

Showcasing the TAI AO Project: Providing Resources for Machine Learning from Images of New Zealand’s Natural Environment

Nick Lim nlim@waikato.ac.nz
and Albert Bifet abifet@waikato.ac.nz
and Daniel Bull danbull1@yahoo.com
and Eibe Frank eibe@waikato.ac.nz
and Yunzhe, Jia ajia@waikato.ac.nz
and Jacob Montiel jmontiel@waikato.ac.nz
and Bernhard Pfahringer bernhard@waikato.ac.nz

Department of Computer Science, University of Waikato, 3216 Hamilton, New Zealand

ARTICLE HISTORY

Compiled September 20, 2022

ABSTRACT

Proper management of the earth’s natural resources is imperative to combat further degradation of the natural environment. However, the environmental datasets necessary for informed resource planning and conservation can be costly to collect and annotate. Consequently, there is a lack of publicly available datasets, particularly annotated image datasets relevant for environmental conservation, that can be used for the evaluation of machine learning algorithms to determine their applicability in real-world scenarios. To address this, the Time-evolving Data Science and Artificial Intelligence for Advanced Open Environmental Science (TAIAO) project in New Zealand aims to provide a collection of datasets and accompanying example notebooks for their analysis. This paper showcases three New Zealand-based annotated image datasets that form part of the collection. The first dataset contains annotated images of various predator species, mainly small invasive mammals, taken using low-light camera traps predominantly at night. The second provides aerial photography of the Waikato region in New Zealand, in which stands of Kahikatea (a native New Zealand tree) have been marked up using manual segmentation. The third is a dataset containing orthorectified high-resolution aerial photography, paired with satellite imagery taken by Sentinel-2. Additionally, the TAI AO web platform also contains a collated list of other datasets provided and licensed by our data partners that may be of interest to other researchers.

RESEARCH RESOURCE

1. Introduction

Environmental data science is strategically essential to New Zealand because it supports research on climate change impacts, adaptation to concomitant changes in the environment, and conservation of the biosphere [9]. Effective data science can take an essential role in the New Zealand government’s goals of improving the quality of freshwater and reaching zero carbon by 2050.

While the technical means for data collection in environmental data science are becoming ever more accessible, the sheer volume of these data can be challenging to

clean, annotate, and verify. This is compounded by the fact that environmental datasets, like other specialized datasets, tend to be niche and require expert domain knowledge to accurately label.

From a machine learning point-of-view, datasets that accurately reflect the variability in a natural environment are useful to evaluate the applicability of machine learning algorithms in real-world settings. Images in many popular image datasets tend to be taken in ideal conditions and not be reflective of real-world variability [28]. One of the objectives of the TAI AO project on environmental data science in New Zealand that funds the work presented in this paper is to provide datasets that are suitable for the application and evaluation of machine learning techniques, with a special emphasis on data collected on the New Zealand environment.

We first present a brief overview of the TAI AO project before discussing the three datasets that form the primary contribution of this paper, including results of machine learning techniques applied to them. The first dataset involves species identification from camera trap imagery taken in New Zealand’s native forest. The second dataset consists of aerial photography of the primarily pastoral Waikato region in New Zealand that has been manually annotated to segment stands of Kahikatea, a tree that is native to New Zealand. The third dataset links high-resolution aerial photography of New Zealand land cover to low-resolution satellite imagery and can be used to evaluate the accuracy of machine learning-based super-resolution techniques.

2. Time-Evolving Data Science / Artificial Intelligence for Advanced Open Environmental Science (TAI AO)

TAI AO [25] is a New Zealand government-supported multi-year programme aimed at improving the data science capabilities for environmental researchers in New Zealand. In addition to upskilling New Zealand researchers, the programme aims to help advance the state-of-the-art in environmental data science by enabling the development of new machine learning algorithms with a special emphasis on methods that are suitable for processing data collected on the New Zealand environment. In line with this aim, an important objective of TAI AO is to build an open-source framework to implement machine learning on environmental data and to provide an openly available repository with datasets to improve reproducibility in environmental data science.

The TAI AO programme is a multi-institute and multi-domain programme. It includes data scientists, data engineers, environmental scientists, and machine learning researchers from undergraduate to post-graduate levels. Moreover, collaboration extends beyond technical aspects to include regional councils, iwi¹, and co-governance entities to implement the methods developed by the project to support governance and management decisions.

2.1. Reproducible Notebooks

Figure 1 provides a conceptual view of the TAI AO software platform and how the corresponding components of the platform interact. It consists of a catalogue of datasets, a library of transformations that can potentially be applied to a dataset (denoted in Figure 1 as Transf.), and a catalogue of so-called “notebooks” containing program code for processing the data along with documentation of the code and the purpose of the

¹These are the largest social units in Aotearoa Māori society.

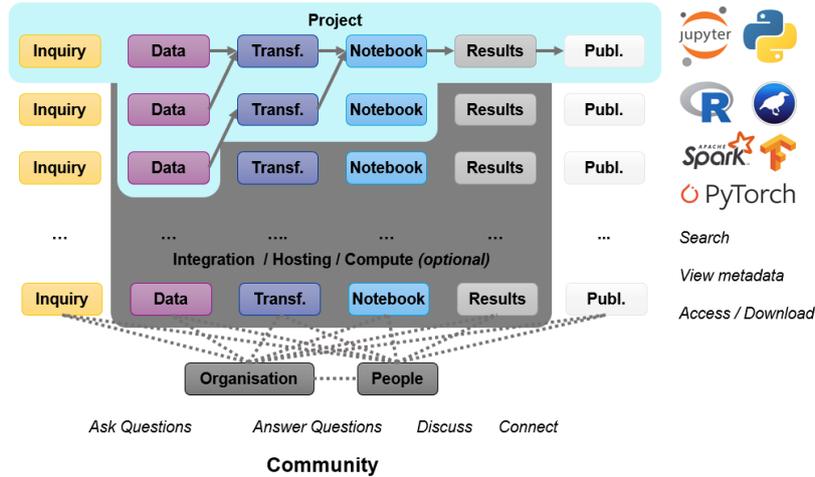


Figure 1.: Conceptual view of the TAI AO platform and the interaction of the corresponding components.

notebook.

The TAI AO software platform uses Jupyter notebooks to ensure reproducibility and sufficient documentation because the Jupyter software is open-sourced and lightweight. Each notebook is associated with a task and includes a description of how the data can be accessed. The notebooks also document the application for which the dataset was gathered as well as the questions the data provider and researcher seek to answer with the data. We envision that as more notebooks in the platform are developed, the project will improve the accessibility of data science research to environmental scientists by providing examples of the application of state-of-the-art techniques.

2.2. Indigenous Data Science

TAI AO and the New Zealand government are committed to “Vision Mātauranga” [13], which aims to unlock the potential of traditional indigenous knowledge and recognize the value of the tradition and knowledge passed down through the people living in the land. As part of this commitment, TAI AO recognizes the community as partners in science and innovation, and guardians of natural resources and indigenous knowledge.

TAI AO is also committed to the principles of “Te Mana Raraunga” [24], which recognizes indigenous data rights and sovereignty, practising regular dialogue with the owners of the data on how it is being used. It works closely with local co-governance entities and iwi to understand the interests and issues of the community. Additionally, TAI AO holds regular dialogues and meetings to review the environmental findings and review the capability developed.

3. Featured Datasets

This paper features three annotated datasets provided by the TAI AO project. The first dataset contains annotated images of various predator species present in New Zealand, mainly small invasive mammal species, taken using night vision cameras. The second dataset contains aerial photography of the Waikato region, annotated with

the segmentation of stands of Kahikatea, a native New Zealand tree. The third is a dataset containing orthorectified high-resolution aerial photography, paired with satellite imagery taken by the Sentinel-2 satellites. Users may access the datasets through the following DOI identifiers:

- Karioi Predator Camera: DOI: 10.5281/zenodo.5276593 [14]
- Kahikatea Dataset: DOI: 10.5281/zenodo.5059769 [11]
- Waikato Satellite Super-resolution: DOI: 10.5281/zenodo.5296368 [3]

3.1. Predator Camera Dataset

The Karioi Predator Camera dataset is derived from videos originally provided by a New Zealand regional council. These videos were recorded from 2018-2020 in the Mount Karioi region by motion-activated cameras. The raw video files consist of 2,101 thirty-second clips of videos captured using motion-triggered cameras deployed in the native forests located at Mount Karioi in New Zealand. These videos exhibit between thirty and sixty frames per second, at a 1920x1080 resolution, and are predominantly taken at night. Due to the nature of the predator animals considered, they tend to hide in the foliage, rendering this a particularly challenging machine learning problem.

A large proportion of the videos (approximately 85%) are false positives and do not contain predator species of interest. The main challenge to building an image classification dataset from these videos comes from the large number of frames (approximately 2.5 million) that need to be processed to find images of the predator species. In order to reduce the amount of manual work required for this, the background subtraction function provided by the Matlab software, namely *vision.ForegroundDetector* with 5-frame subtraction, was applied to help detect the movement of the predator species in the video. Subsequently, the blob detector function *vision.BlobAnalysis*, with `minimumBlobArea` set to 49, was used to further filter out frames solely exhibiting movement of leaves and plants. Through the combination of these two processes, the number of frames needed for manual processing was reduced to a manageable ≈ 90000 frames. The frames of the video clips were then manually annotated with the predator species as well as bounding boxes using the background-subtracted videos as a guide. Finally, the bounding boxes of the predator species were used to crop and resize the images to 224×224 pixels. For the negative class, containing images that do not exhibit predators, random 224×224 crops were taken from the corresponding video frames. Figure 2 summarizes the process used to generate the image dataset from the raw video files.

For our publicly available dataset, 1200 images from three of the predator species in the data (cats, possums, and rats) were randomly selected, but all 123 images available for stoats are included. Additionally, 2400 images that do not contain a predator are provided. The images are organized by folders, with the folders representing the class of the image. Table 1 and Figure 3 summarize the breakdown of the classes and present representative images respectively.

The Karioi Predator Camera dataset differs significantly from other publicly available camera trap datasets. Datasets such as Snapshot Safari [23] and WCS Camera Trap [22] focus on large mammals in an open field, and images in datasets that focus on small mammals such as the LILA Wellington Camera Trap data [1] are usually taken during daylight hours, away from the foliage.

To provide benchmark results for this challenging data, we present results for several image classification networks that were pretrained on ImageNet and fine-tuned on the

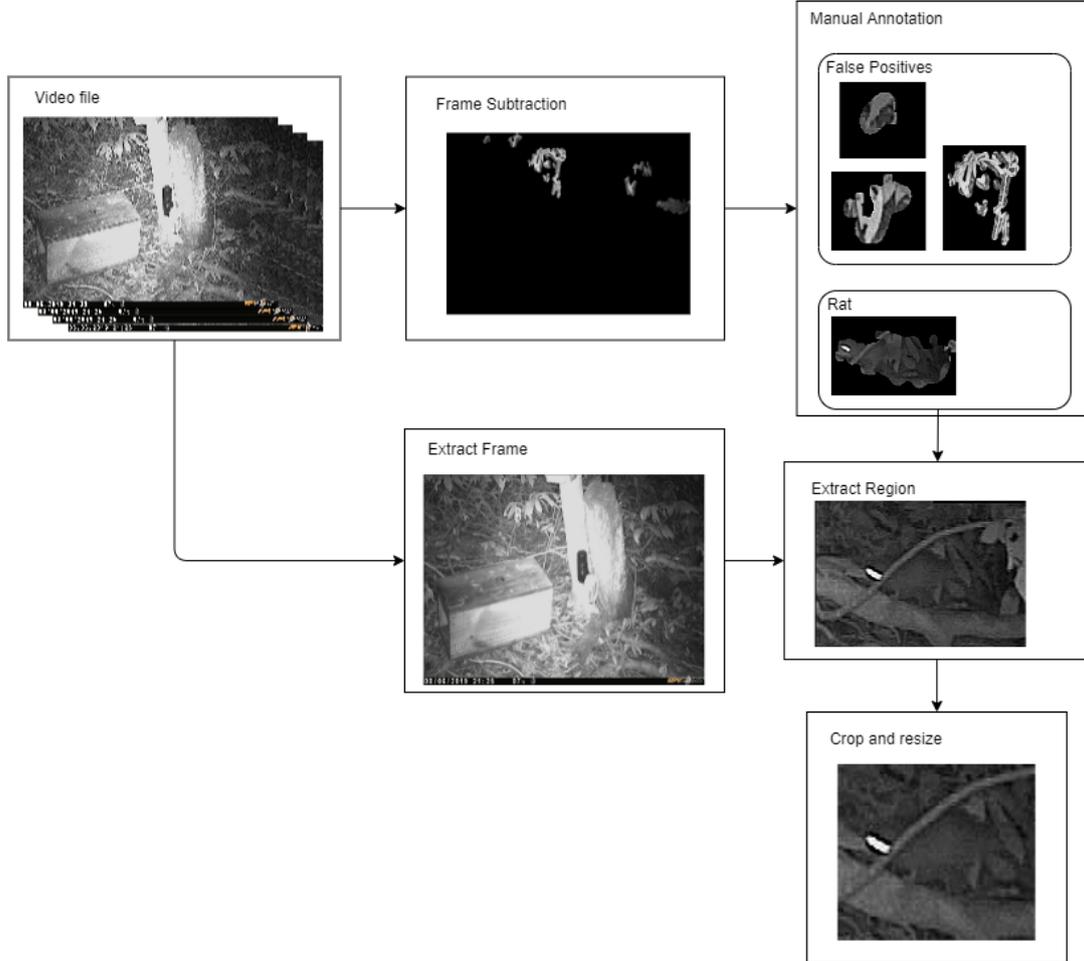


Figure 2.: Process used to generate the image dataset from the raw video files.

training data of the Karioi Predator Camera dataset. The following data augmentation was applied during training: random flips in both horizontal and vertical orientations, brightness and contrast jitter with a strength of 0.5, randomized rotation, and random crops of 60% to 80% of each image. We trained for 25 epochs in each run and picked the model with the lowest classification error on a separate internal unseen validation data set (a subset of the training set). The neural networks (the convolutional neural networks EfficientNet-B0, EfficientNet-B4, MobileNetV3, ResNet-18, ResNet-50, and the vision transformer network VIT-S) were trained using cross-entropy loss and the ADAM optimizer, with a batch size of 64 and learning rate of 0.001. The backbone of the vision transformer network VIT-S used pretrained weights provided by Facebook AI Research [6] and used a linear classifier as the classification head. The codes used to fine-tune the models as well as the Matlab codes used to process the videos are available at <https://github.com/nlim-uow/predatorcam>. We also include baseline results for a linear support vector machine (SVM).

Table 2 summarizes the classification performance of the various image classification networks on the evaluation set. The table shows the means and standard deviations obtained from five runs. Note that while we did not use methods to handle the imbalance of the classes during the fine-tuning, we report the macro-F1 score to give

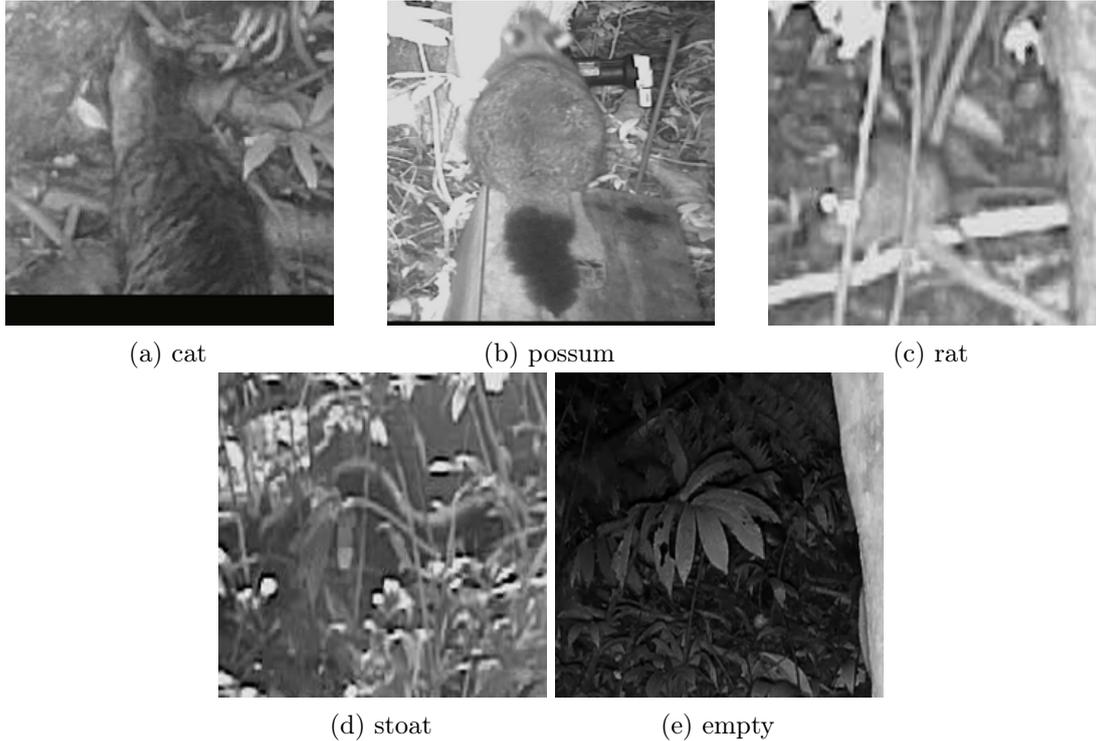


Figure 3.: Sample images in the Karioi Predator Camera dataset with the corresponding classes.

Table 1.: Breakdown of the classes in the dataset.

	Training	Evaluation
Cats	1000	200
Possums	1000	200
Rats	1000	200
Stoats	100	23
Empty	2000	400

equal weight to the minority class (stoats). Users may obtain better performance in the minority classes through methods such as random over-sampling during training.

3.2. Aerial Photography of Kahikatea Trees

The kahikatea is a protected native tree in New Zealand, and there is strong interest from regional councils to identify and catalogue such trees from aerial photography. The aerial photographs in the second dataset featured in this paper, the Kahikatea dataset, were taken at two different times, namely in June 2016 and March 2017, using an undercarriage camera at heights between 200m and 400m above sea level. One of the challenges in creating this dataset was the difficulty in identifying and distinguishing species of the tree from the approximately 80,000 aerial photographs obtained from one of New Zealand’s regional councils. To address this challenge, the New Zealand Land Cover Database 4.1 [27] was used to identify the known approximate locations

Table 2.: The average image classification performance of various image classification networks across five runs. Standard deviations are listed in parenthesis.

	Top-1 Accuracy	Macro F1
EfficientNet-B0	0.9501 (0.0038)	0.9465 (0.0027)
EfficientNet-B4	0.9371 (0.0033)	0.9285 (0.0050)
MobileNetV3	0.9609 (0.0018)	0.9625 (0.0025)
ResNet-18	0.9423 (0.0016)	0.9465 (0.0034)
ResNet-50	0.9499 (0.0028)	0.9505 (0.0024)
VIT-S	0.9609 (0.0052)	0.9594 (0.0119)
SVM (linear kernel)	0.4213 (0.1212)	0.4145 (0.1017)

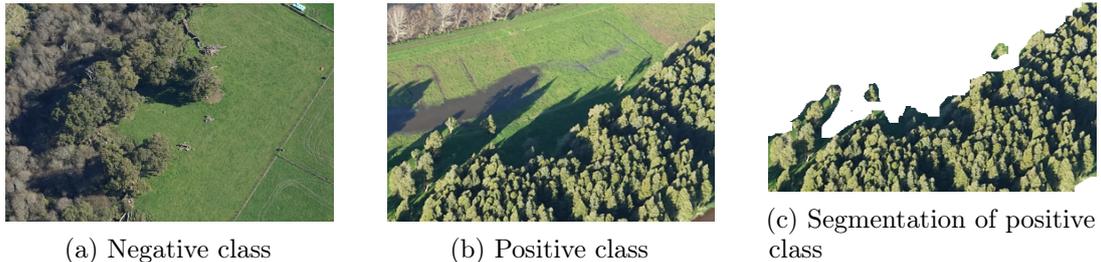


Figure 4.: Sample images from the Kahikatea dataset.

of kahikatea trees and cross-reference the locations with the GPS tags in the aerial photographs. After cross-referencing, images were manually separated based on whether or not they contained kahikatea. Finally, to enable training of image segmentation models, the kahikatea trees in the positive examples were manual segmented.

The final dataset contains 232 positive and 287 negative examples of images with kahikatea. Figure 4 shows some representative images from the dataset. There appears to be relatively few such publicly available datasets of segmented aerial photography. The INRIA [15] dataset is one example, but it focuses on classifying man-made buildings versus non-buildings and is thus not as fine-grained as the Kahikatea Dataset. The resisc45 [7] dataset contains more classes (45 in total) but does not have segmentation-level annotations.

The kahikatea dataset was used to produce the results in [12] detailing the correlation between the explainability of deep learning models and their test accuracy. The paper shows that there is a positive correlation between explanation quality and test accuracy and a negative correlation between explanation quality and test loss. The codes for reproducing the results can be found at <https://github.com/Alvence/AccuracyAndExplanationQuality>. Additional details and the process of how the explanation quality is defined can be found in the cited paper.

3.3. Paired High-resolution Aerial Photography and Sentinel-2 Satellite Imagery

The third dataset featured in this paper contains high-resolution orthorectified aerial photography provided by the Waikato Region Aerial Photography initiative (WRAPS) [26] paired with low-resolution satellite images. These images were collected as part of a

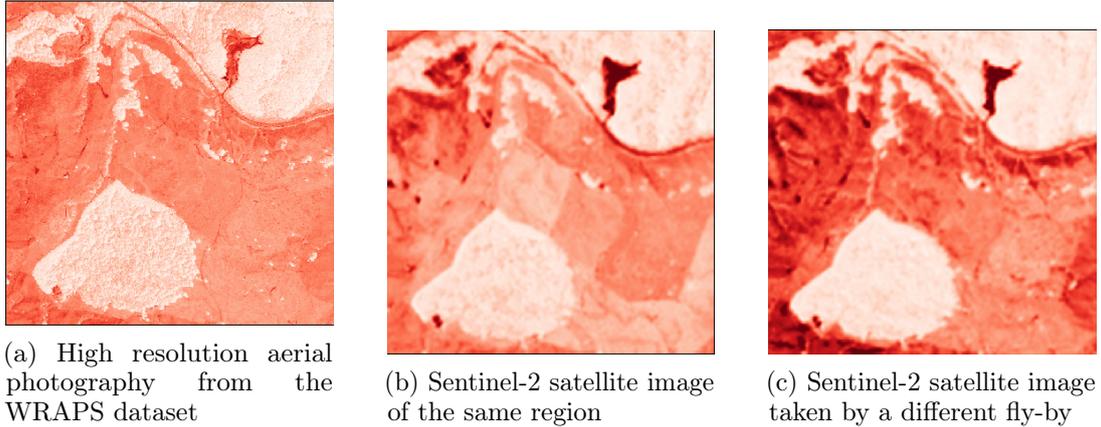


Figure 5.: Paired images from the WRAPS (aerial photography) and Sentinel-2 (satellite) data. Images shown here are from the red-channel dataset. Also note that there are six Sentinel-2 images paired with each of the WRAPS images in the actual data, not just two as shown here.

satellite imagery super-resolution research project. Each aerial photograph was down-sampled to a spatial resolution of 2.5m per pixel, while the satellite images were taken at a spatial resolution of 10m per pixel. The satellite images were taken from eight fly-bys of the Sentinel-2 satellites [30] between 16th January to 2nd March 2019 so that this data is temporally consistent with the aerial photographs taken in February of 2019.

Experiments with variants of the data showed that super-resolution algorithms do not perform well when the reference satellite images contain high cloud coverage, producing spurious results. In order to address this, the IdePix plug-in provided by ESA [8] was used to identify satellite images with high cloud coverage and generate a cloud mask of the images. Regions were then filtered such that each region covered in the data contains at least six relatively cloud-free satellite images. Subsequently, the aerial photographs were cropped to 512×512 pixels, and the satellite images were cropped to 128×128 pixels. These images were stored as NumPy .npy files and organized by their respective visible channel (red, green, and blue). Figure 5 shows some representative images from the dataset.

The final dataset comprises of 16,266 training scenes and 71 evaluation scenes of the three visible bands. The dataset also includes binary cloud masks generated by IdePix representing the presence of clouds in the satellite images.

Considering related work, the ProbaV Super-Resolution Challenge dataset [19] is also a paired satellite image dataset to be used for super-resolution, but there are some major differences between the dataset presented here and the ProbaV dataset. First, the ProbaV dataset uses images from the same satellite taken with two different resolutions (namely 100m and 300m), with the 100m resolution image as the ground truth. The dataset presented here instead uses more costly aerial photography as the ground truth. This setting is more challenging and has more wide-reaching remote-sensing applications: ultimately, the goal is to provide high-resolution aerial photography without the need for costly aerial photography using planes.

The target resolution and target magnification rate (2.5m and $4 \times$ respectively) is also substantially higher than for the ProbaV dataset (100m and $3 \times$) making this dataset more challenging for super-resolution as there are more fine details to be resolved.

Table 3.: The effect of various super-resolution methods on the accuracy and perceptual metrics. The bolded values indicate the best performing methods.

	PSNR	SSIM	LPIPS
Bicubic Upsampling	19.1	0.30	0.34
DeepSum	20.1	0.31	0.35
ESRGAN	19.5	0.22	0.21
DeepSum+ESRGAN	19.7	0.25	0.20

Finally, it contains substantially more scenes than the ProbaV dataset, which contains 1460 scenes.

In our experiments with super-resolution, the dataset was used to train DeepSum [18], a convolutional neural network for super-resolution, as well as ESRGAN, a generative-adversarial network approach for super-resolution. Figure 6 illustrates some results. Row 1 contains up-sampled satellite images obtained using bicubic upsampling, a common traditional method used in image upsampling, row 2 is the output of super-resolution using ESRGAN, row 3 is generated using DeepSum, and finally, row 4 is a fusion method combining the output of DeepSum and ESRGAN and row 5 contains the high-resolution ground truth from the aerial photography. Table 3 is a summary of the peak signal to noise ratio (PSNR), Structural Similarity Index (SSIM), and Learned Perceptual Image Patch Similarity (LPIPS) between the aerial photography and the super-resolution output on this dataset. Considering the results in the table and the example imagery provided, it appears that the hybrid approach provides the best trade-off between classic metrics and perceptual quality. Additional details, an ablation study, and further results are available in [4] and [5]. The codes used to generate the results as well as the codes used to generate the image set are available at <https://github.com/danbull1/superresolution>

4. Additional Collated Datasets

One of the challenges for environmental researchers is finding a list of available datasets that may be pertinent to answer the research questions relevant for their research. To address this, we have begun to curate a repository of New Zealand-specific environmental data provided by our data partners. At the time of writing this paper, the TAIAO environmental dataset repository contains links to, or mirrors, over 30 datasets of various types, including images, videos, textual data, and multi-variate tabular data.

A partial list of the datasets available through the platform, including the original owners of the datasets, is given below; note that these datasets are not prepared by us and ownership of these datasets belongs to our data partners. In addition, some of these datasets may require credentialed access.

- Aggregated wave and atmospheric forecast derived from GFS guidance & Hindcast and forecast wave and atmosphere model - MetOcean [16]
- Automatic weather station & Nowcast precipitation, temperature and barometric for Ardmore, Ashburton, Auckland, Birchwood, Flat Hills, and Haast - MetServices [17]
- Lightning feed & archive of lightning records for Blitzen, GPATS Oceania, and TOA - MetOcean [16]

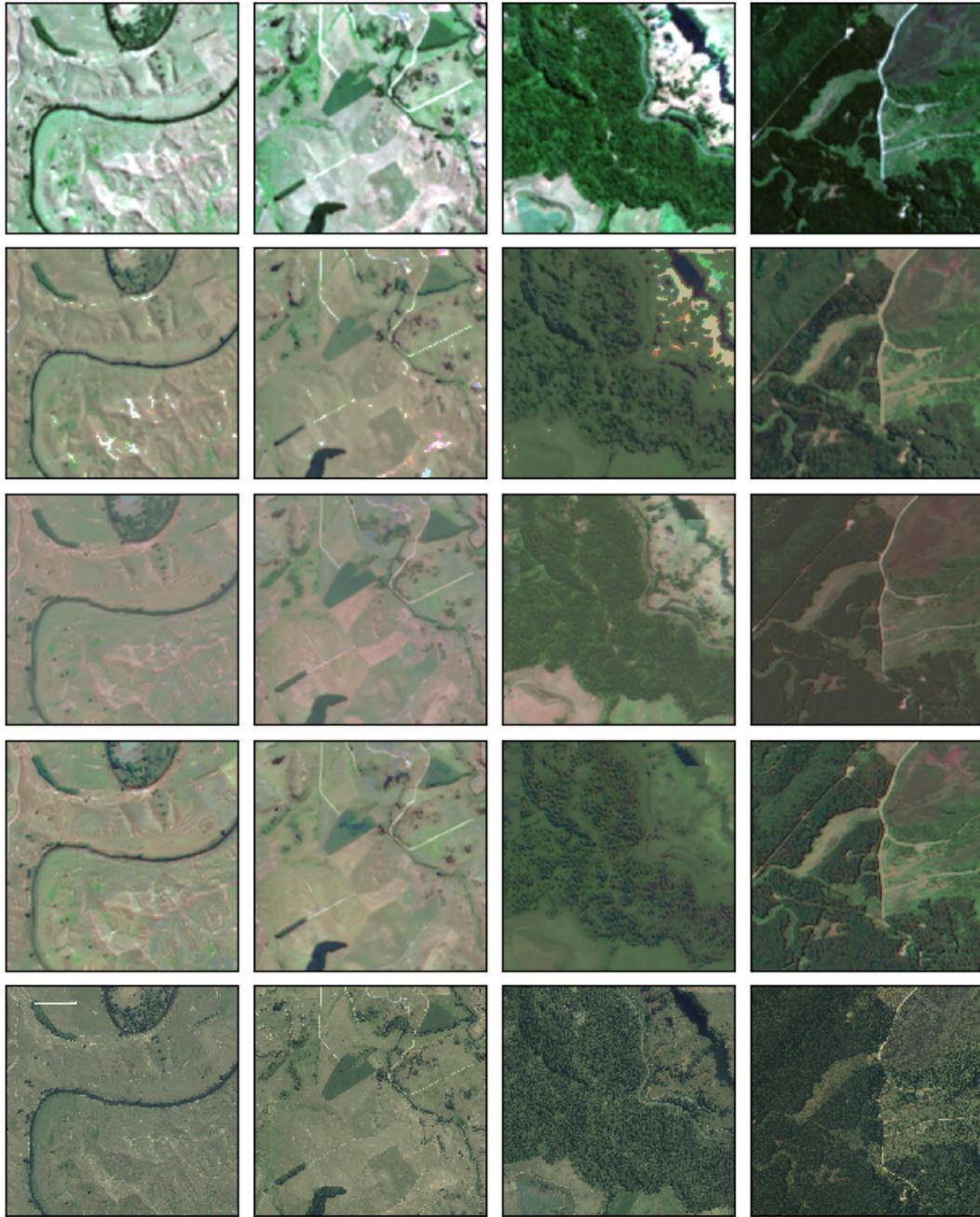


Figure 6.: Result of the super-resolution of the satellite images. Row 1 contains up-sampled satellite images obtained using bicubic upsampling. Row 2 is the output of super-resolution using ESRGAN. Row 3 is the output of super-resolution using DeepSum. Row 4 is generated using a fusion method combining the output of DeepSum and ESRGAN. Row 5 is the high-resolution ground truth from the aerial photography.

- Time series of rivers and rain gauges in Coromandel Peninsular dated from 2010-2020 at 5 minutes resolution - Waikato Regional Council [29]
- NASA Global topography (elevation, slope, and aspect) for the New Zealand region [20]
- Himawari-8 500m and 2km half-disc archive [2]
- Full resolution archive of Himawari-8 satellite from 2020 [2]
- Multispectral satellite data from Landsat 8 OLI sensor [30]
- LILA Wellington Camera Trap [1]
- Moana New Zealand Hydrodynamics Re-analysis v1.9 - MetOcean [16]
- Archive of MetServices Doppler radar network - MetService & MetOcean [16]
- Regional Council Water quality and Discharge - Hawke's Bay Regional Council [10]
- Tropical cyclone trajectory archive - MetOcean [16]
- PM10 and PM100 Air Quality particulate for Horizon and Hawke's Bay - Hawke's Bay Regional Council [10]
- Historical precipitation forecast of the Coromandel region from 2015-2018 - Waikato Regional Council [21]

5. Conclusion

In this paper, we have detailed three image datasets provided by the TAI AO project on environmental data science in New Zealand. The first dataset contains images of small invasive mammals taken in challenging conditions. The second dataset contains aerial photography of the Waikato region in New Zealand in which stands of kahikatea (a native New Zealand tree) have been marked up using manual segmentation. The third consists of paired aerial photography and satellite imagery that is spatially and temporally consistent. We hope that these datasets help in the evaluation of machine learning algorithms to determine their applicability in real-world scenarios.

As the TAI AO project progresses, we plan to extend the platform to provide greater coverage of the set of relevant available data. We also plan to increase the number of notebooks and include more complex, cross-domain, multi-dataset examples showing the application of cross-domain machine learning.

References

- [1] V. Anton, S. Hartley, A. Geldenhuis, and H. U. Wittmer. Monitoring the mammalian fauna of urban areas using remote cameras and citizen science. *Journal of Urban Ecology*, 4(1):juy002, 2018.
- [2] K. Bessho, K. Date, M. Hayashi, A. Ikeda, T. Imai, H. Inoue, Y. Kumagai, T. Miyakawa, H. Murata, T. Ohno, et al. An introduction to himawari-8/9—japan's new-generation geostationary meteorological satellites. *Journal of the Meteorological Society of Japan. Ser. II*, 94(2):151–183, 2016.
- [3] D. Bull. Paired Waikato Region Aerial Photography and Sentinel-2 images. <https://doi.org/10.5281/zenodo.5296368>, Aug. 2021.
- [4] D. Bull. Super-resolution of satellite images. Master's thesis, University of Waikato, University of Waikato, 2021.
- [5] D. Bull, N. Lim, and E. Frank. Perceptual improvements for super-resolution of satellite imagery. In *2021 36th International Conference on Image and Vision*

- Computing New Zealand (IVCNZ), pages 1–6, 2021.
- [6] M. Caron, H. Touvron, I. Misra, H. Jégou, J. Mairal, P. Bojanowski, and A. Joulin. Emerging properties in self-supervised vision transformers. In Proceedings of the International Conference on Computer Vision (ICCV), 2021.
 - [7] G. Cheng, J. Han, and X. Lu. Remote sensing image scene classification: Benchmark and state of the art. Proceedings of the IEEE, 105(10):1865–1883, Oct 2017.
 - [8] European Space Agency. SNAP - ESA’s SentiNel Application Platform. <https://step.esa.int/main/download/snap-download/>, 2020. Accessed: 2020-10-02.
 - [9] K. Gibert, J. S. Horsburgh, I. N. Athanasiadis, and G. Holmes. Environmental data science. Environmental Modelling & Software, 106:4–12, 2018. Special Issue on Environmental Data Science. Applications to Air quality and Water cycle.
 - [10] Hawkes Bay Regional Council. Hawkes Bay Regional Council, Te Kaunihera ā Rohe o Matau-a-Māui. Retrieved from <https://www.waikatoregion.govt.nz/services/regional-services/river-levels-and-rainfall/rainfall-latest-reading/>, 2022.
 - [11] Y. Jia. Kahikatea dataset with human-annotated true explanations for positive examples. <https://doi.org/10.5281/zenodo.5059769>, July 2021.
 - [12] Y. Jia, E. Frank, B. Pfahringer, A. Bifet, and N. Lim. Studying and exploiting the relationship between model accuracy and explanation quality. European Conference on Machine Learning & Principles & Practice of Knowledge Discovery in Databases, 2021.
 - [13] L. H. Kaiser and W. S. A. Saunders. Vision mātauranga research directions: opportunities for iwi and hapū management plans. Kōtuitui: New Zealand Journal of Social Sciences Online, pages 1–13, 2021.
 - [14] N. Lim. Karioi Predator Camera Trap. <https://doi.org/10.5281/zenodo.5276593>, Aug. 2021.
 - [15] E. Maggiori, Y. Tarabalka, G. Charpiat, and P. Alliez. Can semantic labeling methods generalize to any city? the inria aerial image labeling benchmark. In IEEE International Geoscience and Remote Sensing Symposium (IGARSS). IEEE, 2017.
 - [16] MetOcean. MetOcean Solutions. Retrieved from <https://metocean.co.nz/metocean-api>, 2022.
 - [17] MetService. MetService, Te Ratonga Tiorangi. Retrieved from <https://metservice.com>, 2022.
 - [18] A. B. Molini, D. Valsesia, G. Fracastoro, and E. Magli. DeepSUM: Deep neural network for super-resolution of unregistered multitemporal images. IEEE Transactions on Geoscience and Remote Sensing, 58(5):3644–3656, 2019.
 - [19] M. Märten, D. Izzo, A. Krzic, and D. Cox. Super-resolution of proba-v images using convolutional neural networks. Astrodynamic, 8 2019.
 - [20] NASA JPL. NASADEM. Merged DEM Global 1 arc second V001. 2020, distributed by NASA EOSDIS Land Processes DAAC,, 2021.
 - [21] National Centers for Environmental Prediction, National Weather Service, NOAA, U.S. Department of Commerce. Ncep gfs 0.25 degree global forecast grids historical archive, 2015.
 - [22] T. O’Brien. An annotated bibliography of camera trap literature: 1991-2013. Wildlife Conservation Society Working Paper, 2016.
 - [23] L. E. Pardo, S. P. Bombaci, S. Huebner, M. J. Somers, H. Fritz, C. Downs, A. Guthmann, R. S. Hetem, M. Keith, A. le Roux, N. Mgqatsa, C. Packer, M. S. Palmer, D. M. Parker, M. Peel, R. Slotow, W. M. Strauss, L. Swanepoel, C. Tambling, N. Tsie, M. Vermeulen, M. Willi, D. S. Jachowski, and J. A. Venter. Snapshot safari: A large-scale collaborative to monitor africa’s remarkable biodiversity. South

- African Journal of Science, 117(1/2), Jan. 2021.
- [24] T. M. Raraunga. Principles of Māori Data Sovereignty. Retrieved from <https://www.temanararaunga.maori.nz/nga-rauemi>, 2018.
- [25] TAI AO. TAI AO, Time-Evolving Data Science / Artificial Intelligence for Advanced Open Environmental Science. Retrieved from <https://taiao.ai>, 2020.
- [26] Toitū Te Whenua Land Information New Zealand. WRAPs aerial photography. <https://data.linz.govt.nz/layer/104600-waikato-03m-rural-aerial-photos-2016-2019/>, 2017. Accessed: 2020-10-02.
- [27] Toitū Te Whenua Land Information New Zealand. LCDB v4.1 (Deprecated) - Land Cover Database version 4.1, Mainland New Zealand, 2015.
- [28] D. Tsipras, S. Santurkar, L. Engstrom, A. Ilyas, and A. Madry. From imagenet to image classification: Contextualizing progress on benchmarks. Thirty-eighth International Conference on Machine Learning, 2020.
- [29] Waikato Regional Council. Waikato Regional Council, Te Kaunihera ā Rohe o Waikato. Retrieved from <https://www.waikatoregion.govt.nz/services/regional-services/river-levels-and-rainfall/rainfall-latest-reading/>, 2022.
- [30] M. A. Wulder, T. R. Loveland, D. P. Roy, C. J. Crawford, J. G. Masek, C. E. Woodcock, R. G. Allen, M. C. Anderson, A. S. Belward, W. B. Cohen, J. Dwyer, A. Erb, F. Gao, P. Griffiths, D. Helder, T. Hermosilla, J. D. Hipple, P. Hostert, M. J. Hughes, J. Huntington, D. M. Johnson, R. Kennedy, A. Kilic, Z. Li, L. Lyburner, J. McCorkel, N. Pahlevan, T. A. Scambos, C. Schaaf, J. R. Schott, Y. Sheng, J. Storey, E. Vermote, J. Vogelmann, J. C. White, R. H. Wynne, and Z. Zhu. Current status of landsat program, science, and applications. Remote Sensing of Environment, 225:127–147, 2019.