in which ortho-aerial photographs and ALB datasets are cut into small panels using the above scales for pre-processing.

2.1.4.2 Processing of LCC mapping using the modified DeepLabv3+ model

The original DeepLabv3+ module extracts feature from ortho-aerial photographs using an "encoder–decoder" structure. The model's parameters are then optimized using TL. Subsequently, these parameters are saved as a "trained model." This training procedure determines the relations among input data, such as photographs, and the TL. Throughout this study, this processing was designated as the RGB-based LCC method. Later, this technique was upgraded by including an additional module with a "decoder" function using the ALB dataset. To combine with the RGB-based "trained model," the ALB dataset was expanded twice by factors of 2 and 5 in the additional module. Subsequently, we performed upsampling using an imaging technique called "nearest-neighbor interpolation." Then we chose n and l as input data for the additional module for the ALB dataset. The parameters were optimized with the same TL as the RGB-based LCC method. This processing method was designated for this study as the RGBnl-based LCC method. The upgraded method's goal is to incorporate ALB data into the model to improve the accuracy of the inference results. **Figure 2.10** depicts the workflows used for the processing of LCC mapping with the modified DeepLabv3+ module. The RGB-based method used for this processing is traced roughly as (a) training phase – [RGB image as input] → [DeepLabv3+ model] → [trained model], and (b) inference phase – [RGB image as input] → [trained model] → [LCC as output class 1]. By contrast, the RGBnl-based method image processing is represented as (a) training phase – [RGB image and ALB dataset as input] → [upgraded model] → [trained upgraded model], and (b) inference phase → [RGB image and ALB dataset as input] → [trained upgraded model] → [LCC as output class 2]. Finally, **Table 2.4** presents a summary of the all training environment parameters used for programming.

Table 2.4 Data training environment parameters and model setting of the DL methods.

Developing Environment							Model Setting	
OS	GPU	GPU memory	GPU driver	CUDA	cuDNN	Framework	Epoch	Batch size
Ubuntu 20.04	GeForce RTX 3090	24GB	Ver. 460.39	Ver. 11.2	8.04	Tensorflow Ver.2.4.0	400	8

2.1.5 Application

2.1.5.1 Comparisons of LCC mapping

Table 2.5 Analysis conditions of the train, valid and test data in March, July and November 2017 using RGB- (Case 1) and RGB*nl*-based (Case 2, Case 3) methods.

	train data valid data	test data		train data valid data	test data
Case 1-1		Mar. 2017 RGB	Case 2-1		Mar. 2017 RGB+ALB
Case 1-2	Mar. 2017 RGB	Jul. 2017 RGB	Case 2-2	Mar. 2017 RGB	Jul. 2017 RGB+ALB
Case 1-3		Nov. 2017 RGB	Case 2-3		Nov. 2017 RGB+ALB
Case 1-4		Mar. 2017 RGB	Case 2-4		Mar. 2017 RGB+ALB
Case 1-5	Jul. 2017 RGB	Jul. 2017 RGB	Case 2-5	Jul. 2017 RGB	Jul. 2017 RGB+ALB
Case 1-6		Nov. 2017 RGB	Case 2-6		Nov. 2017 RGB+ALB
Case 1-7		Mar. 2017 RGB	Case 2-7		Mar. 2017 RGB+ALB
Case 1-8	Nov. 2017 RGB	Jul. 2017 RGB	Case 2-8	Nov. 2017 RGB	Jul. 2017 RGB+ALB
Case 1-9		Nov. 2017 RGB	Case 2-9		Nov. 2017 RGB+ALB
			Case 3-1	Mar. 2017 RGB+ALB	Mar. 2017 RGB+ALB
			Case 3-2	Jul. 2017 RGB+ALB	Jul. 2017 RGB+ALB
			Case 3-3	Nov. 2017 RGB+ALB	Nov. 2017 RGB+ALB

To compare ALB-based LCC mapping, designated as Case 0, to DL method-based LCC mapping, both the whole area and the test area must be evaluated. We herein set the datasets into three cases to assess the RGB-based and RGBnl-based results, as presented in **Table 2.5**. Case 1 employs RGB data, whereas Cases 2 and 3 use combined data, including RGB data and the ALB dataset. The difference between Cases 2 and 3 is the input-to-output data ratio, with Case 3 having a larger input dataset. Especially, we aim at confirming two points: (a) whether or not more training data improve inference results, and (b) versatility in training and inferring data from different periods. Finally, the confusion matrix (CM) and some indexes were used to assess the relative performance of the RGB-based and RGBnl-based methods.

2.1.5.2 **Case 0: ALB-based method**

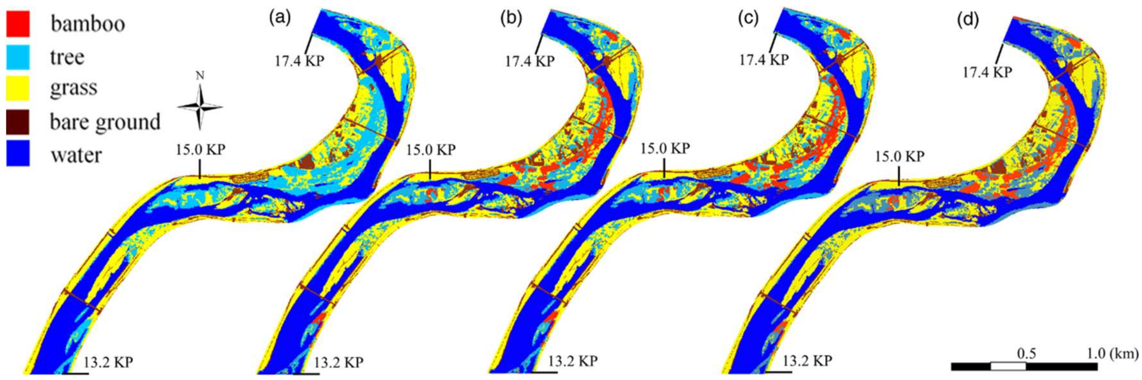


Figure 2.11 Comparison of ALB-based results and DL method results for the whole area in November 2017: (a) Case 0 result, (b) Case 1-9 result, (c) Case 2-9 result, and (d) true label.

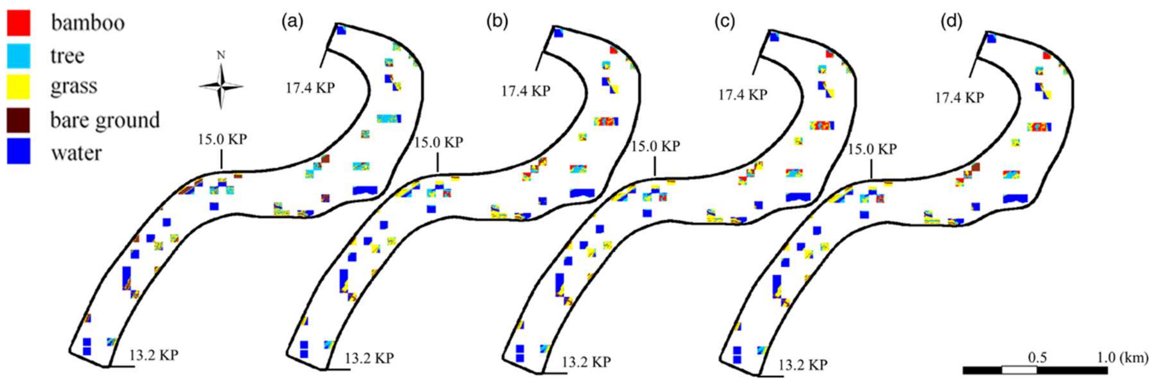


Figure 2.12 Comparison of ALB-based result and DL methods results for the test area in November 2017: (a) Case 0 result, (b) Case 1-9 result, (c) Case 2-9 result, and (d) true label.

First, as presented in **Figure 2.11**, we visually compared the LCC mapping based on the ALB-based method result, the DL method results, and the TL in November 2017. Because the ALB-based method can only segment five labels without "road" and "clutter," DL methods must also adhere to this rule, with "road" and "clutter" being treated as "bare ground." Based on **Figure 2.12**, we compared the ALB-based and DL method results obtained using the CM valuation index, as shown in **Table 2.6**. In the case of the comparison index, we chose overall accuracy (OA) and macro-F1 score as our targets. Correspondingly, **Table 2.7** presents a sample of the CM valuation index. Finally, **Tables 2.8**, **2.9**, and **2.10** present the CM results obtained using the ALB (Case 0), RGB (Case 1-9), and RGB*nl*-based (Case 2-9) methods.

Table 2.6 CM (test area) of ALB- (Case 0), RGB- (Case 1-1) and RGB*nl*-based (Case 1-9) result.

Confusion matrix valuation index		
Symbol	Definition	Formula
Precision (X)	The ratio of the pixels for correctly predicted as X to all the pixels predicted as X	TP-X / PR-X
Recall (X)	The ratio of the pixels for correctly predicted as X to all the pixels true label as X	TP-X / TL-X
F1-score (X)	F1-score is the weighted average of Precision and Recall	2 * Precision (X) * Recall (X) / (Precision (X) + Recall (X))
OA	Overall accuracy value of the confusion matrix	∑ TP-X / Amount of total pixels
Macro-F1	Macro-F1 is the average of all F1-score	∑ F1-score (X)/amounts of labels

[1] X or Y includes 5 labels (B: Bamboo, T: Tree, G: Grass, BG: Bare Ground, W: Water);

[2]TP-X is the amount of the pixels where true label and prediction are all X;

[3]PR-X is the amount of the pixels where prediction is X;

[4]TL-X is the amount of the pixels where true label is X

Table 2.7 Sample of confusion matrix valuation indices.

Sample of CM valuation index							
	B	T	G	BG	W	Total	Recall (%)
B	TP-B	E-B/T	E-B/G	E-B/BG	E-B/W	TL-B	TP-B/TL-B
T	E-T/B	TP-T	E-T/G	E-T/BG	E-T/W	TL-T	TP-T/TL-T
G	E-G/B	E-G/T	TP-G	E-G/BG	E-G/W	TL-G	TP-G/TL-G
BG	E-BG/B	E-BG/T	E-BG/G	TP-BG	E-BG/W	TL-BG	TP-BG/TL-BG
W	E-W/B	E-W/T	E-W/G	E-W/BG	TP-W	TL-W	TP-W/TL-W
Total	PR-B	PR-T	PR-G	PR-BG	PR-W	Total pixels	
Precision (%)	TP-B/PR-B	TP-B/PR-T	TP-B/PR-G	TP-B/PR-BG	TP-B/PR-W		
OA, Macro-F1							

E-X / Y: Amount of the pixels where true label is X, prediction is Y.

Table 2.8 Accuracy valuation for Case 0 LCC (2 m resolution test area ALB-based result).

Case 0 (2 m resolution test area ALB-based result)						
	B+T	G	BG	W	Total	Recall (%)
B+T	2216	486	247	61	3010	73.62
G	656	1998	1301	178	4133	48.34
BG	81	276	793	79	1229	64.52
W	16	46	315	5045	5422	93.05
Total	2969	2806	2656	5363	13794	
Precision (%)	74.64	71.20	29.86	94.07		
OA = 0.73, Macro-F1 = 0.67						

Table 2.9 Accuracy valuation for Case 0 LCC (2 m resolution test area ALB-based result).

Case 1-9 (2 m resolution test area RGB-based result)							
	B	T	G	BG	W	Total	Recall (%)
B	998	12	0	214	5	1229	81.20
T	16	1692	29	198	11	1946	86.95
G	0	142	875	47	0	1064	82.24
BG	113	153	46	3805	16	4133	92.06
W	11	40	1	15	5355	5422	98.76
Total	1138	2039	951	4279	5387	13794	
Precision (%)	87.70	82.98	92.01	88.92	99.41		
OA = 0.92, Macro-F1 = 0.89							

Table 2.10 Accuracy valuation for Case 2-9 LCC (2 m resolution test area RGBnl-based result).

Case 2-9 (2 m resolution test area RGB*nl*-based result)							
	B	T	G	BG	W	Total	Recall (%)
B	925	7	0	294	3	1229	75.26
T	20	1608	49	242	27	1946	82.63
G	0	119	893	52	0	1064	83.93
BG	133	123	40	3815	22	4133	92.31
W	9	25	0	19	5369	5422	99.02
Total	1087	1882	982	4422	5421	13794	
Precision (%)	85.10	85.44	90.94	86.27	99.04		
OA = 0.91, Macro-F1 = 0.88							

Findings revealed that LCC mapping using DL methods can achieve higher accuracy than when using ALB-based methods, with DL methods improving by nearly 25% in terms of the OA and macro-F1 score. The CM shows that these three LCC results are generally diagonally dominant, and demonstrate that LCC can be achieved to some degree, even with only ALB point cloud data. However, using only *n* and *l* values, distinguishing between "bamboo" and "tree" is impossible when using the ALB-based approach. In addition, because of the ALB-based LCC method's mapping rule for the targeted site (**Yoshida *et al.*, 2020**), grasses less than 30 cm tall are regarded as bare ground. For that reason, distinguishing "grass" from "bare ground" might be difficult. Furthermore, because of the "salt and pepper effect," ALB-based method LCC mappings were not highly accurate in reproducing the corresponding TL mapping.

2.1.5.3 Case 1: RGB-based method

Case 1-1 to Case 1-9 from **Figure 2.13** shows the confusion matrix relevant evaluation index (i.e. OA and macro F1-score) for the results obtained using the RGB-based method. The indexes are more prominent when the data from the same period are trained and inferred (Cases 1-1, 1-5, and 1-9). In contrast, when training and inferring data from different periods (Cases 1-2, 1-3, 1-4, 1-6, 1-7, and 1-8), the classification performance deteriorated, possibly because of differences in coloration between periods. For example, when training using March data and inferring on July data, the "tree" in March appears brown, with only branches, whereas "bamboo" seems green. In addition, the effects of solar radiation, water quality, and wind waves might affect the classification performance. **Figure 2.14** presents some examples of misclassification: (a) while a "tree" in July has leaves and appears green, a "tree" in July is incorrectly classified as a "bamboo"; (b) this reason also applies to the case of November data; and (c) when training with July data and inferring March data, the "bare ground" in July data looks brown, and the "tree" and "grass" in March data are inferred as "bare ground" because they are dead.

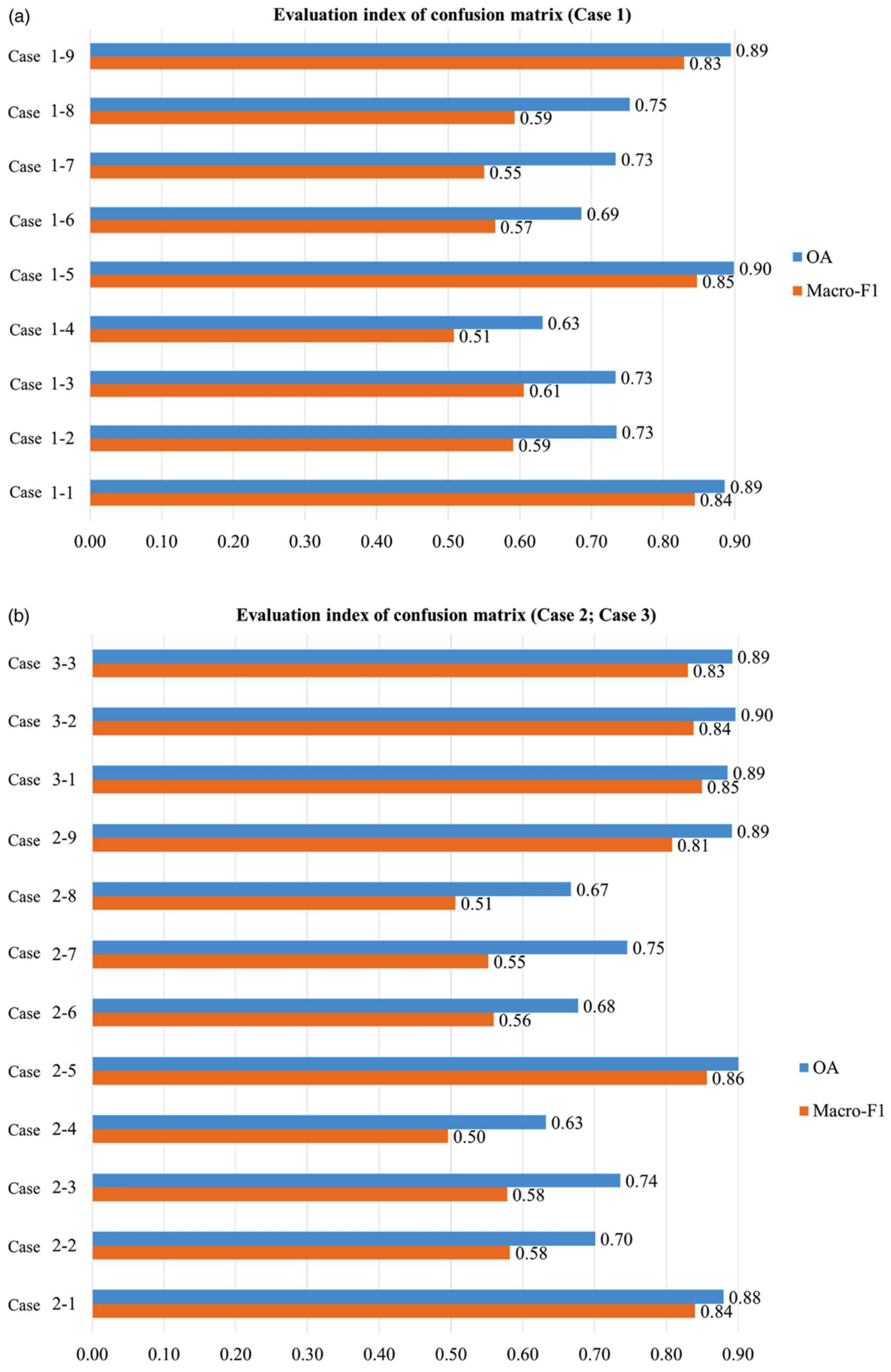


Figure 2.13 Relevant evaluation index of the confusion matrix.

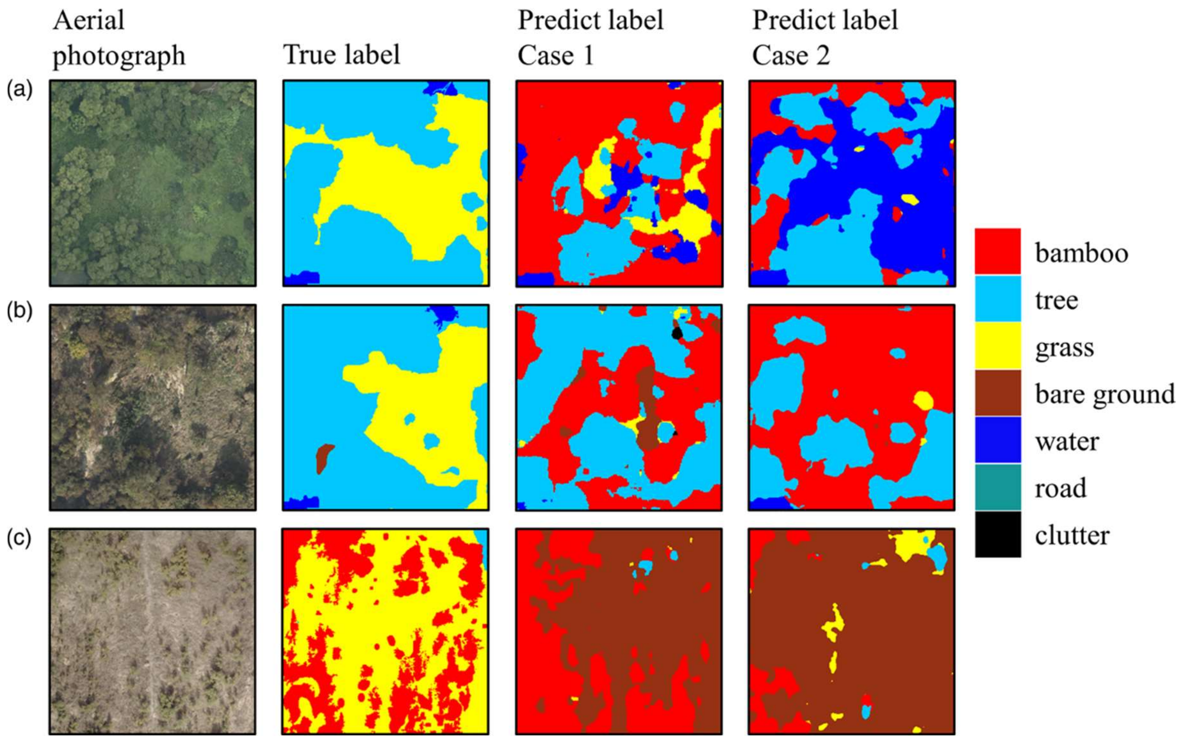


Figure 2.14 Predicted label results for the specified area in targeted periods: (a) July, (b) November, and (c) March 2017.

2.1.5.4 Case 2 and Case 3: RGBnl-based method

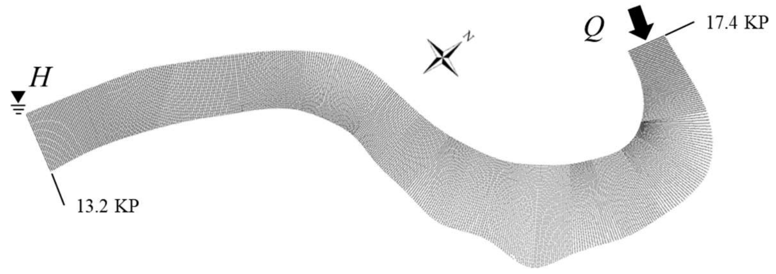
Based on results in Case 2, compared to the RGB-based method, the RGB*nl*-based approach is less effective at improving accuracy (OA, macro F1-score). The findings imply that ALB data in use do not contribute as much to classification performance as RGB data when using an additional module. Finally, Case 3-1, Case 3-2, and Case 3-3 demonstrated results of inferring data from March, July, and November using a combination of data from the three targeted periods. Results show a slight decrease in accuracy when compared to Case 2-1, Case 2-5, and Case 2-9, which were trained and inferred during the same period. Therefore, it is preferable to train and infer using data from the same period rather than combining data from different periods to improve the classification performance. For reference, Cases 2-1 / Case 2-4 / Case 2-7, Case 2-2 / Case 2-5 / Case 2-8, and Case 2-3 / Case 2-6 / Case 2-9 showed training results and inferred data from the same or different periods for each of the three periods. When comparing these results to Cases 3-1, 3-2, and 3-3, it is apparent that we can improve classification performance when we have data from multiple periods by limiting it to a specific period and by using only data from that period. The findings also imply that using all available data can reduce the risk of degrading classification performance if data for the detailed period cannot be specified.

2.1.6 Application of inferred LCC results for 2018 Asahi River Flood Simulation

To examine the applicability and efficacy of LCC predictions in estimating spatially distributed hydrodynamic roughness parameters (i.e. vegetation density values for different species), inferred LCC results based on the ALB-based and DL-based approaches were used for 2018 Asahi River flood modelling. The targeted flood records of observed water levels and the estimated discharge (based on a stage–discharge relation) at different hydraulic stations in the Asahi River were presented in an earlier report by **Yoshida *et al.*, (2021)**, revealing two peaks in the hydrograph observed during the flooding event, with peak discharge of 4,512 m3 s-1. According to the lower Asahi River flooding history, such flooding occurs approximately once every 40 years. For this study, we used a depth-averaged numerical approach with a steady-state flow condition for the peak flood simulation using a boundary-fitted coordinate system (**Yoshida *et al.*, 2021**). In the earlier study (**Yoshida *et al.*, 2021**), researchers revealed that simulated findings were reasonably consistent with observation results when the roughness parameters attributable to distributed vegetation were derived from ALB data, resulting in no significant uncertainty in longitudinal water level predictions. The researchers also demonstrated that distinguishing between the dominant species (e.g., woody vegetation and bamboo grove) in the river studied herein was challenging using an unsupervised LCC method based on ALB datasets alone. Consequently, such a misclassification could significantly impact flow resistance parameterization estimation, affecting the spatial distribution of water levels and depth-averaged flow velocities.

Furthermore, because the previous study (**Yoshida *et al.*, 2021**) demonstrated that no substantial deformation occurred during the targeted flooding, we did not consider transient changes in bed elevation in hydrodynamic modelling. The time increment in the current study was 0.05 s, and the computational mesh for the Asahi River was composed of 434 × 57 cells with average size of 10 m, representing 434 cross-sections and 57 nodes in each cross-section. The upstream boundary condition was determined using the estimated river discharge at Shimomaki Hydraulic Station (19 KP), whereas flood marks at the peak stage defined the downstream boundary condition at 13.2 KP. Based on earlier research by **Maeno *et al.* (2005)**, Manning's roughness coefficient values were set as 0.028 and 0.026, respectively, for the main channel and floodplains. For this simulation, the drag forces for the targeted vegetation species were estimated using the term of 0.5ρ□$C_D l_{\mathrm{min}} u^2$, where ρ stands for water density, □□represents the vegetation density, C_D is the drag coefficient, $l_{\mathrm{min}} = \min\{h, l\}$ denote the minimum value of vegetation height l and local flow depth h, and u expresses represents the local flow velocity. Additionally, we assigned the drag coefficient value of 1 (Yoshida *et al*. 2021) for the current flood flow simulation. **Table 2.11** presents the computational conditions used in the current 2-D flood flow simulation and the density values of the targeted vegetation species in this study. Furthermore, during the field survey, only herbaceous species were observed under bridges crossing the targeted Asahi River. The presence of such vegetation might have a negligible effect on flow resistance parameterization. Consequently, areas with bridges were treated as bare ground for this study.

Table 2.11 Parameters used in the flood simulation 2-D shallow water model

Simulation mesh			
	Mesh Number	Longitudinal: 434	
		Cross-sectional: 57	
Discrete Interval	Time step	Δt = 0.05 s	
	Spatial interval	$\Delta x = \Delta y$ = 10 m	
Vegetation	Vegetation Density λ (m^{-1})	Tree (trunk)	0.013 ($l > 5$)[a], 0.023 ($0 < l \leq 0$)[b]
		Bamboo	0.286[a]
		Grass	0.031[a]
	Drag coefficient C_D		1.0
Manning roughness coefficient ($m^{-1/3}$ s^{-1})			Low water (main channel): 0.028, floodplain: 0.026
River discharge (at peak stage)			Q = 4251 (m^3 s^{-1})
Downstream water Level (at peak stage)			Asahi River 13.2KP: H = 10.67 m

a: Values proposed by Maeno et al. (2005). b: Values suggested by Shimizu et al. (2000).

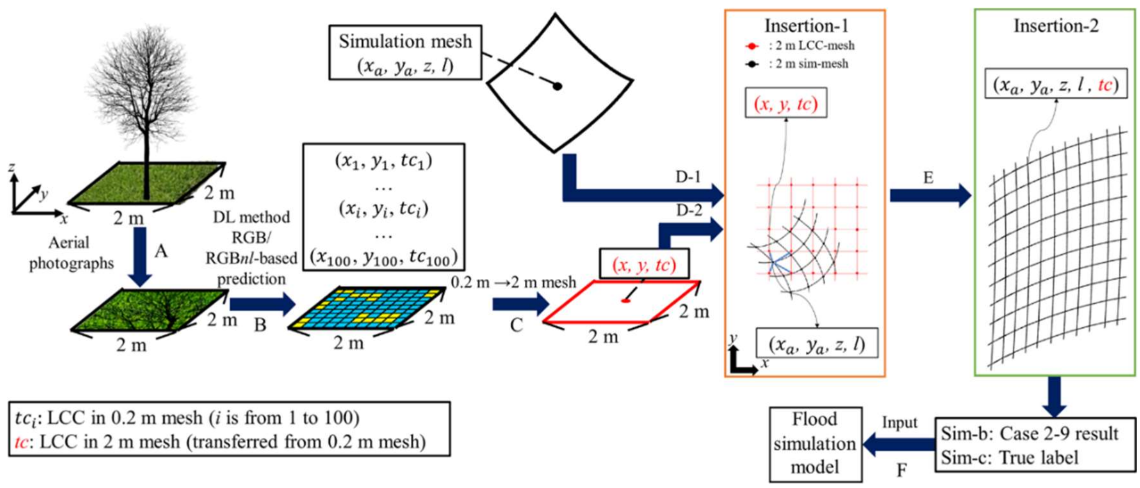


Figure 2.15 Processing of transferring inferred LCC results for flood simulation.

2.1.6.1 Processing of inferred LCC results for flood simulation

As shown in **Figure 2.15**, after the pre-processing of step A and inference of step B, we obtained the 0.2 m pixel-based mesh LCC mapping from 13.2 to 17.4 KP. Simultaneously, mesh transformation is necessary if the inferred results in a square mesh are to be used as flooding simulation parameters in a boundary-fitted mesh. In step C, the RGB-based and RGB*nl*-based methods inference can be transformed from a 0.2 m pixel-based mesh to a 2 m pixel-based mesh by considering the most frequently appearing labels. Then, using a 2 m simulation mesh that includes the LCC information (proceeding with steps D-1 and D-2), we transformed the information into 10 m simulation mesh via step E. Herein, for the flow resistance parameterization, Sim-a was created using all the simulation mesh LCC information (ALB-based method), whereas Sim-b and Sim-c were generated respectively using RGB*nl*-based results and true label. Subsequently, as shown in step F (**Figure 2.15**), the inferred LCC results were transferred as input data for the 2018 Asahi River flood simulation (**Figure 2.16**).

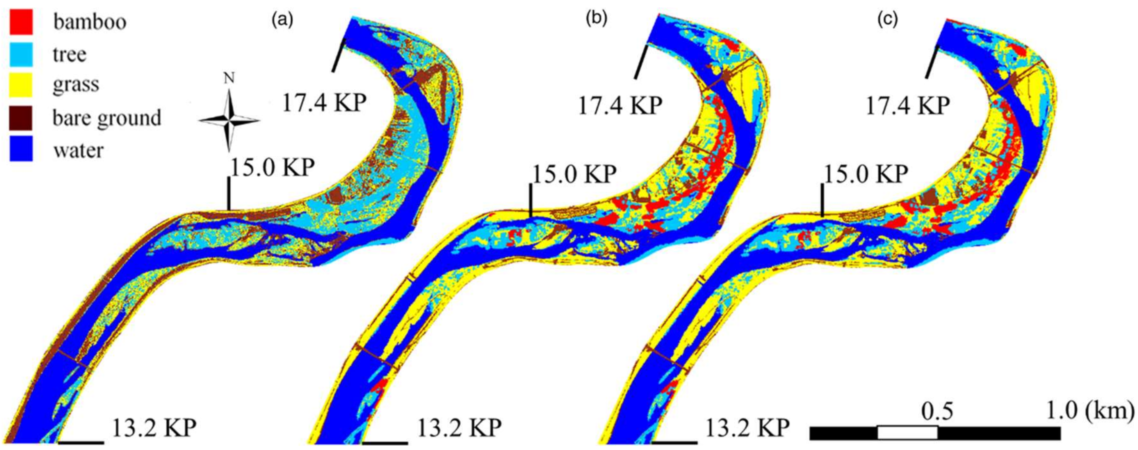


Figure 2.16 Inferred LCC results for parameterization in flood simulation model: (a) Sim-a – ALB-based, Case 0 result, (b) Sim-b – RGB*nl*-based, Case 2-9 result, and (c) Sim-c – true label.

2.1.6.2 Flood simulation using LCC inferences results

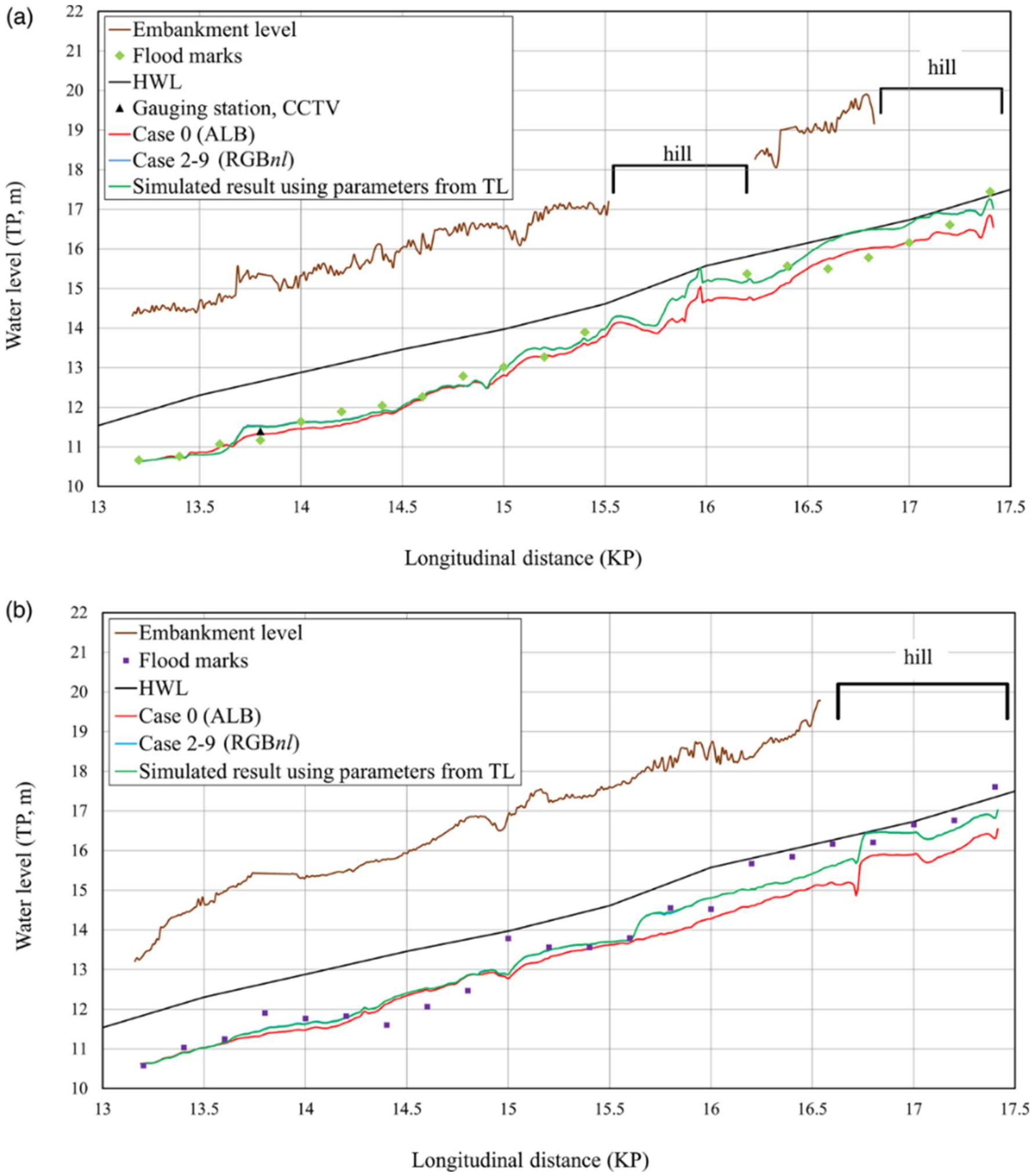


Figure 2.17 Water level estimated from flood simulation results obtained using parameters of different LCC methods: (a) left-bank side and (b) right-bank side. HWL and TL stand for high water level and true label, respectively.

Figure 2.17 presents a comparison of simulated and observed water levels along the Asahi River's left-bank and right-bank reaches during the peak stage of the 2018 flooding. As benchmarked points, the **Figure 2.17** also includes flood marks at the peak stage, the high water level (HWL), and the river embankment level along the targeted reaches. For the left-bank case (**Figure 2.17a**), simulated water levels using both the ALB-based and DL-based parameters were reasonably consistent with the referenced flood marks and the water level estimates based on images from closed-circuit television (CCTV). In contrast, in terms of residual sum of squares (RSS) values (**Table 2.12**), the DL-based simulation reproduced flood marks that were markedly better than the ALB-based simulation for the right-bank case (**Figure 2.17b**), thereby implying that the DL results outperformed the ALB results. Furthermore, **Figure 2.18** depicts the flow velocity and water depth results estimated from the current flood simulation using flow resistance parameters derived from the ALB-based and RGB*nl*-based LCC results. Those findings revealed that the simulated flow velocity and depth have differed considerably in both cases because of differences in land cover between the targeted ALB-based and RGB*nl*-based LCC results. For example, at location b (**Figure 2.18**), the RGB*nl*-based water velocity varied markedly from the ALB-based results because the DL method correctly distinguished the dominant bamboo grove from woody species in the targeted area. Overall, the numerically simulated results have demonstrated the importance of high-accuracy LCC mapping in hydraulic engineering tasks.

Table 2.12 RSS of the different LCC method results comparing with flood marks

	RSS (m^2)		
	$\sum(h_{ALB}-h_{FM})^2$	$\sum(h_{RGBnl}-h_{FM})^2$	$\sum(h_{TL}-h_{FM})^2$
Left-bank-side	1.58	1.77	1.73
Right-bank-side	8.49	3.68	3.70

RSS: Residual sum of squares; FM: Flood marks;
h_{ALB}: Water level at flood mark using ALB-based result;
h_{RGBnl}: Water level at flood mark using RGB*nl*-based result;
h_{TL}: Water level at flood mark using True label;
h_{FM}: Field observation of water level at flood mark

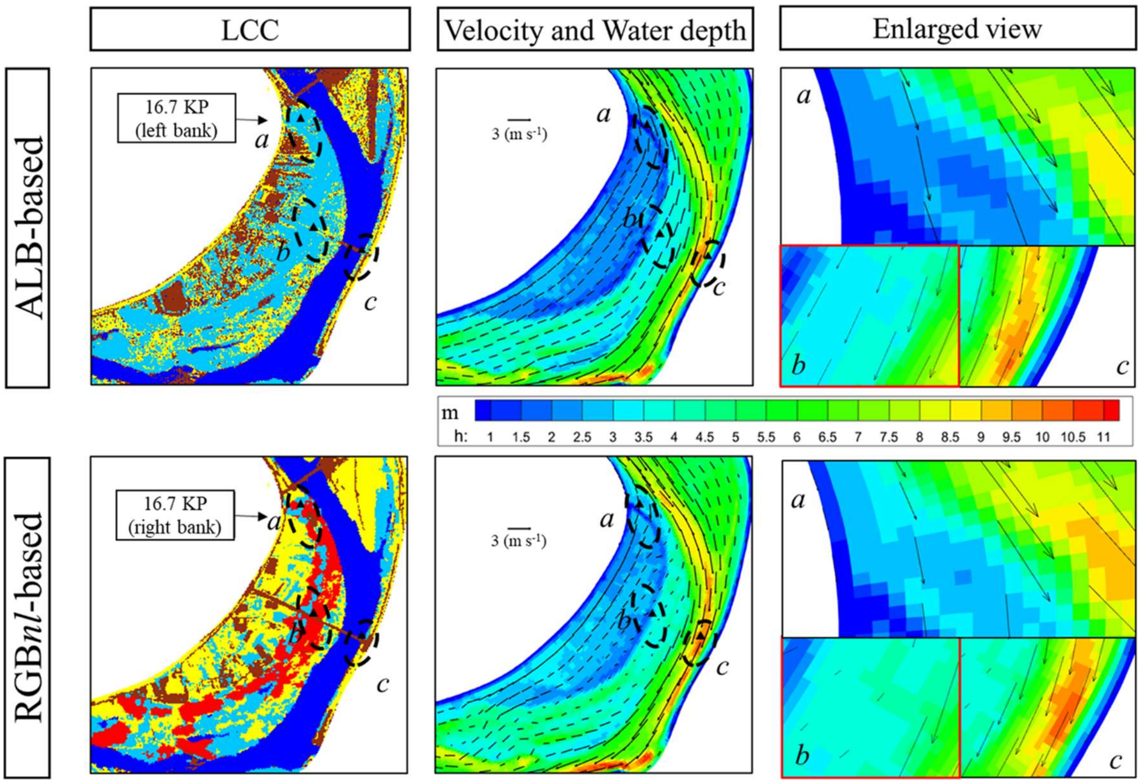


Figure 2.18 Velocities and water levels inferred from flood simulation using parameters of ALB-based and RGB*nl*-based results with LCC.

2.1.7 Conclusions

Results revealed that the DL methods outperformed the ALB-based method in terms of typical valuation indexes. In this study, three seasonal datasets with different leafy conditions (i.e. no–leaf and leaf–on) significantly influenced LCC results, demonstrating that use of the same period datasets for the different trained and test areas yielded higher accuracy. Currently used datasets with shorter time variations of around three months might limit our results because longer periods datasets supposedly provide better predictions using the DL approach. Furthermore, the depth-averaged flood simulation model showed that the water level inferred using the DL-based method much more closely approximated the observed water level than the conventionally used ALB-based approach did. In addition, the flow velocity and water depth inferred from the DL method results differed from those inferred from ALB-based LCC results because of changes in the classification of the most dominant riparian vegetation species in the targeted region: trees and bamboo. In addition, although the RGB*nl*-based method includes ALB-derived voxel-based laser points and vegetation height information, the LCC mapping accuracy has not improved markedly over the RGB-based approach. Overall, these findings might compel us to revise our model for use in future studies, considering additional processed attributes from ALB datasets (i.e. reflection intensity from DTM and DSM) and using a few more inputs to the original DL model in addition to RGB. To conclude, the results of this study are expected to support reasonable engineering measures for flood control in vegetated rivers. Finally, based on our current findings, we recommend conducting comprehensive research investigating balanced riparian ecosystems conservation. Furthermore, because of higher cost in ALB data acquisition and recent advances in remote sensing technologies, we intend to use cost-effective unmanned aerial vehicle-borne lidar-derived data (**Islam *et al.*, 2021**) for future relevant research due to its convenience of more detailed point density and concurrently captured high spatial resolution aerial images. In addition, because sedimentation can change the LCC and roughness of rivers (**Pinho *et al.*, 2020**), it is recommended to identify such factors using RGB analysis and the well-proven ALB technique, which can aid in identifying potential uncertainty in hydrodynamic-numerical modelling.

2.2 UAV-derived Orthophotograph aided with UAV-borne LiDAR on LCC

2.2.1 Introduction

River environmental information includes crucially important data such as topographic bathymetry and vegetation attributes that are necessary to develop balanced river management measures, addressing issues such as flood control (**Yoshida *et al.*, 2021**) and ecosystem management (**Mandlburger *et al.*, 2015**). When confronted with environmental system difficulties, informing the specific distribution of vegetation, including but not limited to location, quantity, and species, can help researchers in their assessments of environmental changes over time (**Carbonneau *et al.*, 2020**). In recent years, researchers have become increasingly interested in changes in the numbers of original dominant species as a result of exotic species invasion (Mooney & Cleland, 2001). In this regard, field inspections have been conducted primarily of limited river sections, requiring personnel to enter the site to conduct measurements with a total station or a real-time dynamic global positioning system (**Legleiter, 2013**). These ground-truth surveys are daunted by several limitations: they are expensive, time-consuming, and non-repeatable (**Campbell-Palmer *et al.*, 2020)**. Furthermore, field surveys might encounter extreme situations (e.g., flood-affected areas) that can endanger the investigators (**Wei *et al.*, 2020; Zzaman *et al.*, 2021**), or areas that are difficult to access for the researchers (e.g., densely vegetated areas), resulting in ineffective data collection (**Anjum and Tanaka, 2020**). In addition, when a larger study area must be observed, satellite observations are ideal, particularly when regular revisits are available, which can provide more surface information in different seasons (**Rodriguez-Galiano & Chica-Rivas, 2014**). However, spaceborne platforms have usually collected imagery with low temporal and spatial resolution. Although spaceborne platforms with high spatial resolution are very expensive, a few public datasets are available for free. Nevertheless, they cover only limited areas. Because of those shortcomings of satellite-based platforms, most researchers find it difficult to extract detailed features for accurate land cover mapping (**Fisher *et al.*, 2018**). Furthermore, the public satellite data resolution ranges between 10 and 60 m (**Chandler *et al.*, 2021**). However, such a coarse mesh size will be insufficient to achieve our targeted goals (i.e., flood control and ecosystem management) because, for instance, flow regimes and related vegetation dynamics (**Sanjaya & Asaeda, 2017**) will be difficult to predict near water edges and different hydraulic structures.

To overcome limitations imposed by traditional methods, digital surface models (DSMs) and digital terrain models (DTMs) have become indispensable tools for describing terrain conditions (i.e., using DSM minus DTM to infer the height of the land cover). These remotely sensed data are used in various fields, including not only flood simulation (Hou et al., 2021), but also hydrological models (e.g., **Veeck *et al.*, 2020; Xu *et al.*, 2021**). Light Detection and Ranging (LiDAR) technology generates high-resolution and accurate DSMs and DTMs by halving the time between the emitted pulse and detection of the reflected echo (**Baltensweiler *et al.*, 2017, Yan *et al.*, 2015**). Based on LiDAR technology characteristics, improved airborne LiDAR topo-bathymetry (ALB) (**Wieser *et al.*, 2016**), which has been fully validated over the years, has been shown to elucidate the vegetation distribution in riverine areas. In addition to the technologies described earlier, a more cost-effective unmanned aerial vehicle (UAV)-borne green LiDAR system (GLS) (**Islam *et al.*, 2020**) has recently become available as a reliable tool for use in high-resolution surveys. The GLS can acquire both laser point clouds and high spatial resolution aerial photographs simultaneously, making GLS suitable for land cover classification (LCC) (**Islam *et al.*, 2022**). Additionally, compared to ALB's low-resolution aerial photographs and less uniform point density (**Mandlburger *et al.*, 2015**), UAV-borne GLS can produce more uniform and higher point cloud densities of about 100–200 pts m^{-2}. Furthermore, because of its UAV-integrated operation during the GLS campaign, its flight altitude can be varied depending on the target, thereby facilitating better acquisition of complicated land cover data. Although aircraft operated during ALB measurement can fly at higher altitudes with higher speed of approximately 220 km h^{-1} (**Islam *et al.*, 2020**) and collect laser point clouds efficiently over larger reaches, the major shortcomings of aircraft are noteworthy: higher data acquisition costs, platform management difficulties, specialized personnel for operations, and weather and flight conditions. Consequently, in comparison to the more portable and cost-effective GLS, which acquires data at a shorter scale, ALB systems are often challenging when they are expected to provide instantaneous and flexible measurements. However, despite these challenges in airborne platforms, a more recent study (**Yoshida *et al.*, 2020**) used ALB-derived attributes in LCC mapping to parameterize flood modeling for vegetated river reaches, thereby revealing difficulties in distinguishing major species (e.g., woody vegetation and bamboo grove in our targeted vegetation).

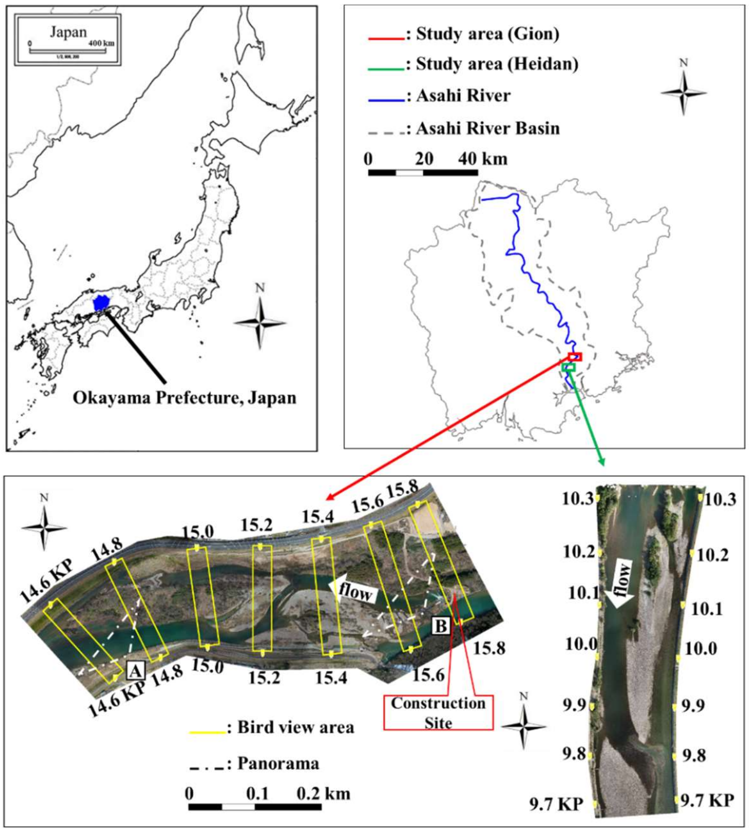
Several approaches to addressing LCC-related difficulties have been tested over the years, including decision tree algorithm (**Yu *et al.*, 2006; Gauci *et al.*, 2018; Fehérváry *et al.*, 2020**), manual setting of thresholds (**Do *et al.*, 2019; Yoshida *et al.*, 2020**), and deep learning (**Carbonneau *et al.*, 2020; Pourmohammadi *et al.*, 2020**). The decision tree algorithm and threshold-setting approaches necessitate that researchers modify several parameters to adapt to region-related differences and that they discover data features on their own. The two methods described above demand considerable time to investigate data relations. In contrast, among the methods discussed here, the deep learning method, particularly with the atrous convolution module (**Chen *et al.*, 2017**), has the most effective feature extraction for LCC. Therefore, it was chosen for this study. In an earlier deep learning-based study (**Chen *et al.*, 2017**), researchers used DeepLabv3+ model with the atrous convolution module to perform semantic segmentation, which has since become used widely in remote sensing-related research (**Erfani *et al.*, 2022; O'Neil *et al.*, 2020**). Furthermore, in a more recent study (**Yoshida *et al.*, 2022**), fluvial researchers used the DeepLabv3+ model in conjunction with ALB-derived attributes for the first time to infer LCC mapping, demonstrating that hydraulic parameters derived from LCC results were used reasonably for flood simulation. They also revealed that recognition accuracy was high when training and validation sets were selected from the same location and period. However, when imagery data (aerial photographs) for deep learning in the validation set were chosen from the same location at different periods, similar recognition accuracy to that obtained in the case with the same location and period was not achieved, thereby severely limiting the widely used LiDAR between targeted seasons.

Given those findings, the current study was conducted as an attempt to use data from UAV-borne photographs and GLS, because of their inherent benefits, to achieve higher-accuracy LCC (**Islam *et al.*, 2022**). The difficulty of heterogeneous seasonal photographs being unable to predict one another is caused primarily by leaf-on or leaf-off situations and by sunlight-angle-derived luminance difference of imagery data. Based on these considerations, this study was conducted to achieve mutual predictability among data from the same location in different periods. Especially, a coupling methodology of stable and reliable GLS-derived LiDAR data with UAV-based distinctive image features was proposed for this study: High Contrast Superimposed (HC-Si) method for combining input data for the existing RGB (3-channel)-based DeepLabv3+ model. Furthermore, for complete comparison of the performance of our model, we examined an additional input channel for the existing 3-channel DeepLabv3+ model, i.e., 4-channel DeepLabv3+ model (i.e., RGBL; channel L represents GLS-derived height of land cover, *l* as described in detail hereinafter) for channel entry extension. For reference, the 3-channel-based model using input data derived solely from aerial photographs or LiDAR were also included in the comparison groups. Finally, the approaches proposed for this study are expected to be used in potential riverine studies to distinguish the most dominant riparian vegetation species accurately (i.e., woody vegetation and bamboo grove in our targeted area as described hereinafter), which can reasonably estimate spatially distributed hydrodynamic roughness in streamflow modeling and also can be useful in proper ecosystem management tasks.

2.2.2 Study site and methods

2.2.2.1 Study site

(a)



(b)

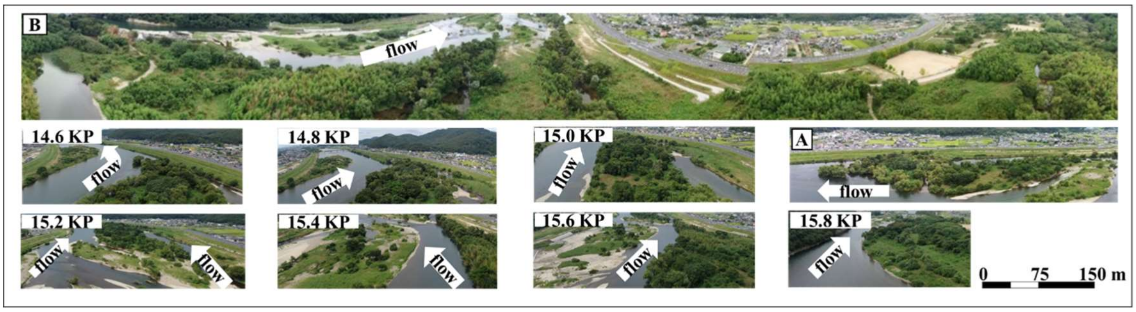


Figure 2.19 Perspective of Green LiDAR measurement area: (a) location of the Asahi River in Japan with kilo post (KP) values representing the longitudinal distance (km) from the river mouth, and (b) drone-captured photographs based on the marked positions in (a).

Figure 2.19 depicts our study site, which is located in a downstream area of the Asahi River, a first-class (state-controlled) river in Japan which flows through Okayama Prefecture into the Seto Inland Sea. During 1965–2005, the average river discharge at the Makiyama hydraulic station, which is located 20 KP upstream of the targeted domain, was 57.12 m^3 s^{-1} (**MLIT 2007**). For this study, the kilo post (KP) value represents the longitudinal distance (km, kilometers) from the target river mouth. The main target domain, known locally as the "Gion" area, is 14.6–15.8 KP (1.2 km long), as shown in **Figure 2.19**. Another test domain, known locally as the "Heidan" area, is 9.7–10.3 KP (0.6 km). The river section specifically examined herein has a mean bed slope of about 1:600, with about 300 m channel width. Moreover, extensive and diverse vegetation has been observed at the target site recently, raising concern among researchers about effective flood control and ecosystem management measures. During GLS campaigns, we recently observed a few river management tasks involving the cutting down of bamboo groves upstream of the study site (**Figure 2.19**). Based on the flood control aspects, the riparian vegetation at this study site was classified roughly into three types based on flow resistance characteristic (Yoshida et al., 2017), which include bamboo forests (bamboo), herbaceous species (grass), and woody species (tree). Along with riverine vegetation, this study added another four labels (*i.e.,* water, bare ground, road, and clutter) to represent local surface environmental changes better.

2.2.2.2 Data collection and processing

For this study, we used GLS (TDOT Green; Amuse Oneself Inc.) along the targeted section of the lower Asahi River (**Figure 2.19**). Several flight operations were conducted under leaf-off (March 2020, Gion) and leaf-on (October 2020 and April 2021, Gion; and October 2021, Heidan) conditions to achieve overlapping coverage of the target area. The GLS-device simply uses green lasers to scan the study area, which includes both the underwater and the terrain surfaces, as shown in **Figure 2.20**. Furthermore, for each GLS measurement, a digital camera was mounted directly beneath the platform to take aerial photographs of the target river. Table 1 presents the equipment specifications, measurement parameters, and river water quality at the time of measurement. Because turbidity strongly influences the amount of light incident into the water column, the turbidity values were checked carefully before each GLS measurement. During the research periods, the water quality in the GLS-target area was suitable for measuring the underwater topographic surface, although we recorded a few missing data points because the current green laser's power was insufficient to penetrate deeper waters (**Islam *et al*., 2022**).

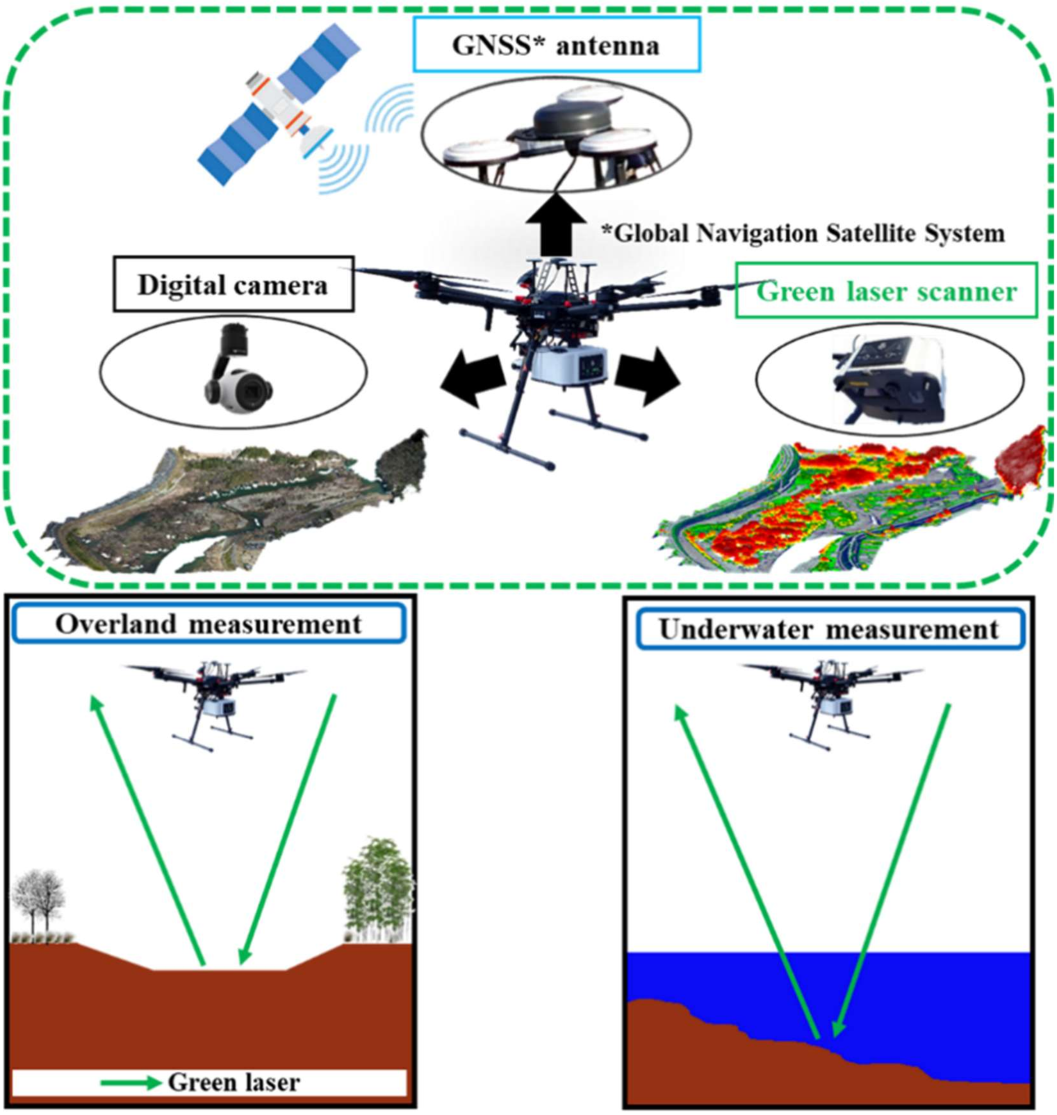


Figure 2.20 GLS using a green laser for overland and underwater surveys.

Table 2.13 Current Green LiDAR System (GLS) specifications and measurement conditions in the targeted river reach

	Items	Measurement date of GLS and Ortho-photo Mar. 2020	Oct. 2020	Apr. 2021
Equipment specifications	Laser wavelength range (nm)	532	532	532
Measurement specifications	Number of laser beams (s^{-1})	60,000	60,000	60,000
	Ground altitude (m)	50	50	GND: 100 WTR: 50
	Flight speed (km h^{-1})	9	9	GND: 14.4 WTR: 9.0
	Density of measurement points (m^{-2})	100	100	GND: 50 WTR: 100
Photograph specifications	Resolution (cm $pixel^{-1}$)	3	3	5
Water quality	Turbidity (FTU*: Surface level)	0.8	3.12	2.45
	Turbidity (NTU**: Intermediate level)	-	3.7	2.6

GND, Overland area; WTR, Underwater area;

*, Formazin Turbidity Unit; **, Nephelometric Turbidity Unit

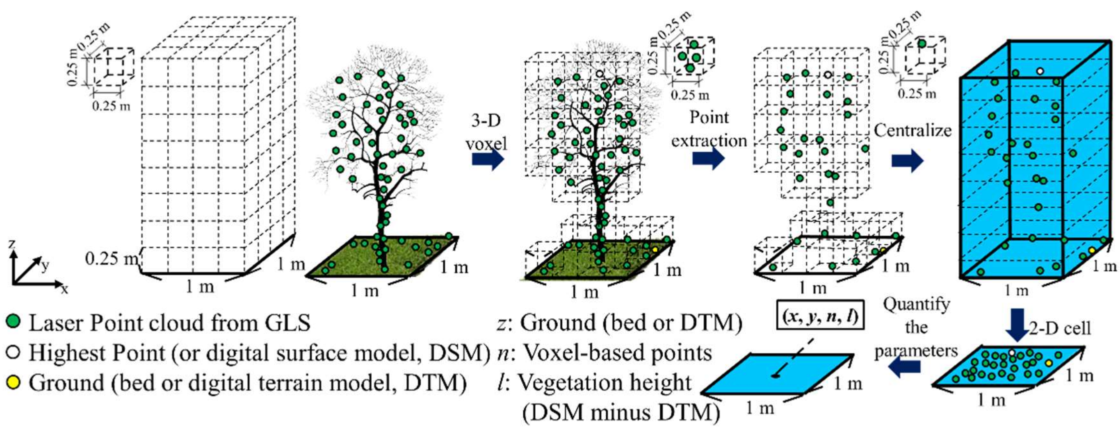


Figure 2.21 Voxel-based GLS data processing.

Both river topographic mapping and vegetation attribute measurements were taken using GLS at normal water levels. Data from GLS measurements were not recorded using the traditional waveform approach applied for ALB surveys. The current green LiDAR typically employs a reflected pulse recording technique, with a maximum of four echoes per laser pulse. However, because the use of effective recording intensity with four echoes per laser pulse for this study was not possible, we only recorded reflection intensity (int) values with a maximum of two echoes. In the case of GLS data processing, as shown in **Figure 2.21**, a Cartesian grid of cubic voxels with 0.25 m per side was initially developed to filter out noise in the laser point cloud. Following preprocessing, a horizontal two- dimensional (2-D) cell was developed with 1 m width, with the points in each 2-D cell designated as voxel-based points (*n*). The ground height is defined by filtering the point cloud near the bottom of the 2-D cell, known as DTM. In addition, the terrain surface was evaluated using the highest point in each 2-D cell: DSM. Finally, the height of land cover (*l*) was estimated using the DSM value minus the DTM value.

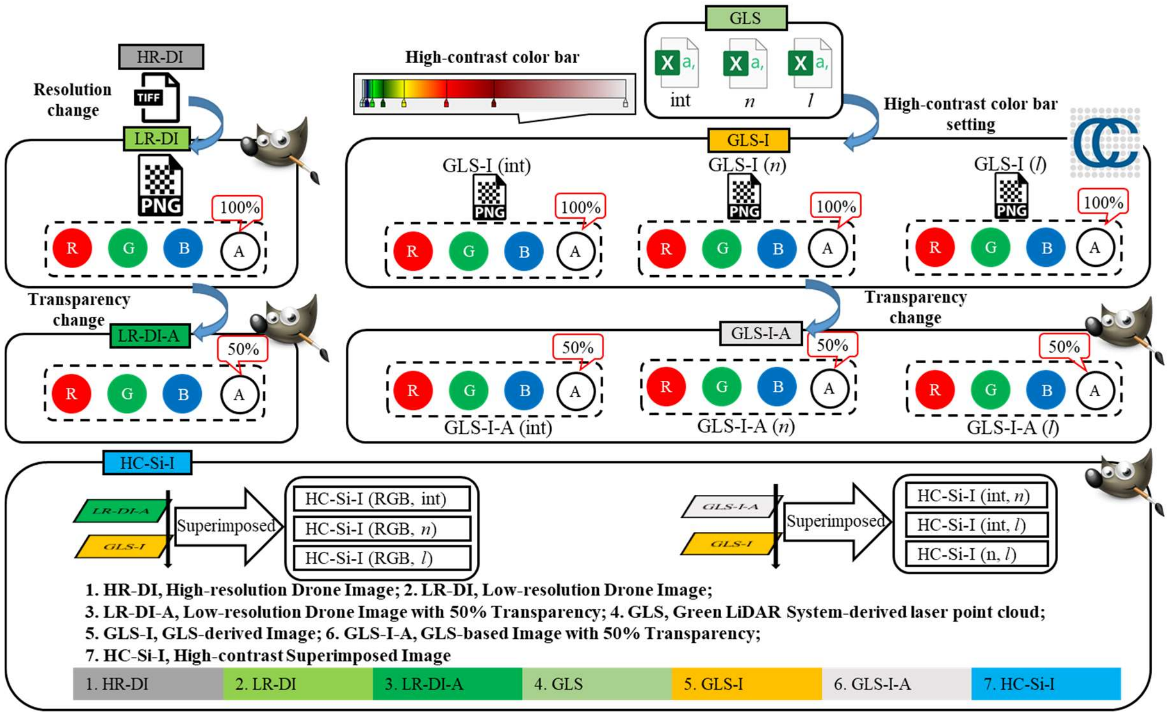


Figure 2.22 Preprocessing of imagery-based input data from drone images (aerial photographs) and LiDAR dataset (GLS).

Table 2.14 List of Input Imagery for the 3-channel-based and 4-channel-based model

Type-based Source	Elements	Method	Input Imagery for DL
3-channel			
Aerial Photographs	HR-DI	—	LR-DI (RGB)
LiDAR Data	int	HC	GLS-I (int)
	n	HC	GLS-I (*n*)
	l	HC	GLS-I (*l*)
	l	Gray	GLS-I (*l*)-Gray
Aerial Photographs & LiDAR	LR-DI (RGB) & GLS-I (int)	HC-Si	HC-Si-I (RGB, int)
	LR-DI (RGB) & GLS-I (*n*)	HC-Si	HC-Si-I (RGB, *n*)
	LR-DI (RGB) & GLS-I (*l*)	HC-Si	HC-Si-I (RGB, *l*)
	GLS-I (int) & GLS-I (*n*)	HC-Si	HC-Si-I (int, *n*)
	GLS-I (int) & GLS-I (*l*)	HC-Si	HC-Si-I (int, *l*)
	GLS-I (*n*) & GLS-I (*l*)	HC-Si	HC-Si-I (*n*, *l*)
	LR-DI (RGB) & GLS-I (*l*)	Gray-Si	Gray-Si-I (RGB, *l*)
4-channel			
RGBA Format	LR-DI (RGB) & GLS-I (*l*)(Gray)	AC	RGBL

HR-DI, High-Resolution Drone Image; int, Intensity of GLS; *n*, Voxel-based points of GLS; *l*, Vegetation height of GLS; RGBL, 4-channel data including R-, G-, B-, *l*-based information in each single channel; HC, High-Contrast Color bar-based method; Gray, Gray-Scale Color bar-based method; HC-Si, High-Contrast Color bar-based Superimposed method, a kind of image fusion approach; Gray-Si, Gray-Scale Color bar-based Superimposed method; AC, Alpha Layer Change-based method; DL, Deep Learning

Figure 2.22 depicts imagery input data of three types for this study: low-resolution drone images (LR-DI), images derived from GLS-based point cloud using a high-contrast color bar (GLS-I), and images created by superimposing the above images (HC-Si-I). The following sections describe processing of the input datasets of the three types (**Table 2.14**).

In this case, we used typical input imagery of aerial photographs for deep learning. Because the LiDAR data resolution was 1 m pixel^{-1} and because that of high-resolution drone images (HR-DI) was about 0.03 m pixel^{-1}, size conversion was necessary to unify the size of the input imagery. Consequently, to achieve 1 m pixel^{-1} LR-DI (i.e., changed from 0.03 m pixel^{-1} HR-DI), the interpolation-free method (i.e., a method that extracts only the color of pixels at fixed interval positions) was used to maintain the color of each pixel in the newly generated imagery as unchanged from the original imagery.

Raw GLS data include laser point cloud coordinates, DSM, DTM, DSM minus DTM (l), voxel-based laser points (n), and intensity (int). The visualization software, CloudCompare (i.e., 3-D point cloud and mesh processing software with open source project), was used to convert the laser point cloud information, which includes only numerical values (i.e. csv files), into visual imagery. Furthermore, different visualization effects were achieved by changing the type of color scale or the distribution of the selected slider (i.e., the point to set specific color in the color scale). Because deep learning models have strong learning ability for data with more visible features, one challenge of this study was determining how to present the features of LiDAR data better using visualization software. Finally, after comparing the visual effects of various default color scales, the "high-contrast color scale" was chosen for this study. Furthermore, the selected color scale displayed the LiDAR data as imagery after dividing the GLS data (from min to max) into 256 steps. In addition, the "high contrast color scale" included six selected slides with different colors in the 0–50% section of the color scale (i.e., 1%, 2%, 4%, 8%, 16%, and 32%).

1%: R = 158, G = 158, B = 158
2%: R = 0, G = 0, B = 127
4%: R = 0, G = 255, B = 0
8%: R = 0, G = 85, B = 0
16%: R = 255, G = 255, B = 0
32%: R = 148, G = 97, B = 97

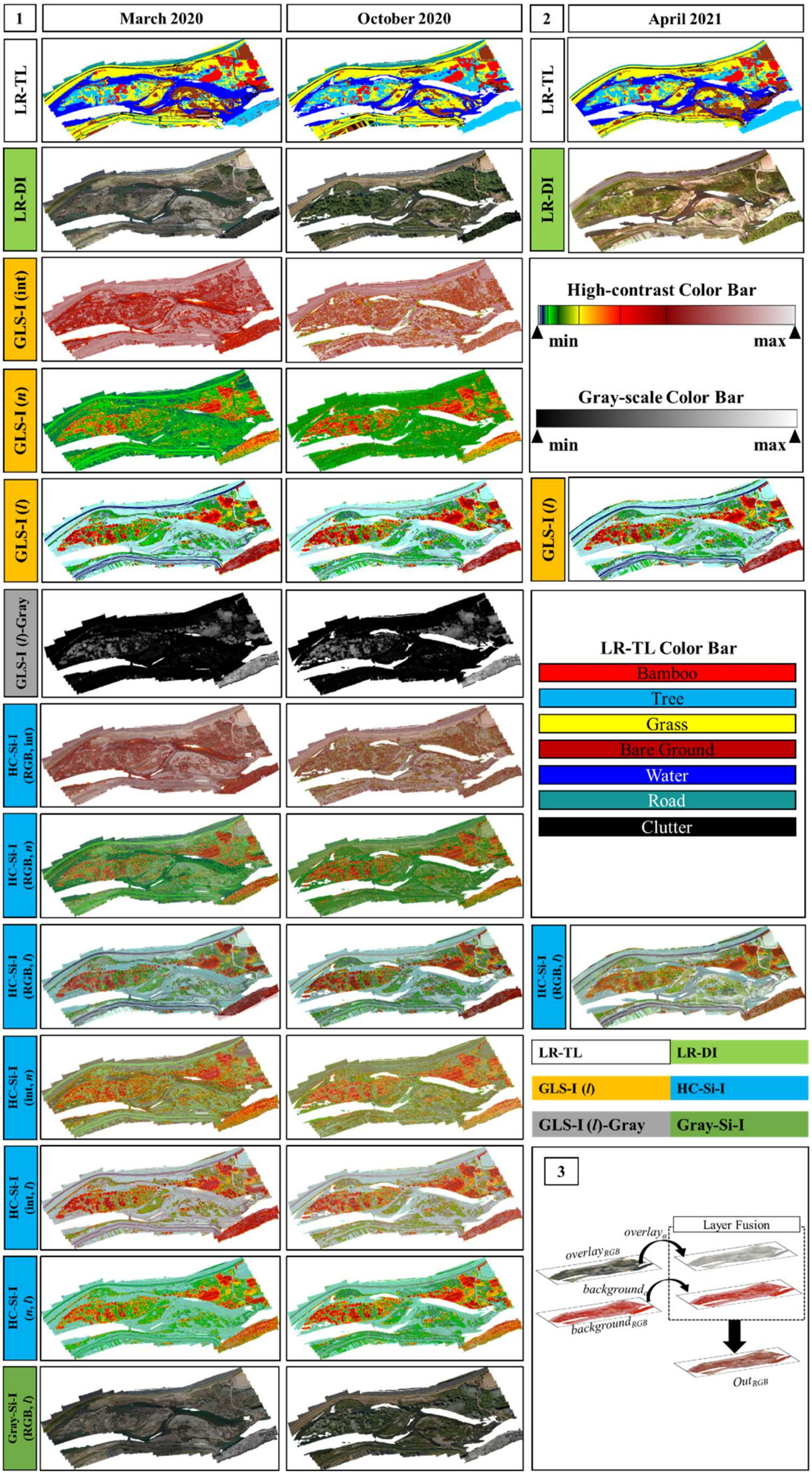


Figure 2.23 Imagery-based input and HC-Si-I processing: LR-TL, low-resolution true label.

This color scale was helpful for distinguishing the laser point cloud better and for characterizing the data better by emphasizing even weakly distinguished image features. Because of the use of multiple split color bands, the "high-contrast color scale" provided better-distinguished data for the *n* and *l* in the first half of the value. However, this was the minor exception for int, which can be attributable to its non-uniform distributions. Following the selection, *n*, *l* and int were used as transformed data with the default color scale. Finally, the transformed imagery was designated as GLS-I (*n*), GLS-I (*l*), and GLS-I (int). Subsequently, the visualized images were rendered to files using zoom-free method. Consequently, the output imagery maintained the same resolution of 1 m pixel^{-1}. In addition, another default color setting, grayscale, was used as a reference in this study for comparison with the 4-channel input RGBL in the following section with the same standard. Unlike the high-contrast color scale, which has a selected slide that is not distributed evenly across the entire color scale, grayscale has no selected slide: all the colors are distributed evenly across the entire color scale.

Because data collected using a single LiDAR sensor remain insufficient, the amount of information for imagery was increased by incorporating data from another sensor, such as a digital camera. Accordingly, we used image fusion, which is effective in the discipline of remote sensing, to create a fused image that includes clearer, more accurate, and more comprehensive information than from any single image (**Jiahuan *et al.*, 2018**). Image fusion of various dataset types was conducted using GIMP software using the gamma-corrected algorithm. Gamma correction has been used widely in remote sensing research for shallow removal (**Yavari *et al.*, 2020**) and haze removal (**Ju *et al.*, 2018**). The combined high-contrast color scale-based images were treated as High Contrast Superimposed Images (i.e., HC-Si-I). Actually, HC-Si-I is a two-by-two combination of four elements: LR-DI, GLS-I (i.e., reflection intensity (int), voxel-based laser points (*n*), and GLS-based (*l*)). Because of the 50%

transparency overlay layer, the order of the overlay and background layers need not be considered, as shown in **Figure 2.23**. Finally, HC-Si-I was created and processed using image processing software (i.e., GIMP) as the following.

1. Reduce the transparency of the overlay layer to 50%.
2. Maintain the original transparency of the background layer.
3. Merge these two layers into a single image with gamma correction (default method of layer combination in GIMP).

This operation combined features of the two layers to generate an image with new features, supplementing the information derived from single-type data. Equation (1) represents the blending layer calculation formulation (i.e., the HC-Si-I producing process).

For comparison with the subsequent 4-channel model under the same data conditions, the high-contrast color scale in GLS-I (*l*) and HC-Si-I (RGB, *l*) used above were replaced with new grayscale-based input imageries: GLS-I (*l*)-Gray and Gray-Si-I (RGB, *l*) (**Figure 2.23**).

Preparation of 4-channel imagery included three steps: disassembly, replacement, and synthesis. LR-DI (png format) was first disassembled into four grayscale images, one for each of the four channels, i.e., R, G, B, and A. In the second step, the original gray image of channel A was replaced by the gray image transformed from the GLS-derived *l* (i.e., channel L). The final step was to combine the original R-, G-, B-, and L-channel-based grayscale images into a single image (RGBL).

We considered early field observation experience during the preprocessing stage of mapping the LCC to obtain support for mapping the high-resolution true label (HR-TL). Earlier field observation photographs, as shown earlier in **Figure 2.19**, provide crucially important information such as texture and color difference for identifying target area characteristics. Consequently, based on earlier field observations of land cover attributes of our targeted river (**Yoshida *et al*., 2022**) and flow-resistance characteristics (**Green, 2005; Nepf, 2012**), we sketched classification of the natural objects in the study area with five labels: bamboo, tree, grass, bare ground, and water. Furthermore, in the HR-TL mapping, we attempted to produce a rough distinction between natural and artificial areas. Therefore, in addition to the five labels described above, we considered two additional labels in the HR-TL mapping: "roads" and "clutter" (i.e., all artificial objects except roads). Based on the earlier judgment standard (**Yoshida *et al.*, 2022**), land cover of seven types for HR-DIs from three target periods were labeled (**Figure 2.23**). The 1 m pixel^{-1} LR-DI and low-resolution true label (LR-TL) were extracted from the HR-DI and HR-TL using an interpolation-free method. As an example of **Figure 2.23**, **Figure 2.24** depicts a dataset for October 2020 campaign, including LR-DI, GLS-I (*i.e.,* int, *n*, *l*), and LR-TL.

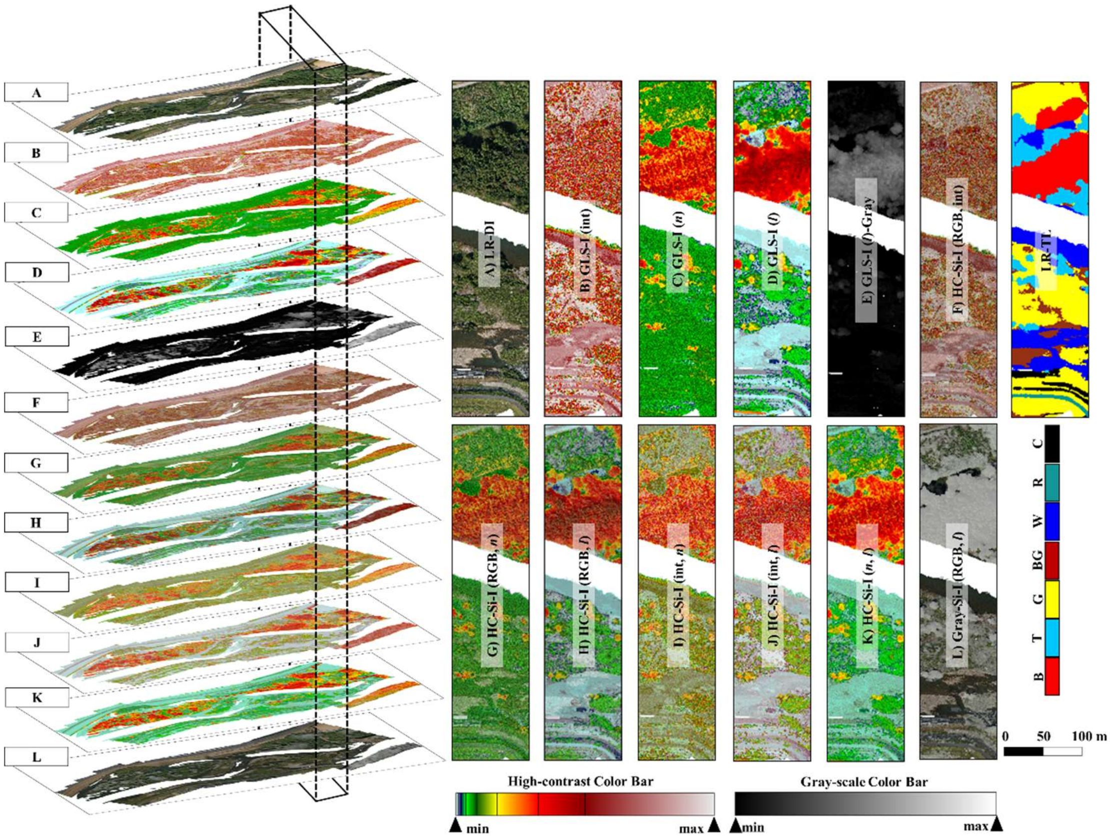


Figure 2.24 Samples of the 3-channel imagery input (*i.e.,* LR-DI, GLS-I (int), GLS-I (*n*), GLS-I (*l*), GLS-I-Gray, HC-Si-I (RGB, int), HC-Si-I (RGB, *n*), HC-Si-I (RGB, *l*), HC-Si-I (int, *n*), HC-Si-I (int, *l*), HC-Si-I (*n*, *l*), Gray-Si-I (RGB, *l*), LR-TL); (*i.e.*, Bare Ground, BG; Tree, T; Bamboo, B; Grass, G; Water, W; Road, R; Clutter, C).

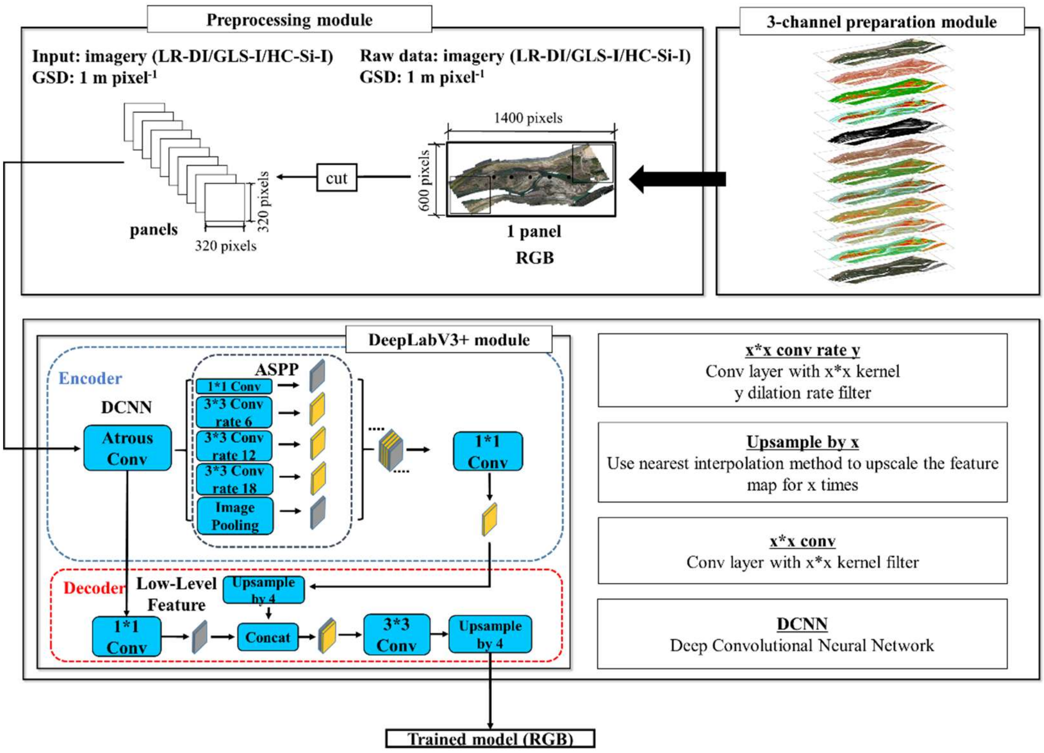


Figure 2.25 Processing of mapping LCC with 3-channel DeepLabv3+ model; GSD: Ground Sample Distance.

2.2.3 Preprocessing Module of LCC mapping

Chen *et al.* (2018) demonstrated that the DeepLabv3+ model's backbone (i.e. Xception) minimized the original input data size (width, height, and channels) as (299, 299, 3). After considering the overall size of the input imagery (1400, 600, 3), we settled on (320, 320, 3) as the uniform input data size. The preprocessing module's primary function is to convert the prepared input imagery into multiple (320, 320, 3) panels. As **Figure 2.25** shows, the crop method (i.e., dividing the larger imagery into panels) employs a sliding (320, 320, 3) window with a (320, 320, 3) window-size stride, and saves the name of the imagery with the position of the most upper-left pixel. It is noteworthy that the 4-channel input format (RGBL) should be transformed as (320, 320, 4) from the 3-channel-based format (320, 320, 3). Following the inference, the operator can combine all the inferred panels with this position into a single image.

2.2.4 Models of producing LCC mapping

2.2.4.1 RGB (3-channel)-based DeepLabv3+ model

To extract features from the input RGB (3-channel)-based imagery (e.g., LR-DI, GLS-I, HC-Si-I), the DeepLabv3+ model initially employs an "encoder–decoder" structure (Chen et al., 2018). The encoder of this model is assisted by the atrous spatial pyramid pooling module (**Chen *et al.*, 2018**) to capture multi-scale features via atrous convolution with different dilation rates (e.g., 6, 12, 18) and image-level feature pool. These multi-scale feature maps are then concatenated together, upsampled by a factor of 4, and then concatenated with high-resolution feature maps extracted from the bottom of the encoder module. As a final point, category prediction is performed on this merged feature map. The predictions are then upsampled to produce the output. The model's parameters are then optimized using true label (TL). Subsequently, these parameters are saved as a "trained model (RGB)". This training process determines the relation between the input data: input imagery and TL. The workflow for processing LCC mappings using the DeepLabv3+ model is depicted in **Figure 2.25**. The DeepLabv3+ processing is represented as (a) training phase: [Input imagery] → [DeepLabv3+ model] → [Trained model] and (b) inference phase: [Input imagery] → [trained model] → [LCC as output].

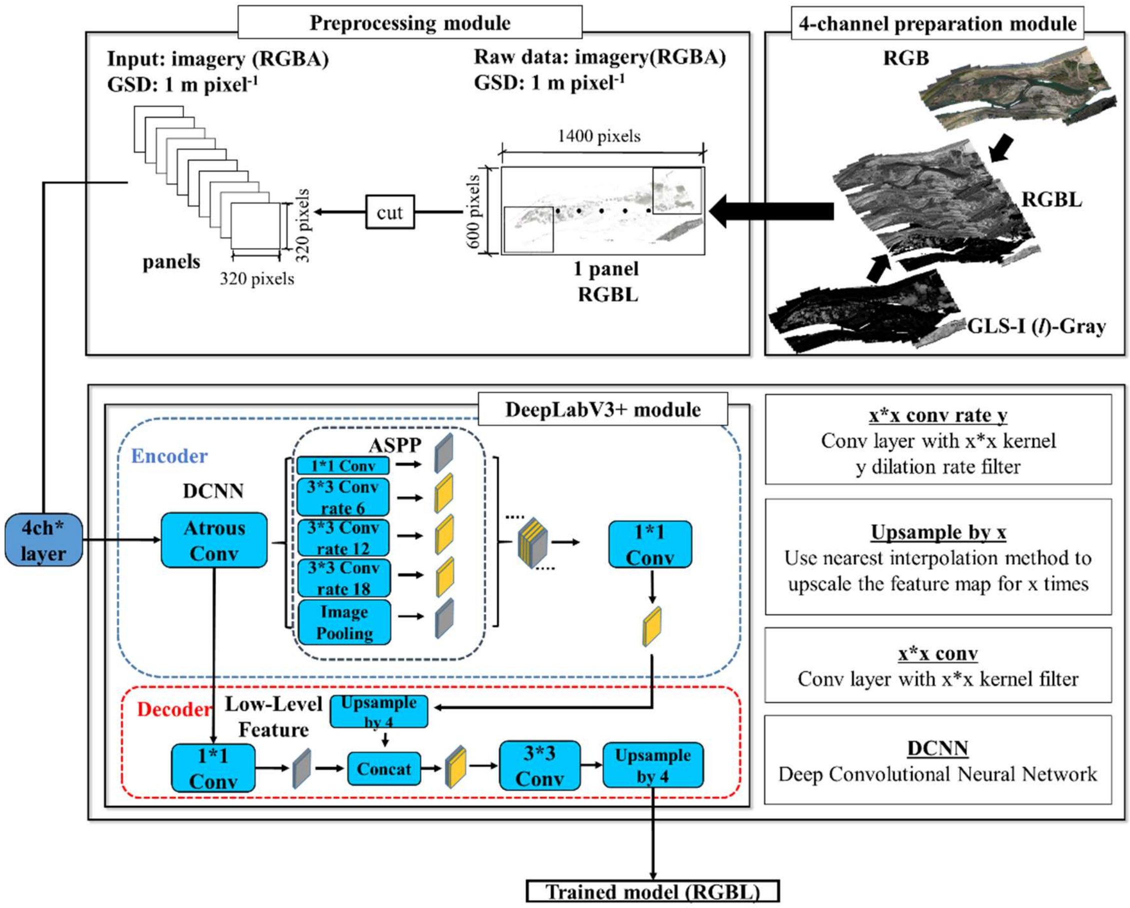


Figure 2.26 Processing of mapping LCC with 4-channel DeepLabv3+ model (*i.e.*, LiDAR cooperated RGB method, LiCR method).

2.2.4.2 4-channel DeepLabv3+ model

As shown in **Figure 2.26**, the transformed RGBL imagery is no longer compatible with the 3-channel-based DeepLabv3+ model, which has only three channels. For adapting to this newly generated 4-channel imagery, modification of other parts of the DeepLabv3+ model is unnecessary, except for modification of the entrance by adjusting the number of channels in the input layer.

2.2.5 Comparison of LCC mapping

The input imageries for deep learning (**Table 2.14**) are divided into four comparison groups

(**Table 2.15**). In addition, the main analysis processing is divided into the following four groups.

Table 2.15 Comparison conditions of the training and validation datasets in the three targeted periods

	Input Imagery	Periods	Sites	Models
Group 1	LR-DI (RGB)	Mar. & Oct. 2020 (T & P)	Gion	3-channel
	GLS-I (int)	Mar. & Oct. 2020 (T & P)		
	GLS-I (*n*)	Mar. & Oct. 2020 (T & P)		
	GLS-I (*l*)	Mar. & Oct. 2020 (T & P)		
	HC-Si-I (RGB, int)	Mar. & Oct. 2020 (T & P)		
	HC-Si-I (RGB, *n*)	Mar. & Oct. 2020 (T & P)		
	HC-Si-I (RGB, *l*)	Mar. & Oct. 2020 (T & P)		
	HC-Si-I (int, *n*)	Mar. & Oct. 2020 (T & P)		
	HC-Si-I (int, *l*)	Mar. & Oct. 2020 (T & P)		
	HC-Si-I (*n*, *l*)	Mar. & Oct. 2020 (T & P)		
Group 2	GLS-I (*l*)-Gray	Mar. & Oct. 2020 (T & P)	Gion	3-channel
	Gray-Si-I (RGB, *l*)	Mar. & Oct. 2020 (T & P)		
	RGBL	Mar. & Oct. 2020 (T & P)		4-channel
Group 3	HC-Si-I (RGB, *l*)-Apr	Mar. or Oct. 2020 (T), Apr. 2021 (P)	Gion	3-channel
	HC-Si-I (RGB, *l*)	Mar. & Oct. 2020 (T & P)		
Group	GLS-I (*l*)	Mar. & Oct. 2020 (T), Oct. 2021 (P)	Gion & Heidan	3-channel
	HC-Si-I (RGB, *l*)	Mar. & Oct. 2020 (T), Oct. 2021 (P)		

T, Train; P, Predict; T & P, Mutual train and predict

Group 1: Compare the DeepLabv3+ model's performance based on different input datasets and choose the best solution among them. The comparison indices include the average and absolute difference values of both overall accuracy (OA) and Macro-F1 for the two predicted results. Mutual results were based on the one solution-trained dataset and were then predicted with the other for the data (i.e., GLS-derived datasets in March and October 2020).

Group 2: Based on the results of Group 1, compare methods using RGB and *l* as the best input dataset in different models with three cases (GLS-I (*l*)-Gray for LiDAR-derived input data, Gray-Si-I for input data combination with 3-channel DeepLabv3+ model, and 4-channel DeepLabv3+ model for channel entry extension). This group was used to compare the difference in accuracy when using the same data (Gray-Si-I (RGB, *l*) and RGBL) with different input channels (3-channel and 4-channel) and different methods (Superimposed method and Alpha layer change). Furthermore, GLS-I (*l*)-Gray was compared to Gray-Si-I (RGB, *l*) and RGBL as a reference.

Group 3: Results of the earlier two groups clarify that the HC-Si-I method has some advantages. Then again, the performance and stability of this method must be tested when inferred from new data. Therefore, we tried to test the data collected using GLS in April 2021. This group was trained with March and October 2020 datasets and was predicted with April 2021 data. Results were compared to HC-Si-I (RGB, *l*) in the first group using the same processing rules as those used for Group 1.

Group 4: After testing data from heterogeneous periods, another location was chosen to test the generalizability of HC-Si-I, especially for the LCC labels of water and bare ground.

2.2.5.1 Group 1: Best combination of input data type for LCC mapping

Data source-based comparison

Depending on the quantity of information or data source, this group can be divided roughly into three main sections as presented below (**Figure 2.27**):

Part 1: Only low-resolution UAV images (i.e., LR-DI (RGB))

Part 2: GLS-I using only single LiDAR-derived data (i.e., GLS-I (int), GLS-I (*n*), GLS-I (*l*))

Part 3: HC-Si-I that combines LR-DI and GLS-I (i.e., HC-Si-I (RGB, int), HC-Si-I (RGB, *n*), HC-Si-I (RGB, *l*)), or GLS-I and GLS-I superimposed into a new image (i.e., HC-Si-I (int, *n*), HC-Si-I (int, *l*), or HC-Si-I (*n*, *l*))

In terms of accuracy comparison, increasing the amount of information improves accuracy (i.e., the mean value of Part 3 was improved compared to Part 1 (HC-Si-I) or the mean value of Part 2 (GLS-I)).

Data-type-based comparison

Furthermore, depending on the data types and their combination, the accuracies of LR-DI (RGB), GLS-I (int), HC-Si-I (RGB, int), HC-Si-I (RGB, *n*), and HC-Si-I (int, *n*) have some advantages over GLS-I (*n*) using *n* alone, but the cases with *l* have much higher accuracy (i.e., GLS-I (*l*), HC-Si-I (RGB, *l*), HC-Si-I (int, *l*), and HC-Si-I (*n*, *l*)).

The following are reasons for these results: RGB, *n*, and int are the most vulnerable to seasonal environmental and insolation changes, leading to variations in land cover conditions. However, to minimize the effects of these changes, cases including *l* are expected to improve accuracy while maintaining stability. **Figure 2.27** shows that the only four cases in Group 1 with an average Macro-F1 score close to 0.7 were GLS-I (*l*), HC-Si-I (RGB, *l*), HC-Si-I (int, *l*), and HC-Si-I (n, *l*), all of which involved *l*. The highest average OA value among them was 0.78 for HC-Si-I (RGB, *l*). All absolute difference values were less than 0.02. The results also indicated that GLS-I (*l*) can reflect LCC to some degree.

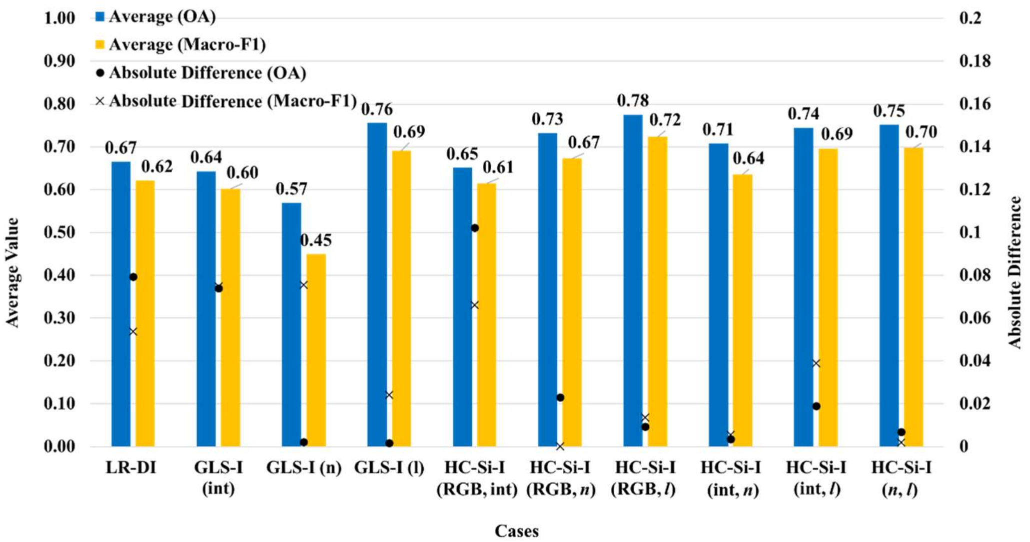


Figure 2.27 Group-1 (*i.e.* best imagery data type for DeepLabv3+ model-based LCC mapping) with typical comparative indices, including averaged and absolute values of overall accuracy (OA) and Macro-F1.

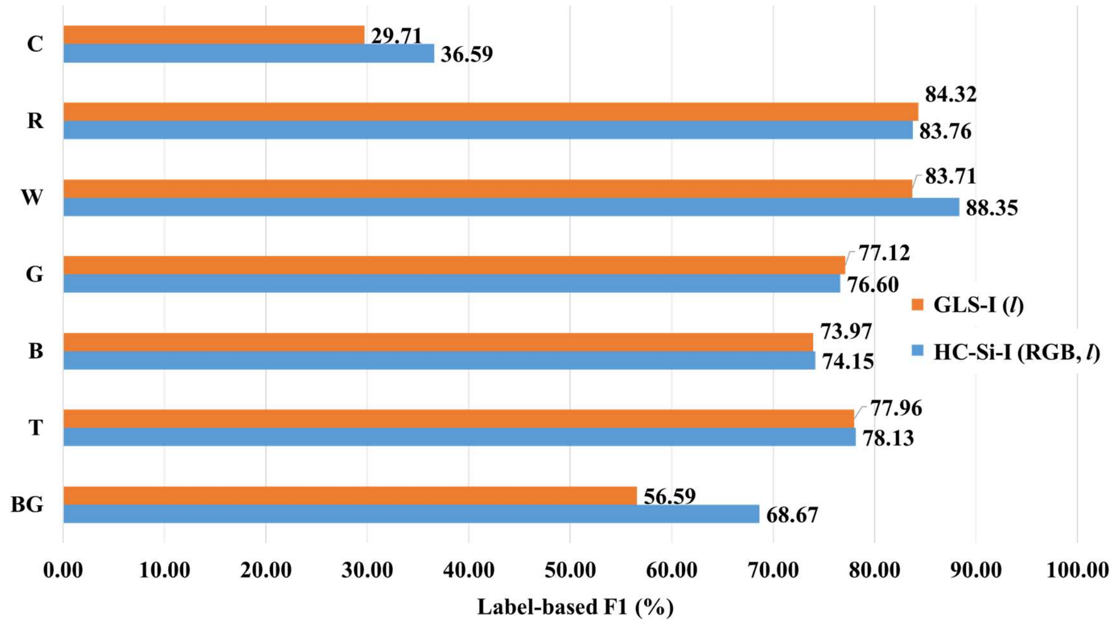


Figure 2.28 Label-based comparison of GLS-I (*l*) and HC-Si-I (RGB, *l*) in Group-1 with targeted seven labels.

Comparison of the high-performance l-based method: GLS-I (l) and HC-Si-I (RGB, l)

Label-based comparison results are classifiable into three parts for consideration (**Figure 2.28**).

(1) Vegetation parts (Grass, Bamboo, Tree)

The accuracy of HC-Si-I (RGB, *l*) in classifying targeted vegetation (i.e., Grass, Bamboo, Tree) was not found to be significantly different from that of the GLS-I (*l*). The marginal variation is attributable to the fact that the local planar distribution information provided by GLS-I (*l*) was sufficient for the DeepLabv3+ model to learn its features (including but not limited to the undulations of vegetation-to-vegetation distribution, the position of vegetation communities, and the vegetation communities fixed at a certain value interval).

(2) Vegetation-off homogeneous pattern part (Road, Water)

Considering the disparity in data provided by GLS-I (*l*) and HC-Si-I (RGB, *l*), the color of the road and water remains largely unchanged (the water might change because of sunlight and water quality immediately after minor flooding). In terms of height (*l*), both values are nearly identical (i.e., 0 m), resulting in difficulties in learning features locally. However, because of the fact that the study location remains unchanged, the extent and contours of the presence of road and water sections in GLS-I (*l*) were sufficient to extract features (i.e., the planar distribution from the general view). Therefore, if the test area is changed, then the possibility exists that the accuracy will be lowered, as described later for reference in the section 5.4. In contrast, when using HC-Si-I (RGB, *l*), accuracy is guaranteed because of the amount of information supplemented when compared to GLS-I (*l*).

(3) Vegetation-off heterogeneous pattern part (Bare Ground, Clutter)

Distinguishing between grass and bare ground labels solely on GLS-I (*l*) would be extremely difficult because both have similar heights. Moreover, the features distinguished in height are not noticeable. That fact might lead to less accuracy in LCC for these two labels. Furthermore, because clutter is an artifact, parts of it can be changed over time, resulting in a complete change in character (i.e., civil engineering constructions site). Moreover, the amount of data available was insufficient for the model to learn the features (less pixel amount of clutter).

Tables 4–7 present the accuracy valuation indices (OA and Macro-F1) calculated using pixel-based confusion matrix in comparison to corresponding TL for the GLS-I (*l*) and HC-Si-I (RGB, *l*) cases. Because the number of grass (G) labels is approximately three times that of bare ground (BG) labels, the accuracy of grass is higher, although their distinguishing features are nearly identical. In Tables 4 and 5, the results demonstrated that GLS-I (*l*) misclassified from bare ground to water areas. Furthermore, the GLS-I (l)-derived labels were based on both the pattern's outline and color, with the bare ground label and the water label having similar colors in GLS-I (*l*). In contrast, whereas the water label had similar outlines in March and October 2020, the bare ground label's outline varied greatly because of river management tasks (**Figure 2.19**) and grass growth and decay. The water label had high accuracy under the condition of similar outlines (water level) in different seasons for the same panel where features are learned (**Figure 2.25**). In contrast, because of large changes in the hetero-seasonal conditions, the bare ground label had low accuracy. However, as a result of combining LR-DI (RGB) and GLS-I (*l*) to increase features, HC-Si-I (RGB, *l*) was found to have significantly improved accuracy. The study revealed that when the HC-Si-I (RGB, *l*) approach was used, the recall of bare ground for the GLS-I (*l*) case using March 2020 *l* as the training dataset and October 2020 *l* as the validation dataset improved from 59.07% (**Table 2.16**) to 75.81% (**Table 2.18**). Furthermore, in the case of October 2020 *l* as the training dataset and March 2020 *l* as the validation dataset, the recall of the bare ground was improved from 50.74% (**Table 2.17**) to 66.86% (**Table 2.19**). Moreover, with the exception of the bare ground and clutter parts, the recall of other labels was maintained with an absolute difference of less than 10% (**Tables 2.16–2.19**).

Table 2.16 Confusion matrix of GLS-I (*l*) using March 2020 *l* as training dataset, October 2020 *l* as validation dataset (*i.e.*, T03P10 GLS-I (*l*)).

T03P10 GLS-I (*l*)									
	BG	T	B	G	W	R	C	Total	RC (%)
BG	25362	860	121	6361	9421	384	426	42935	59.07
T	1482	74060	4267	11877	3030	96	589	95401	77.63
B	407	4785	20512	3259	251	0	122	29336	69.92
G	14279	6401	764	111294	3629	3758	1972	142097	78.32
W	6587	1081	24	1895	64990	0	540	75117	86.52
R	1590	10	0	3328	4	28204	332	33468	84.27
C	1823	270	133	6977	976	1256	2142	13577	15.78
Total	51530	87467	25821	144991	82301	33698	6123	431931	
PR (%)	49.22	84.67	79.44	76.76	78.97	83.70	34.98		
OA = 0.76, Macro-F1 = 0.68									

RC, Recall; PR, Precision; OA, Overall Accuracy; Macro-F1, Macro-averaged F1 score

Table 2.17 Confusion matrix of GLS-I (l) using October 2020 l as training dataset, March 2020 l as validation dataset (*i.e.*, T10P03 GLS-I (l)).

T10P03 GLS-I (l)									
	BG	T	B	G	W	R	C	Total	RC (%)
BG	33075	2778	338	13606	13537	845	1001	65180	50.74
T	842	61464	2342	6833	906	46	90	72523	84.75
B	188	8142	22660	2259	65	0	26	33340	67.97
G	5982	14356	2535	110491	1747	4678	3350	143139	77.19
W	4365	3978	224	4007	86333	32	733	99672	86.62
R	758	103	0	2202	2	29743	993	33801	87.99
C	824	745	171	5582	1228	1116	4792	14458	33.14
Total	46034	91566	28270	144980	103818	36460	10985	462113	
PR (%)	71.85	67.13	80.16	76.21	83.16	81.58	43.62		
OA = 0.75, Macro-F1 = 0.70									

Table 2.18 Confusion matrix of HC-Si-I (RGB, l) using March 2020 l as training dataset, October 2020 l as validation dataset (*i.e.*, T03P10 HC-Si-I (RGB, l)).

T03P10 HC-Si-I (RGB, l)									
	BG	T	B	G	W	R	C	Total	RC (%)
BG	33767	1175	140	6022	2403	535	500	44542	75.81
T	1127	78859	5458	9109	2996	209	366	98124	80.37
B	209	5531	20366	2915	226	0	119	29366	69.35
G	15320	9000	1692	105951	3183	3691	4757	143594	73.79
W	5028	1372	21	1410	65364	8	794	73997	88.33
R	1193	50	2	2812	83	29397	454	33991	86.48
C	2470	549	148	5526	799	1354	4807	15653	30.71
Total	59114	96536	27827	133745	75054	35194	11797	439267	
PR (%)	57.12	81.69	73.19	79.22	87.09	83.53	40.75		
OA = 0.77, Macro-F1 = 0.72									

Table 2.19 Confusion matrix of HC-Si-I (RGB, *l*) using October 2020 *l* as training dataset, March 2020 *l* as validation dataset (*i.e.*, T10P03 HC-Si-I (RGB, *l*)).

T10P03 HC-Si-I (RGB, *l*)									
	BG	T	B	G	W	R	C	Total	RC (%)
BG	43664	2251	398	10503	5943	1135	1410	65304	66.86
T	1086	58507	3389	8647	872	22	121	72644	80.54
B	202	5805	25263	2018	43	0	59	33390	75.66
G	6404	12182	2686	109774	2132	7293	2753	143224	76.64
W	2420	3613	341	4599	88338	14	647	99972	88.36
R	541	84	0	1618	197	30901	887	34228	90.28
C	1342	427	88	5485	1032	1286	4795	14455	33.17
Total	55659	82869	32165	142644	98557	40651	10672	463217	
PR (%)	78.45	70.60	78.54	76.96	89.63	76.02	44.93		
OA = 0.78, Macro-F1 = 0.73									

2.2.5.2 Group 2: Model comparison (3-channel or 4-channel) with RGB and l

Although GLS data are stable across seasons (**Islam *et al.*, 2022**), digital photographs can provide more details and features than LiDAR, such as the difference between bamboo forest and trees during the lush vegetation period, the color difference between bare ground and water, and the feature difference between the grass and bare ground. In addition, when confronted with the difficulty of overcoming seasonal differences, simply using LiDAR or digital photographs as input might necessitate a large amount of data to be trained to find the relations between them, which is one limitation of this study. Consequently, with limited data, ascertaining how to characterize the data effectively and how to allow the model to learn its features have come to pose new challenges.

Therefore, whether the model can learn the features more effectively by transforming LiDAR and digital imagery into 3-channel input data with new features or by directly feeding the unprocessed 4-channel data is a topic that must be discussed. To validate the 4-channel part of the idea above, the DeepLabv3+ model's input layer must be changed from a 3-channel-based format to a 4-channel-based format while retaining the DeepLabv3+ model's internal structure. According to Group 2 results (**Figure 2.29**), 3-channel-based Gray-Si-I (RGB, *l*) outperforms the 4-channel-based LiDAR cooperated RGB method (i.e., LiCR (RGB, *l*)) slightly in terms of both the average and absolute difference estimates, despite using the same amount of data.

Based on this comparison group, it can also be inferred under the condition of the limited input dataset that the internal structure of the original model remains unchanged, that the input data can be processed more directly to be visualized more easily, and that much higher accuracy is attainable. In contrast, the input dataset for the 4-channel DeepLabv3+ model might be insufficient for training to extract features of the RGBL data. Furthermore, irrespective of the fact that the 3-channel Gray-Si-I (RGB, *l*) used more data than the 3-channel GLS-I (*l*)-Gray, no significant improvement was found in the valuation indices (**Figure 2.29**), implying that simply collecting more data is unlikely to help the model learn features more effectively using the gray-scale color bar.

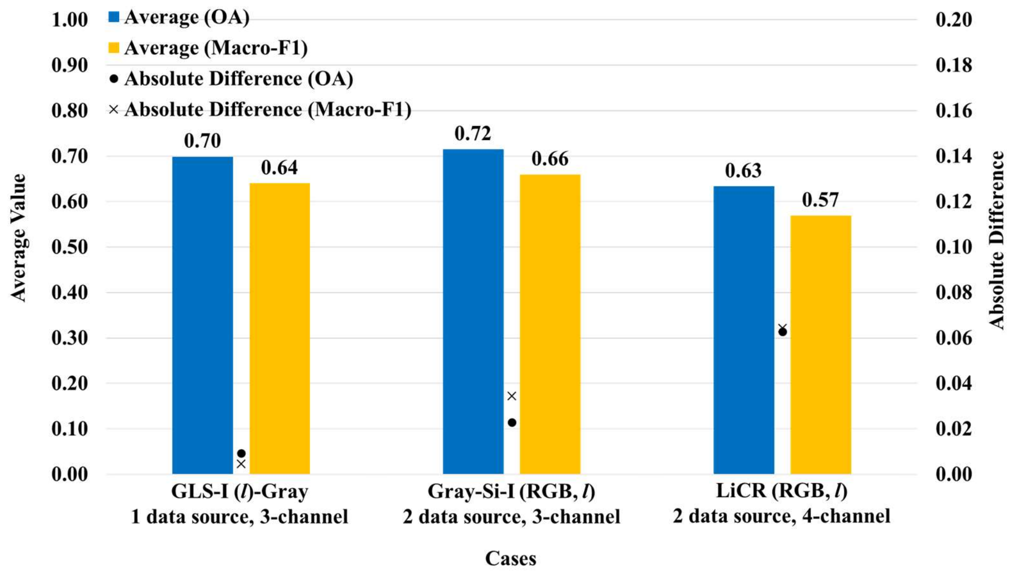


Figure 2.29 Group-2 (*i.e.*, comparison of 3-channel-based and 4-channel-based methods using RGB and *l*) with typical comparative indices.

2.2.5.3 Group 3: Seasonal change test of HS-Si-I (RGB, l) method

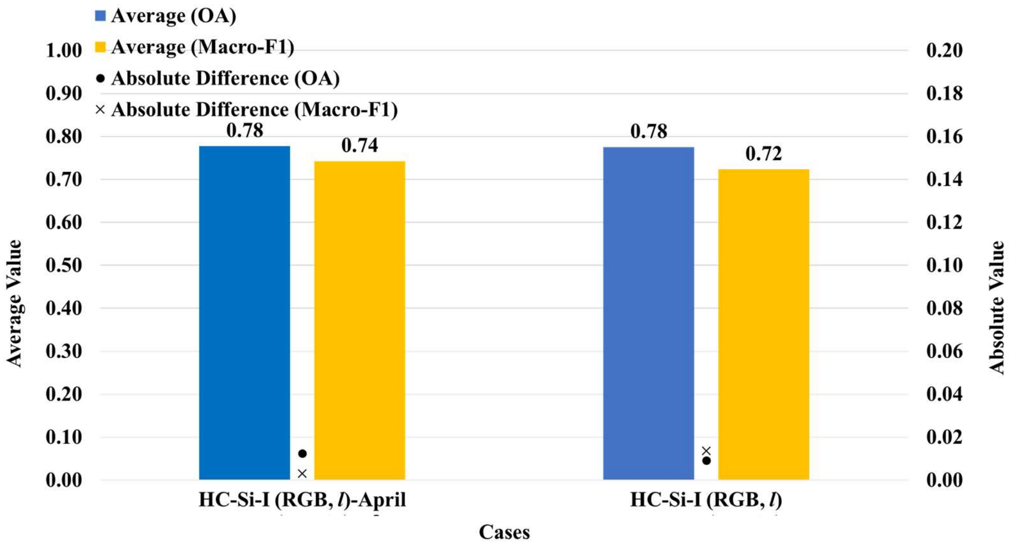


Figure 2.30 Group-3 (*i.e.*, test the HC-Si-I method with April 2021 dataset, comparing with the HC-Si-I (RGB, *l*) result in Group-1) with typical comparative indices.

From group 3 results, as shown in **Figure 2.30**, it is apparent that the HC-Si-I method maintains high performance and stability even after validation data were changed to data taken in April 2021. The fact that HC-Si-I (RGB, *l*)-April maintained comparable accuracy to that of HC-Si-I (RGB, *l*) was attributable to a shorter data collection interval between the valid and the training datasets (i.e., from March 2020 to April 2021; Table 3). Furthermore, despite a few upstream tree cutting areas and minor flooding between the targeted periods, no significant topographic changes were found during this interval. However, seasonal changes were found in foliage conditions with no significant change in GLS-based *l* estimations. Therefore, based on results of comparisons of the three groups described above, this method can achieve more accurate mutual prediction of GLS-based imagery input between different periods in the targeted area.

2.2.5.4 Group 4: Position change test of GLS-I (l) and HC-Si-I (RGB, l) method

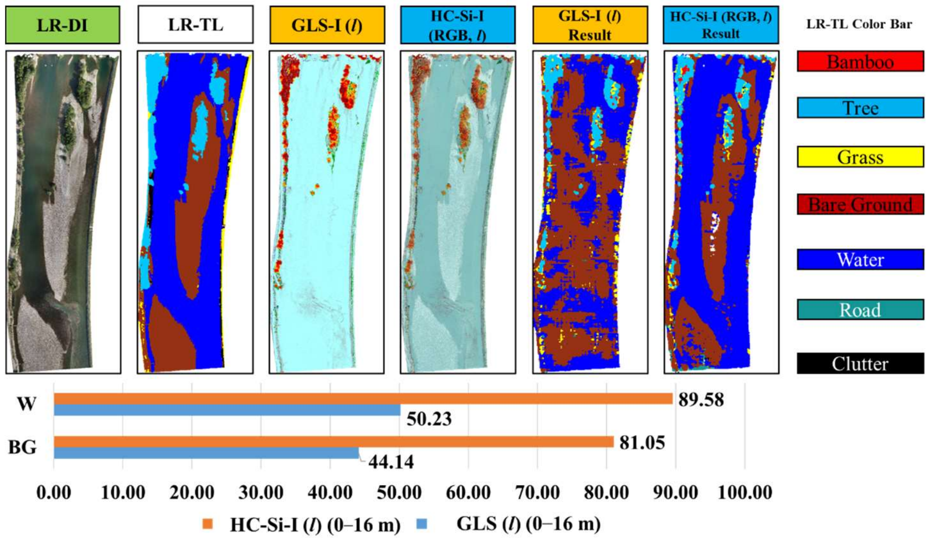


Figure 2.31 Group-4 (*i.e.*, test the GLS-I (*l*) and HC-Si-I (RGB, *l*) method with Heidan October 2021 dataset) with typical comparative indices of W-labels and BG-labels.

As stated previously for Group 1 (**Figure 2.28**), the accuracy of bare ground and water labels differed greatly when only GLS-I (*l*) was used. The absolute difference was nearly negligible when compared to the HC-Si-I (RGB, *l*) method. Because of their nearly identical *l* values, the possibility exists of lowering GLS-I (l) accuracy, particularly in water areas with changing test sites. To verify whether the method used for this study is applicable to other study sites, the two best-performing cases in Group 1 were chosen for comparison: GLS-I (*l*) and HC-Si-I (RGB, *l*). Using the same method as that shown in **Figure 2.25**, data from March and October 2020 measures from the Gion site were trained to predict data from October 2021 measures from the Heidan site. Test site findings (**Figure 2.31**) revealed that the accuracy of the water label declined drastically when using GLS-I (*l*) in comparison to the result demonstrated earlier (**Figure 2.28**), as discussed in section 5.1. Furthermore, the HC-Si-I (RGB, *l*) approach maintained higher accuracy in both water and bare ground labels. Therefore, test results for the specific two labels indicate that HC-Si-I (RGB, *l*) is more generalizable than GLS-I (*l*).

2.2.6 Conclusion

Based on the original DeepLabv3+ model for single camera drone images as input data in the LCC problem, the new 3-channel input dataset was proposed for this work by including LiDAR data transformed with default color scales (i.e., high-contrast color scale, gray scale) and data superimposing data of multiple types, i.e., RGB + LiDAR, LiDAR + LiDAR. The newly generated 3-channel input data were used to classify riverine land cover, which was divided into four groups for comparison. In Group 1, the LCC accuracy was improved by increasing the amount of source data, whereas cases using *l* data were beneficial in terms of both accuracy (average indices) and stability (absolute difference indices). The GLS-I (*l*) and HC-Si-I (RGB, *l*) approaches outperformed in all cases using *l* data. In addition, HC-Si-I (RGB, *l*) improved the accuracy of the bare ground and water labels, whereas GLS-I (*l*) cannot differentiate with high indices. In group 2, the 3-channel model with our input dataset outperforms the 4-channel approach with the gray scale color bar, despite using the same amount of data. Such a drop in performance, even after upgrading the channel, might be attributed to an insufficient training dataset, which should be considered in future studies with inclusion of more investigations. Furthermore, despite considering more data in the same 3-channel-based model, no significant improvement was achieved in the average and absolute difference indices, implying that simply adding more information with simple gray scaling is unlikely to help the model learn features more effectively. Furthermore, the generalizability of HC-Si-I (RGB, *l*) was well demonstrated in both Group 3 and Group 4 using heterogeneous seasonal and site data. To conclude, when compared to the earlier study conducted by Yoshida et al. (2022) using RGB and ALB datasets for the same river and vegetation species targeted herein, our results obtained for different seasons (Macro-F1 was close to 0.70) showed improvement of about 15%, particularly in terms of Macro-F1 estimate differences. Although earlier flood simulation results based on the same seasonal-based LCC findings were reasonable (Macro-F1 was close to 0.84), their hetero-seasonal results were insufficient (Macro-F1 was close to 0.55). Therefore, using only hetero-seasonal GLS-derived *l*, the current LCC findings demonstrated more accurate monitoring of riparian vegetation, particularly tree species, implying that the current LCC findings might be adequate for flow modeling purposes and proper ecosystem management tasks.

2.3 Interchangeability of Cross-platform Orthophotograph applied on LCC

2.3.1 Introduction

River environmental information includes crucially important data such as topographic bathymetry and vegetation attributes that are necessary to develop balanced river management measures, addressing issues such as flood control (**Yoshida *et al.*, 2021**) and ecosystem management (**Mandlburger *et al.*, 2015**). At past time, to get these mentioned data, field surveys that require personnel to entry to the site are necessary. These years, to overcome this limitation, digital surface models (DSMs) and digital terrain models (DTMs) have become indispensable tools for describing terrain conditions (i.e., using DSM minus DTM to infer the height of the land cover). Incidentally, Light Detection and Ranging (LiDAR) technology generates high-resolution and accurate DSMs and DTMs by halving the time between the emitted pulse and detection of the reflected echo (**Yan *et al.*, 2015**). Based on LiDAR technology characteristics, airborne LiDAR bathymetry (ALB) and unmanned aerial vehicle (UAV)-borne green LiDAR system (GLS) have been already applied in the riverine environment measurements in Japan (**Islam *et al.*, 2020**). In accordance with the mentioned measurements results (i.e., DSM minus DTM), how to classify the land cover using these results is becoming one of the tasks in the river engineering-related research.

These years, several approaches for overcoming land cover classification (LCC) difficulties have been tested, including decision tree algorithm (**Yu *et al.*, 2006**), manual setting of thresholds (**Do *et al.*, 2019**), and deep learning (**Yoshida *et al.*, 2022**). Considering the time cost on discovering and extracting the feature of data, the deep learning method, particularly with the atrous convolution module has some advantage in effective feature finding. Therefore, DeepLabv3+ model with atrous convolution module was chosen for this study.

Subsequently, with the help of the DeepLabv3+ model, the high accuracy of LCC producing has been well proven from past studies (**Yoshida *et al.*, 2022**). These studies utilized ortho-photographs with LiDAR data to classify the land cover using DeepLabv3+ model with additional module for adding LiDAR dataset with photographs. It also revealed that, because of the similar feature, recognition accuracy of LCC is high (i.e., averaged overall accuracy is almost 90%) when training and validation sets were selected from the similar location and period. However, this mentioned study was based on the same platform (e.g., training ALB dataset, predicting ALB dataset). If the data collected by different platforms, the impacts on the LCC mapping results derived by this operation has not been demonstrated yet. Thus, one more step, to mutual-predict cross-platform LCC to verify the interchangeability (i.e., train ALB dataset, predict GLS dataset; train GLS dataset, predict ALB dataset) becomes the target in this study. Alternatively, instead of using additional module for accuracy improvement, this study tried to increase the feature by image fusion derived from imagery and LiDAR. And how to transform the LiDAR data into imagery in an expressive way is also required in this study.

To observe the impact of cross-platform on LCC in this study as most as possible, the resolution and data styles (i.e., digital imagery, LiDAR) of ALB and GLS in mutual-prediction are chosen as same. On the other side, because of the data limitation, instead of totally same season, the similar season-related ALB and GLS dataset are selected. Eventually, the comparisons in this research obtained from cross-platform mutual-predictions can facilitate understanding of cross-platform data features, and determine a reasonable method to retain the robustness using deep learning method with cross-platform data in LCC mapping producing.

2.3.2 Study site and methods

2.3.2.1 Study site

Figure 2.32 depicts our study site. As a consequence of that, as shown in **Figure 2.33 (*a*), (*b*),** both of ALB and GLS measurement include overland and underwater area, and the land cover species all include the mentioned seven labels.

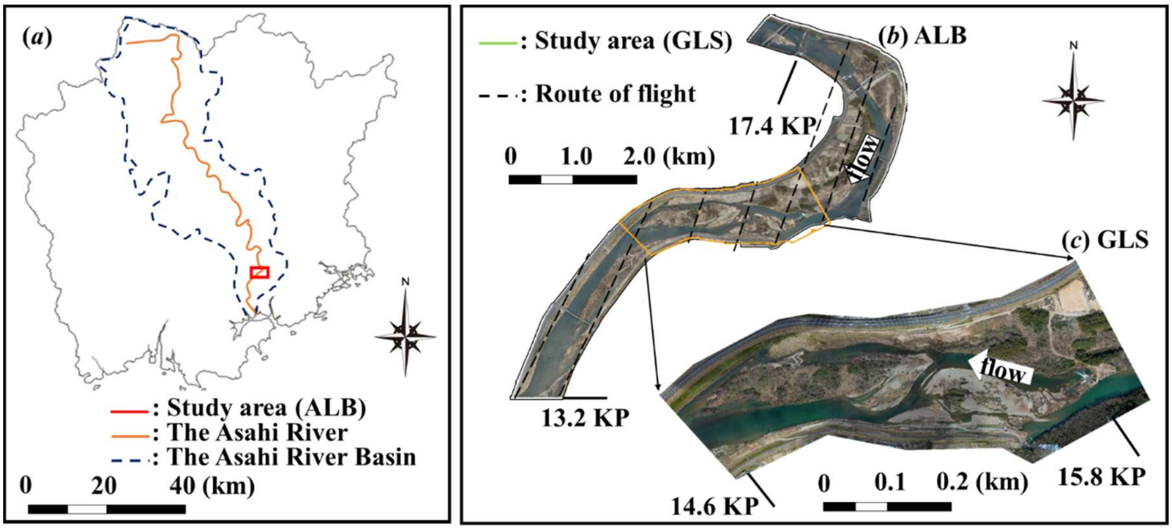


Figure 2.32 Perspective of Airborne LiDAR Bathymetry and Green LiDAR measurement area: (*a*) location of the Asahi River in Japan with kilo post (KP) values representing the longitudinal distance (km) from the river mouth, (*b*) aerial-captured photographs based on the marked positions in (*a*), and (*c*) drone-captured photographs based on the marked positions in (*b*).

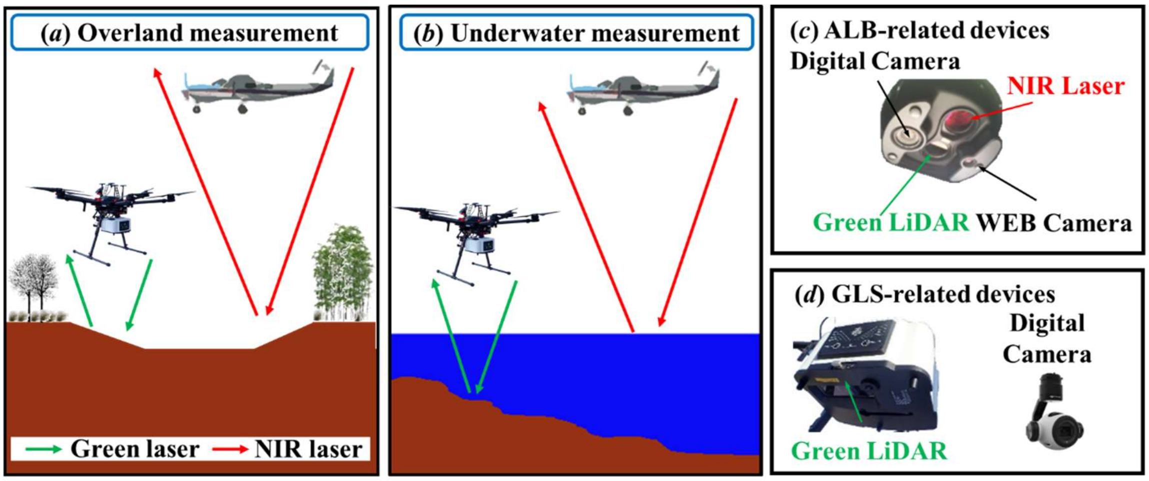


Figure 2.33 In (*a*) overland and (*b*) underwater surveys, Light Detection and Ranging **(**LiDAR) using a Near InfraRed (NIR) from (*c*) and green laser from (*d*), for ALB and GLS, respectively.

2.3.2.2 Data collection

For this study, we conducted ALB (Leica Chiroptera II; Leica Corp.) surveys in March and November 2017 along a 4.2 km reach of the lower Asahi River (13.2–17.4 KP) controlled by the national government. As shown in **Figure 2.32 (*b*)**, multiple flight operations were conducted in leaf-off (i.e., March 2017) condition to achieve overlapping coverage of the target area. The current system scanned the river channel for LCC using aircraft-mounted Near InfraRed (NIR) and green lasers as presented in **Figure 2.33 (*a*)**. The device commonly uses the green laser to detect underwater (bottom) surfaces because green light can penetrate the water column to some degree. Conversely, the NIR laser is used to detect terrain surfaces, including vegetation, because it is readily reflected by the air-water interface. In the case of ALB, only NIR was used to calculate DSM and DTM.

Moreover, during each ALB measurement, a digital camera as shown in **Figure 2.33 (*c*)**, mounted directly beneath the aircraft took aerial photographs of the target river. Among other things, to remove tilt and relief effects, the aerial photographs were converted to orthophotos. Herein, the aerial photograph over-lap and side-lap ratios were respectively greater than 60% and 30%.

Table 2.20 Specifications of the present GLS and ALB system and measurement conditions in the targeted river reach

Items		ALB 2017 Mar.	Nov.	GLS 2020 Mar.	Oct.
Laser wavelength range (nm)	Green	515	515	532	532
	NIR	1,064	1,064	-	-
Number of laser beams (10^3 s^{-1})	Green	35	35	60	60
	NIR	148	148	-	-
Ground altitude (m)		500	500	50	50
Flight speed (km h^{-1})		220	110	9	9
Density of measurement points (m^{-2})	Green	2	4	100	100
	NIR	9.0	9.0	-	-
Resolution of raw imagery (cm $pixel^{-1}$)		10	10	3	3
Resolution of raw LiDAR (m $pixel^{-1}$)*		2	2	1	1
FTU**		-	-	0.8	3.12
NTU***		-	-	-	3.7
Degree****		2.9	3.2	-	-

*: Based on the LiDAR-I.

**: Formazin Nephelometric Unit.

***: Nephelometric Turbidity Unit.

****: One degree of Japan Industrial Standard (JIS K0101) is the same as when 1 mg of standard substance (kaolin or formazine) is contained in 1 L of purified water.

Table 2.20 shows specifications of the equipment, measurement parameters, and river water quality at the time of measurements. Because the magnitude of turbidity in a river can strongly affect the amount of light incident into the water column, its value was confirmed before each ALB measurement. The water quality of the three target periods was reasonable for measuring the underwater terrain surface.

Both river topo-bathymetry and vegetation attribute measurements (i.e., LiDAR data, drone imagery) were performed on March and October 2020 with a normal water level utilizing digital drone-mounted GLS, respectively, through several flight operations. **Figure 2.33 (*b*)** illustrates a typical view of GLS during overland and underwater measurements. **Table 2.20** shows specifications of the equipment in **Figure 2.33 (*d*)**, measurement parameters, and river water quality at the time of measurements.

2.3.2.3 Data processing of imagery-based input

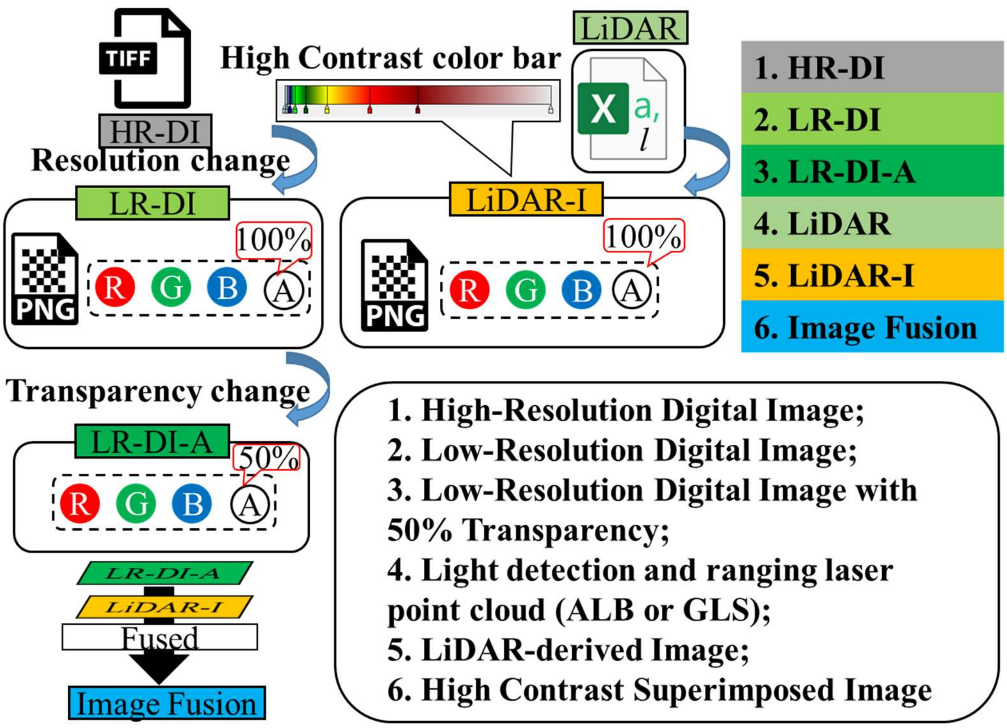


Figure 2.34 Preprocessing of imagery-based input data from drone images (aerial photographs) and LiDAR dataset (ALB or GLS).

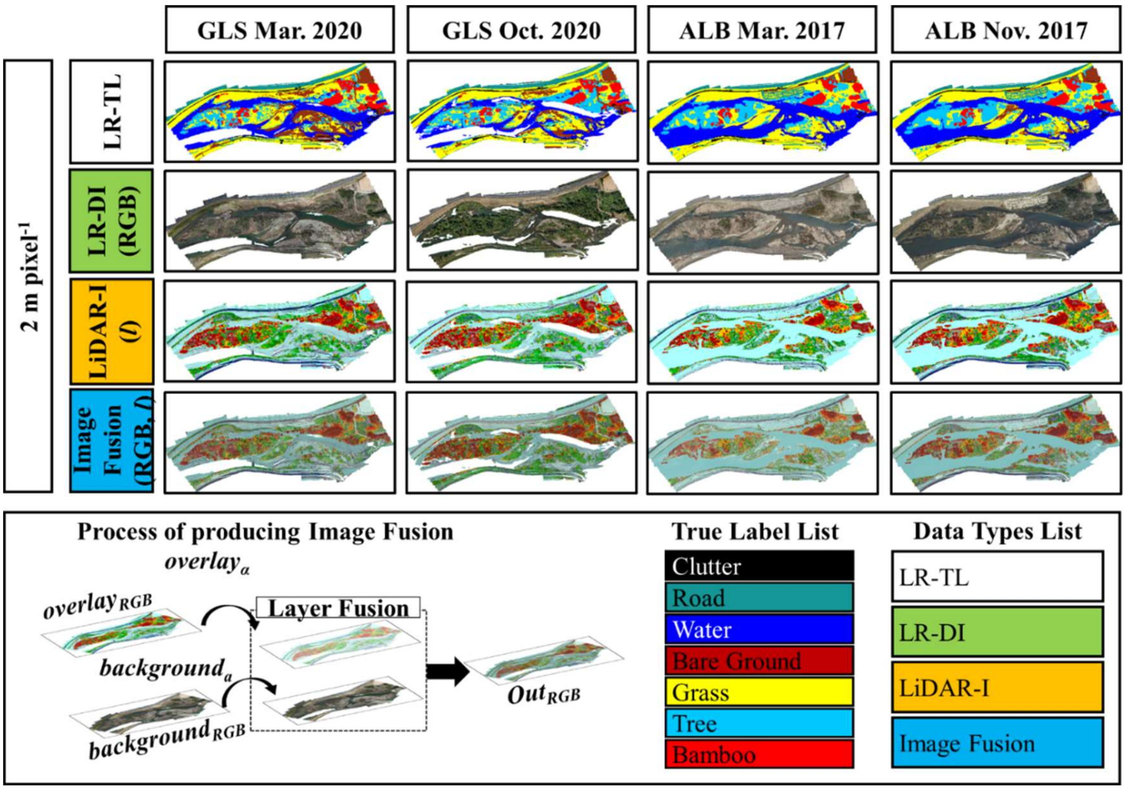


Figure 2.35 2 m pixel^{-1} Imagery-based input (LR-TL, LR-DI, LiDAR-I, Image Fusion), Image Fusion processing, True Label List and Data Types.

Figure 2.34 depicts the processing of producing imagery input data for this study. It includes three types of data, i.e., LR-DI (low-resolution digital images derived from high-resolution digital images), LiDAR-I (images derived from LiDAR-based point cloud using a high-contrast color bar), and Image Fusion (images created by superimposing the above images). In mapping the LCC pre-processing stage, we considered earlier field observation experience8) to obtain critical assistance in mapping the true label (TL). Earlier field observation photographs can provide information about the study target LCC characteristics such as texture and color difference. **Figure 2.35** presents an example of ALB- and GLS-based TL mapping (i.e., LR-TL).

The resolution conversion operation is necessary to ensure a consistent resolution of these input imagery. Incidentally, in order to test the cross-platform interchangeability of several data styles, the 2 m pixel^{-1} resolution was chosen as shown in **Figure 2.35**. The resolution transformation in this study is utilizing a non-interpolation approach (i.e., a method that extracts only the color of pixels at fixed intervals) to keep the color of each pixel in the new generated image the same as that of the original image.

Raw GLS data includes laser point cloud coordinate, DSM minus DTM (*l*). To convert the *l* information in csv file into visual imagery, the visualization software, CloudCompare (i.e., 3-D point cloud and mesh processing software with open source project) was exploited. What's more, different visualization effects were achieved by changing the types of color scale or the distribution of the selected slider (i.e., the point to set specific color in the color scale).

Briefly, after comparing the visual effects of various default color scales, the "high-contrast color scale" was chosen for this study. Furthermore, the selected color scale displayed the LiDAR data as imagery after dividing the GLS data (from min to max) into 256 steps. Moreover, the "high-contrast color scale" included six particular slides with different colors in the 0–50% segment of the color scale (i.e., 1%, 2%, 4%, 8%, 16%, and 32%).

This color scale was supportive for differentiating the laser point cloud better and for characterizing the data better by emphasizing even weakly distinguished image features. Because of the use of multiple split color bands, the "high-contrast color scale" provided better-distinguished data for the *l* in the first half of the value. Subsequently, the visualized imagery was rendered to files using zoom-free method. Consequently, the output imagery maintained the same resolution of 1- (GLS) and 2- (ALB) m pixel^{-1}. Eventually, transform the 1- (GLS) to 2- (GLS) m pixel^{-1} to keep same resolution.

Because data collection using a single LiDAR sensor remain insufficient, the amount of information for imagery was augmented by integrating data from another sensor, such as a digital camera. Accordingly, image fusion approach, which is effective in the discipline of remote sensing, to create a fused image that includes clearer, more accurate, and more comprehensive information than from any single image. Actually, Image Fusion is a combination of two elements: LR-DI, LiDAR-I. Conclusively, Image Fusion was created and processed, as shown in **Figure 2.35**, using image processing software (i.e., GIMP) as the following steps: 1. Reduce the transparency of the overlay layer to 50%. 2. Maintain the original transparency of the background layer. 3. Merge these two layers into a single image with gamma correction (default method of layer combination in GIMP).

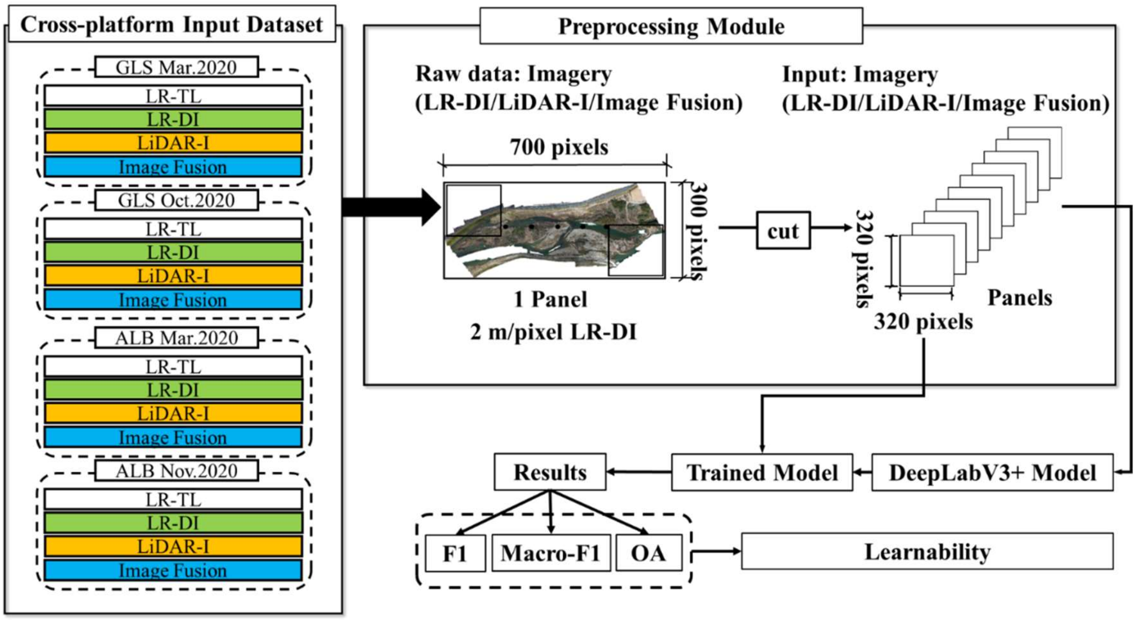


Figure 2.36 Processing of mapping LCC with DeepLabv3+ model; Overall accuracy (OA) is an accuracy measure that indicates how many of the total pixels are classified correctly; The macro-averaged F1 score (Macro-F1) is computed by taking the arithmetic mean (i.e., unweighted mean) of all the per-class F1 scores.

2.3.2.4 DeepLabv3+ Model

In the previous research[8)], an approach of 3-channel DeepLabv3+ with an additional module to use LiDAR dataset as supplement was tried. And in this study, a 4-channel DeepLabv3+ model with modified input layer was also challenged for the higher accuracy. Nevertheless, because of the data limitation, it is very difficult for the 4-channel DeepLabv3+ model to extract the feature. Therefore, in this research, we just force on the effect of changing the input types and resolution of cross-platform dataset, without changing the internal of DeepLabv3+ model. As presented in **Figure 2.36**, to extract features from the input RGB-based imagery (i.e., LR-DI, LiDAR-I, Image Fusion), the processing of training and inference is represented as following: 1. Trimming the raw data from cross-platform input dataset with preprocessing module to attain imagery-based input data (320 px × 320 px), then training the input data with DeepLabv3+ model to achieve trained model; 2. Predicting the imagery-based input data with trained model. To end with attaining the results (i.e., OA and Macro-F1 from mutual-prediction, separately); 3. Comparing the averaged and absolute difference value for confirming interchangeability of cross-platform dataset.

2.3.3 Results and discussion

Three data types of 2 m pixel^{-1} input data were used to test the interchangeability as shown in **Figure 2.36**. The following four main comparison parameters, were utilized to quantify the interchangeability, i.e., Average (OA), Average (Macro-F1), Absolute Difference (OA) and Absolute Difference (Macro-F1), that were derived from 6 groups-based results in **Table 2.21**. Overall accuracy (OA) is an accuracy measure that indicates how many of the total pixels are classified correctly, subsequently, the macro-averaged F1 score (Macro-F1) is computed by taking the arithmetic mean (i.e., unweighted mean) of all the per-class F1 scores. Furthermore, lower Absolute Difference value means the stability of using cross-platform is much better. As shown in **Figure 2.37**, Image Fusion has improved in terms of the interchangeability rather than LR-DI and LiDAR-I. Training GLS Predicting ALB has much higher OA and Macro-F1 value using three data styles, separately. The reason of this phenomenon is the data unbalance derived from lack of the water area in GLS as shown in **Figure 2.38**. And without considering the water area, as shown in **Table 2.22**, OA and Macro-F1 derived from cross-platform become much similar rather than the results derived from **Table 2.22.**

Table 2.21 Comparison of 2 m pixel^{-1} resolution cross-platform interchangeability using multiple method (LR-DI, LiDAR-I, Image Fusion).

Data styles	Groups	Train		Predict		Average (OA)	Average (Macro-F1)
		GLS	ALB	GLS	ALB		
LR-DI (RGB)	Group-1	○			○	0.73	0.66
	Group-2		○	○		0.61	0.59
LiDAR-I (*l*)	Group-3	○			○	0.72	0.63
	Group-4		○	○		0.60	0.56
Image Fusion (RGB, *l*)	Group-5	○			○	0.73	0.65
	Group-6		○	○		0.64	0.60

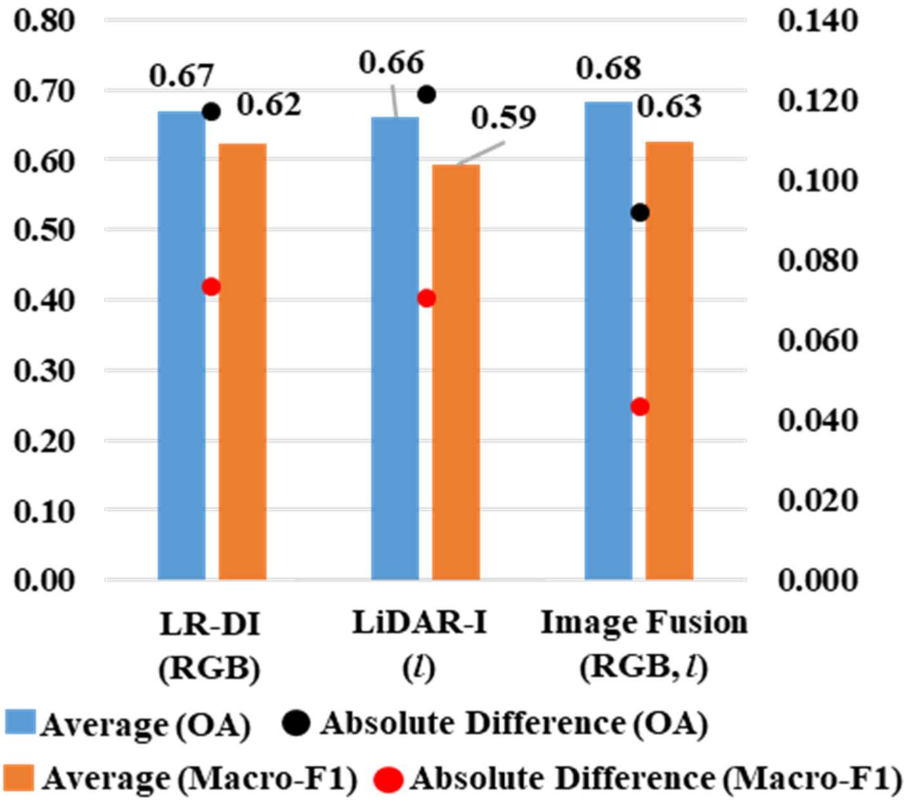


Figure 2.37 Comparison of data style-based averaged 2m pixel^{-1} resolution cross-platform interchangeability derived from **Table 2**; Left vertical axis: the reference of OA and Macro-F1 value; Right vertical axis: the reference of Absolute Difference value.

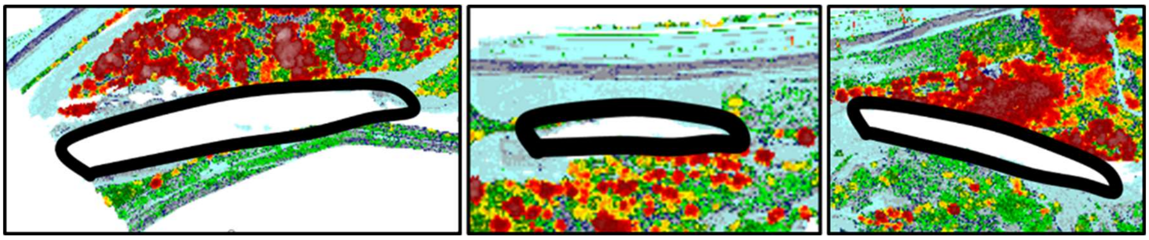


Figure 2.38 Water area that cannot be extracted with GLS only.

Table 2.22 Results without considering the water area

Data styles	**Groups**	**Train** GLS	**Train** ALB	**Predict** GLS	**Predict** ALB	**Average (OA)**	**Average (Macro-F1)**
LR-DI (RGB)	Group-1	○			○	0.66	0.66
	Group-2		○	○		0.62	0.59
LiDAR-I (*I*)	Group-3	○			○	0.61	0.63
	Group-4		○	○		0.64	0.56
Image Fusion (RGB, *I*)	Group-5	○			○	0.63	0.65
	Group-6		○	○		0.66	0.60

2.3.4 Conclusions

This study used the cross-platform LiDAR and photograph to prove the interchangeability between the data derived from different platforms in the term of performing LCC. Compared with LR-DI and LiDAR-I, Image Fusion approach improved the performance of cross-platform LCC. And all the approaches have the results of over 0.65 OA and around 0.6 Macro-F1. To put it another way, to some content, cross-platform data can be used for inter-predicting each other. Noteworthy, digital imagery solely is not sufficient for producing TL mapping under multiple weather conditions because of the sensitive camera senor in the different weathers. Responsibly, at that moment, LiDAR becomes a considerable reference in assisting to recognize the targets.

References (Chapter 2)

Anjum, N., Tanaka, N., 2020. Hydrodynamics of longitudinally discontinuous, vertically double layered and partially covered rigid vegetation patches in open channel flow. River Research and Applications 36(1), 115–127. https://doi.org/10.1002/rra.3546

Baltensweiler, A., Walthert, L., Ginzler, C., Sutter, F., Purves, R.S., Hanewinkel, M., 2017. Terrestrial laser scanning improves digital elevation models and topsoil pH modelling in regions with complex topography and dense vegetation. Environmental Modelling & Software 95, 13–21. https://doi.org/10.1016/j.envsoft.2017.05.009

Campbell-Palmer, R., Puttock, A., Wilson, K.A., Leow-Dyke, A., Graham, H.A., Gaywood, M.J., Brazier, R.E., 2021. Using field sign surveys to estimate spatial distribution and territory dynamics following reintroduction of the Eurasian beaver to British river catchments. River Research and Applications 37(3), 343–357. https://doi.org/10.1002/rra.3755

Carbonneau, P.E., Dugdale, S.J., Breckon, T.P., Dietrich, J.T., Fonstad, M.A., Miyamoto, H., Woodget, A.S., 2020. Adopting deep learning methods for airborne RGB fluvial scene classification. Remote Sensing of Environment 251, 112107. https://doi.org/10.1016/j.rse.2020.112107

Chandler, C.J., van der Heijden, G.M., Boyd, D.S., Foody, G.M., 2021. Detection of Spatial and Temporal Patterns of Liana Infestation Using Satellite-Derived Imagery. Remote Sensing 13(14), 2774. https://doi.org/10.3390/rs13142774

Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L., 2017. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. IEEE Transactions on Pattern Blot Analysis and Machine Intelligence 40(4), 834–848. https://doi.org/10.1109/TPAMI.2017.2699184

Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018. Encoder–decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European Conference on Computer Vision (ECCV) (pp. 801-818).

Chow V. T. 1959 Open-Channel Hydraulics. McGraw-Hill, New York, NY.

Dargan S., Kumar M., Ayyagari M. R. & Kumar G. 2019 A survey of deep learning and its applications: a new paradigm to machine learning. Archives of Computational Methods in Engineering 1–22. https://doi.org/10.1007/s11831-019-09344-w.

Dimitriadis P., Tegos A., Oikonomou A., Pagana V., Koukouvinos A., Mamassis N., Koutsoyiannis D. & Efstratiadis A. 2016 Comparative evaluation of 1D and quasi-2D hydraulic models based on benchmark and real-world applications for uncertainty assessment in flood mapping. Journal of Hydrology 534, 478–492. https://doi.org/10.1016/j.jhydrol.2016.01.020.

Do, H.T., Raghavan, V., Truong, L.X., Yonezawa, G. Multi-scale object-based fuzzy classification for LULC mapping from optical satellite images. *Spatial Information Research* 27(2), 247–257, 2019. https://doi.org/10.1007/s41324-019-00240-w

Erfani, S.M.H., Wu, Z., Wu, X., Wang, S., Goharian, E., 2022. ATLANTIS: A benchmark for semantic segmentation of waterbody images. Environmental Modelling & Software 105333. https://doi.org/10.1016/j.envsoft.2022.105333

Fehérváry, I., Kiss, T., 2020. Identification of Riparian Vegetation Types with Machine Learning Based on LiDAR Point-Cloud Made Along the Lower Tisza's Floodplain. Journal of Environmental Geography 13, 53–61. https://doi.org/10.2478/jengeo-2020-0006

Fisher, J.R.B., Acosta, E.A., Dennedy-Frank, P.J., Kroeger, T., Boucher, T.M., 2018. Impact of satellite imagery spatial resolution on land use classification accuracy and modeled water quality. Remote Sensing in Ecology and Conservation 4(2), 137–149. https://doi.org/10.1002/rse2.61

Gauci, A., Abela, J., Austad, M., Cassar, L.F., Adami, K.Z., 2018. A Machine Learning approach for automatic land cover mapping from DSLR images over the Maltese Islands. Environmental Modelling & Software 99, 1–10. https://doi.org/10.1016/j.envsoft.2017.09.014

Global floods 2021 IAHR experts call for science-informed action! International Association for Hydro-Environment Engineering and Research. https://www.iahr.org/index/detail/482.

Green, J.C., 2005. Modelling flow resistance in vegetated streams: review and development of new theory. Hydrological Processes: An International Journal 19(6), 1245–1259. https://doi.org/10.1002/hyp.5564

Hou, J., Ma, Y., Wang, T., Li, B., Li, X., Wang, F., ... & Ma, H., 2021. A river channel terrain reconstruction method for flood simulations based on coarse DEMs. Environmental Modelling & Software 140, 105035. https://doi.org/10.1016/j.envsoft.2021.105035

Islam, M.T., Yoshida, K., Nishiyama, S., Sakai, K., Tsuda, T., 2022. Characterizing vegetated rivers using novel unmanned aerial vehicle-borne topo-bathymetric green lidar: Seasonal applications and challenges. River Research and Applications 38(1), 44–58. https://doi.org/10.1002/rra.3875

Islam, M.T., Yoshida, K., Sakai, K., Nishiyama, S., Tsuda, T., Nagata, K., Maeno, S., 2020. Model-induced comparative study of ALB and UAV-borne green LiDAR systems on vegetated lower Asahi river assessment in Japan. Journal of Japan Society of Civil Engineers, Ser. B1 (Hydraulic Engineering) 76(2), I_529–I_534. https://doi.org/10.2208/jscejhe.76.2_I_529

Jiahuan, L., Jian, Z., Yunfei, D., 2018. A fusion method of multi-spectral image and panchromatic image based on NSCT transform and adaptive gamma correction. In 2018 Third International Conference on Information Systems Engineering (ICISE) (pp. 10-15). IEEE.

Ju, M., Ding, C., Guo, Y.J., Zhang, D., 2018. Remote sensing image haze removal using gamma-correction-based dehazing model. IEEE Access 7, 5250–5261.

Legleiter, C.J., 2013. Mapping river depth from publicly available aerial images. River Research and Applications 29(6), 760–780. https://doi.org/10.1002/rra.2560

Maeno S., Watanabe S. & Fujitsuka Y. 2005 Improvement of modeling of flow analysis using easily obtained vegetation characteristics. Journal of Hydraulic, Coastal and Environmental Engineering 803, 91–104 (in Japanese with English abstract). https://doi.org/10.2208/jscej.2005.803_91.

Mandlburger, G., Hauer, C., Wieser, M., Pfeifer, N., 2015. Topo-bathymetric LiDAR for monitoring river morphodynamics and instream habitats—A case study at the Pielach River. Remote Sensing 7(5), 6160–6195. https://doi.org/10.3390/rs70506160

Mason D. C., Cobby D. M., Horritt M. S. & Bates P. D. 2003 Floodplain friction parameterisation in two-dimensional river flood models using vegetation heights derived from airborne scanning laser altimetry. Hydrological Processes 17 (9), 1711–1732. https://doi.org/10.1002/hyp.1270.

MLIT 2007 River Flow Situation of the Asahi River and Water Quality, Ministry of Land, Infrastructure, Transport and Tourism, Japan. Available from: https://www.mlit.go.jp/river/basic_info/jigyo_keikaku/gaiyou/seibi/asahi_index.html.

Mooney, H.A., Cleland, E.E., 2001. The evolutionary impact of invasive species. Proceedings of the National Academy of Sciences 98(10), 5446–5451. https://doi.org/10.1073/pnas.091093398

Nepf H. M. 2012 Hydrodynamics of vegetated channels. Journal of Hydraulic Research 50 (3), 262–279. https://doi.org/10.1080/00221686.2012.696559.

Nikora V., Larned S., Nikora N., Debnath K., Cooper G. & Reid M. 2008 Hydraulic resistance due to aquatic vegetation in small streams: field study. Journal of Hydraulic Engineering 134 (9), 1326–1332. https://doi.org/10.1061/(ASCE)0733-9429(2008)134:9(1326).

O'Neil, G.L., Goodall, J.L., Behl, M., Saby, L., 2020. Deep learning using physically informed input data for wetland identification. Environmental Modelling & Software 126, 104665. https://doi.org/10.1016/j.envsoft.2020.104665

Pinho J. L., Vieira L., Vieira J. M. P., Venâncio S., Simoes N. E., Sa Marques J. A. & Santos F. S. 2020 Assessing causes and associated water levels for an urban flood using hydroinformatic tools. Journal of Hydroinformatics 22 (1), 61–76. https://doi.org/10.2166/hydro.2019.019.

Pourmohammadi, P., Adjeroh, D.A., Strager, M.P., Farid, Y.Z., 2020. Predicting developed land expansion using deep convolutional neural networks. Environmental Modelling & Software 134, 104751. https://doi.org/10.1016/j.envsoft.2020.104751

Rodriguez-Galiano, V.F., Chica-Rivas, M., 2014. Evaluation of different machine learning methods for land cover mapping of a Mediterranean area using multi-seasonal Landsat images and Digital Terrain Models. International Journal of Digital Earth 7(6), 492–509. https://doi.org/10.1080/17538947.2012.748848

Ronneberger O., Fischer P. & Brox T. (2015, October) U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241. Springer, Cham. https://doi.org/10.1007/978-3-319-24574-4_28.

Sanjaya, K., Asaeda, T., 2017. Application and assessment of a dynamic riparian vegetation model to predict the spatial distribution of vegetation in two Japanese river systems. Journal of Hydro-environment Research 16, 1–12. https://doi.org/10.1016/j.jher.2017.05.002

Shimizu Y., Kobatake S. & Arafune T. 2000 Numerical study of the flood-flow stage in gravel-bed river with the excessive riverine trees. Annual Journal of Hydraulic Engineering 44, 819–824. (in Japanese with English abstract).

Shih S. S. & Chen P. C. 2021 Identifying tree characteristics to determine the blocking effects of water conveyance for natural flood management in urban rivers. Journal of Flood Risk Management, e12742. https://doi.org/10.1111/jfr3.12742.

Straatsma M. W. & Baptist M. J. 2008 Floodplain roughness parametrisation using airborne laser scanning and spectral remote sensing. Remote Sensing of the Environment 112, 1062–1080. https://doi.org/10.3390/rs9111101.

Sun X., Shiono K., Rameshwaran P. & Chandler J. H. 2010 Modelling vegetation effects in irregular meandering river. Journal of Hydraulic Research 48 (6), 775–783. https://doi.org/10.1080/00221686.2010.531101.

Tian L., Qu Y. & Qi J. 2021 Estimation of forest LAI using discrete airborne LiDAR: a review. Remote Sensing 13 (12), 2408. https://doi.org/10.3390/rs13122408.

Veeck, S., da Costa, F.F., Lima, D.L.C., da Paz, A.R., Piccilli, D.G.A., 2020. Scale dynamics of the HIDROPIXEL high-resolution DEM-based distributed hydrologic modeling approach. Environmental Modelling & Software 127, 104695. https://doi.org/10.1016/j.envsoft.2020.104695

Vetter M., Höfle B., Hollaus M., Gschöpf C., Mandlburger G., Pfeifer N. & Wagner W. 2011 Vertical vegetation structure analysis and hydraulic roughness determination using dense ALS point cloud data – a voxel based approach. International Archives of

Photogrammetry, Remote Sensing and Spatial Information Sciences 38 (5). https://doi.org/10.5194/isprsarchives-XXXVIII-5-W12-265-2011.

Wei K., Ouyang C., Duan H., Li Y., Chen M., Ma J., An H. & Zhou S. 2020 Reflections on the catastrophic 2020 Yangtze River Basin flooding in southern China. The Innovation 1 (2), 100038. https://doi.org/10.1016/j.xinn.2020.100038.

Wieser, M., Hollaus, M., Mandlburger, G., Glira, P., Pfeifer, N., 2016. ULS LiDAR supported analyses of laser beam penetration from different ALS systems into vegetation. ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences 3(3). https://doi.org/10.5194/isprsannals-III-3-233-2016

Xu, Z., Wang, S., Stanislawski, L.V., Jiang, Z., Jaroenchai, N., Sainju, A.M., Shaversc, E., Usery, E.L., Chen, L., Li, Z.Y., Su, B., 2021. An attention U-Net model for detection of fine-scale hydrologic streamlines. Environmental Modelling & Software 140, 104992. https://doi.org/10.1016/j.envsoft.2021.104992

Yan, W.Y., Shaker, A., El-Ashmawy, N., 2015. Urban land cover classification using airborne LiDAR data: A review. Remote Sensing of Environment 158, 295–310. https://doi.org/10.1016/j.rse.2014.11.001

Yavari, S.M., Amiri, H., 2020. Effect of shadow removal by gamma correction in SMQT algorithm in environmental application. Environment, Development and Sustainability 22(7), 7057–7074.

Yoshida K., Maeno S., Mano K., Iwaki T., Ogawa S. & Akoh R. 2017 Determination method for vegetation species distribution in rivers using airborne laser bathymetry. Journal Japan Society of Civil Engineers, Series A2 (Applied Mechanics) 73 (2), I_607–I_618. (in Japanese with English abstract). https://doi.org/10.2208/jscejam.73.I_607.

Yoshida, K., Maeno, S., Ogawa, S., Mano, K., Nigo, S., 2020. Estimation of distributed flow resistance in vegetated rivers using airborne topo-bathymetric LiDAR and its application to risk management tasks for Asahi River flooding. Journal of Flood Risk Management 13(1), e12584. https://doi.org/10.1111/jfr3.12584

Yoshida, K., Nagata, K., Maeno, S., Mano, K., Nigo, S., Nishiyama, S., Islam, M.T., 2021. Flood risk assessment in vegetated lower Asahi River of Okayama Prefecture in Japan using airborne topo-bathymetric LiDAR and depth-averaged flow model. Journal of Hydro-Environment Research 39, 39–59. https://doi.org/10.1016/j.jher.2021.06.005

Yoshida, K., Pan, S., Taniguchi, J., Nishiyama, S., Kojima, T., & Islam, T. Airborne LiDAR-assisted deep learning methodology for riparian land cover classification using aerial photographs and its application for flood modelling. *Journal of Hydroinformatics* 24 (1), 179–201, 2022. https://doi.org/10.2166/HYDRO.2022.134

Yu, Q., Gong, P., Clinton, N., Biging, G., Kelly, M., Schirokauer, D., 2006. Object-based detailed vegetation classification with airborne high spatial resolution remote sensing imagery. Photogrammetric Engineering & Remote Sensing 72(7), 799–811. https://doi.org/10.14358/PERS.72.7.799

Zzaman, R.U., Nowreen, S., Billah, M., Islam, A.S., 2021. Flood hazard mapping of Sangu River Basin in Bangladesh using multi-criteria analysis of hydro-geomorphological factors. Journal of Flood Risk Management 14(3), e12715. https://doi.org/10.1111/jfr3.12715

CHAPTER 3

Application of Deep Learning and Drone Camera in Riparian Area Monitoring (Riverbed Waste Detection)

3.1 Garbage Detection in Riparian Area Monitoring

3.1.1 Introduction

In recent years, plastic litter pollution in the oceans has become a global environmental problem, and its impact on ecosystems is also becoming obvious (**Thompson *et al.*, 2004**). More recently, the dumping of garbage in the river, one main resource of ocean garbage, has been a common occurrence and has gradually started to affect the normal flow of the river (**MLIT, 2022**), which has added a lot of work for the river patrol staff.

Facing with these problems, river authorities instantly need a reasonable, better price-performance method to help Japanese river patrol staff to investigate the situation of garbage in rivers. These years, drones and Artificial Intelligent (AI) technology have been widely used in the civil engineering research (**Lu *et al.*, 2012**), specially for the case of work that is lacking of professional personnel. In this research, AI technology with drone showed its advantages in the riverine garbage-related detection. Apart from improving the recognition accuracy of riverine garbage detection, this research also can be an experimental reference for the researchers who determine to use the proved model- and drone-related parameters in the practical experiment.

3.1.2 Study site and methods

3.1.2.1 Study site

Figure 3.1 illustrates our study site, which is located downstream of the Asahi River, a first-class river (state-controlled) in Japan that flows through Okayama Prefecture into the Seto Inland Sea. In this study, the KP value demonstrates the longitudinal distance (km) from the target river mouth. Future more, the slope of the river bed is about 1:600. the channel width of the target river section is about 300 m.

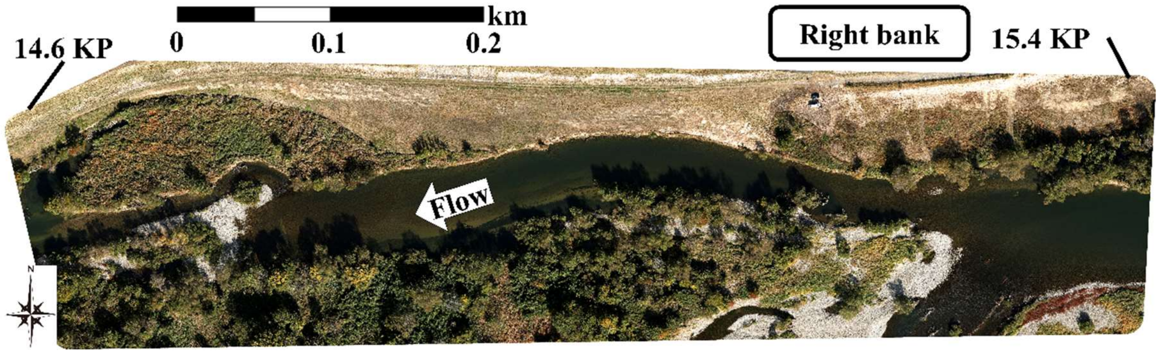


Figure 3.1 Perspective of drone photographs collection area, the right bank area of the Asahi River, Okayama in Japan with kilo post (KP) values representing the longitudinal distance (km) from the river mouth.

3.1.2.2 Specification of the devices

Table 3.1 Specifications of the devices (drone and camera).

UAV	MATRICE 600 PRO (DJI)	
Camera model	Zenmuse Z3	
Size (Original)	Width (pixel)	4000
	Height (pixel)	3648
Size (Trimming)	Width (pixel)	900
	Height (pixel)	600

For the photograph collection of the Asahi River, as shown in **Table 3.1**, a MATRICE600PRO drone manufactured by DJI was used as an aerial platform carrying a ZenmuseZ3 camera manufactured by the same company for continuous photography taking in the study area. The original size of the captured images is 4000 pixels × 3648 pixels. In order to allow the detection model (i.e., ReinaNet) to be trained properly (the original images cannot be trained adequately), the original size images are cutted with a smaller uniform size of 900 pixels × 600 pixels.

3.1.2.3 Drone-related parameters determination

In order to be able to adjust the objects in the drone imagery simply by visual inspection, relatively small Ground sample distance (i.e., GSD, including 2-, 3-, 4- cm) as displayed in **Table 3.2** was chosen as the relevant setting parameter for the drone imagery in this study. Notwithstanding, in order to obtain a larger amount of information per unit of time, in this study, besides the camera angle of 90° from the normal ortho-images, other three angles, i.e., 45°, 60°, and 75°, were made an effort.

Table 3.2 The drone-related parameters determined for the accuracy verification; Ground sample distance (GSD) is defined as the distance between the centers of two adjacent pixels measured on the ground.

Items	Value
GSD (cm)	2, 3, 4
Camera Angle (°)	45, 60, 75

3.1.2.4 Objects of the garbage detection

Considering the types of garbage that can be easily abandoned in the river environment, as presented in **Figure 3.2**, bicycle, PET, cardboard, plastic bag are chosen as the objects in this study. It is noteworthy that all these four types are placed on the floodplain and the background is grass. From the perspective of pixels in the drone imagery with same height, PET is smaller than the other three objects. In the object detection model, the small objects are more difficult to be recognized, therefore this study desires to improve the recognition accuracy of PETs with the aid of the Public Dataset or self-made supplementary datasets (i.e., Random PET Dataset).

Figure 3.2 Object-related samples from Original Dataset in the study site.

3.1.2.5 Public and random PET dataset

As showed in **Figure 3.3**, the Public Dataset (**Wang *et al.*, 2018**) used a drone to maintain low altitude and to take the photographs of PETs with eight different outdoor backgrounds (i.e., sand, lawn, bush, land, step, mixture, ground and playground). On the contrary, due to the low flight altitude, the GSD of this Public Dataset differs significantly from that of the targeted dataset in this study. Therefore, in order to use imagery with the same GSD as test set, an image from the web was also collected, annotated and data-augmented in this study (i.e., Random PET Dataset). This Random PET Dataset is a set that plastic bottles with different sizes are seen in a ditch where floodwaters have receded following heavy rainfall in Xinxiang, Henan province, China July 25, 2021 (**Song, 2021**). It should note that, instead of using manual labeling for this dataset, the annotations of Random PET Dataset are totally generated by predicting from a model that previously was trained with multiple garbage classification dataset. Continuely, to increase the amounts of annotations in this dataset, that using data augmentations (i.e., Flip, Rotate and Shear) without changing the quality of imagery is necessary.

Figure 3.3 Samples of Public and Random PET Dataset, (1) to (8) represents Public Dataset, include 8 types of backgrounds (i.e., sand, lawn, bush, land, step, mixture, ground and playground) with 342 pixels × 342 pixels, (9) is part of Random PET Dataset with 1365 pixels × 1365 pixels.

3.1.2.6 Model for object detection

RetinaNet (**Sultana *et al.*, 2020**) is a two-stage detection model that solves this problem in two steps, first extracting regions of interest and then classifying them using the model, and although it achieves high detection accuracy, it requires a long learning time. YOLOv5 (**Jiang *et al.*, 2022**) is the latest version of the YOLO series (a kind of one-stage detection model). One-stage detection model is object classification and bounding-box regression are done directly without using pre-generated region proposals (candidate object bounding-boxes). In this research, balance the speed and accuracy performance, YOLOv5l is chosen as the model.

Table 3.3 Analysis conditions of comparison groups.

			Train dataset (amounts of annotation)				Model-related		
Cases	**Models**	**Dataset**	**Garbage**	**PET**	**Cardboard**	**Bicycle**	**Parameters**		**Content**
Group-1: Comparison-Asahi with PET from Original									
A	RetinaNet	Trimming	3423	468	194	348	Batch Epochs	4 50	Raw Input Comparison
B	YOLOv5l	Original	3423	468	194	348	Batch Epochs	4 50	
C	YOLOv5l	Original	3423	468	194	348	Batch Epochs	15 500	
Group-2: Comparison-Asahi with mixed PET data									
D	YOLOv5l	Original, Public (8 types)	3423	31866	194	348	Batch Epochs	4 50	Best Performance of Original + Alpha
E	YOLOv5l	Original, Random PET	3423	28909	194	348	Batch Epochs	4 500	
F	YOLOv5l	Original, Public (8 types)	3423	468	194	348	Batch Epochs	4 50	Performance of Original + Public
G	YOLOv5l	Original, Public (8 types)	3423	468	194	348	Batch Epochs	15 500	
H	YOLOv5l	Original, Random PET	3423	468	194	348	Batch Epochs	4 50	Performance of Original + Random PET
I	YOLOv5l	Original, Random PET	3423	468	194	348	Batch Epochs	15 500	
Group-3: Comparison-Asahi with no PET data from Original									
J	YOLOv5l	Original (no PET), Random PET	3423	1011	194	348	Batch Epochs	4 500	Performance of Original (no PET) + Alpha
K	YOLOv5l	Original (no PET), Public (8 types)	3423	977	194	348	Batch Epochs	4 500	
L	YOLOv5l	Original (no PET), Public (8 types)	3423	457	194	348	Batch Epochs	4 50	
M	YOLOv5l	Original (no PET), Public (Lawn)	3423	532	194	348	Batch Epochs	4 50	

3.1.2.7 Comparison of models using multiple dataset and model-related parameters

Broadly speaking, the comparison cases in this study can be divided into three groups **(Table 3.3)**: **1.** PET annotations from Original Dataset only (i.e., Group-1); **2.** PET annotations from Original Dataset mixed with additional datasets (i.e., Group-2); **3.** Only non-Original PET annotations (i.e., Group-3). In Group-1, RetinaNet (i.e., Case-A) is a default model-related parameter reference for YOLOv5l (i.e., Case-B, Batch and Epochs are 4, 50, respectively), thus comparison of these two models with same model-related parameters is required. Subsequently, to inspect the changes of YOLOv5l with larger parameters (i.e., Case-C), enlarging Batch and Epochs from 4, 50 to 15, 500 (the values depended on the NVIDIA RTX 3090 24GB GPU memory to try the largest one). Conseqeuntly, Group-2 can be divided into three parts: **1.** Case-D and Case-E demonstrate the best performance of this research without considering the model-related parameters; **2.** Case-F and Case-G use the same amount of PET annotations as Group-1 with mixed PET annotations (Original + Public); **3.** Case-H and Case-

I just change the PET annotation style (Original + Random PET) from Case-F and Case-G. Conclusively, all the model-related parameters in Group-3 use PET annotations without ones from Original Dataset as a reference.

3.1.3 Results and discussion

In AI tasks, Recall value is an indicator of the percentage of correctly detected objects, i.e., understanding the number of the missed objects. Due to the application to patrols, this indicator was adopted.

3.1.3.1 Group-1

First of all, as shown in **Figure 3.4**, from RetinaNet of Case-A, although plastic bag, bicycle and cardboard can achieve high recognition accuracy (almost close to 100%), the recognition of PET still has a big defect, only when the GSD is 2 cm, the Recall value of each angle is more than 50%. When cases in Case-A have GSD larger than 2cm, that are completely unable to improve the Recall value more than 20%. Compared to RetinaNet, YOLOv5l in Case-B not only can't recognize PET, nevertheless also can't recognize cardboard properly. This also proves that YOLOv5l does not perform as well as RetinaNet in the case of relatively small model-related parameters. After enlarging the model-related parameters in Case-C, the prediction performance of the YOLOv5l model is greatly improved. Recognizing PET with a Recall value of more than 50%, subsequently, the Recall value of cardboard is almost 100%, that both PET and cardboard performed much better than the small model-related parameter cases (i.e., Case-A and -B).

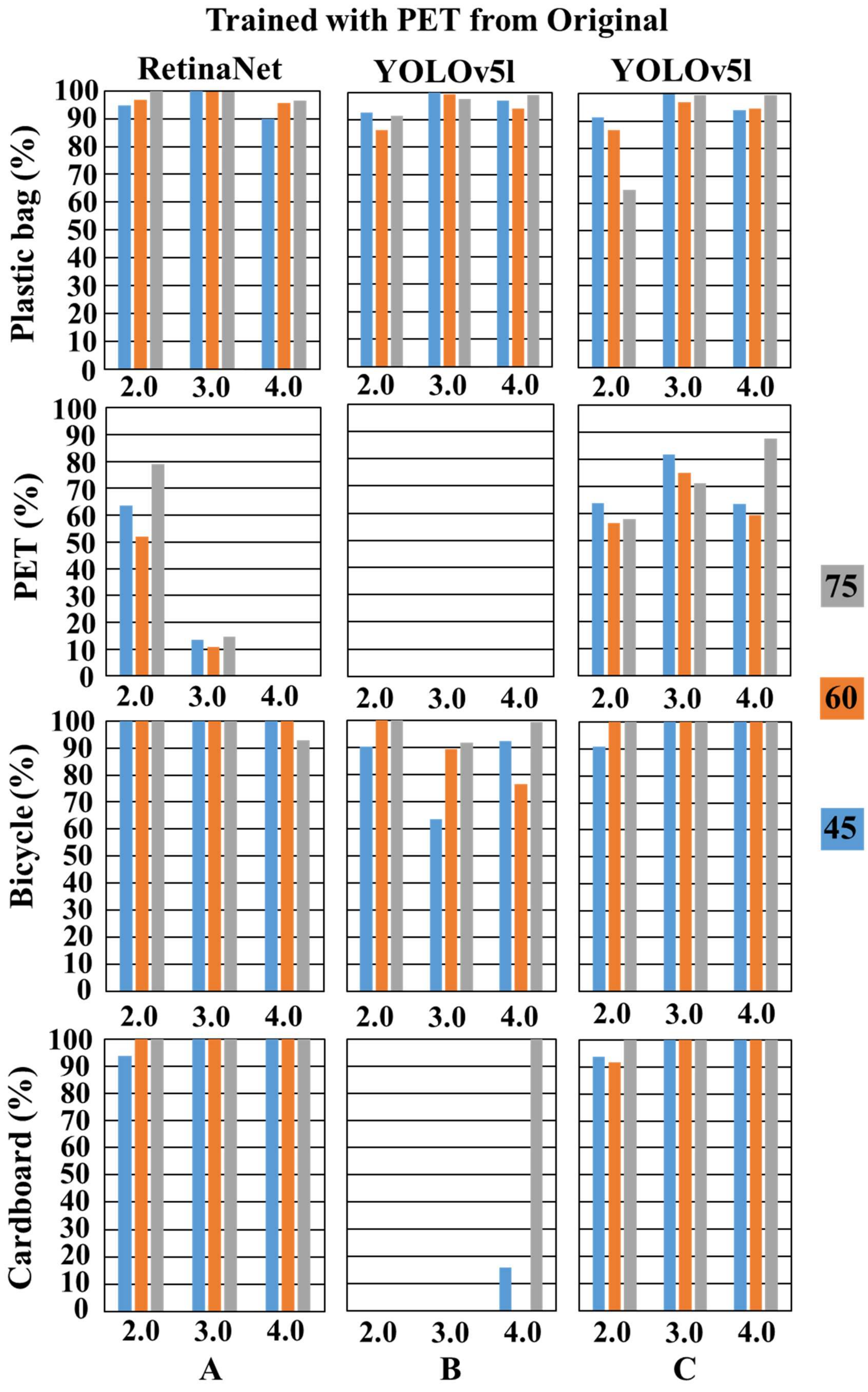


Figure 3.4 Results of Group-1.

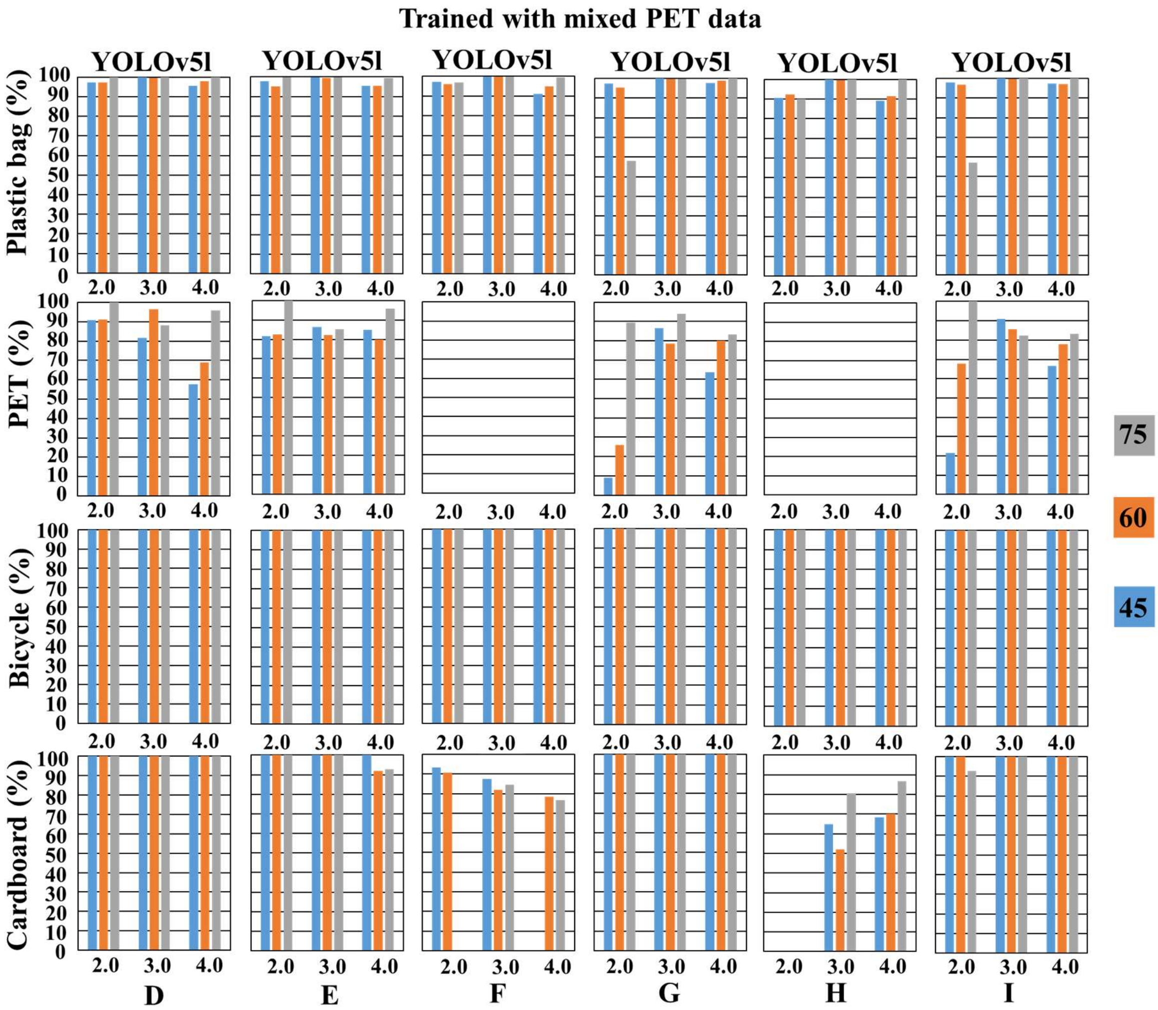


Figure 3.5 Results of Group-2.

3.1.3.2 Group-2

As displayed in **Figure 3.5,** after changing the model-related parameter as presented in Case-C, although Recall value of all four objects have been improved, Recall value of PET still lags far behind the other objects (i.e., bicycle, cardboard, plastic bag). In order to be able to improve the Recall value of PET, the dataset related to PET was increased to challenge the best performance of PET recognition in Case-D and Case-E, considering reasonable model-related parameters. From the results of Case-D and Case-E, the addition of datasets greatly improved the recognition accuracy of PET, especially when the GSD is 2 cm and camera angle is 75°, both of Case-D and Case-E attained almost 100% Recall value. In contrast, when

GSD is 4 cm, the Recall value of Case-E with 45° and 60° have large advantage rather than Case-D with same angle. In consequence of that, the data augmentation of Case-E (i.e., Flip, Rotate and Shear) have some effort on improving the Recall value. From the general perspective of Case-D, the GSD of additional dataset has an effort on the Recall value, that the cases have similar resolution, i.e., Case-D attains relatively higher Recall value when GSD of test dataset is 2.0 cm.

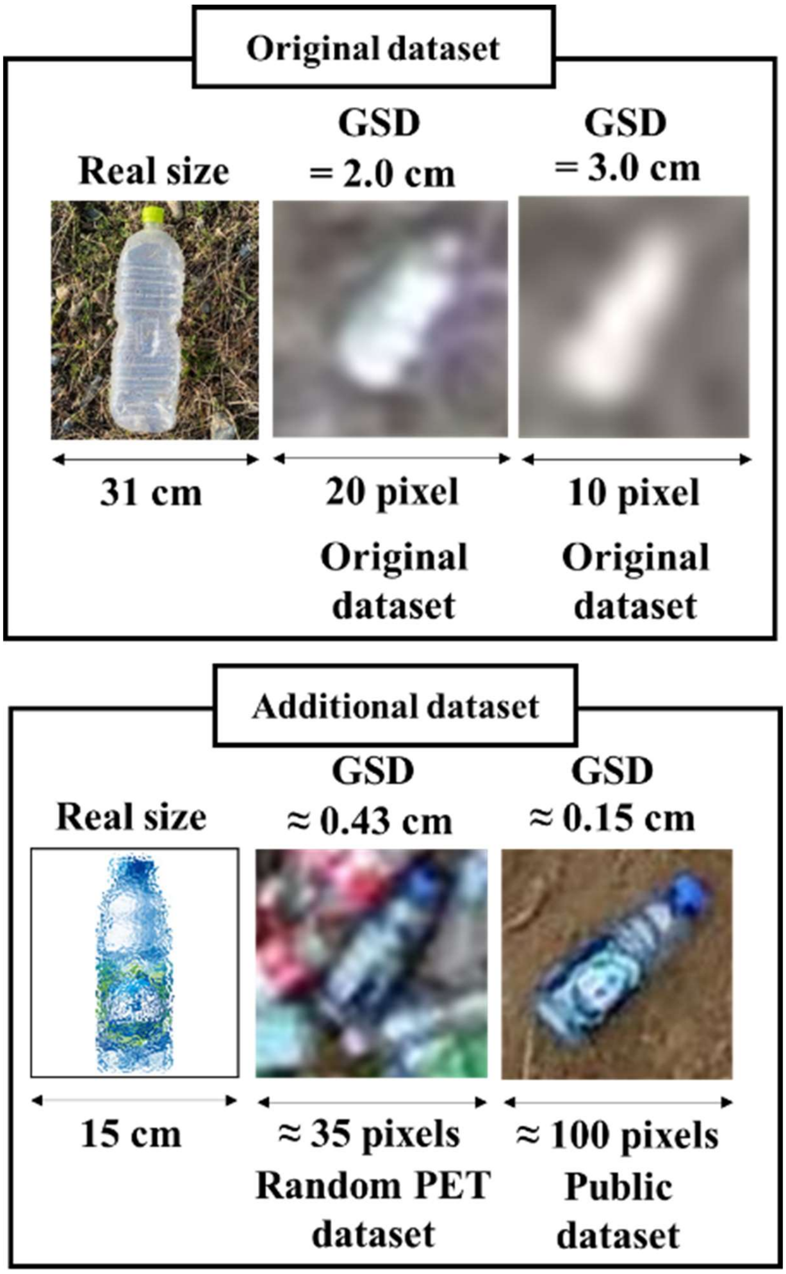


Figure 3.6 Sample of PET at real size and in different resolution.

Figure 3.7 Samples of PET that has similar size with plastic bag (left-side, PET in additional dataset; right-side, plastic bag in Original Dataset).

After proving the ability of additional dataset on improving Recall value in Case-D and Case-E, successively, to observe the PET annotations-based Recall value change, the PET annotations in Case-F, Case-G, Case-H and Case-I were component with Original and Public (or Random PET) dataset as same annotations amounts as Group-1. Then, in Case-F and Case-H, PETs have all failed to be identified. Oppositely, with same Batch, Epochs, annotations amount, the Recall values of cardboard in Case-F and Case-H have partly improved from Case-B with mixed PET data.

Consecutively, Case-G and Case-I with the unchanged dataset from Case-F and Case-H, enlarge the model-related parameters, i.e., Batch and Epochs, from 4 and 50 to 15 and 500, individually. Obviously, larger model-related parameters increase the ability of YOLOv5l to identify the PET-associated objects. Nevertheless, correspond, because of increasing the amounts of PET annotations with smaller GSD (i.e., 0.43- and 0.15- cm from additional dataset as shown in **Figure 3.6**), simultaneously, on account of apparent outlook and similar size as performed in **Figure 3.7** with alike camera angle (i.e., 75° for plastic bag and 90° for PET), plastic bag and PET with smaller GSD are too similar to be misclassified. As a consequence, the plastic bag from Case-F and Case-H all reduced the Recall value from around 90% to less than 60%.

In comparison of results from Group-1 and Group-2, as shown in **Figure 3.8**, the Recall value of each object can reach high values when the data are relatively sufficient and the model-related parameters are as large as possible, i.e., Case-D and Case-E.

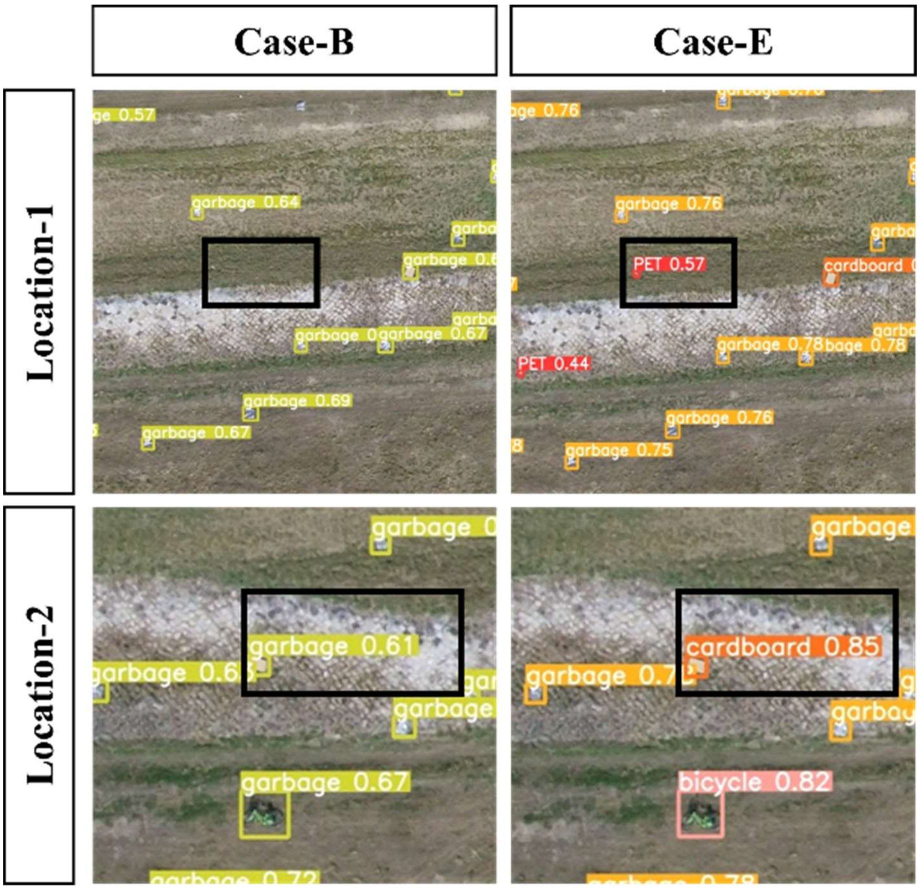


Figure 3.8 Comparison of results trained by YOLOv5l with same model-related parameters using Original Dataset (i.e., Case-B) and Original + Public Dataset (i.e., Case-E) at same locations (i.e., lacation-1 and -2), respectively. Case-E has improved the Recall value of PET and cardboard from Case-B in the black square.

3.1.3.3 Group-3

In Group-3 as presented in **Figure 3.9**, the PETs are almost definitely impossible to be recognized because of using only PET annotations from Public (or Random PET) dataset. Compared with the results in Case-D, training the dataset including data with the same background as test dataset is the better choice for attaining higher Recall value. If the Original Dataset is not enough, additional dataset can support to improve the results (i.e., Group-2). Nevertheless, if using additional dataset only to train the model, the Recall value decrease immediately. Accordingly, how to combine and choose the ratio of the Original and additional dataset, has become an important subject in the future research.

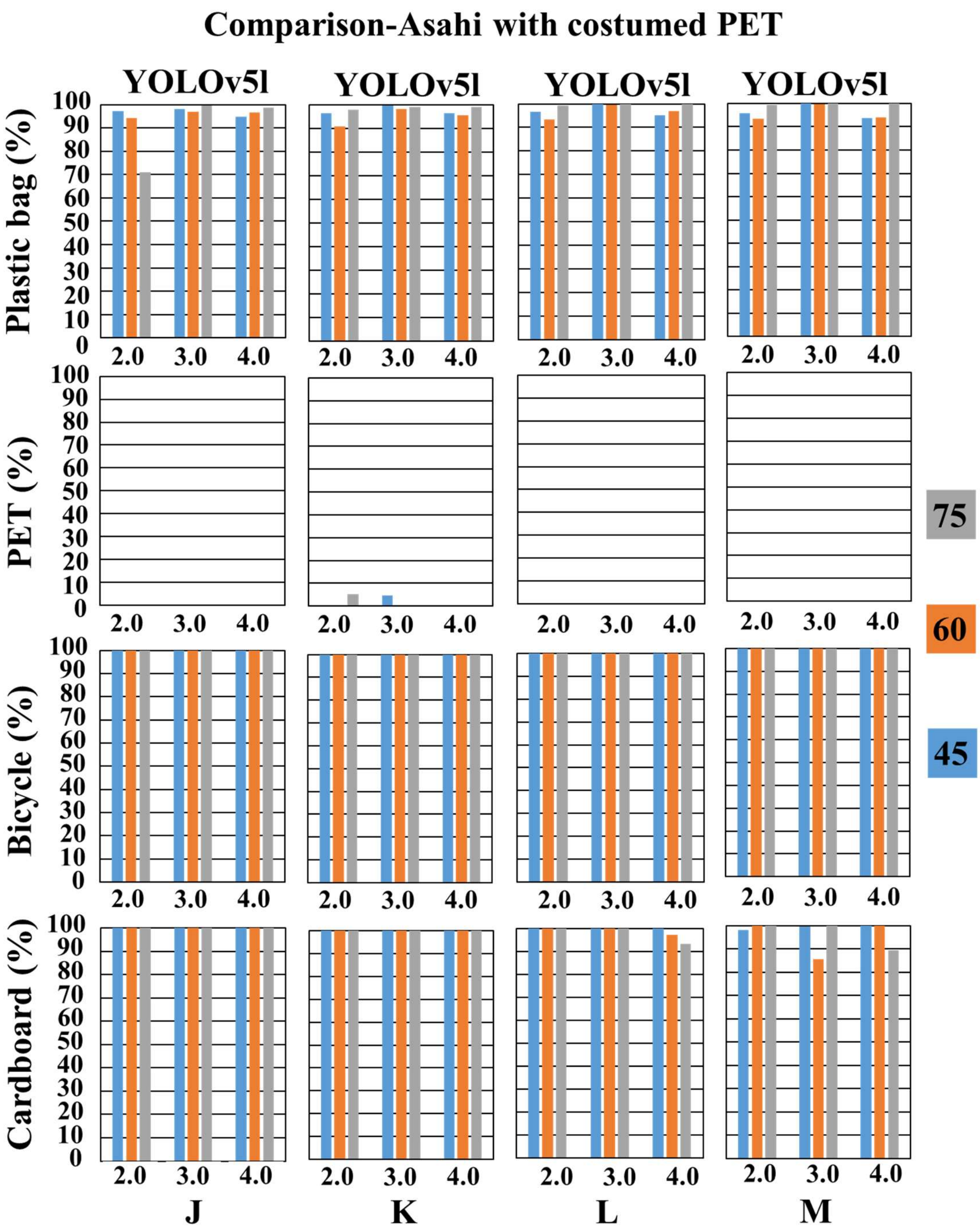


Figure 3.9 Results of Group-3.

3.1.4 Conclusion

From the comparison of results in the Group-1, YOLOv5l has no advantage than RetinaNet using small model-related parameters. Alternatively, when the model-related parameters have been enlarged, the YOLOv5l showed its advantage in improving the Recall value. Compared the Group-1 with the Group-2, on condition that the objects (i.e., PET) are too small to identify, additional dataset with similar size objects can increase the robustness of the models. Considering the Group-3 results, provided that the backgrounds of train dataset are far from test dataset, the identification of the objects are difficult.

3.2 AIGC-aided Garbage Detection in Riparian Area Monitoring

3.2.1 Introduction

Recently, waste pollution in water ecosystems has emerged as a global environmental problem. One of the primary factors contributing to its occurrence is the phenomenon of indiscriminate waste dumping. And monitoring waste pollution along riverbanks with a better cost-performance way is an emergency need for the riparian management. In response to this issue, drones and artificial intelligence (AI) technologies, including You Only Look Once version 5 (YOLOv5), have been employed for the study of waste pollution monitoring at riverbanks (**Pan *et al.*, 2022; Ultralytics**). These technologies have provided valuable insights into the extent of waste pollution. Nonetheless, certain challenges remain, such as the scarcity of data required for training the YOLOv5 due to difficulties in collecting high-quality drone images featuring specific waste targets. In particular, as depicted in **Figure 3.10 (left)**, the collection of Real World Dataset necessitates a significant amount of equipment, such as UAVs, and the placement of specific targets on the site, such as Bikes, Cardboards, PET Bottles, and Plastic Bags. Due to the limitations of only using the Real World Dataset for model training, the YOLOv5 in **Figure 3.10 (right)** can just focus on the features present in the limited dataset, which may lead to misclassification of other targets that have been not included in the training (i.e., non-universal training). As shown in **Figure 3.10 (top)**, one of the open-source image-based generative AI models, Stable Diffusion model (**AUTOMATIC1111**), that uses deep learning (DL) text-to-image technology. It is designed to generate detailed images based on text descriptions (i.e., prompts) and can also be utilized for tasks like image to image (i.e., img2img) translation guided by prompts.

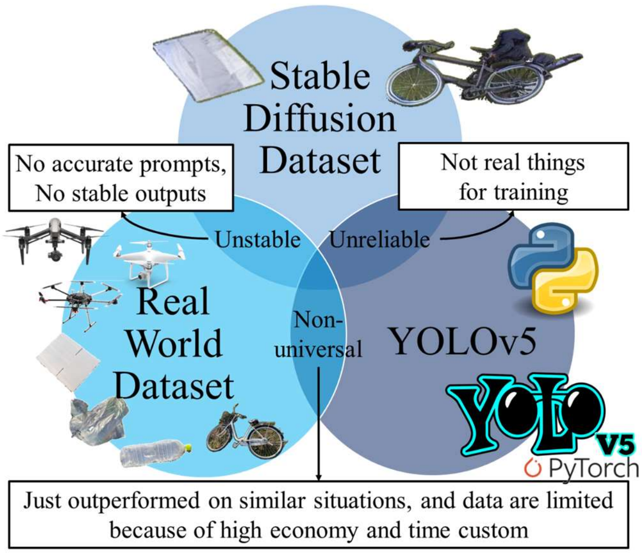


Figure 3.10 Existing problems among the current datasets and YOLOv5.

In this research, the Stable Diffusion Dataset was generated based on the features of the targets present in the Real World Dataset. It is important to note that the quality of the Stable Diffusion Dataset primarily depends on the well-performed and accurate prompts. And it cannot be stable-generated just based on the Real World Dataset directly using img2img function. These unstable outputs may have contributed to lead to the unreliable trained model. Although the aforementioned issues have existed in the practical application, there has been no comparison conducted to assess the trained YOLOv5 derived from the Real World Dataset and Stable Diffusion Dataset using a benchmark-based evaluation approach. In this research, the authors focus on the possibility of replacing or enhancing the Real World Dataset with the Stable Diffusion Dataset during the training of the YOLOv5 for the detection of real targets in practical waste pollution detection.

3.2.2 Study site and methods

3.2.2.1 Study site

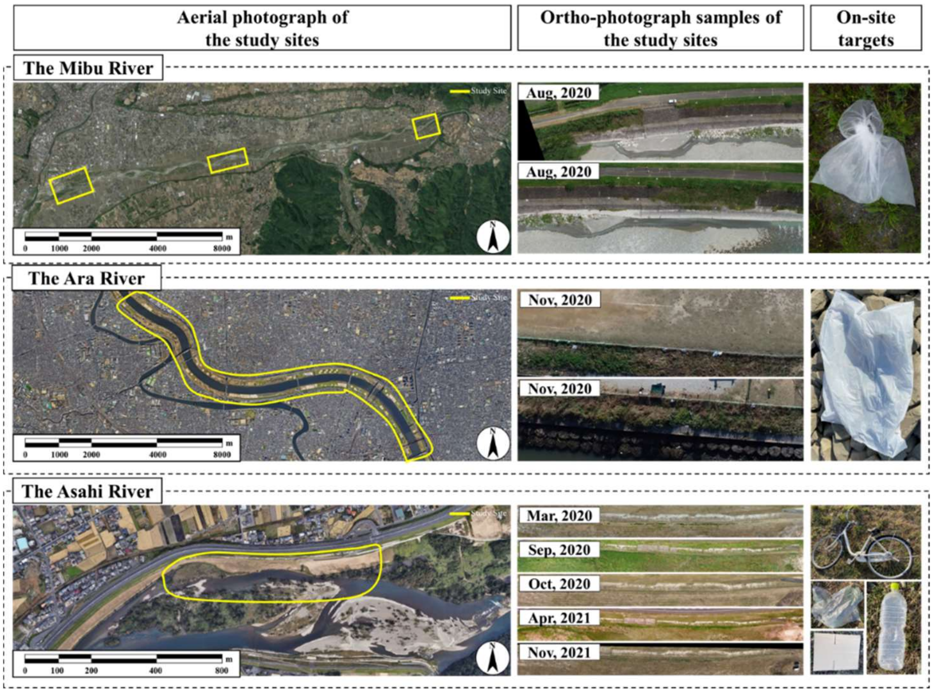


Figure 3.11 Aerial-, ortho-photograph and on-site targets of the study sites from up to down side (i.e., the Mibu River, the Ara River and the Asahi River). Noteworthy, among the Ortho-photograph in the Asahi River, only the Nov, 2021 consists the On-site targets. Except of the Nov, 2021, the other data in the Asahi River are prepared for the background change operation (i.e., Background Images). Aerial photographs are from Google Map; Ortho-photographs are from original.

Figure 3.11 (left) displayed the aerial photographs of the study sites, which are located in the Mibu River, the Ara River and the Asahi River, from up to down sides, individually. And these three state-controlled first-class rivers in Japan that flows through Nagota, Tokyo and Okayama Prefecture. To understand the detailed situations of these sites, thus in the **Figure 3.11 (middle),** ortho-photograph samples are also performed in this work. The **Figure 3.11**

(right) showed the on-site targets in this research of individual location. And the targets are mainly around Bikes, Cardboards, PET Bottles, and Plastic Bags.

3.2.2.2 Flow chart of research process

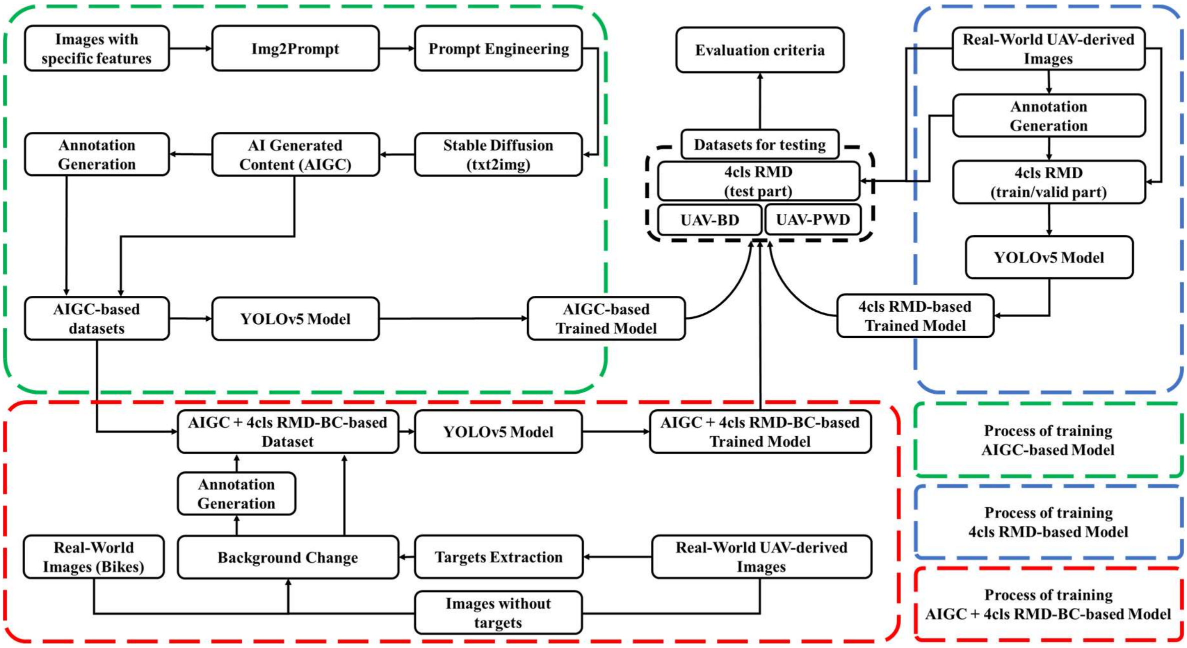


Figure 3.12 Process of assessing the AIGC and Real World Dataset-based models with benchmark datasets (i.e., AIGC, AI Generated Content or Stable Diffusion Dataset; 4cls RMD, River Monitoring Dataset with 4 classes waste pollution; BC, Background Change).

Generally, this research is separated into four main sections in the **Figure 3.12**, i.e., process of training AIGC-based Model **(left)**, 4cls RMD-based Model **(right)**, AIGC + 4cls RMD-BC-based Model **(down)** and the evaluation criteria **(middle)**.

In the **Figure 3.12 (left)**, AIGC-based Model is mainly derived from the Stable Diffusion Dataset that is generated by the txt-based prompts (i.e., txt2img). The first step of generating the images with features in need is to capture the images that are matching the requirements. Then applying the website with img2prompt function to extract the information of the images (i.e., CLIP Interrogator online version in this research). CLIP (Contrastive Language-Image Pre-training) is a neural network trained on a variety of (image, text) pairs, that can predict the most relevant text snippet given an image (**Radford *et al*., 2021**). CLIP can be instructed in natural language to predict the most relevant text snippet, given an image, without directly optimizing for the task, similarly to the zero-shot capabilities of GPT-2 and 3. Worth mentioning, these prompts derived from the CLIP Interrogator can just provide approximate information. For the AIGC with more detailed information, Prompt Engineering is necessary. In this research, after the comparison of several AIGC samples derived from the Prompt Engineering, the key words that can indicate the reasonable results have been confirmed (e.g. UAV, 8k, super detailed and high resolution).

Based on the AIGC derived from the Stable Diffusion (i.e., txt2img function), the annotation Generations are also important for the model training. All the annotation generations for the AIGC are based on the public dataset-based garbage reorganization standard (**Maharjan *et al*., 2022**) with similar feature. After collecting the AIGC and corresponding annotation generations, the authors used the Roboflow (i.e., an online platform to pre-process the dataset) to preprocess the AIGC-based Dataset.

Continually in the **Figure 3.12 (right)**, Real-World UAV-derived Images are separated into three parts (i.e., train/valid and test part). And 4cls RMD-based Trained Model is derived from the train/valid part in this dataset. Remarkably, annotation generations were mainly based on the practical situation of the on-site targets.

Shown in the **Figure 3.12 (down)**, several targets were extracted from the Real-World UAV-derived Images, and combined the images without targets to generate the images with the Background Change. And the Bikes targets are not enough, the supplement of the Real World Dataset (i.e., Bikes) are necessary. After the generation of the annotations, the AIGC + 4cls RMD-BC-based Model can be trained based on the dataset combination of the AIGC and 4cls RMD-BC-based Dataset. In general, as performed in the **Figure 3.12 (middle)**, the mentioned trained models need to be evaluated by the following three datasets: UAV-BD

(**Wang *et al.*, 2018**), UAV-PWD (**Han *et al.*, 2021**) and 4cls RMD (test part) for the evaluation criteria, individually.

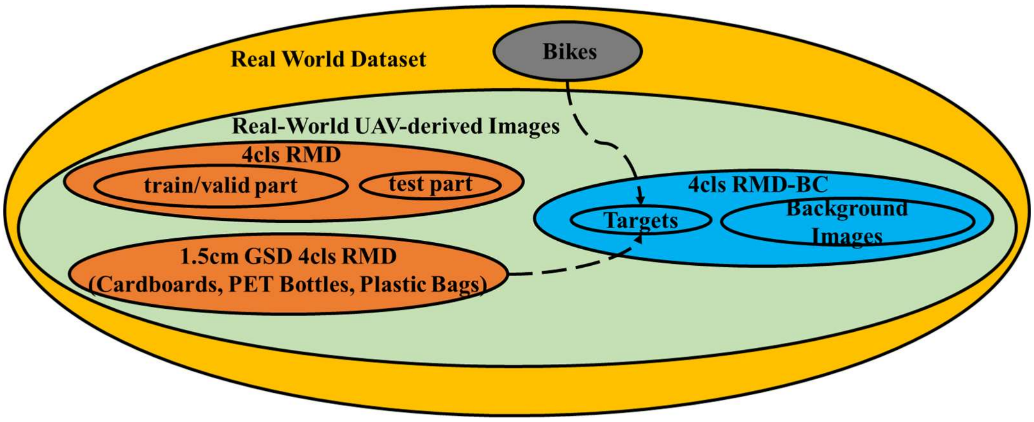


Figure 3.13 Composition of the Real World Dataset.

For understanding the sections in the Real World Dataset, the **Figure 3.13** explained the relationship among each section in the dataset. Firstly, Real-World UAV-derived Images have two sections (i.e., 4cls RMD with train/valid/test parts, 4cls RMD-BC with background images and targets). Except of the mentioned images, there are also Targets (i.e., Bikes) existing for the supplements.

3.2.2.3 Methods

Mainly the Stable Diffusion consists of three main components: the variational autoencoder (VAE), U-Net, and an optional text encoder. The Stable-Diffusion-v1-5 checkpoint used in this research was initialized with the weights of the Stable-Diffusion-v1-2 checkpoint (**Rombach *et al.*, 2022**) and subsequently fine-tuned on 595k steps at resolution 512px ×512px on "laion-aesthetics v2 5+" (i.e., 600M image-text pairs with predicted aesthetics scores of 5 or higher in the LAION 5B dataset) and 10% dropping of the text-conditioning to improve classifier-free guidance sampling.

The You Only Look Once (YOLO) version 5 model (i.e., YOLOv5), which is an open-source software based on convolutional neural networks (CNNs) with optimal detection accuracy and reasonable computational complexity. Based on the mentioned issues, YOLOv5 was chosen as the model for object detection training model in this work.

3.2.2.4 Datasets for training/validation

Figure 3.14 Samples of the AI Generative Content (AIGC).

Table 3.4 Components of the prompts in this work.

Components	Samples
Subjects	Bikes, Cardboards, PET Bottles, Plastic Bags
Resolution	High resolution, 8k camera
View angle	Bird view, UAV view
Area	Riparian area

Quality and quantity of the AIGC were mainly controlled by the model-related parameter setting in the Stable Diffusion web UI. The model-related parameters setting were mainly adjusted derived from the total computational time-consuming and VRAM (i.e., GPU memory). The generated samples are performed in the **Figure 3.14** derived from the specified prompts. As performed in **Table 3.4**, the prompts used in this research include three main components: subject, resolution, view angle, and area.

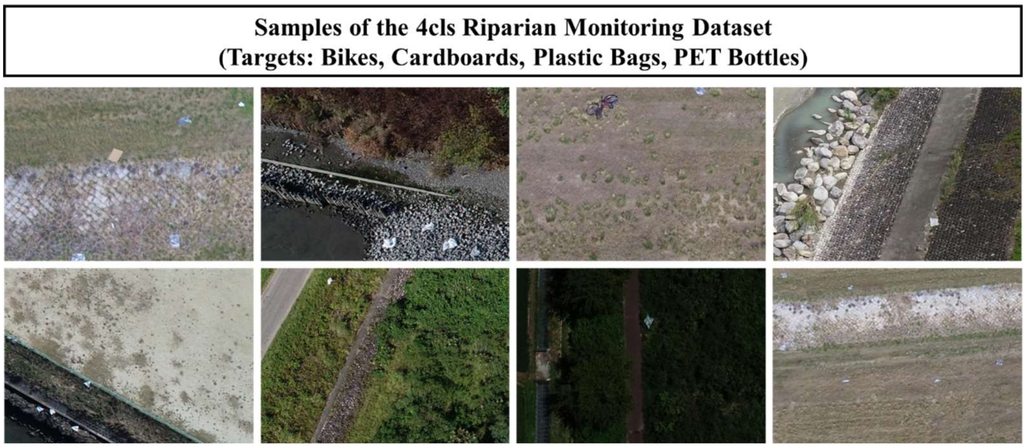


Figure 3.15 Samples of 4cls RMD (i.e., River Monitoring Dataset).

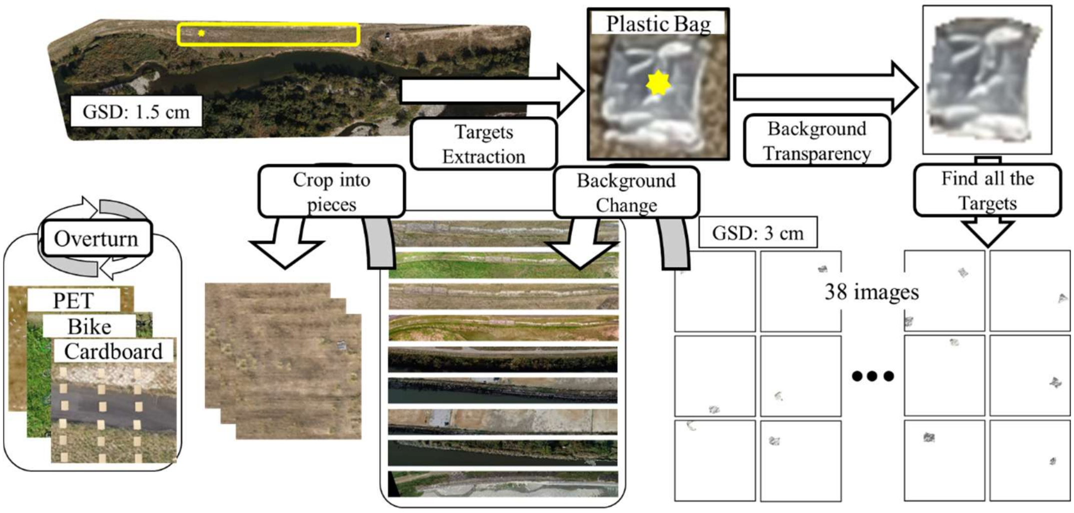


Figure 3.16 Process of generating 4cls RMD-BC (i.e., River Monitoring Dataset-Background Change).

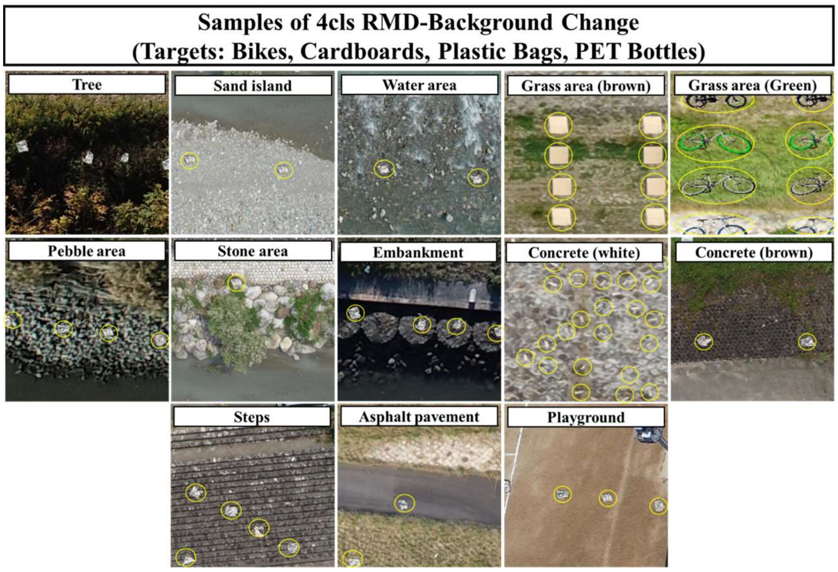


Figure 3.17 Samples of 4cls RMD-BC.

The images of the 4cls RMD were taken by multiple drones (i.e., Inspire2, Phantom4 Pro, Zenmuse X4s) with different sensors (i.e., Zenmuse X4s and Z3) on three riparian areas using multiple camera angles (i.e., 45°, 60°, 75°) and GSDs (i.e., 2-, 3-, 4- cm). As performed in the **Figure 3.15,** the before-mentioned four garbage are all concluded in the sample images.

As the supplement of the AIGC, the 4cls RMD-BC followed the steps in **Figure 3.16**. Extracting all the Plastic Bags and replacing the background using anther UAV-derived image without Plastic Bags. As a final point, cropping the background-changed images into pieces, and overturning the same operation on the other targets. Shown in the **Figure 3.17**, there are thirteen kinds of backgrounds have been collected for supplement. Worth mentioning, not only natural also artificial environment has been collected in the dataset.

Table 3.5 Dataset-based composition of each case.

	AIGC	**4cls RMD-BC**	**4cls RMD**
Case-1	○	×	×
Case-2	○	○	×
Case-3	×	×	○

(1)

Image Numbers	**Case-1**	**Case-2**	**Case-3**
Bikes	500	1042	155
Cardboards	500	986	452
PET Bottles	500	997	309
Plastic Bags	500	928	2605

(2)

As displayed in the **Table 3.5 (1) & (2)**, three cases with specified image numbers have been considered in this research for confirming the effect of the AIGC in detecting the Real World Dataset. Case 1 and Case 2 consist the Stable Diffusion Dataset, and Case 3 is totally derived from Real World Dataset.

3.2.2.5 Model-related parameter setting

Table 3.6 Model-related parameter setting.

Parameters	Configuration (Stable Diffusion)	Configuration (YOLOv5)
Operating system (Version)	Windows (Win 11)	Linux (Ubuntu 20.04.4 LTS)
Model version	v1-5-pruned-emaonly.safetensors	v 6.0
Image-size (pixel)	608×608	1024×1024
Initial learning rate	-	0.01
Final learning rate	-	0.1
Optimizer	-	SDG
Momentum	-	0.937
Batch size	1	12
Batch count	100	-
Epochs	-	500
Patience	-	100

The details of the parameters setting derived from the Stable Diffusion and YOLOv5 have been performed in the **Table 3.6**. The Stable Diffusion is using the pre-trained model that was downloaded from the Hugging face (i.e., v1-5-pruned-emaonly.safetensors), is an American company that develops tools for building applications using machine learning.

3.2.2.6 Evaluation method

Table 3.7 Performance measurement TP, TN, FP, FN are the parameters used in the evaluation of Recall (R), Precision (P), F1.

True \ Predicted	Positive	Negative
Positive	True Positive (*TP*)	True Negative (*FN*)
Negative	False Positive (*FP*)	True Negative (*TN*)

As shown in the **Table 3.7,** the binary confusion matrix has four entries: the number of true positive (TP) and true negative (TN) samples, which are respectively those that are correctly detected as positive and negative, and the two error categories of false positive (FP) and false negative (FN) samples, which represent the number of negatives incorrectly detected as positives.

$$(1)\ \ Recall\ \ (R) = \frac{TP}{TP + FN}$$

$$(2)\ \ Precision\ \ (P) = \frac{TP}{TP + FP}$$

$$(3)\ \ F1 = \frac{2 \times R \times P}{R + P}$$

$$(4)\ \ mAP_{\ IoU} = \frac{1}{N} \sum_{\substack{k=1 \\ k \in (1,\, 2 \ldots, N)}}^{N} AP_{k_{IoU}}$$

$AP_{k_{IoU}}$: AP of class k under the IoU threshold.

N: Number of all the classes (class is 1 in this study).

When using the YOLOv5 to detect the garbage, it is important to choose evaluation measures for this object detection task. Here, as shown in the **Equation 3.1 (1) & (2)**, both Precision and Recall should be considered as the measure that the model can accurately detect the garbage or not, Precision and Recall value depend on the factors from the **Table 3.7** basically. And **Equation (3)** performed the harmonic mean of Precision and Recall, that is main evaluation criteria in this research.

The mean Average Precision (mAP) in **Equation (4)** provides an overall assessment of the YOLOv5's performance in detecting the garbage accurately and consistently derived from Precision and Recall. mAP50 and mAP50-95 are two variants of the mAP metric, where the numbers indicate the IoU threshold used for evaluating the model. The mAP50 uses an IoU threshold of 0.5, while mAP50-95 uses a range of IoU thresholds from 0.5 to 0.95.

3.2.2.7 Datasets for testing

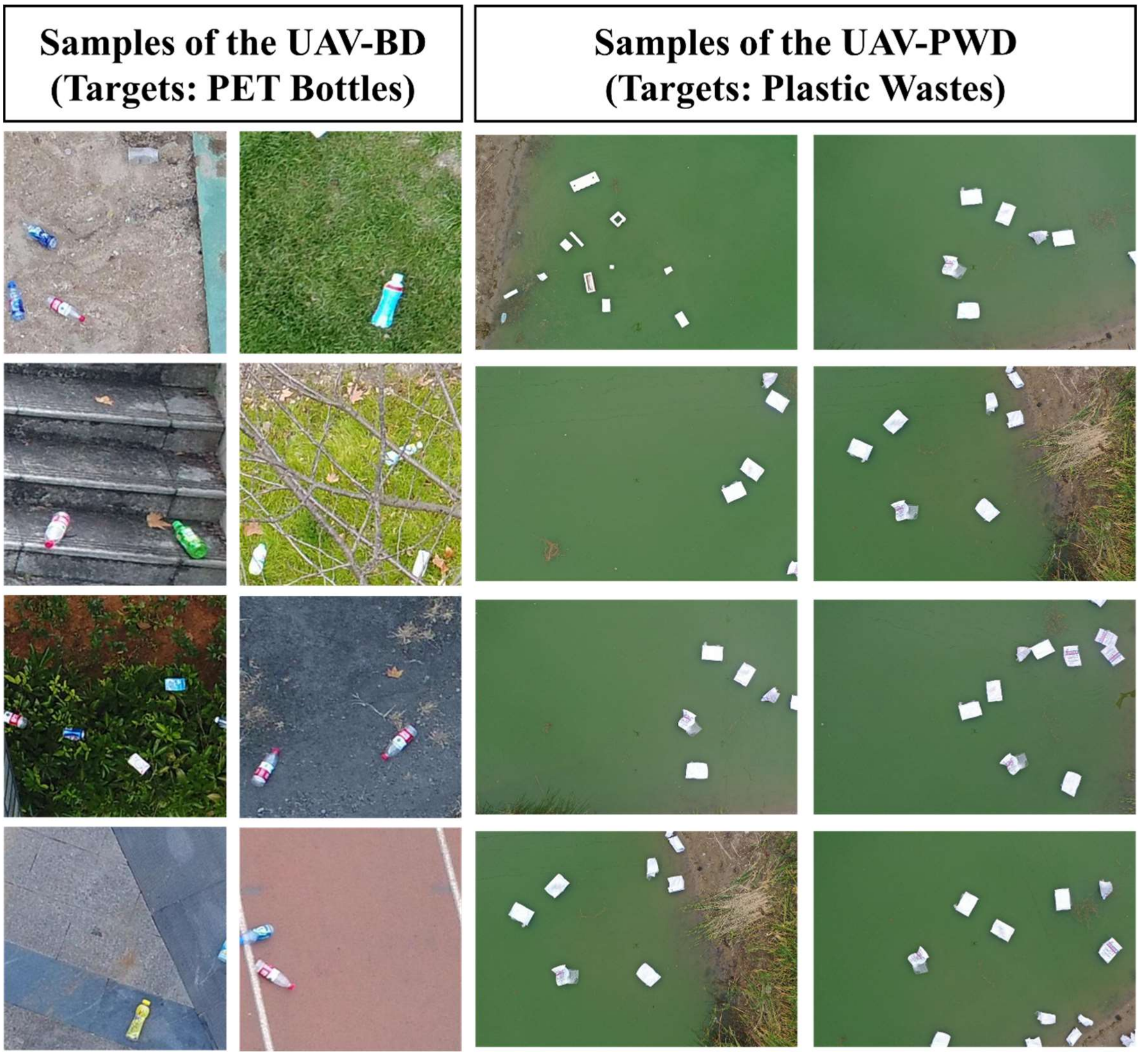


Figure 3.18 Samples of the images derived UAV-BD and UAV-PWD, mainly bottles and plastic waste pollution.

Except for the 4cls RMD (test part), two public datasets have been prepared for testing. **Figure 3.18** performed the samples of the images derived from UAV-BD and UAV-PWD. UAV-BD has eight types of backgrounds to be selected to collect the images (i.e., Ground, Step, Bush, Land, Lawn, Mixture, Sand, and Playground). And UAV-PWD has just one type of background (i.e., water area) without the complex feature. Compared with the complicated color and textures of the backgrounds in UAV-BD, UAV-PWD is comparably much simpler than UAV-BD. In other words, UAV-PWD has a simple background than UAV-PWD. Based on the results derived from these two test datasets, this work can measure the ability of the AIGC-based models to detect the targets both in simple and complex backgrounds.

3.2.3 Results and discussion

Table 3.8 Dataset-based composition of each case.

Test Datasets	Cases	Image size	P	R	F1	mAP50	mAP50-95
4cls RMD (test part)	**1**	1024	0.727	0.721	0.724	0.744	0.377
	2		0.823	0.716	0.766	0.8	0.402
	3		0.952	0.893	0.922	0.966	0.783
UAV-PWD	**1**	1024	0.807	0.864	0.835	0.87	0.508
	2		0.834	0.861	0.847	0.891	0.535
	3		0.554	0.657	0.601	0.421	0.176
UAV-BD	**1**	342	0.771	0.746	0.758	0.764	0.328
	2		0.744	0.689	0.715	0.727	0.31
	3		0.575	0.533	0.553	0.46	0.179

Table 3.9 4cls RMD (test part)-derived class-based results using Case 2.

Class	Image	Instances	P	R	F1	mAP50	mAP50-95
Bikes	38	38	0.317	0.395	0.352	0.213	0.060
Cardboards	114	138	0.647	0.551	0.595	0.656	0.342
PET Bottles	68	73	0.518	0.691	0.592	0.589	0.241
Plastic Bags	713	1615	0.775	0.736	0.755	0.767	0.401

This study is mainly discussing waste pollution detection using UAVs aided with deep learning algorithms. And the authors also explored the challenges of collecting and labeling training data for waste pollution detection models and introduce AIGC as a potential data source. The Stable Diffusion, a text-to-image model, is used to generate images based on specified prompts.

The prompts are derived from the existing images, and the AIGC is automatically labeled using a pre-trained object detection model. The generated dataset is then utilized to train object detection models for the detection of the waste pollution. In summary, this study compares the performance of the AIGC-based Dataset with Real World Datasets using benchmark datasets for evaluation.

Performed the results of using 4cls RMD (test part) for testing in the **Table 3.8**, Case 3 showed the dominant high accuracy (i.e., F1 value) than Case 1 and 2 derived from AIGC. And Case 2 has improved from Case 1 because of using the Real World Dataset with background change. As shown in **Table 3.9**, because of the limited additional targets-based colors/shapes (i.e., Bikes), Bikes have not been detected with comparably low F1 value using Case 2. On the other hand, the results in the **Table 3.8** derived from UAV-PWD and UAV-BD indicate that the AIGC-based Dataset (i.e., Case 1, 2) showed superior accuracy in detecting waste pollution on the simple backgrounds (i.e., water area) compared to the Case 3. In the case of UAV-BD, even Case 2 has increased the data amount, Case 1 also outperformed both in Precision and Recall value. The increased background-change images in Case 2 have almost the same targets (i.e., cropped images including Bikes, Cardboards, Plastic Bags, PET Bottles), which reduced the F1 score of the trained model in detecting the targets with complex features (i.e., different colors, complicated shapes). Generally speaking, if the background of the test dataset is simple, more targets for training even similar could improve the F1 score. On the contrast side, the more complex features the targets of test datasets have, the more data with complex features need to be added to the training dataset.

3.2.4 Conclusion

In this study, to some content, using the AIGC can support (i.e., replacing or enhancing) the UAV-based Real World waste pollution detection tasks. Especially with the assistance of the Prompt Engineering, the images with specified targets can be generated with purposes. But there are also some limitations that cannot be solved yet. The pre-trained model for generating annotations for the AIGC is just one dataset with specified features, and the generated annotations are totally derived from the features of this dataset. Alternatively, if the pre-trained model changed, the generated annotations can also be an unstable factor for training a model derived from the AIGC. In the near future, the more detailed and accurate prompts that can increase the accuracy of detecting the targets in complex backgrounds are looking forward to being applied in practical riparian monitoring tasks.

3.2.5 Future work

Figure 3.19 Samples of the results derived 4cls RMD using Case 2.

Figure 3.19 has misclassified the rocks, concrete blocks, and electric wires protectors as wastes. This phenomenon has indicated the limitations of the AIGC-based Dataset, that if the non-waste targets with waste-similar-outlines in the test datasets have not been trained in the model, it is difficult for the trained model to separate the wastes and non-waste-targets. Based on the mentioned issues, in the future works, dataset supplements of the images with waste-similar-outlines are necessary.

Table 3.10 1.5 cm GSD 4cls RMD-derived class-based results using Case 1.

Class	Image	Instances	P	R	F1	mAP50	mAP50-95
Bikes	2	2	0	0	0	0	0
Cardboards	2	2	0.978	1	0.989	0.995	0.566
PET Bottles	2	2	1	1	1	0.995	0.224
Plastic Bags	3	3	0.972	1	0.986	0.995	0.714

Considering the possibility of improving the F1 value derived from the AIGC using a lower GSD value, 1.5 cm GSD 4cls RMD with detailed information has been utilized for confirmation. As shown in **Table 3.10**, waste pollution samples in 1.5 cm GSD 4cls RMD with 90° camera angle have been inferred by Case 1. Except for the Bikes class, all the other targets were detected with almost 1.0 F1 value using 0.45 IoU and 0.1 Confidence threshold. The reason of mis-detecting the Bikes is mainly based on the prompts. The results can be improved if prompts with more details are used.

Figure 3.20 Samples of the results derived 1.5 cm GSD 4cls RMD using Case 1.

As performed in **Figure 3.20**, although the Bike as a whole target has not been detected using the mentioned IoU and confidence threshold, the tire part has been seen with 0.3 Confidence. Based on this information, the Bike class can be considered to be annotated part by part to increase the accuracy, and if the IoU value can be changed from the default value used in this study (i.e., 0.45) to a lower value, the accuracy can also be improved.

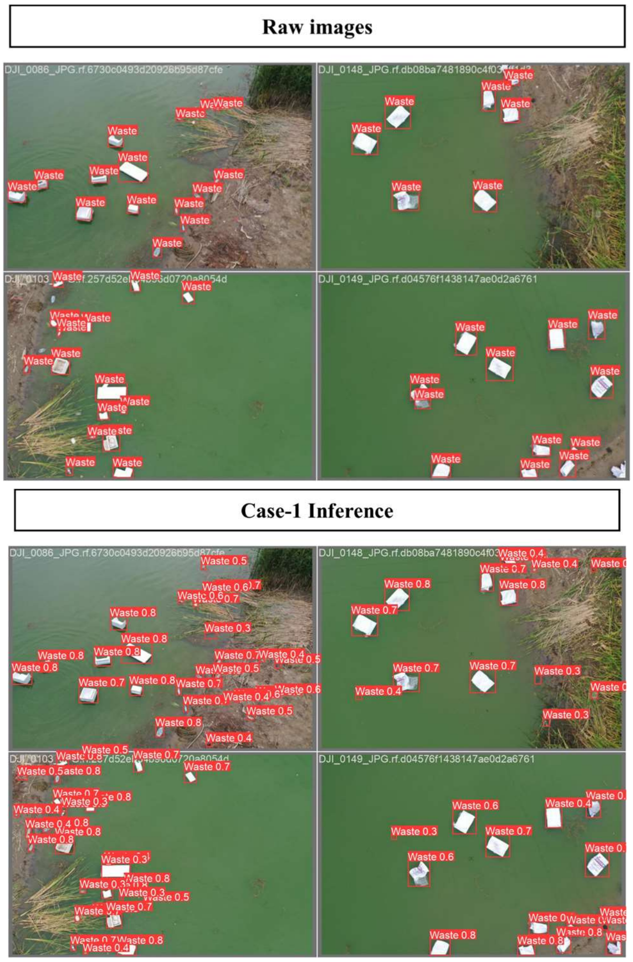


Figure 3.21 Samples of the results derived UAV-PWD using Case 1.

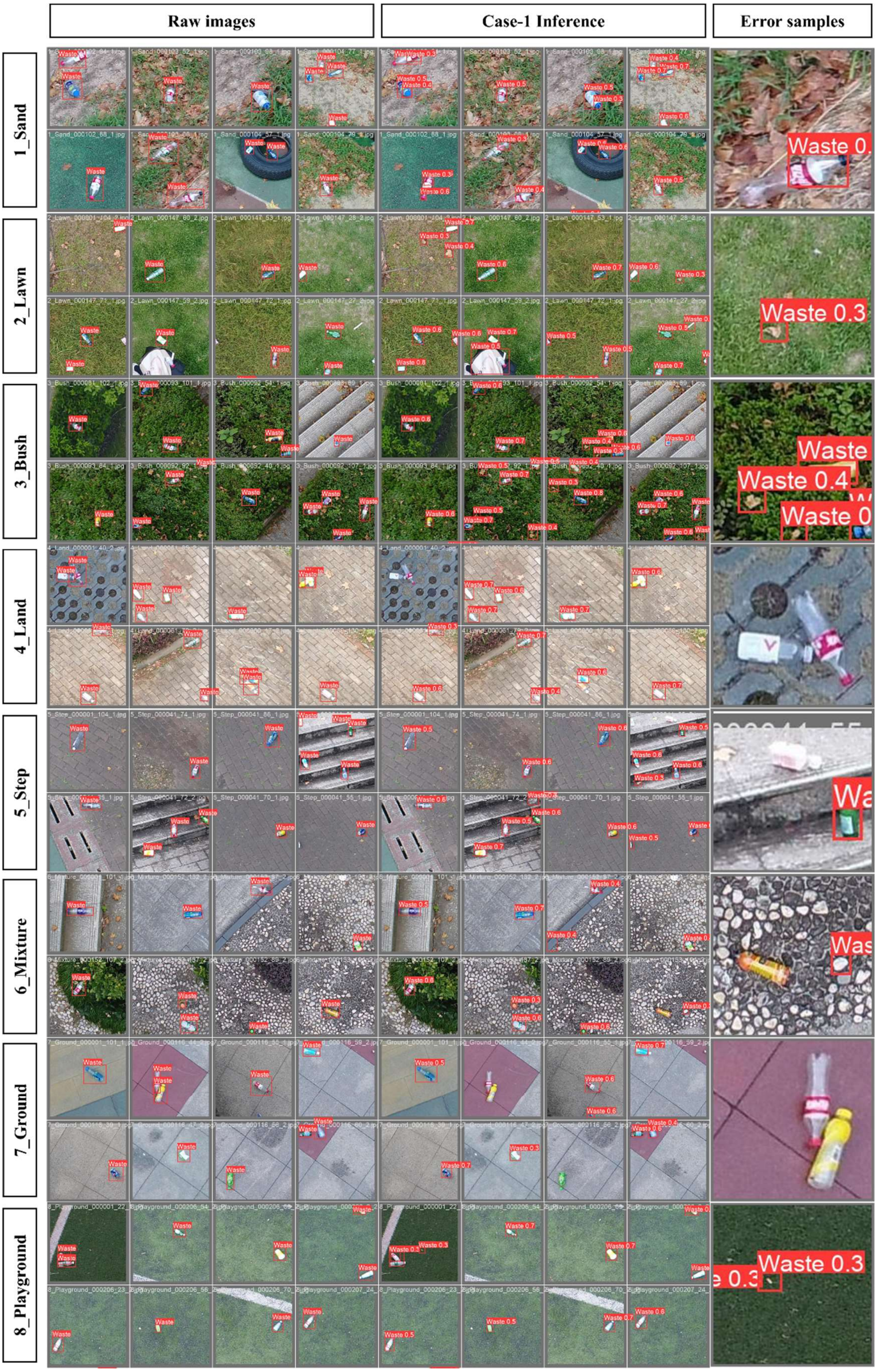


Figure 3.22 Samples of the results derived UAV-BD using Case 1.

Table 3.13 UAV-BD-derived background-based results using Case 1.

Background	Images	Instances	P	R	F1	mAP50	mAP50-95
1_Sand	2704	4630	0.758	0.800	0.778	0.806	0.385
2_Lawn	5778	8424	0.860	0.904	0.881	0.911	0.480
3_Bush	1812	3254	0.721	0.774	0.747	0.724	0.326
4_Land	1538	2365	0.709	0.587	0.642	0.636	0.269
5_Step	1325	2198	0.737	0.648	0.690	0.716	0.315
6_Mixture	3702	5205	0.698	0.625	0.659	0.658	0.272
7_Ground	4355	6246	0.692	0.644	0.667	0.664	0.262
8_Playground	4180	5178	0.889	0.862	0.875	0.907	0.442

As shown in **Figure 3.21**, all the plastic wastes have been detected, on the other hand, **Figure 3.22** performed several left-unnoticed wastes in the groups of 4_Land, 5_Step, 6_Mixture, 7_Ground, individually. Respond, as displayed in the **Table 3.11**, F1 value of all the groups with left-unnoticed bottles are lower than 0.7. The wastes in all the natural or similar-natural background can be detected with comparatively high F1 value derived from the prompt in this study (i.e., riparian area). In the future, it is necessary to expand the scope of the prompts in the AIGC systematically for expanding the application.

References (Chapter 3)

Han, W.L. UAV data monitoring plastic waste dataset. V1. Science Data Bank. 2021. https://doi.org/10.11922/sciencedb.01121

Jiang, P., Ergu, D., Liu, F., Cai, Y., & Ma, B. A Review of Yolo Algorithm Developments. Procedia Computer Science, 199, 1066-1073, 2022. https://doi.org/10.1016/j.procs.2022.01.135

Lu, P., Chen, S., & Zheng, Y. Artificial intelligence in civil engineering. Mathematical Problems in Engineering, 2012. https://doi.org/10.1155/2012/145974

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A. C., Lo, W. Y., Dollár, P & Girshick, R. Segment anything. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 4015-4026). 2023.https://openaccess.thecvf.com/content/ICCV2023/papers/Kirillov_Segment_Anything_ICCV_2023_paper.pdf

Maharjan, N., Miyazaki, H., Pati, BM., Dailey, MN., Shrestha, S., Nakamura, T. Detection of River Plastic Using UAV Sensor Data and Deep Learning. Remote Sensing. 14(13):3049. 2022. https://doi.org/10.3390/rs14133049

MLIT: Water Management and Land Conservation Bureau, River Environment Division Aiming for Litter-Free Water-front (Japanese version), 2022.

Pan, S., Yoshida, K., Boney, A. S., & Nishiyama, S. The Application of Drone-Assisted Deep Learning Technology in Riverbank Garbage Detection. Journal of Japan Society of Civil Engineers, Ser. B1 (Hydraulic Engineering), 78(2), I_133-I_138. 2022. https://doi.org/10.2208/jscejhe.78.2_I_133

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., & Sutskever, I. 2021. Learning transferable visual models from natural language supervision. In International conference on machine learning (pp. 8748-8763). PMLR. https://proceedings.mlr.press/v139/radford21a.html

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. High-resolution image synthesis with latent diffusion models. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 10684-10695). 2022. https://doi.org/10.48550/arXiv.2112.10752

Song, A. China vows to join fight against global plastic waste as UN talks loom. REUTERS. www.reuters.com/news/picture/china-vows-to-join-fight-against-global-idUSKBN2KS0AY, 2021.

Sultana, F., Sufian, A., & Dutta, P. A review of object detection models based on convolutional neural net-work. Intelligent computing: image processing based applications, 1-16, 2020. https://doi.org/10.1007/978-981-15-4288-6_1

Thompson, R. C., Olsen, Y., Mitchell, R. P., Davis, A., Rowland, S. J., John, A.W.G., McGonigle, D. and Russell, A. E.: Lost at sea: Where is all the plastic? Science, Vol.34, 838p, 2004. https://doi.org/10.1126/science.1094559

Wang, C. Y., Bochkovskiy, A., & Liao, H. Y. M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for realtime object detectors. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp.7464-7475). 2023. https://doi.org/10.48550/arXiv.2207.02696

Roboflow, https://universe.roboflow.com/university-bswxt/crackbphdr

Ultralytics, https://github.com/ultralytics/yolov5.

AUTOMATIC1111, https://github.com/AUTOMATIC1111/stable-diffusionwebui.

CHAPTER 4

Application of Deep Learning and Drone Camera in Riparian Area Monitoring (Riverbank Topped-paved Crack Detection)

4.1 Introduction

Riparian crack monitoring involves observing and recording changes in cracks. This practice is crucial in assessing the stability of riparian construction and identifying potential hazards to human safety and the environment. And considering the situations of the crack transformation process shown in the **Figure 4.1**, the phenomenon of grass grown in step d is the most remarkable features of the riparian area cracks. Thus to avoid the pothole generations in step e (not included in this study), earlier locating and mesh-based extracting the cracks in step b and d are also necessary. The crack survey is time-consuming and professional-expertise in need. Until now, the vehicle-platformed AI-based asphalt paved road crack monitoring methods have been developed and applied in the practical road crack-based management (**Asada *et al.*, 2020; Emoto *et al.*, 2023; Numata *et al.*, 2023**). Although the mentioned technology can support the administrators to improve maintance and management, there are also some of the top-surfaced asphalt paved road in the riparian area, that are too dangerous or difficult for the viechcle and personnel to entry, and there is also no general standard for the crack detection. Based on the mentioned limitation, the integration of drones with computer vision algorithms has seen an increase in its usage for infrastructure inspections in the field of civil engineering (**Qiu *et al.*, 2023; Zhu *et al.*, 2022; Chun *et al.*, 2021**). This study presents a method for crack detection in riparian road asphalt pavement using a drone equipped with a divided digital camera for custom data collection, the You Only Look Once version 7 (YOLOv7) object detection model (**Wang *et al.*, 2023**) for data analysis, and the public datasets for data supplement, individually. The YOLOv7 model, a state-of-the-art object detection model, was trained in this study on the datasets of asphalt-paved or concrete cracks. The trained models were achieved to detect the cracks in new images of riparian road asphalt pavement, that were derived from the drone collection. Experimental results of the crack detection attended to show that the combination of the drone technology and the YOLOv7 model has the potential in enhancing the efficiency and accuracy of crack.

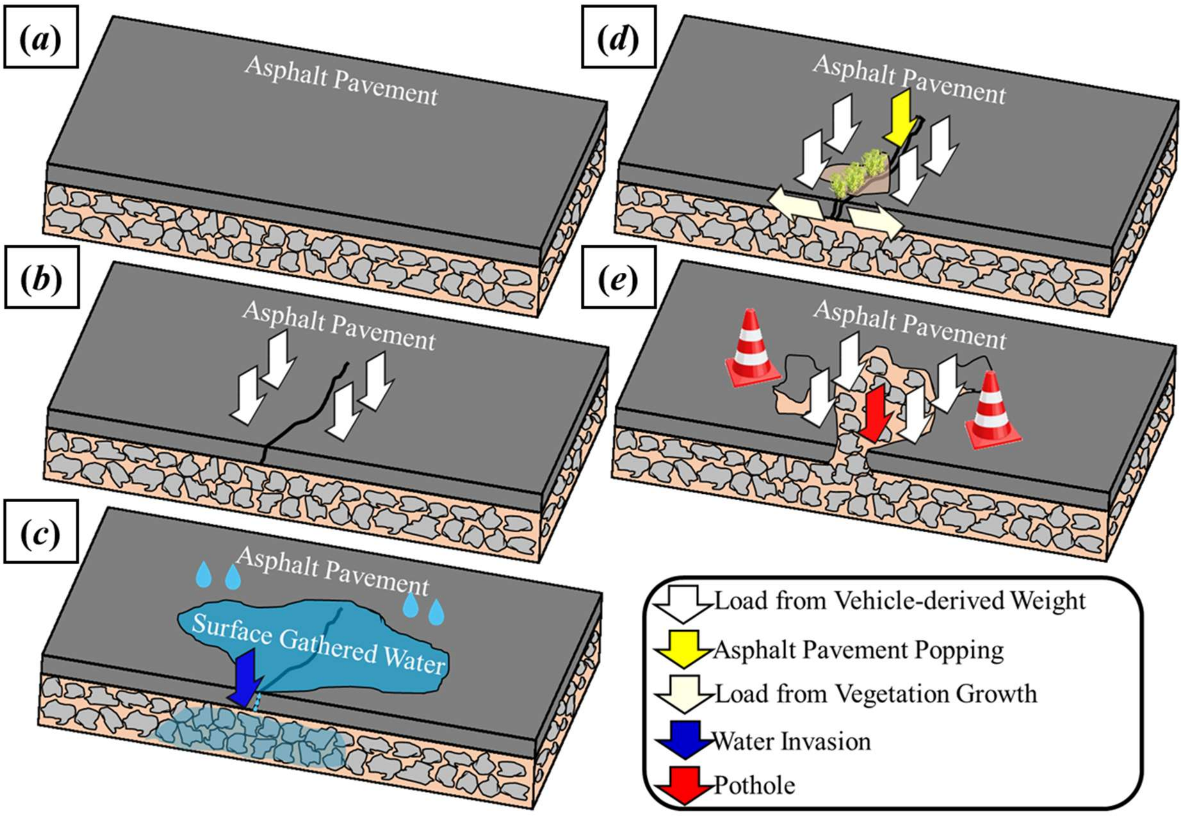


Figure 4.1 Process of transferring the asphalt pavement from no crack to pothole.

4.2 Study site and methods

4.2.1 Study site

Figure 4.2 illustrates the study site, which is located upstream of the Chikuma River, a state-controlled first-class river in Japan that flows through Nigata and Nagano Prefecture into the Sea of Japan. In this site, the riparian asphalt pavements are suffering from the deformation derived from the cracks in the road. To understand the detailed situation of this site, thus one of the asphalt-paved road section is selected as the study site of this work.

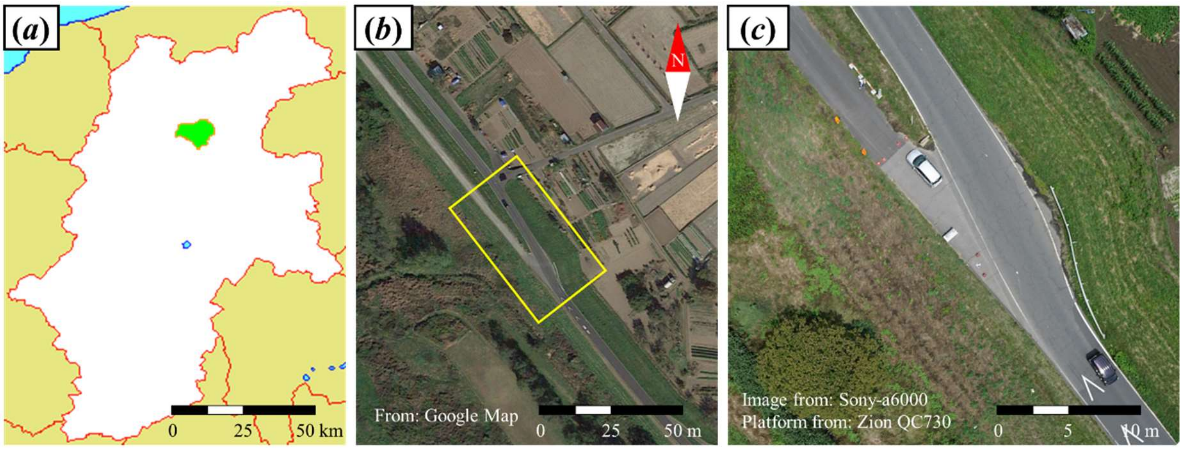


Figure 4.2 The study site is located in the north of the Nagano Prefecture (i.e., **Fig. 1-1**), where a sample position in a yellow square with several crack species (i.e., **Fig. 1-2**) was selected for the model test, as shown in **Fig. 1-3** derived from the drone dataset collection.

4.2.2 Drone-related parameters determination

In order to be able to assess the cracks in the images derived from the drones simply by visual inspection, a relatively small Ground Sample Distance (i.e., 2.5 cm GSD with around 100 m fight height) was chosen as the drone-related parameter in this study. Notwithstanding, as shown in **Figure 4.3**, the camera lens is settled vertically to the ground. The device was used to collect the custom train/valid dataset and images for accuracy assessment.

Figure 4.3 UAV Platform used Zion QC730, and data collection used the camera Sony-a6000.

4.2.3 Objects of this work

Considering the crack situations in the **Figure 4.4** for accuracy assessment, there are several species of the cracks chosen as the objects in this study, including alligator, lateral and longitudinal cracks. In the object detection models, if the objects for inference that are too much different from the objects in training dataset, recognitions (i.e., location, classification) are always difficult. In this study, considering of the accuracy improvement on these objects, additional implements of the dataset are necessary. But the generation of the asphalt-paved crack dataset with annotations in several weather-, light-, and road-situation is time-consuming. Thus instead of annotating the additional images, and solving the mentioned data-lack problem using the published dataset is in consideration. And alternatively, individual-crack-related method is not enough for assessing all the cracks, especially when facing with the complex crack species (i.e., alligator cracks), the individual-crack-related method has some limitation. Therefore, this study desires to improve the recognition accuracy on these complex cracks. Furthermore, recognizing the widths of the cracks using object detection approach also has some limitations, a method that can recognize the instance-based cracks is in the consideration after the crack detection.

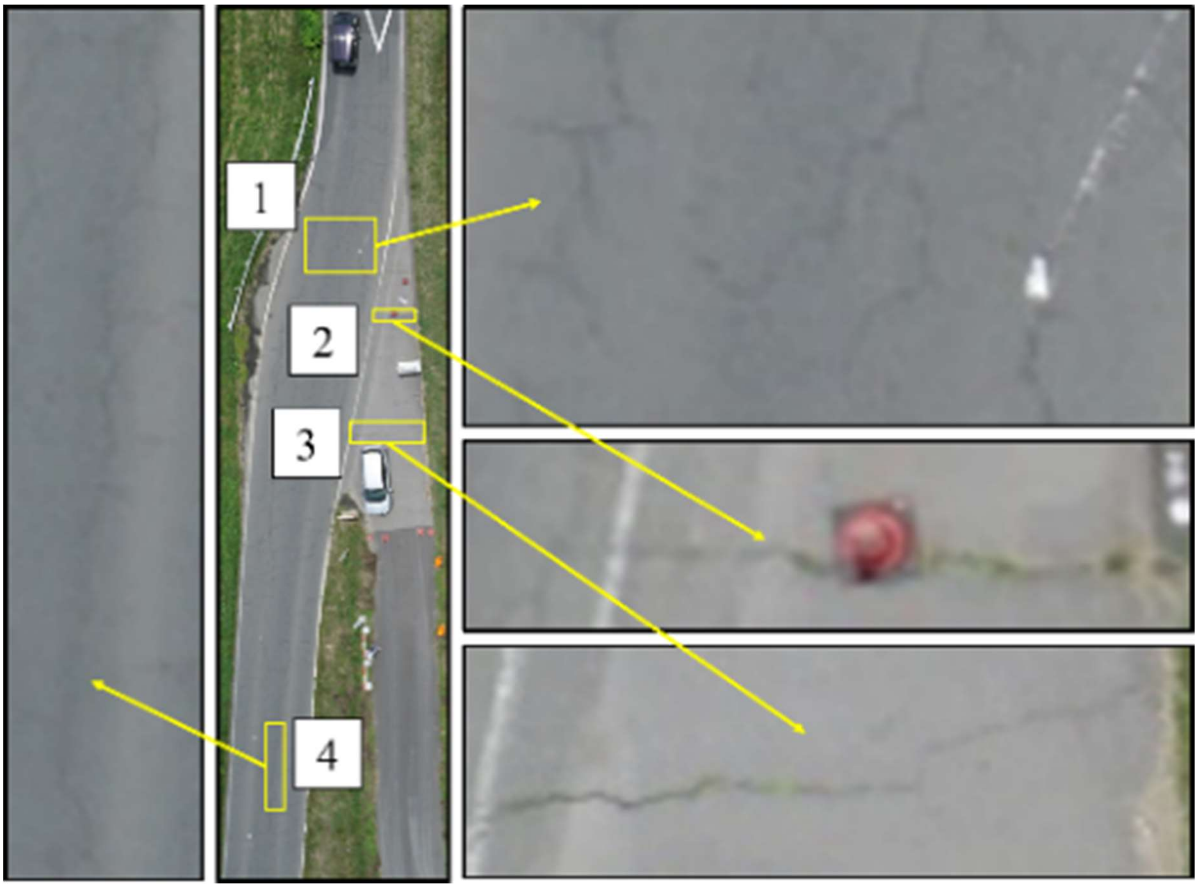


Figure 4.4 Object-related samples from Original Dataset in the study site (i.e., sample-1 is alligator crack, sample-2, -3 are lateral cracks, sample-4 is a longitudinal crack).

4.2.4 Crack datasets in this work

The crack datasets in this work mainly comprise of two parts, custom- and public-one. As performed in upper part of the **Figure 4.5**, the custom one mainly comprised of the uniformed bounding box sizes, which include information on the location and size of the cracks in each image (i.e., the bounding box size-based crack dataset).

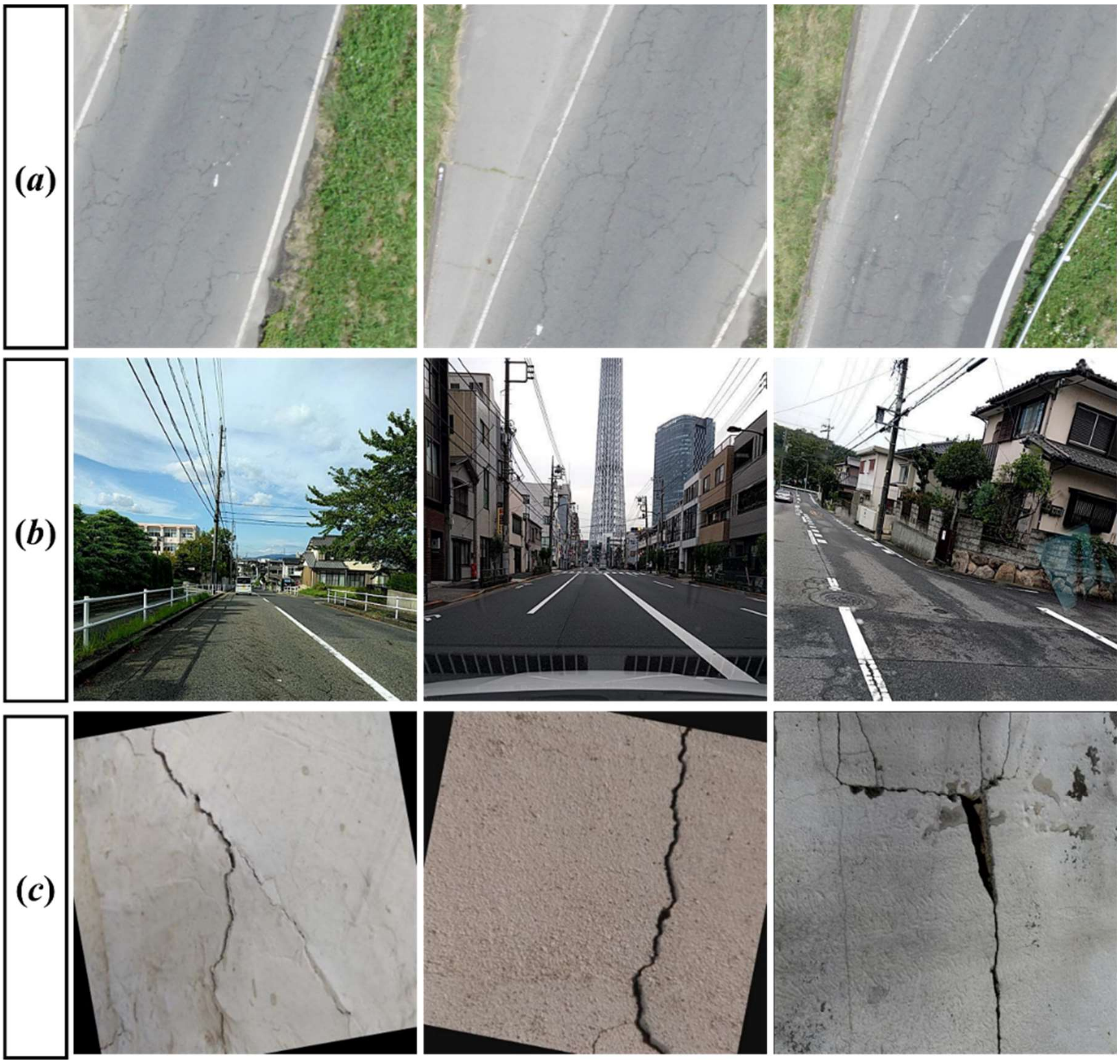


Figure 4.5 Samples of the images in custom and public datasets.

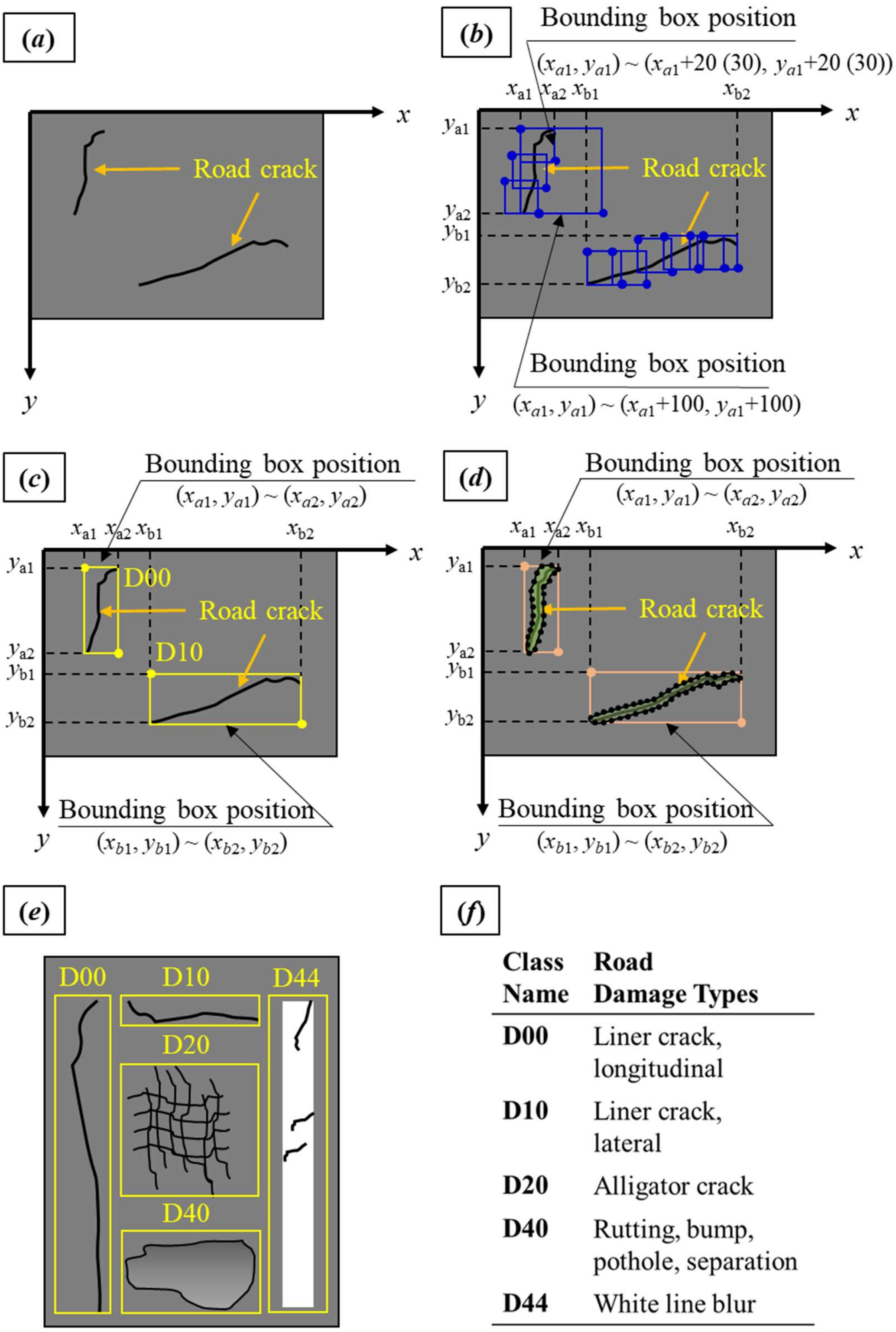
The public dataset with several crack types, is called "Road Damage Dataset 2020" (**Arya *et al*., 2021**) or "RDD 2020" in the publication, annotations depend basically on the crack species (i.e., the crack species-based dataset in this research). And the dataset with instance cracks is derived from the "Top Transportation Datasets" project in the Roboflow Universe (**Roboflow**) for visualizing detected cracks on concretes using instance segmentation.

The images in RDD 2020 were collected from various road or concrete types, such as highways, city roads, and rural roads in different weather (i.e., sunny, cloudy, and rainy as displayed in the middle part of the **Figure 4.5**). And capturing a range of crack species, such as linear cracks, alligator cracks. The annotations in the crack species-based dataset include information on the location, size, and the species of the cracks in each image.

The crack-specie-based classes provide a systematic and detailed representation of road damage and its potential impact on road safety. And the class names and identifications are mainly based on the road maintenance and repair guidebook 2013 design and subcommittee in Japan.

Remarkably, the public dataset used a vehicle-platformed smartphone with a tilted camera angle to collect the dataset. And this dataset was collected in several countries, e.g. Japan, India. In this study, to keep the information between training and testing dataset as similar as possible, just the images taken from Japan are chosen as the additional supplement for the custom dataset.

The images in the instance segmentation dataset as displayed in the underneath part of the **Figure 4.5** are mainly derived from the concrete with more detailed position points rather than the bounding box positions. And the images were taken very close to the cracks, so the details can be observed clearly that are much easier for the model to learn the features. Noteworthablely, the images for accuracy assessment are taken from 100 m height, that the features are different from the close distance, thus the resolution adjustments are necessary in the instance segmentation inference.



Class Name	Road Damage Types
D00	Liner crack, longitudinal
D10	Liner crack, lateral
D20	Alligator crack
D40	Rutting, bump, pothole, separation
D44	White line blur

Figure 4.6 Object-related samples for the crack annotation: (*a*) raw image, (*b*) annotations derived from multi-uniform sizes, (*c*) annotations derived from crack species, (*d*) annotations derived from instance cracks with closeser borders than bounding boxes, (*e*) and (*f*) performed the class names and road damage types of the cracks appeared in this study.

Based on the raw image in the **Figure 4.6 (*a*)**, as shown in the **Figure 4.6 (*b*)**, Multi-uniform size-based crack dataset is mainly derived from the bounding box size (i.e., 20-, 30-, 100-px), and 100 images were collected from a bird view using the UAV-platformed digital camera. These 100 images with the size of 6000px × 4000px collected by the mentioned drone were cropped into the same size, i.e., 600px × 600px each image. From the cropped images, 1564- and 224- images with the cracks are selected for training and validation, individually. Worth mentioning, the overlapped bounding boxes with different sizes are existed like the sample in the **Figure 4.6 (*b*)**.

On the contrast-side, as displayed in the **Figure 4.6 (*c*)**, bounding box sizes of the species-based dataset are totally derived from the actual sizes of each individual crack. The 4- crack species samples and class names shown in the **Figure 4.6 (*e*)** and **(*f*)** are the targets for annotating. 7744- and 880- images with the cracks are selected for training and validation, separately. In the **Figure 4.6 (*d*)**, except of the bounding box position as same as the other ones in **Figure 4.6 (*b*)** and **(*c*)**, instance segmentation annotations have tighter boundaries around objects and fewer missing detections overall, and provides more precise and detailed information. To put it differently, the annotations for the instance segmentation include the class label for each pixel in the image, rather than just for the object as a whole target, that has reduced the effect of the background on the accuracy as much as possible. 3717- and 200- images with the cracks are selected for training and validation, separately.

4.2.5 Model

YOLOv7 shown in the **Figure 4.7**, is a state-of-the-art object detection model that utilizes a single convolutional neural network to perform object detection and classification on an input image. This model leverages the latest advancements in computer vision and deep learning to deliver accurate and fast object detection results. YOLOv7 employs an anchor-based approach to object detection, utilizing anchor boxes to detect and classify objects within an image.

The model uses multiple parallel layers to process an input image in a hierarchical manner, allowing it to learn fine-grained features and perform object detection with high accuracy. Additionally, YOLOv7 utilizes techniques such as cross-stage partial connections and mosaic data augmentation to enhance its ability to generalize to new data and improve its accuracy on a variety of tasks. YOLOv7 released the instance segmentation module (i.e., YOLOv7-seg), the data preparation and usage are derived from YOLOv5, and the algorithm is interlinked with the original YOLOv7 object detection weights.

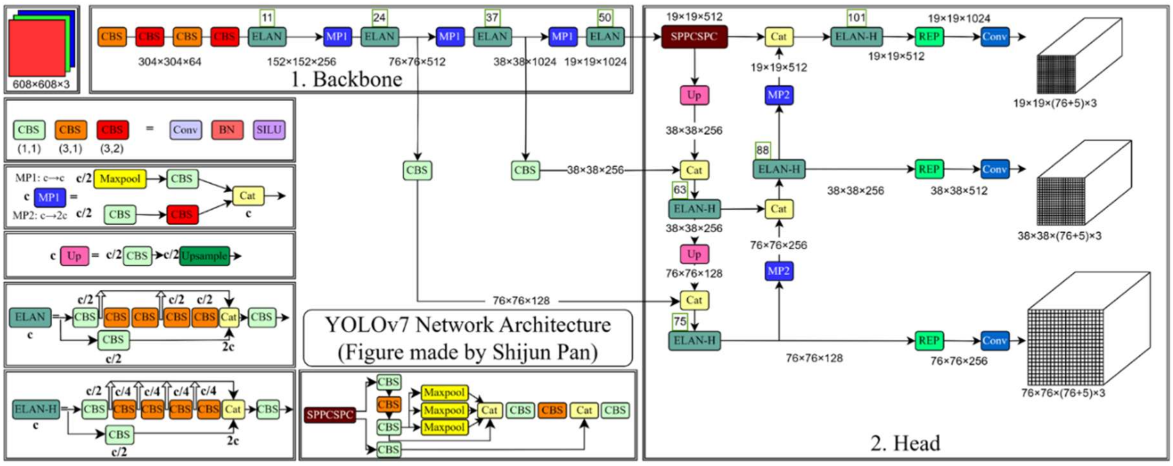


Figure 4.7 YOLOv7 network architecture.

4.2.6 Model-related parameters setting

As shown in the **Table 4.1**, YOLOv7 and YOLOv7-seg model training are basically related to following model parameter setting, i.e., batch size, epochs, learning rate, optimizer, and image size for input. Batch size refers to the number of images that are processed at once during training. A larger batch size can result in faster training, also require more memory on the Graphical Processing Unit (i.e., GPU). Epochs refer to the number of times the entire dataset is passed through the network during training. A larger number of epochs can improve the accuracy of the model.

Table 4.1 Parameter settings

	YOLOv7	YOLOv7-seg
GPU memory	21 GB	23 GB
Batch size	64	64
Epochs	300	300
Lr0	0.01	0.01
Optimizer	SGD	SGD
image size (pixel)	608 (custom) 450 (RDD)	416 (public)
Confidence Threshold	0.1	0.0001

4.2.7 Model-based evaluation method

Broadly speaking, the evaluation metric for object detection models are mainly based on Precision and Recall values. And how to identify the individual crack for a fair evaluation is comparably difficult, thus it is challenging and strict to prepare a test dataset with annotations for all the individual crack correctly with a uniformed size. So in this study, instead of the index for evaluating the crack one by one, the area with the crack detection are in consideration. The minimum unit for the evaluation of the area with cracks are one mesh, and the mesh size are 10- and 50-px.

As shown in the **Figure 4.8**, the trained YOLOv7 models derived from the custom and public datasets predict the raw images in the test dataset, separately. And the inferred bounding-box parts of the results derived from the trained models (i.e., size-based and species-based results) are being selected without considering the species classes.

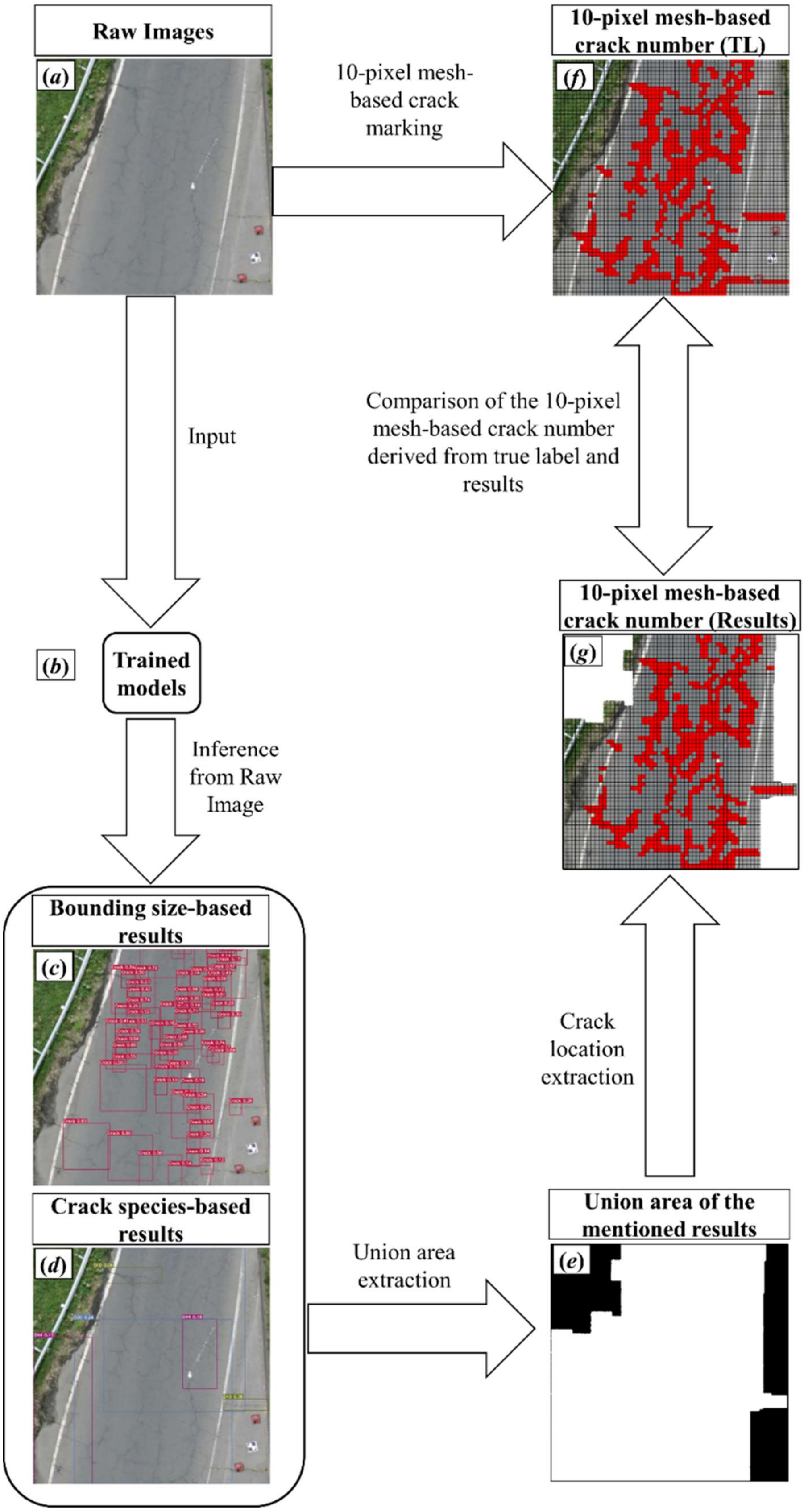


Figure 4.8 Flow chart of 10-px mesh-based crack number comparison for the YOLOv7 result.

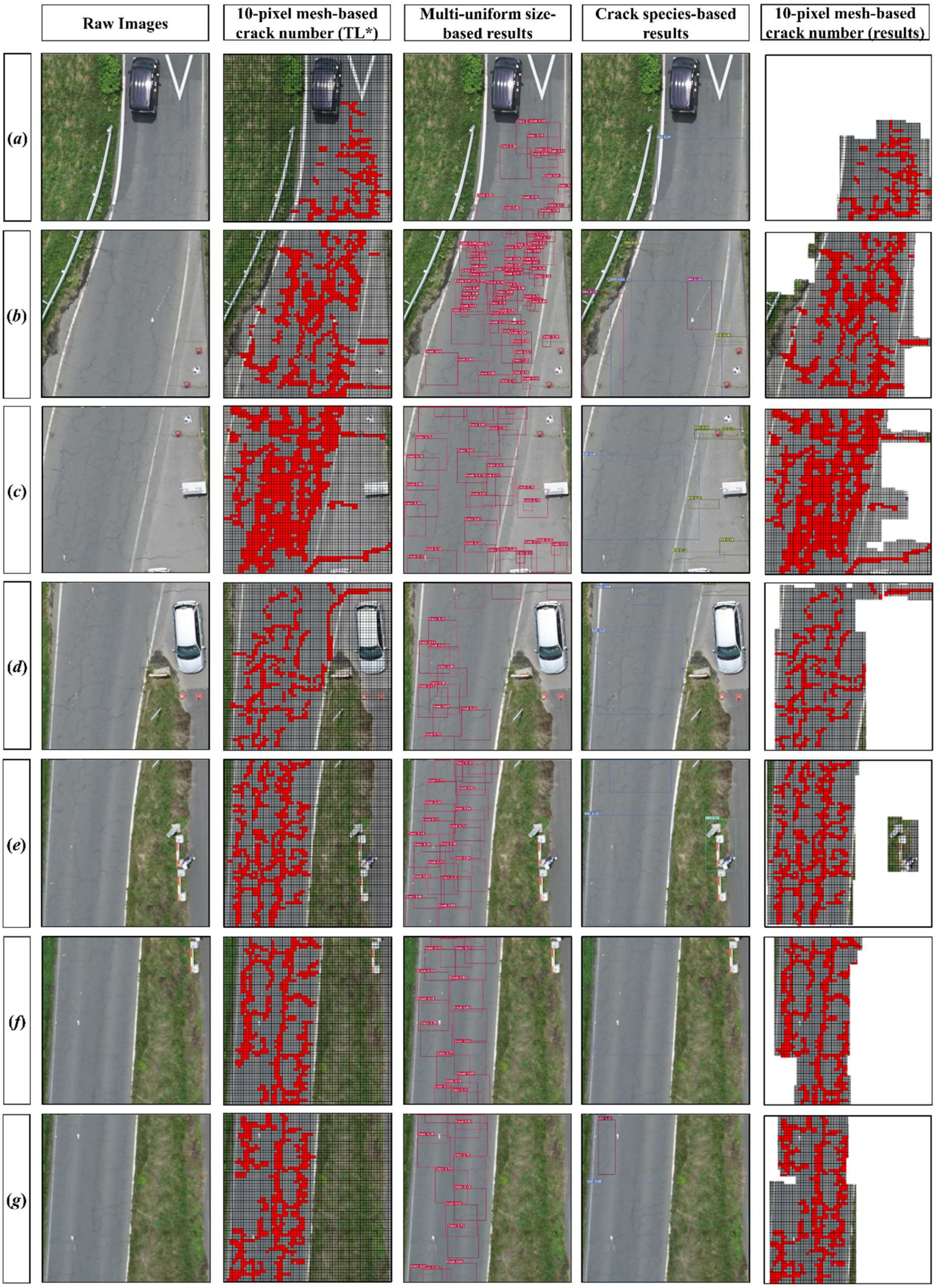


Figure 4.9 YOLOv7-based results. (*: True Label)

Then the numbers of the meshes with cracks in the union area derived from the inference result are being used to compare with the true label. After the comparison shown in the **Figure 4.9**, seven samples in the **Figure 4.10** are selected for observing the miss-detected cracks using multi-uniform size model. After extracting the 10-px mesh-based crack numbers from the true label and results in the **Figure 4.9 (a~g)**, a scatter plot in the **Figure 4.11** performed the relationship between the crack-based numbers in true label and result.

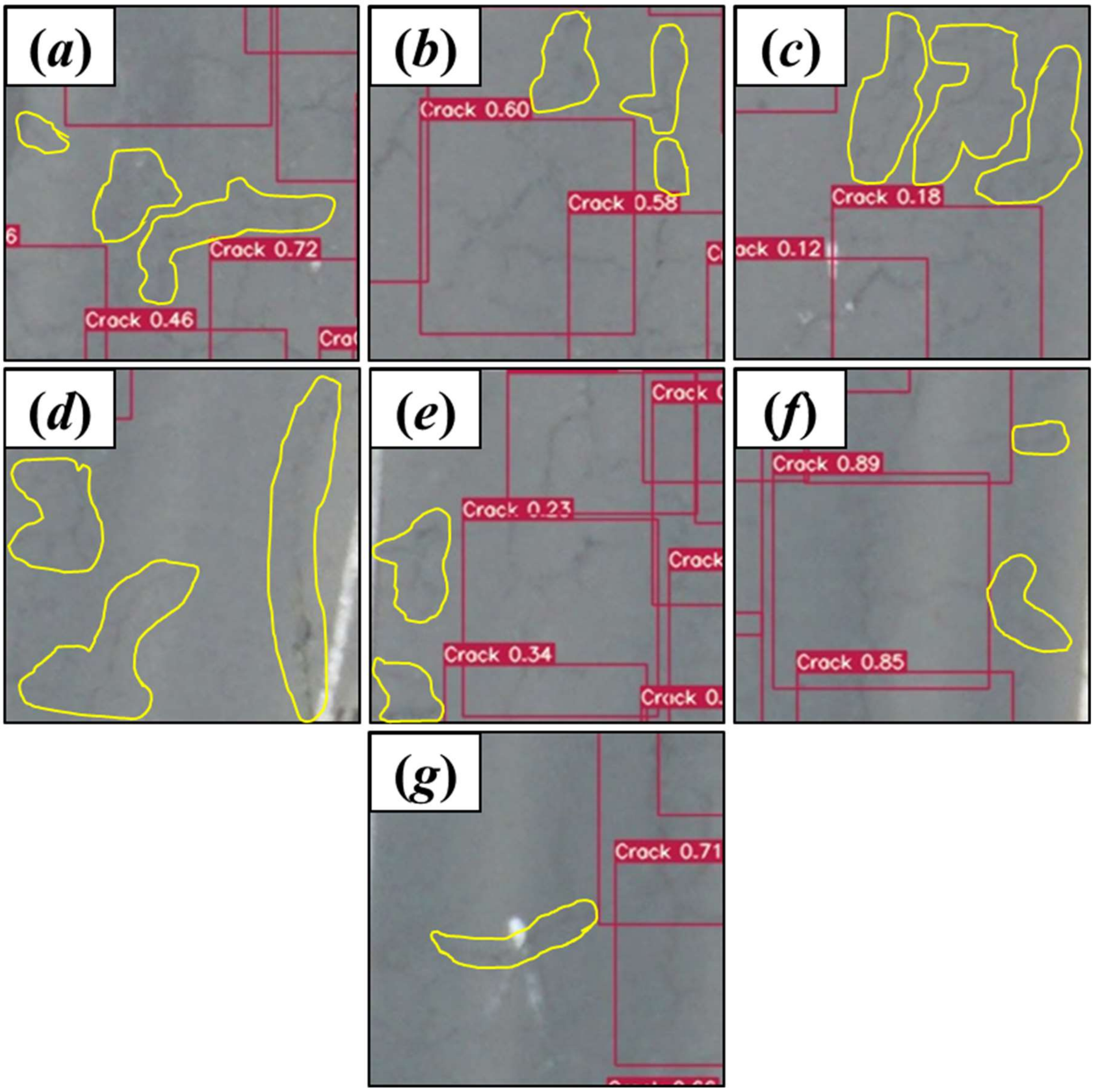


Figure 4.10 Cracks that cannot be detected in the multi-uniform size results.

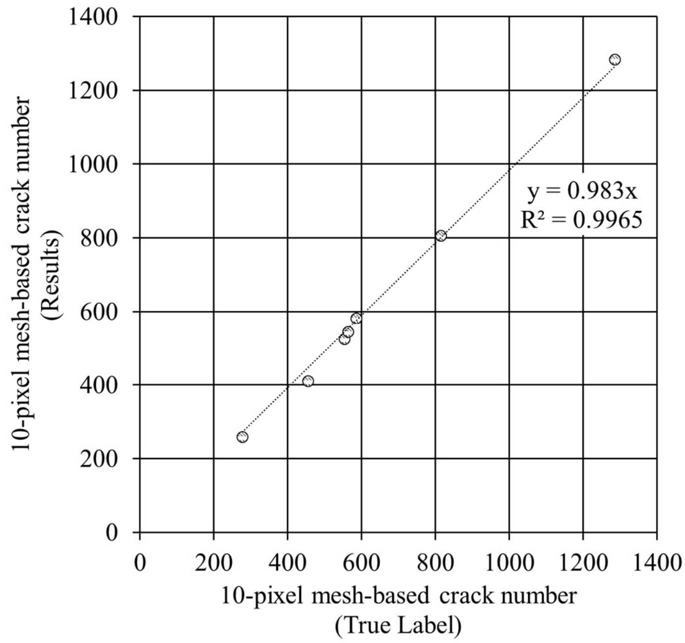


Figure 4.11 Crack numbers derived from the 10-px mesh-based result (horizontal axis) and True Label (vertical axis).

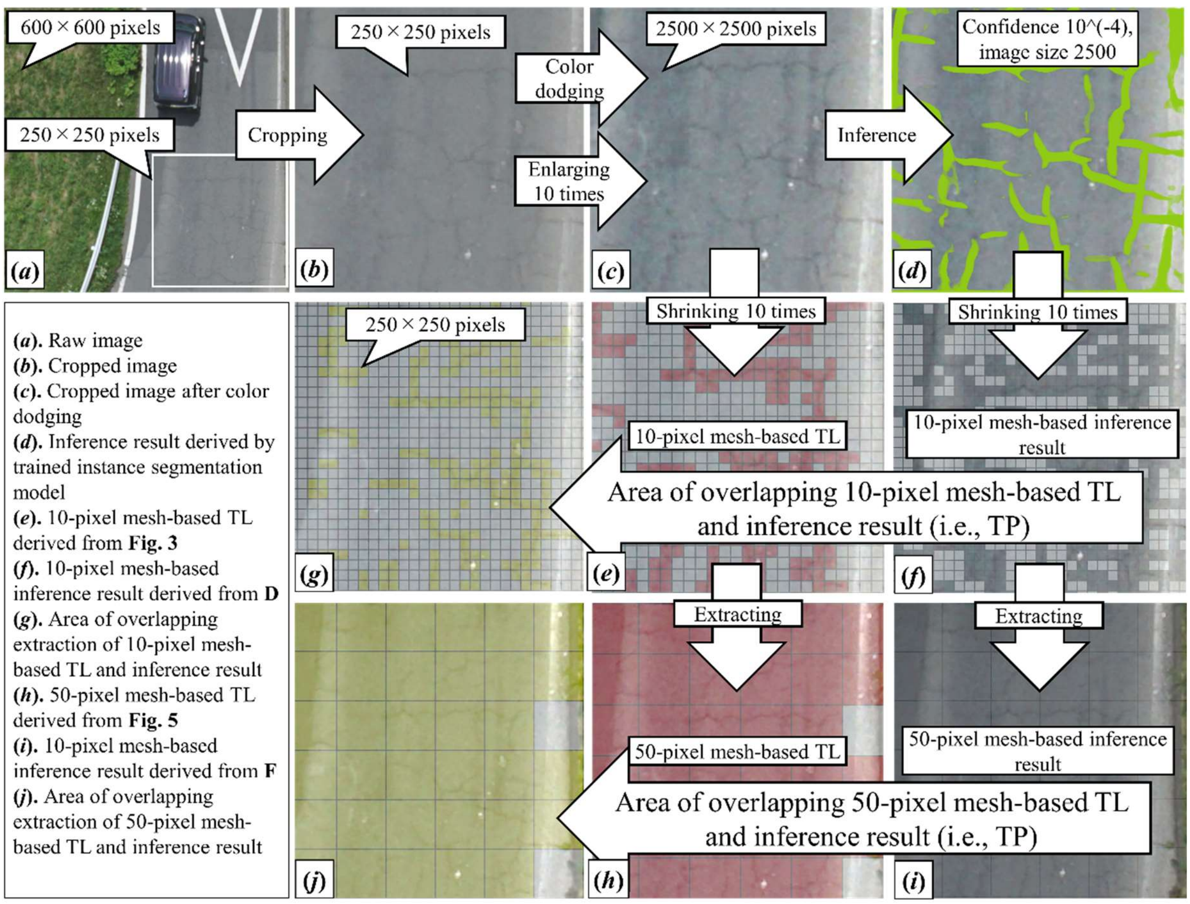


Figure 4.12 Flow chart of 10- and 50-px mesh-based crack number comparison for the YOLOv7-seg results.

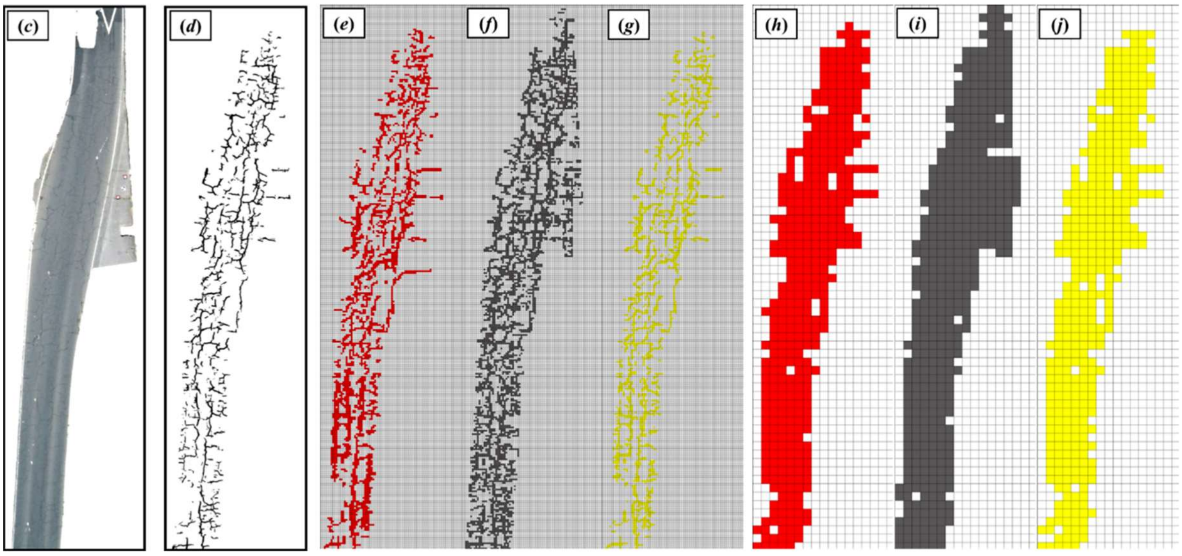


Figure 4.13 Images after color dodging and YOLOv7-seg results.

Figure 4.12 displayed the process of transferring the segmentation result derived from the YOLOv7-seg model to the 10- and 50-px mesh-based result. And in the **Figure 4.13**, more details for the results in the whole study site have been shown. Remarkably, comparing with the YOLOv7 model, YOLOv7-seg model need more details to be much tighter to the borders around the targets, that the image size enlargement is necessary.

4.3 Results and discussion

4.3.1 Multi-uniform size-based results (custom dataset, YOLOv7)

The reason of choosing the uniform size bounding box for annotating the cracks is totally based on the randomness of the practical cracks in the asphalt pavement. For including all the sizes of the cracks, several uniform sizes are chosen basically on the same crack. The inferred bounding boxes are overlapped by each other, and some smaller ones (i.e., 20-, 30-px) are also included in the bigger one (i.e., 100-px). Noteworthy, some of the connections between the bigger ones without 100-px bounding box, as shown in the **Figure 4.9 (b)**, the smaller ones have covered these blanks. But the **Figure 4.10** also showed that some blanks part which need to be improved.

From the result, there are some crack parts with specified features that cannot be recognized by the trained model. As shown in the **Figure 4.9 (b~e)**, the cracked asphalt pavement has some familiar features that some grasses have grown in the blank space in the crack. Because of lacking the similar images with grass grown in the crack crevices for the training/validation dataset (i.e., with-grass images), the recognition of these cracks in the test dataset are also difficult until now.

Another point that needs to be mentioned that each of the **Figure 4.9 (a~g)** has the alligator crack in the image. But not all the alligator cracks are detected individually. That is because of lacking the images with alligator cracks in the training dataset. And it is an issue that need to be solved for improving the accuracy.

4.3.2 Crack species-based results (public dataset, YOLOv7)

For supplying the custom dataset that lack of robustness, RDD 2020 dataset has also been trained to infer the cracks in the study site. Several alligator and linear cracks were detected in **Figure 4.9 (a~g)**. Comparing with the multi-uniform size-based results, the crack species-based results can replenish the shortage of the with-grass and alligator crack detection ability. But because of the camera angle and SGD used in RDD 2020 dataset are too different from the images for testing, the confidence of the recognition targets is all less than 0.3.

In the **Figure 4.9 (e)**, the tools on the grassland has been misclassified as the D40 (i.e., pothole). This phenomenon has performed that the usage of public dataset also has some limitations, especially when the inference images have too many differences from the images for training.

From the combination of the multi-uniform size-based and the crack species-based results, to some degree, the possibility of using the different GSD, camera angle dataset for the supplement has been proved. But to understand the accuracy of the results after supplement, only inferred bounding boxes are not enough, especially when the true labels for the test dataset are not prepared with the reasonable standards. In this research, for the reasonable assessment, the extraction using mesh is necessary.

4.3.3 10-px mesh-based crack numbers (YOLOv7)

Based on the multi-uniform size (i.e., 20-, 30-, 100-px) derived from the custom dataset, 10-px mesh is matching the need for assessing the crack detection accuracy and the annotation labor is also comparably reasonable. Because the crack detection is one of the yearly monitoring activities in the riparian area, roughly locating the crack and marking the distribution map are most considerable for the policy makers. Derived from the mentioned points, **Figure 4.11** has showed the results that the crack species-based supplied results can detect and locate over 90% of the cracks (i.e., $y = 0.983x$ in **Figure 4.11**). But the result is also very rough and without detailed size information.

4.3.4 Instance-based results (YOLOv7-seg)

Following the needs of more detailed crack information, the model using more comprehensive annotations has been trained using a public dataset. This public dataset is mainly derived from the images taken with the close distance to the targets, that included but not limited to concrete walls and asphalt pavement. Shown in the **Figure 4.12**, there are 10 steps for getting the 10- and 50-px mesh-based segmentation results on the crack targets.

Cropping the 600px × 600px raw images used in the **Figure 4.9** into 250px × 250px is the first step. And comparing with the object detection methods more details are in need, the image enlarging and color dodging are necessary for the inference. So the images in the step B have been enlarged 10 times from 250px × 250px to 2500px × 2500px in the step C. The trained YOLOv7-seg model inferred the enlarged images in the step D, and the inferred masks overlapped on the enlarged images for the visualization. Noteworthy, the inference results are derived from the 0.0001 confidence value.

After step C and D, the enlarged images were shrunk 10 times back to the original sizes, and the 10-px mesh-based TL and inference results were made in the step E (red color) and F (grey color). Area of overlapping the step E and F is called TP (i.e., true positive) using yellow color in the step G, that means the meshes have been correctly identified. For recognizing the larger mesh-based results, 50-px mesh-based TL, inference, and TP have been extracted in the step H, I and J, individually.

In the **Figure 4.13**, an ortho-photograph after color dodging was showed in the step C, then following the flow chart shown in the **Figure 4.12** from step D to J, the results in the study site are ready for the mesh-based extraction.

Table 4.2 Crack numbers and evaluation criteria. derived from the 10- and 50-px mesh-based results in **Figure 4.12** (Precision, Recall and F1).

Numbers	**10-pixels mesh**	**50-pixels mesh**	**Equations**
Yellow (TP)	2125	360	
Red (TL)	2774	378	
Grey (Inference)	3873	440	
Evaluation criteria	**10-pixels mesh**	**50-pixels mesh**	$\boldsymbol{Recall\ (R) = \frac{TP}{TL}}$
Precision (P)	0.55	0.82	$\boldsymbol{Precision\ (P) = \frac{TP}{Inference}}$
Recall (R)	0.77	0.95	$\boldsymbol{F1 = \frac{2 \times R \times P}{R + P}}$
F1	0.64	0.88	

R: Ratio of yellow to red mesh numbers;
P: Ratio of yellow to grey mesh numbers;
F1: Harmonic mean value of R and P.

4.3.5 10- and 50-px mesh-based crack numbers (YOLOv7-seg)

Focusing on the results in the **Figure 4.13,** crack numbers are extracted from 10- and 50-px mesh-based results. As performed in the **Table 4.2**, 10-px mesh-based Recall is over 0.75, that means just around one quarter of all the labelled cracks are not detected correctly in the 0.25 m ground mesh. If the mesh size has been enlarged 5 times, the F1 score can be increased from 0.64 to 0.88 in the 1.25 m ground mesh. For the yearly river monitoring, 1.25 m ground mesh is detailed enough in the road asphalt pavement assessment. Especially considering the long distance images collections and high quality image standard for the crack detection, until now, 0.025 m GSD is fitting for this riparian monitoring mission.

Derived from the 50-px cracked mesh distributions, the policy maker can also make the asphalt pavement renew plan without time-consume vehicle driving for the data collection, at the same time that the distribution mapping can also have a believable Recall (i.e., 0.95).

4.3.6 Discussion

From the above results, the abilities of solving the object detection and instance segmentation tasks using YOLOv7 and YOLOv7-seg algorithms have been well-proved individually. Located crack numbers using YOLOv7 in the 10-px mesh have been 90% correctly counted, that means unnoticed cracks are very less. Based on the detailed annotations, YOLOv7-seg algorithm in the 50-px mesh has supported to detect the mesh-based cracks with the 0.88 F1 score. These two approaches have all proved that the cracks can be located and detected with a reasonable standard. But these methods also showed the limitation of the riparian crack monitoring, i.e., the quality of the training/validation dataset need to be comparable high both in the images and annotations, which take lots of time.

Generally speaking, the trained models in this research have proved with the comparable high assessment criteria, but also existing some limitations, like lacking of the robustness in weathers and crack species, the training/validation dataset are not enough, assessment standards for cracks are difficult for the large-scale area. Then this research also showed the possibility of using the YOLOv7 and YOLOv7-seg models to support the UAV-based riparian asphalt-paved crack monitoring. In the future work, if the UAV-based images can be taken by the camera with higher resolution zoom-in function, the current accuracy can be more improved.

4.4 Conclusion

This study presented methods for crack detection and segmentation in riparian road asphalt pavement using a drone equipped with a divided digital camera for data collection, the YOLOv7 model for data analysis, and public datasets for data supplement. The combination of drone technology and the YOLOv7 model showed potential in enhancing the efficiency and accuracy of crack detection and segmentation. However, there were limitations in detecting and segmenting the certain cracks due to complex target shapes under the random contrast and brightness.

Various crack datasets, including uniformed bounding box size-, species-, and instance-based, were used for training and validation. The YOLOv7 (YOLOv7-seg) model with its anchor-based approach and advanced techniques, achieved reasonable crack detection and segmentation, with both around 0.9 F1 value derived from the mesh-based assess approach in the detected crack areas around. Noteworthy, the YOLOv7-seg model, which performed with segmentation result, required tighter boundaries of the cracks for accurate results. Overall, the study highlighted the potential of UAVs and computer vision algorithms for efficient and accurate crack detection in riparian road asphalt pavement.

4.5 Future work

The application of the YOLOv7-seg model in asphalt-paved crack segmentation has demonstrated that the close-distance dataset can be effectively used in remote sensing tasks. Typically, remote sensing tasks involve analyzing and interpreting data collected from a distance, such as satellite or aerial imagery. The close-distance dataset in this study has inspired the authors to consider it a valuable supplement to the crack feature in future work, mainly due to its relatively more straightforward obtainability. But the generations of the annotations also need a large amount of the human-labeling. Recently, as performed in **Figure**

4.14, a new AI model called Segment Anything Model (i.e., SAM) has empowered the generations of Instance Segmentation annotations, that can reduce the burden of the time-consuming human-labeling for the researchers. SAM (**Kirillov *et al.*, 2023**) is a prompt-able segmentation system with zero-shot generalization to unfamiliar objects and images, without the need for additional training.

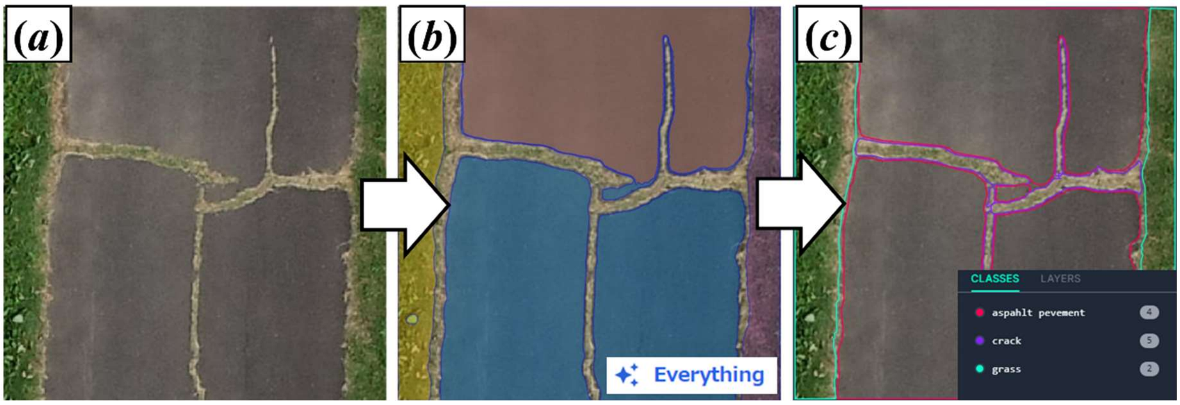


Figure 4.14 Instance Segmentation annotations can be generated by the AI-powered web browser-based annotation tool "Smart Polygon" efficiently: (*a*) Raw Image; (*b*) Mask generated by Segment Anything Model (i.e., Everything function); (*c*) Annotations generated by a cloud-hosted Segment Anything model, that can apply the accurate polygon annotations with a fast speed in the Roboflow UI using just one-click (i.e., Smart Polygon function).

The limitations of the trained YOLOv7 model derived from the random brightness and contrast also need to be in consideration. To some degree, the technologies like data augmentation can improve the accuracy of detecting and segmenting the cracks under these random situations. But then it is difficult to adjust the thresholds or the possibility of the parameters without considering the weather conditions. If reasonable Parameters setting is necessary, considering the standard evaluation approach of the brightness and contrast value is an acceptable opinion.

References (Chapter 4)

Arya, D., Maeda, H., Ghosh, S. K., Toshniwal, D., Omata, H., Kashiyama, T., Seto, T., Mrax, A & Sekimoto, Y. RDD2020: an image dataset for smartphone-based road damage detection and classification. Mendel. Data, 1. 2021. https://doi.org/10.1016/j.dib.2021.107133

Asada, T., Ikoma, S., Nagaya, H., & Kameyama, S. Development of pavement inspection system by using carmounted camera and U-net model. Journal of Japan Society of Civil Engineers, Ser. E1 (Pavement Engineering), 76(2). 2020. https://doi.org/10.2208/jscejpe.76.2_I_123

Chun, P. J., Yamane, T., & Tsuzuki, Y. Automatic detection of cracks in asphalt pavement using deep learning to overcome weaknesses in images and GIS visualization. Applied Sciences, 11(3), 892. 2021. https://doi.org/10.3390/app11030892

Emoto, H., Numata, M., & Shiga, A. Application of Aibased Deformation Extract Function from a Road Surface Video to a Road Pavement Condition Assessment System. In Proceedings of The 17th East Asian-Pacific Conference on Structural Engineering and Construction, 2022: EASEC17, Singapore (pp. 1503-1514). Singapore: Springer Nature Singapore. 2023. https://doi.org/10.1007/978-981-19-7331-4_119

Numata, M., Ota, R., Emoto, H., & Yoshitake, T. Practical study of a road surface evaluation system using extracting degradation automatically and map system. Journal of Infrastructure Maintenance Practice and Research (Web), 2(1), 66-75. 2023. https://doi.org/10.11532/jsceim.2.1_66

Qiu, Q., & Lau, D. Real-time detection of cracks in tiled sidewalks using YOLO-based method applied to unmanned aerial vehicle (UAV) images. Automation in Construction, 147, 104745. 2023. https://doi.org/10.1016/j.autcon.2023.104745

Zhu, J., Zhong, J., Ma, T., Huang, X., Zhang, W., & Zhou, Y. Pavement distress detection using convolutional neural networks with images captured via UAV. Automation in Construction, 133, 103991. 2022. https://doi.org/10.1016/j.autcon.2021.103991

Roboflow, https://universe.roboflow.com/university-bswxt/crackbphdr

Ultralytics, https://github.com/ultralytics/yolov5.

CHAPTER 5

Improvements and Suggestions on River Patrolling Methods

6.1 Manual of River Patrolling Methods on AI and UAV

Application of UAVs (Unmanned Aerial Vehicles) and AI for river patrol operations in Japan involves a multi-faceted approach.

Until now, long-range drones, such as hybrid multi-copter models, are deployed to cover extensive river areas from a single launch point. These drones are equipped with all-weather capabilities to operate in various conditions, including rainy and strong-windy. Real-time data transmission systems using Wi-Fi, radio, LTE, 5G, or satellite communication allow remote operators to monitor the patrol processing.

AI, particularly deep learning techniques, is employed to automate anomaly detection in the collected imagery. Object detection algorithms identify specific items like waste pollution or infrastructure damage (i.e., asphalt paved cracks on the top-surface in this research), while semantic segmentation analyzes land cover classification changes in the riparian area over time.

In December 2022, Japan amended its aviation laws to promote drone utilization. The new regulations allow for "Level 4" flights, and to implement this system effectively, a comprehensive flight management system is crucial for planning, executing, and monitoring beyond visual line of sight (i.e., BVLOS) flights in populated areas., which are now permitted under Japan's updated aviation laws. This regulatory change is crucial for enabling more extensive and efficient river patrol operations using drones. Additionally, a robust data management system is necessary to handle the large volumes of data generated during patrols, including standardized file formats, naming conventions, and detailed metadata for efficient searching and analysis.

Manual of river patrolling methods identifies four key technical elements necessary for effective drone-based river patrols: Long-range Flight Capability, All-weather Operation, Communication and Data Transmission, Flight Management Systems. Drones must be able to cover the entire river area under management from a single launch point. This capability is essential for improving work efficiency in river patrols. River patrols are required in various weather conditions, including rain and strong winds after disasters. Drones need to be dust-proof and waterproof to operate in these diverse conditions. Real-time transmission of drone-captured imagery to remote operators is crucial. The document mentions the use of WiFi, radio, LTE, 5G, and satellite communication technologies. These systems are necessary for implementing Level 4 (BVLOS) flights. Key functions include drone registration and management, flight planning and approval, tracking and operation execution, monitoring and detection, and emergency response capabilities.

The manual outlines the use of AI, particularly deep learning techniques, for automating anomaly detection in river patrols: Object Detection, which was used for identifying specific objects like garbage or cracks in river infrastructure. And the AI model outputs bounding boxes, object names, and confidence scores; Semantic Segmentation, was used for analyzing changes in riparian areas over time, such as sandbar formation or vegetation growth, and this AI model can classify each pixel in the image into predefined categories.

Derived from the mentioned content, there are also several challenges that need to be solved, i.e., Lack of sufficient training data, especially for rare events or anomalies; Need for large amounts of labeled images across various patrol item categories. Considering of the challenges, data Augmentation is a reasonable solution to increase the amount of the dataset. Techniques like image flipping, rotation, and cropping are suggested to increase the amount of training data. And these methods can help compensate for the lack of real-world data for certain patrol items.

This manual also emphasizes the importance of efficient data management for drone-based river patrols: Data Formats, i.e., specifies file formats for different types of data (e.g., .jpg for drone images, .tif for orthophotos, .csv for flight logs); File Naming Conventions, providing a standardized naming system including river name, section, date, etc.; Metadata, which recommends creating detailed metadata for each data type to facilitate efficient searching and management. Metadata includes information like river name, coordinates, capture date, etc.

6.2 Improvements and Suggestions on current River Patrolling Methods

Enhance the feature/amount of data for AI

- Implement the long-term metadata accumulation for the AI models.
- Continuously improving AI models by expanding rare events.
- Applying the public dataset to supply the current data lack
- Applying the multimodal data (i.e., LiDAR) to supply the current feature lack.
- Applying the AIGC dataset to match the specific data needs.

Collaborate and share knowledge

- Establish partnerships with other river management authorities to share best practices and data.
- Collaborate with technology providers and research institutions to stay at the forefront of drone and AI advancements in river patrol applications.

In conclusion, this research provides a comprehensive overview of how to solve the data feature/quantity lack in river patrolling. It outlines the current state of the technology, identifies challenges, and provides guidelines for implementation while also looking ahead to future developments in this rapidly evolving field. The integration of these technologies promises to significantly enhance the efficiency and effectiveness of river management and monitoring in Japan.

CHAPTER 6

Concluding Remarks

In conclusion, this research has pioneered several innovative methodologies that have significantly advanced the field of civil engineering, remote sensing and AI technology. The development and application of the **LiDAR-assisted DeepLabV3+ Model** have not only enhanced the features of aerial photography, but also improved the accuracy of land cover classification tasks. The successful implementation of this model in the **2018 Asahi Flood Simulation** has optimized water level inference, demonstrating the practical impact of the method derived from this research.

Furthermore, the novel **fusion methodology** of aerial photography and **LiDAR**, utilizing a high contrast color scale, has expanded data features and improved performance. The exploration of the exchangeability between **ALB** and **GLS** for **LCC** tasks has opened new avenues for research and application.

The adaptation of open-source **object detection** and **instance segmentation** models for UAV-based detection tasks has shown promising results in **asphalt-paved crack**s and **waste pollution detection**, leading to the development of a comprehensive manual for the application.

The creation of **AI-based Generative Content (AIGC)** has increased data availability and proven its potential in augmenting object detection tasks, particularly in waste pollution detection. Lastly, the enhancement of river patrolling tasks through AI-powered UAV technology has set a new standard for efficiency and effectiveness in environmental monitoring.

This research not only contributes to the academic community but also holds immense potential for real-world applications, paving the way for future innovations in the domain of civil engineering, remote sensing and AI technology.

In the future, the author is planning that the mentioned technologies can be widely applied in the practical tasks for making out the not-solved engineering problems. And the author also hope that the other researchers can also get inspirations from the works in this research.

Future Works List:

- More study sites (except the Aashi River) are need to use to prove the generalizability of the methods in this research;
- LCC TL mapping should be made with a more efficient way, which reduce the work time and work load (e.g. using the pre-trained model to generate the rough LCC mapping);
- Comparing the advantages and disadvantages of DL and RF approach on the LCC tasks;
- More detailed parameter settings are necessary in the YOLO model training (i.e., batch size, epochs, Lr0, optimizer);
- More Object Detection/Instance Segmentation models should be trained to be as the reference for YOLO;
- An AIGC data bank/platform should be built for the accumulation of the Digital Twin in the near future.
- The AI + UAV concept for the waste detection should be shared with the local community to increase the impact and the change of the practical application.
- The resolution/GSD of the drone images for the detection should link the relationship with the resolution/GSD of the training dataset.

Acknowledgements

I would like to thank my supervisor, Associate Professor Dr. Keisuke YOSHIDA, Graduate School of Environmental and Life Science, Okayama University, Japan, for his kind and scholarly guidance throughout the PhD study.

I would also like to thank my another mentor, Professor Dr. Satoshi NISHIYAMA, Graduate School of Environmental and Life Science, Okayama University, Japan, for his insightful direction throughout my PhD research.

Thanks to Senior Researcher Dr. Takashi KOJIMA, TOKEN C.E.E. Consultants Co.,Ltd. for his valuable direction throughout the PhD research.

Thanks also to my academic co-supervisors, Associate Professor Dr. Kazushi KIMOTO and Professor Dr. Yuji TAKESHITA, Graduate School of Environmental and Life Science, Okayama University, Japan, for their suggestions throughout the course.

Thanks for my family's financial support to complete my PhD study in Japan. Also, thanks to my partner for being the best companions in Japan and for encouraging and supporting my scientific journey.

Special thanks to all members of the Hydraulic Engineering Laboratory, Graduate School of Environmental and Life Science, Okayama University, Japan, for their kind cooperation and for making my student life enjoyable.

I would also like to thank the Chugoku Regional Development Bureau, Ministry of Land, Infrastructure, Transport, and Tourism, TOKEN C.E.E. Consultants Co.,Ltd. and Pasco Corporation, Japan for providing the necessary data to complete this study.

This work was supported by JST SPRING, Grant Number JPMJSP2126.

The author acknowledge the financial support served by Okayama University with the donations from DOWA HOLDINGS Co., Ltd. and TOHO ELECTRIC INDUSTRIAL

Co., Ltd.

This study was supported by the River Fund of the River Foundation, Japan.

This study was supported in part by the Chugoku Kensetsu Kousaikai, and the Wesco Scientific Promotion Foundation.