Deep learning techniques for observing the impact of the global warming from satellite images of water-bodies

Rajdeep Chatterjee1 · Ankita Chatterjee2 · SK Hafizul Islam3

Received: 11 June 2021 / Revised: 1 September 2021 / Accepted: 14 December 2021
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract
Global warming is a threat to modern human civilization. There are different reasons for speed up the global average temperature. The consequences are catastrophic for human existence. Seafloor rise, drought, flood, wildfire, dry riverbed are some of the consequences. This paper analyzes the changes in boundaries of different water bodies such as fresh-water lakes and glacial lakes. Over time, the area covered by a water body has been varied due to human interventions or natural causes. Here, variants of Detectron2 instance segmentation architectures have been employed to detect a water-body and compute the changes in its area from the time-lapsed images captured over 32 years, that is, 1984 to 2016. The models are validated using water-bodies images taken by the Sentinel-2 Satellite and compared based on the average precision (AP), 99.95 and 94.51 at $AP_{50}$ and $AP_{75}$ metrics, respectively. In addition, an ensemble approach has also been introduced for the efficient identification of shrinkage or expansion of water bodies.

Keywords Deep learning · Detectron2 · Global warming · Instance segmentation · Water-body

1 Introduction
Climate change is a real threat to the economic and existential aspects of human civilization. Both nature and human-driven activities cause global warming, and it incurs changes in the weather pattern over a more extended period. According to NASA, “Global warming is the long-term heating of Earth’s climate system observed since the pre-industrial period (between 1850 and 1900) due to human activities, primarily fossil fuel burning, which increases heat-trapping greenhouse gas levels in Earth’s atmosphere” [1, 2]. An increase

---

1 School of Computer Engineering, KIIT Deemed to be University, Bhubaneswar, Odisha 751024, India
2 School of Electrical Sciences, Indian Institute of Technology Bhubaneswar, Khordha 752050, Odisha, India
3 Department of Computer Science and Engineering, Indian Institute of Information Technology Kalyani, West Bengal 741235, India

Published online: 06 January 2022
in the global average temperature triggers a speedy meltdown of glaciers and ice caps across the planet. An estimate suggests that the people who reside in the low line areas near the oceans could be impacted most in the next 50 years. Over the last 800,000 years, there have been ice ages and warmer inter-glacial periods. After the last ice age 20,000 years ago, the average global temperature rose by about 3°C to 8°C, over about 10,000 years. Scientist says that the mother earth goes through the cooling and warming process. However, industrialization in the last two centuries, massive fossil fuel consumption, and deforestation are prominent reasons for the acceleration in global warming. Nature fights back in different ways when Earth’s dominant species exploit it, that is, humans. The outbreak of Ebola and COVID-19 are few recent consequences of human transgressions over natural animal habitats. Irregularity in monsoon impacts the overall crop cycles and eventually hampers the livelihood of millions of farmers. The environmental effects of climate change are broad and far-reaching, affecting oceans, ice, and weather. Changes may occur gradually or rapidly. Another consequence of global warming is either shrinking of rivers and lakes due to increased heat or expanding water-bodies due to the glacial melt [3]. Usually, rivers and lakes are the prime sources of freshwater.

It is observed that the size of a water-body (that is, the area covered by a water-body) changes over a more extended period. Human experts can track it based on satellite images. As the number of such water bodies is massive, an Artificial Intelligence-based approach to automatically detect and monitor a water body’s changes in shape or size from the satellite images is a well-suited and efficient solution. Specifically, the latest Detectron2 [4] Feature Pyramid Network (FPN) architecture with different backbone networks has been used to build the proposed model from Sentinel-2 Satellite water-bodies images. Our main objective is to develop an approach to quantify any water-body changes in shape/size over time. It can be used to monitor the natural or illegal encroachment around a water body. Besides, an ensemble, a model fusion, has been implemented compared to the baseline ResNet50 and ResNet101 backbone networks. The proposed approach provides a more reliable and consistent instance segmentation (that is, detecting a water-body).

The paper is divided into six sections. Few recent and related research articles are discussed in Section 2. In Section 3, the main deep learning architecture has been mentioned. Section 4 describes the used dataset and explains the experiments. The results analysis and the necessary visual presentation are given in Section 5. Finally, the paper has been concluded in Section 6.

2 Related work

Deep learning-based approaches are used for both the semantic and the instance segmentation on the satellite images. Here, few recent and relevant research articles are discussed in brief. In [5], authors have implemented different data augmentation techniques to enlarge both the training and test datasets. A ResNet34 network has been used in the 2D UNet encoder part. The proposed model works well on the wildfire areal datasets. The authors have proposed an attention dilation-LinkNet (AD-LinkNet) neural network. It helps to find the contextual information from the satellite image using an adopting encoder-decoder, dilated convolution, channel-based attention mechanism for semantic segmentation. The results provide multi-scale features for different scaled objects in an image. Their model outperforms the D-LinkNet on the DeepGlobe road extraction competition dataset. In [6], a new hybrid framework has been proposed by the author for semantic image segmentation.
They have used the traditional $k$-means algorithm to segment an image into the $k$ region of interests (ROI) based on pixel similarity. Then their SegNet encoder-decoder deep neural network uses the ROIs for training and tuned feature-map extraction. Finally, a simple multi-layer perceptron network with a backpropagation algorithm is used as a classifier. Their proposed approach has been examined on the aerial images taken from the United States Geological Survey and DeepGlobe datasets. An automatic water-body segmentation has been discussed in [7]. Authors have proposed a restricted receptive field deconvolution network (RRF DeconvNet) to extract a semantic segment of water bodies from areal images. It gives better results than its VGG and ResNet18 variants. One recent research article [8] showcases Mask R-CNN-based instance segmentation for water-bodies detection from satellite images. The authors have discussed the purpose of such a study for automatic flood monitoring. Besides, Lu et al. have introduced a new approach CO-attention Siamese Network (COSNet), to address the object segmentation task [9]. Again in [10], the authors have proposed object segmentation with episodic graph memory networks. However, in both the papers, the focus is on non-areal objects in videos, most human activities, and otherwise. In [11], the authors have proposed a real-time segmentation network, that is, YOLACT. Furthermore, the same authors have also provided in the subsequent year an improved version of YOLACT, that is, YOLACT++ [12]. However, these variants are fast in frame per second but less accurate than the widely accepted Mask R-CNN.

The focus is on instance segmentation of static images rather than video object segmentation in this paper. Therefore, accuracy is more critical than FPS in our study. There is a fundamental difference between semantic and instance segmentation. Let us take an example to understand the difference. Assume an image containing three cats in it. The semantic segmentation gives us two segments- one segment containing all the cats, and the remaining portion has been segmented as background. On the other hand, the three cats are segmented separately in the instance segmentation without the background. Here, the proposed models are implemented on the aerial images for water-body (single object) detection. Most of the images in the used dataset contain only a single water body.

The novelty of this paper is to segment the water bodies captured from the satellite/aerial high-resolution color images with high accuracy using the latest state-of-the-art Detectron2 framework. It is already established that the FPN-based ResNet backbone network provides the best speed versus accuracy trade-off over the tradition RPN+FCN-based Mask R-CNN on COCO dataset [13]. Again, an ensemble instance segmentation framework has been proposed to analyze the change in a water-body shape over a more extended time. It provides efficient and automatic monitoring of shrinkage and expansion of lakes and wetlands across the globe. Thus, it can help the administrators and policy-makers monitor the water-bodies in the real world without human intervention.

3 Background

Detectron2 [14] architecture has been used to implement the Mask R-CNN with Feature Pyramidal Network (FPN), which is a pre-trained-based model in this paper [15]. Detectron2 is a Facebook Artificial Intelligence Research (FAIR) open-source platform for the object detection and segmentation [16, 17]. The complete package contains different algorithms for object detection but with improved speed and scalability. It is better than its first adaptation, that is, Detectron, based on multiple metrics. Distributing training on multiple GPUs is one of the major changes. Detectron2 has been built using an open-source PyTorch
The main novelty of Detectron2 is the panoptic segmentation [18, 19] which is a combination of semantic and instance segmentation. However, the Mask R-CNN [20] with FPN structure with variants of ResNet backbone networks have been used in our study, for instance, segmentation. The default configuration of Detectron2 is compared with our configured models. Detectron2 has three basic building blocks (refer Fig. 1): (a) Backbone Network, (b) Region Proposal Network, and (c) Region of Interest (ROI) Head. The default framework uses Stochastic Gradient Descent (SGD) with a learning rate (lr)=1e−3 and momentum=0.9. In our variation, it has been modified to learning rate=1e−4. Besides, the newly proposed AdaBelief optimization [21] technique has...
also been employed for comparison. The AdaBelief takes $\text{lr}=1e^{-4}$, betas=$(0.9, 0.999)$, $\text{eps}=1e^{-16}$ and a fixed weight decay. Non-maximum suppression (NMS) has been used to eliminate the overlapping boxes [22, 23]. The model is executed for 10000 iterations in all of our experiments.

The complete experiment in this study has been performed on the Google Colab free online cloud-based Jupyter notebook environment with GPU and other installed necessary dependencies for the Detectron2 package. Three primary ResNet backbone networks: R50_FPN_1x, R50_FPN_3x, and R101_FPN_3x have been used without any change. The other important modified configuration data are as follows:

| Modified Configuration |
|------------------------|
| 1. DATASET OPTION      |
| (a) `cfg.DATASETS.PRECOMPUTED_PROPOSAL.L.TOPK_TRAIN=2500` |
| (b) `cfg.DATASETS.PRECOMPUTED_PROPOSAL.L.TOPK_TEST=1000` |
| 2. SOLVER OPTION       |
| (a) `cfg.SOLVER.MAX_ITERS=10000` |
| (b) `cfg.SOLVER.BASE_LR=1e-4` |
| 3. ROI HEAD OPTION     |
| (a) `cfg.MODEL.ROI_HEAD.NUM_CLASSES=1 (water body)` |
| (b) `cfg.MODEL.ROI_HEAD.BATCH_SIZE_PER_IMAGE=512` |
| (c) `cfg.MODEL.ROI_HEADS.SCORE_THRESH_TEST=0.5` |
| 4. FPN OPTION (default) |

4 Proposed Pipeline and Experiments

A straight-forward pipeline for satellite image segmentation has been given in Fig. 2 in which this study aims at the third step of the pipeline shown in a red bounded box.

4.1 Dataset

The dataset has a total of 2811 images, out of which 1087 images have been selected and annotated for the water-body segmentation. The images have been taken from the Kaggle [24] Sentinel-2 Satellite Images of Water Bodies dataset. The dataset has $2699 \times 3771$ high resolution to $54 \times 44$ low resolution images. A single image contains a single water body. However, the raw input images are resized to $448 \times 448$ before feeding to a segmentation model. In practice, a proper image registration step needs to be implemented before image segmentation. However, the used dataset contains already pre-processed satellite images. Therefore, images can be directly used for model building (training). Besides the mentioned dataset, Google time-lapsed videos are also used to examine the proposed instance
The time-lapsed videos are of: Aral lake\(^1\). The high-resolution frames containing different water bodies across the planet are captures from the publicly available Google time-lapse videos. Each frame corresponds to a single year from 1984 to 2016, that is, a total of 32 years of transformation.

The Microsoft open-source Visual Object Tagging Tool\(^6\) (VoTT) has been used for image annotation and cvat.org for transforming it to Detectron2 COCO compatible format.

### 4.2 Experimental setup

The paper has three separate experiments related to a common problem of water-body instance segmentation.

**Experiment-1:** In this experiment, we have implemented a Detectron2 architecture with different backbone networks on 1087 training-set and 103 validation-set, respectively. In this experiment, mAP, i.e., average precision (AP) \([25]\) for a single class with IoU values 50, 75 and 50 : 95 have been computed respectively.

**Experiment-2:** In this experiment, we have randomly selected four water-body images from the validation set to validate the true positive instance segmentation and its accuracy (%). The objective of this experiment is to achieve segmentation accuracy to be as high as possible (the predicted pixels cover the actual annotated pixels).

**Experiment-3:** In this experiment, Google time-lapsed images of five famous lakes for the period 1984 - 2000 and 2000 - 2016 are used to observe automatically whether the shape of the water-bodies has been increased or decreased over time. The objective

---

\(^1\) Aral Lake: https://www.youtube.com/watch?v=YBz_7qTNCQE (between Kazakhstan and Uzbekistan), Lake Mead\(^2\) (southeast Las Vegas, Nevada, USA), Lake Poopó\(^3\) (in a shallow depression in the Altiplano Mountains, Bolivia), Tibet lake\(^4\) (near Tibet, Himalayan mountain range) and South Dead Sea\(^5\) (a landlocked salt lake between Israel and Jordan)

\(^2\) Lake Mead: https://www.youtube.com/watch?v=6db65Pbbe88

\(^3\) Lake Poopó: https://www.youtube.com/watch?v=73r6Y1-XFMg

\(^4\) Tibet Lake: https://www.youtube.com/watch?v=8kmfu0_gnWw

\(^5\) South Dead Sea: https://www.youtube.com/watch?v=ROmKjgocxgY

\(^6\) VoTT: https://github.com/microsoft/vott
of this experiment is to decide whether the water-body shrinks ↓ or expands ↑ over 16 years.

The implementation pipeline (see Algorithm 1) can be described as follows:

(i) It takes input from the aerial/satellite images or video.
(ii) A suitable pre-processing such as image registration is required to stitch different images captured in the same area from multiple views.
(iii) Detectron2 Mask+FPN ResNet backbone has been employed on the training dataset (annotated images).
(iv) Once the model is built, it has been used to generate a mask (object segment) from the test dataset.
(v) The generated masks are compared with the actual human-annotated polygons.
(vi) Finally, the segmentation accuracy (%), mAP has been computed for a particular model to evaluate its performance.

Similarly, an ensemble pipeline (see Fig. 3) has also been proposed in this paper. The difference is that it uses multiple backbone networks to obtain separate masks for a single input image. After that, it combines the mask either based on pixel-wise intersection or union of them. It is observed that the pipeline is more appropriate for evaluating the year-wise shape changes of a water-body.

5 Results and analysis

Detectron2 architecture with ResNet50 and 101 backbone networks has been implemented in different combinations to examine instance segmentation performances. Two distinct types of optimizers, the Stochastic Gradient Descent (SGD) and the AdaBelief [26], have also been used to verify the quick convergence criteria in this study. These optimizers are employed for loss calculation once with 300 epochs and again with 1000 epochs. From Table 1, we observed that AdaBelief outperforms SGD in the early iterations, but SGD matches with AdaBelief with a more significant epoch number (see in Fig. 4).

The results obtained from R50_FPN_1x, R50_FPN_3x and R101_FPN_3x using the 103 test images and different optimization techniques are given in Tables 2 and 3 for (default) SGD and AdaBelief, respectively. Both the bounded box and the segmentation (polygon) metrics have been computed to evaluate the model’s quality. The used models
have been trained for 10000 epochs. The difference between the SGD-based model and AdaBelief is insignificant for R50_FPN_3x; otherwise, SGD provides noticeable improvements over its AdaBelief counterparts. Another important empirical evidence is the frames per second (FPS) comparison. The overall FPS for SGD-based Detectron2 is significantly higher than the same architecture with AdaBelief optimizer, irrespective of the backbone networks.

Here, R101_FPN_3x has the higher average precision values. \( AP_{0.50:0.95} \) and \( AP_{50} \) for the used test images over the remaining models. However, R50_FPN_3x outperforms others due to its highest FPS rate and smaller model size (313MB). On the other hand, the same backbone network outsunds even R101_FPN_3x in terms of \( AP_{0.50:0.95} \), \( AP_{50} \) and FPS. The empirical analysis can conclude that the ResNet50 using Feature Pyramidal Network (3x) is the best performing backbone network among all the employed variants. Therefore, the rest of the study of this paper has been presented based on R50_FPN_3x (SGD) model.

![Fig. 3 Proposed ensemble model workflow diagram](image_url)
Again, four validation images (wb_3.jpg, wb_104.jpg, wb_1421.jpg, and wb_7234.jpg) out of 103 test image-set have been shown here with its actual image, predicted segment, annotated mask, and predicted mask in Figs. 5, 6, 7 and 8. The purpose of the visual representation is to demonstrate the model’s efficiency. The only observable difference is the edges of the annotated and the predicted segments. This is because the hand annotation is done using polygons (multiple straight lines) while the predicted edges are smoother due to the continuous loss function. A Table 4 has also been given to show the instance segmentation performance of different backbones and optimizers.
segmentation accuracy (%) of the best performing Detectron2 architecture. One can observe that the segmentation accuracies are more than 85% for these randomly chosen test images. It has been validated with other test images as well. Here, it must be noted that

| Backbone     | Method | \(AP_{0.50:0.95}\) | \(AP_{50}\) | \(AP_{75}\) | FPS |
|--------------|--------|----------------------|-------------|-------------|-----|
| R50_FPN_1x  | bbox   | 85.95                | 99.98       | 96.73       | 09.68 |
|              | segm   | 84.52                | 99.98       | 96.74       |      |
| R50_FPN_3x  | bbox   | 85.55                | 99.95       | 96.81       | 10.34 |
|              | segm   | 83.10                | 99.95       | 94.51       |      |
| R101_FPN_3x | bbox   | 87.74                | 99.97       | 96.21       | 02.03 |
|              | segm   | 85.58                | 99.93       | 96.75       |      |

| Backbone     | Method | \(AP_{0.50:0.95}\) | \(AP_{50}\) | \(AP_{75}\) | FPS |
|--------------|--------|----------------------|-------------|-------------|-----|
| R50_FPN_1x  | bbox   | 81.85                | 98.67       | 93.12       | 06.86 |
|              | segm   | 81.07                | 97.72       | 95.02       |      |
| R50_FPN_3x  | bbox   | 85.83                | 97.99       | 93.28       | 06.98 |
|              | segm   | 83.97                | 97.99       | 96.02       |      |
| R101_FPN_3x | bbox   | 80.86                | 99.87       | 94.27       | 01.84 |
|              | segm   | 81.65                | 99.87       | 94.74       |      |

Table 2 Results obtained from Average Precision (AP), \(AP_{50}\) and \(AP_{75}\) using different backbones and SGD optimizer in Detectron2 instance segmentation architecture

Table 3 Results obtained from Average Precision (AP), \(AP_{50}\) and \(AP_{75}\) using different backbones and AdaBelief optimizer in Detectron2 instance segmentation architecture

Fig. 5 Visualization of water_body_3.jpg validation image and its predicted instance segmentation using Detectron2 ResNet50 FPN (3x) model

segmentation accuracy (%) of the best performing Detectron2 architecture. One can observe that the segmentation accuracies are more than 85% for these randomly chosen test images. It has been validated with other test images as well. Here, it must be noted that

Fig. 6 Visualization of water_body_104.jpg validation image and its predicted instance segmentation using Detectron2 ResNet50 FPN (3x) model
the segmented accuracy indicates the ratio between the number of actual pixels within the annotation and the predicted number of pixels within the same annotated area.

An ensemble instance segmentation model has also been examined on the same dataset. The ensemble considers a pixel as valid if and only if all the base segmentation models correctly predict it. An example of the proposed model applied on the Guozha Cuo Lake is shown in Fig. 9. The base instance segmentation models provide an overlapping mask indicated by pink, brown, and green segmentation. Finally, all the primary and proposed

Fig. 7 Visualization of water_body_1421.jpg validation image and its predicted instance segmentation using Detectron2 ResNet50 FPN (3x) model

Fig. 8 Visualization of water_body_7234.jpg validation image and its predicted instance segmentation using Detectron2 ResNet50 FPN (3x) model

Fig. 9 The output of Detectron2 ensemble instance segmentation of the Guozha Cuo Lake; Three models provide three different overlapping masks of pink, brown and green segmentation, respectively
ensemble models (both the union and the intersection models) have been tested on the five popular lake images captured by ESA’s Sentinel-2 Satellite. These are compiled as Google Earth 32 years time-lapsed videos from 1984 to 2016. The detailed results of the study have been provided in Table 5. It contains the model-wise and year-based total number of segmented pixels for each lake. This study aims to demonstrate the efficient implementation of the deep learning approach in automated monitoring of shrinkage or expansion of a

Table 4 Results obtained from randomly selected four validation images (wb_3.jpg, wb_104.jpg, wb_1421.jpg and wb_7234.jpg) using Detectron2 ResNet50 FPN 3x instance segmentation model (values are the predicted number of pixels in a segment)

| File Name | wb_3 | wb_104 | wb_1421 | wb_7234 |
|-----------|------|--------|---------|---------|
| Actua     | 36427| 89115  | 38524   | 100244  |
| Predicted | 35862| 88107  | 34362   | 92555   |
| True Positive | 33088| 86681  | 33797   | 88872   |
| Accuracy (%) | 90.83| 97.27  | 87.69   | 88.66   |

Table 5 Results obtained from Detectron2 instance segmentation architectures (values are the predicted number of pixels in a segment)

| Water-body   | Backbone            | 1984  | 2000  | 2016  | Indicator      |
|--------------|---------------------|-------|-------|-------|----------------|
| Lake Mead    | R50_FPN_1x          | 57702 | 55670 | 39660 | Shrinking ↓    |
|              | R50_FPN_3x          | 55363 | 53452 | 38234 | Shrinking ↓    |
|              | R101_FPN_3x         | 54752 | 51499 | 36530 | Shrinking ↓    |
|              | ENS_Union           | 59300 | 56597 | 41132 | Shrinking ↓    |
|              | ENS_Intersection    | 51696 | 49679 | 34650 | Shrinking ↓    |
| Aral Lake    | R50_FPN_1x          | 81130 | 57213 | 38059 | Shrinking ↓    |
|              | R50_FPN_3x          | 79723 | 56980 | 36928 | Shrinking ↓    |
|              | R101_FPN_3x         | 80805 | 57742 | 33668 | Shrinking ↓    |
|              | ENS_Union           | 83255 | 59977 | 39966 | Shrinking ↓    |
|              | ENS_Intersection    | 77137 | 53558 | 31455 | Shrinking ↓    |
| Lake Poopó   | R50_FPN_1x          | 34819 | 22390 | 22791 | Shrinking ↓    |
|              | R50_FPN_3x          | 34950 | 22083 | 22900 | Shrinking ↓    |
|              | R101_FPN_3x         | 34736 | 20563 | 18764 | Shrinking ↓    |
|              | ENS_Union           | 36110 | 23092 | 23876 | Shrinking ↓    |
|              | ENS_Intersection    | 32644 | 19346 | 18096 | Shrinking ↓    |
| South Dead Sea | R50_FPN_1x        | 33992 | 32448 | 27440 | Shrinking ↓    |
|              | R50_FPN_3x          | 34747 | 32574 | 29481 | Shrinking ↓    |
|              | R101_FPN_3x         | 35386 | 31498 | 28491 | Shrinking ↓    |
|              | ENS_Union           | 36958 | 33031 | 29928 | Shrinking ↓    |
|              | ENS_Intersection    | 32197 | 30347 | 25668 | Shrinking ↓    |
| Tibet Lake   | R50_FPN_1x          | 26517 | 32448 | 27440 | Expanding ↑    |
|              | R50_FPN_3x          | 23180 | 26454 | 31592 | Expanding ↑    |
|              | R101_FPN_3x         | 22480 | 27600 | 32191 | Expanding ↑    |
|              | ENS_Union           | 26960 | 35920 | 35911 | Expanding ↑    |
|              | ENS_Intersection    | 21529 | 23809 | 29285 | Expanding ↑    |
water-body from the satellite images. In Table 5, the area covered by these lakes has been shrinking except the Tibet lake, where the amount of water increases over the same time. Aral lake, Lake Mead, Lake Poopó, and South Dead Sea are drying out mostly due to climate change (natural cause) [27–31]. On the other side, the Tibet lake has been expanding its boundary as it is situated in the Himalayan Mountains range and benefited from the glacial melting [32]. This paper plots two prominent examples (Aral Lake and Tibet Lake) with their actual images, segmented images, and predicted masks in Figs. 10 and 11 for the years 1984, 2000 and 2016 (16 years alternatively), respectively. All the used models reflect the same trends of shrinkage or expansion. There is an exception for Lake Poopó and Tibet lake. However, it is noticed that the proposed intersection-based ensemble model gives a consistent descending and ascending segmentation even for the Lake Poopó and the Tibet lake over the remaining models. A comparison bar plot has also been given to visualize the shrinkage and expansion for all the five lakes in Fig. 12.

A rigorous survey has been done based on articles from NASA, National Geographic and BBC, etc., to validate our observations. It is noticed that the outcome of this paper is the same as their analysis and conclusions. For a detailed understanding, one can refer to these articles [33–35].
Multimedia Tools and Applications

Global warming is a significant issue in the 21st century. The reckless and unplanned industrialization causes immense strains on nature. The effects can be observed through naked eyes. However, sometimes the change happens over a longer time. Modern computer vision techniques can track these drastic changes based on the satellite images of different geographical entities. Here, the satellite images of water bodies have been used to build the prediction model using the proposed ensemble Detectron2 instance segmentation architecture. Finally, the Google time-lapsed satellite images of few famous lakes across the planet have been used to understand the grim reality of global warming. It is observed that the area covered by the water bodies across the globe mostly dried out over 32 years. However, the water-body has been grown over the same time in Tibet. It suggests the meltdown of many Himalayan glaciers in that region. The best performing model out of all the used variants is Mask R-CNN with ResNet50+FPN backbone and SGD optimizer. It provides 99.50, 94.51 and \( \sim 10 \) for SGD over 97.99, 96.02 and \( \sim 7 \) for AdaBelief optimizer at \( AP_{50} \), \( AP_{75} \) and FPS, respectively.

Fig. 11 Visualization of time lapsed images of Tibet Lake (1984/2000/2016) and its predicted instance segmentation using Detectron2 ResNet50 FPN (3x) model (Expanding ↑)
In the future, we can extend our work to multi-spectrum or multi-modal satellite image segmentation. A real-world example focuses on a specific metropolitan area where the proposed pipeline can be employed to monitor and automatically identify different water-bodies changes in a single fiscal year.

References

1. NASA Global Climate Change (2018) Vital signs of the planet. https://climate.nasa.gov/vital-signs/global-temperature/(25.12.2019), accessed: 2020-10-6
2. Bates B, Kundzewicz Z, Wu S (2008) Climate change and water. Intergovernmental Panel on Climate Change Secretariat
3. Wolf M, Mooij, Stephan Hülsmann, Lisette N De Senerpont Domis, Bart A Nolet, Paul LE Bodelier, Paul CM Boers, L Miguel Dionisio Pires, Herman J Gons, Bas W Ibelings, Ruurd Noordhuis, et al. The impact of climate change on lakes in the netherlands: a review. Aquatic Ecology, 39(4):381–400, 2005
4. Wu Y, Kirillov A, Massa F, Lo W-Y (2019) Girshick R Detectron2. https://github.com/facebookresearch/detectron2, accessed: 2020-10-17
5. Ming Wu, Zhang Chuang, Liu Jiaming, Zhou Lichen, Li Xiaqi (2019) Towards accurate high resolution satellite image semantic segmentation. IEEE Access 7:55609–55619
6. Barthakur M, Sarma KK (2020) Deep learning based semantic segmentation applied to satellite image. In: Data Visualization and Knowledge Engineering, pages 79–107. Springer
7. Khryashchev V, Larionov R (2020) Wildfire segmentation on satellite images using deep learning. In: 2020 Moscow Workshop on Electronic and Networking Technologies (MWENT), pages 1–5. IEEE
8. Dhyakesh S, Ashwini A, Supraja S, Aasikaa CMR, Nithesh M, Akshaya J, Vivitha S (2020) Mask r-cnn for instance segmentation of water bodies from satellite image. In: 2nd EAI International Conference on Big Data Innovation for Sustainable Cognitive Computing, pages 301–307. Springer
9. Lu X, Wang W, Shen J, Crandall D, Luo J (2020) Zero-shot video object segmentation with co-attention siamese networks. IEEE transactions on pattern analysis and machine intelligence

Fig. 12 Comparison of Detectron2 ensemble (intersection) instance segmentation over 32 years time-span (1984 to 2016) for few popular lakes across the world (values are the predicted number of pixels (px.) in a segment)
10. Lu X, Wang W, Danelljan M, Zhou T, Shen J, Van Gool L (2020) Video object segmentation with episodic graph memory networks. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part III 16, pages 661–679. Springer

11. Bolya D, Zhou C, Xiao F, Lee YJ (2019) Yolact: Real-time instance segmentation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pages 9157–9166

12. Bolya D, Zhou C, Xiao F, Lee YJ (2019) Yolact++: Better real-time instance segmentation. Preprint at arXiv:1912.06218

13. Detectron2 model zoo and baselines (2020) https://github.com/facebookresearch/detectron2/blob/master/MODEL_ZOO.md. accessed: 2020-10-10

14. Detectron2 (2020) https://research.fb.com/wp-content/uploads/2019/12/4.-detectron2.pdf. accessed: 2020-10-17

15. Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L (2009) Imagenet: A large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition, pages 248–255. IEEE.

16. Minaee S, Boykov Y, Porikli F, Plaza A, Kehtrarnavaz N, Terzopoulos D (2020) Image segmentation using deep learning: A survey. Preprint at arXiv:2001.05566

17. Ghosh Swarnendu, Das Nibaran, Das Ishita, Maulik Ujjwal (2019) Understanding deep learning techniques for image segmentation. ACM Computing Surveys (CSUR) 52(4):1–35

18. Kirillov A, He K, Girshick R, Rother C, Dollár P (2019) Panoptic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 9404–9413

19. Kirillov A, Girshick R, He K, Dollár P (2019) Panoptic feature pyramid networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 6399–6408

20. He K, Gkioxari G, Dollár P, Girshick R (2017) Mask r-cnn. In: Proceedings of the IEEE international conference on computer vision, pages 2961–2969

21. Zhuang J, Tang T, Ding Y, Tatikonda S, Dvornek N, Papademetris X, Duncan J (2020) Adabelief optimizer: Adapting stepsizes by the belief in observed gradients. Conference on Neural Information Processing Systems

22. Hosang J, Benenson R, Schiele B (2017) Learning non-maximum suppression. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pages 4507–4515

23. Revaud J, Almazán J, Rezende RS, Roberto de Souza C (2019) Learning with average precision: Training image retrieval with a listwise loss. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pages 5107–5116

24. Kaggle sentinel-2 satellite images of water bodies (2020) https://www.kaggle.com/franciscoescobar/satellite-images-of-water-bodies accessed: 2020-10-6

25. Padilla R, Netto SL, da Silva EAB (2020) A survey on performance metrics for object-detection algorithms. In: 2020 International Conference on Systems, Signals and Image Processing (IWSSIP), pages 237–242. IEEE

26. Adabelief optimizer: Pytorch implementation (2020) https://github.com/juntang-zhuang/Adabelief-Optimizer. accessed: 2020-11-21

27. Shrinking aral sea (2018) https://earthobservatory.nasa.gov/world-of-change/aral_sea.php. accessed: 2020-10-17

28. Landsat top ten - a shrinking sea, aral sea (2016) https://www.bbc.com/news/world-middle-east-36477284. accessed: 2020-10-17

29. Lake mead still shrinking (2014) https://earthobservatory.nasa.gov/images/84105/lake-mead-still-shrin king. accessed: 2020-10-17

30. Bolivias lake poopó disappears (2016) https://earthobservatory.nasa.gov/images/87363/bolivias-lake-poopo-disappears. accessed: 2020-10-17

31. Dead sea drying: A new low-point for earth (2016) https://www.bbc.com/news/world-middle-east-36477284. accessed: 2020-10-17

32. Yiwei Chen, Yongqiang Zong, Bo Li, Shenghua Li, and Jonathan C Aitchison. Shrinking lakes in tibet linked to the weakening asian monsoon in the past 8.2 ka. Quaternary Research, 80(2):189–198, 2013

33. Drying lakes climate change global warming drought (2017) https://www.nationalgeographic.com/magazine/article/drying-lakes-climate-change-global-warming-drought. accessed: 2020-10-17

34. Michael Gross (2017) The worlds vanishing lakes

35. Dan H Shugar, Aaron Burr, Umesh K Haritashya, Jeffrey S Kargel, C Scott Watson, Maureen C Kennedy, Alexandre R Bevington, Richard A Betts, Stephan Harrison, and Katherine Stratman. Rapid worldwide growth of glacial lakes since 1990. Nature Climate Change, 10(10):939–945, 2020

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.