Environmental Modelling and Software 149 (2022) 105333

Contents lists available at ScienceDirect

Environmental Modelling and Software

journal homepage: www.elsevier.com/locate/envsoft

ATLANTIS: A benchmark for semantic segmentation of waterbody images

Seyed Mohammad Hassan Erfani [a], Zhenyao Wu [b], Xinyi Wu [b], Song Wang [b], Erfan Goharian [a,*]

[a] University of South Carolina, 300 Main Street, Columbia, 29201, SC, USA
[b] University of South Carolina, 550 Assembly Street, Columbia, 29201, SC, USA

ARTICLE INFO

Keywords:
Water resources engineering
Computer vision
Waterbody image dataset
Semantic segmentation
Deep learning models

ABSTRACT

Vision-based semantic segmentation of waterbodies and nearby related objects provides important information for managing water resources and handling flooding emergency. However, the lack of large-scale labeled training and testing datasets for water-related categories prevents researchers from studying water-related issues in the computer vision field. To tackle this problem, we present ATLANTIS, a new benchmark for semantic segmentation of waterbodies and related objects. ATLANTIS consists of 5,195 images of waterbodies, as well as high quality pixel-level manual annotations of 56 classes of objects, including 17 classes of man-made objects, 18 classes of natural objects and 21 general classes. We analyze ATLANTIS in detail and evaluate several state-of-the-art semantic segmentation networks on our benchmark. In addition, a novel deep neural network, AQUANet, is developed for waterbody semantic segmentation by processing the aquatic and non-aquatic regions in two different paths. AQUANet also incorporates low-level feature modulation and cross-path modulation for enhancing feature representation. Experimental results show that the proposed AQUANet outperforms other state-of-the-art semantic segmentation networks on ATLANTIS. We claim that ATLANTIS is the largest waterbody image dataset for semantic segmentation providing a wide range of water and water-related classes and it will benefit researchers of both computer vision and water resources engineering.

1. Introduction

Every year, floods claim tens of billions of US dollars losses and thousands of lives globally Hirabayashi et al. (2013). Accurate detection, measurement, and track of the waterbodies can help both the public and decision-makers to take appropriate actions to minimize the risk and losses Huang et al. (2018); Gebrehiwot et al. (2019). With the popularity of smart phones and airborne imagery, various data at flooding sites can be collected rapidly and continuously to provide more useful and heterogeneous information source Eltner et al., 2021; Hosseiny (2021); Eltner et al. (2018); Cervone et al. (2016); Schnebele and Cervone (2013), compared to the conventional gauge sensing and remote sensing Lo et al. (2015). As a fundamental step to leverage such collected images for modeling and decision-making, we need first to conduct a refined semantic segmentation of included waterbodies and related objects in such scenes, which we focus on in this paper.

With the advancement of deep neural networks, semantic segmentation has achieved a great success in recent years on various kinds of images, such as natural images Lin et al. (2014), street images Cordts et al. (2016); Yu et al. (2020), and medical images Bernal et al. (2017); Jha et al. (2020). These successes have motivated the application of deep learning across a wide range of disciplines Razavi (2021). However, waterbody images pose many new unique challenges for semantic segmentation. In some forms, water preserves the intrinsic properties such as reflection, transparency, shapeless and colorless visual features; which in turn, brings difficulties to semantic segmentation of water and related objects. Moreover, in some other forms, these properties can be affected by illumination sources from surroundings, turbidity, and turbulence. Different-labeled waterbodies, such as river and canal, or lake and reservoir, often have similar visual characteristics that make task of semantic segmentation even harder. As shown in our later experiments, these unique challenges may significantly affect the performance of the existing semantic segmentation networks.

Meanwhile, deep-learning based approaches always require large-scale training data with necessary ground-truth annotations. Lack of such a public dataset for waterbody segmentation significantly impedes the research on this problem. The collection and annotation of such a dataset can be very laborious and time-consuming to cover a wide range

* Corresponding author.
E-mail addresses: serfani@email.sc.edu (S.M.H. Erfani), zhenyao@email.sc.edu (Z. Wu), xinyiw@email.sc.edu (X. Wu), songwang@cec.sc.edu (S. Wang), goharian@cec.sc.edu (E. Goharian).

https://doi.org/10.1016/j.envsoft.2022.105333
Received 11 December 2021; Received in revised form 18 January 2022; Accepted 20 January 2022
Available online 24 January 2022

of waterbodies and related objects. There is no specific repository providing relevant images. In addition, team members and annotators are required to have prior knowledge on water resources engineering to be capable of selecting and precisely annotating the images.

In this paper, we present a new benchmark, ATLANTIS (ArTificiaL And Natural waTer-bodIes dataSet). For the first time, this dataset has covered a wide range of natural and man-made (artificial) waterbodies such as sea, lake, river, canal, reservoir, and dam. ATLANTIS includes 5,195 pixel-wise annotated images split to 3,364 training, 535 validation and 1,296 testing images. As shown in Table 1, in addition to 35 waterbody and water-related objects, ATLANTIS also covers 21 general labels. Moreover, we construct ATLANTIS Texture (ATeX) dataset, which consists of 12,503 patches for the water-bodies texture classification, sampled from 15 kinds of waterbodies in ATLANTIS.

In order to tackle the inherent challenges in the segmentation of waterbodies, AQUANet is developed which takes an advantage of two different paths to process the aquatic and non-aquatic regions, separately. Each path includes low-level feature and cross-path modulation, to adjust features for better representation. The results show that the proposed AQUANet outperforms other ten state-of-the-art semantic-segmentation networks on ATLANTIS, and the ablation studies justify the effectiveness of the components of the proposed AQUANet.

2. Related work

2.1. Semantic segmentation dataset

Large-scale annotated datasets, such as COCO Lin et al. (2014), PASCAL Context Mottaghi, Chen, Liu, Cho, Lee, Fidler, Urtasun and Yuille (2014), ADE20K Zhou et al. (2019), Mapillary Vistas Dataset Neuhold, Ollmann, Rota Bulo and Kontschieder (2017) and BDD100K Yu et al. (2020), make it possible for researchers to develop deep-learning based models for real-world applications. Considering the most related dataset to ATLANTIS, Gebrehiwot et al. (2019) collected a small number of top-view waterbody dataset (100 images) using Unmanned Aerial Vehicles (UAVs) which contains only four categories (i.e., water, building, vegetation and road). In addition, Sazara et al. (2019) introduced a larger dataset (253 images) which just focuses on flood region segmentation. More recently, Sarp et al. (2020) provided a dataset that consists of 441 annotated roadway flood images. More recently, Pally and Samadi (2021) built a flood dataset consisting of >9000 flooding images collected from various sources such as Twitter, US department of Transportation (DOT), US Geological Survey (USGS) river cameras and YouTube. However, these datasets have limitations in either the number of annotated images, or the categories they covered, and none of those considers more complex classes of waterbodies such as sea, lake and waterfall. Therefore, we develop a new dataset, ATLANTIS, as the first large-scale annotated dataset to provide a wide-range of waterbodies and water-related objects.

Table 1
List of the ATLANTIS labels.

Artificial	breakwater; bridge; canal; culvert; dam; ditch; levee; lighthouse; offshore platform; pier; pipeline; reservoir; ship; spillway; swimming pool; water tower; water well.
Natural	cliff; cypress tree; fjord; flood; glaciers; hot spring; lake; mangrove; marsh; puddle; rapids; river; river delta; sea; shoreline; snow; waterfall; wetland.
General	road; sidewalk; building; wall; fence; pole; traffic sign; vegetation; terrain; sky; train; person; car; bus; truck; bicycle; parking meter; motorcycle; fire hydrant; boat; umbrella.

2.2. Semantic segmentation network

All of existing semantic segmentation approaches share the same goal to classify each pixel of a given image but differ in the network design, including low-resolution representations learning Long et al. (2015); Chen, Papandreou, Kokkinos, Murphy and Yuille (2017a), high-resolution representations recovering Badrinarayanan et al. (2015); Noh et al. (2015); Lin et al. (2017), contextual aggregation schemes Yuan and Wang (2018); Zhao et al. (2017); Yuan et al. (2020), feature fusion and refinement strategy Lin et al. (2017); Huang et al. (2019); Li, Zhong, Wu, Yang, Lin and Liu (2019a); Zhu et al. (2019); Fu et al. (2019). Typically, method designs are dependent on their respective datasets and all the mentioned networks are developed by training on benchmark datasets such as Cityscapes Cordts et al. (2016), COCO Lin et al. (2014) and VOC Everingham et al. (2010) where the inter-class boundary is clear even for the within-group categories (e.g., car and truck). As mentioned above, waterbody images pose new challenges to semantic segmentation. Previous works on waterbody segmentation mainly use satellite imagery Muñoz et al. (2021); Li, Wang, Zhang, Hu and Meng (2019b); Duan and Hu (2019). In this work, we focus on natural waterbody images terrestrially captured by various cameras, and design AQUANet, a new two-path semantic segmentation network, by including an aquatic branch explicitly for waterbody classes.

3. ATLANTIS dataset

The ATLANTIS dataset is designed and developed with the goal of capturing a wide-range of water-related objects, either those exist in natural environment or the infrastructure and man-made (artificial) water systems (see Fig. 1). In this dataset, labels were first selected based on the most frequent objects, used in water-related studies or can be found in real-world scenes. Aside from the background objects, total of 56 labels, including 17 artificial, 18 natural waterbodies, and 21 general labels, are selected (Table 1). These general labels are considered for providing contextual information that most likely can be found in water-related scenes. After finalizing the selection of waterbody labels, a comprehensive investigation on each individual label was performed by annotators to make sure all the labels are vivid examples of those objects in real-world. Moreover, sometimes some of the water-related labels, e.g., levee, embankment, and flood bank, have been used interchangeably in water resources field; thus, those labels are either merged into a unique group or are removed from the dataset to prevent an individual object receives different labels.

In order to gather a corpus of images, we have used Flickr API to query and to collect "medium-sized" unique images for each label based on eight commonly used "Creative Commons", "No Known Copyright Restrictions" and "United States Government Work" licenses. Downloaded images were then filtered by a two-stage hierarchical procedure. In the first stage, each annotator was assigned to review a specific list of labels and remove irrelevant images based on that specific list of labels. In the second stage, several meetings were held between the entire annotation team and the project coordinator to finalize the images which appropriately represent each of 56 labels.

This sieving procedure has been applied four times in order to meet the limit and to reach the current number of images. The percentage of image acceptance rate for the third and fourth phases are 14.41% and 5.06%, respectively. It means if we want to add 1000 more images to the dataset, we should process at least 20,000 images. Finally, images were annotated by annotators who have solid water resources engineering background as well as experience working with the CVAT Sekachev, Manovich, Zhiltsov, Zhavoronkov, Kalinin, Hoff, TOsmanov, Kruchinin, Zankevich, DmitriySidnev, Markelov, Johannes222, Chenuet, a andre, telenachos, Melnikov, Kim, Ilouz, Glazov, Priya4607, Tehrani, Jeong, Skubriev, Yonekura, vugia truong, zliang7, lizhming and Truong (2020), which is a free, open source, and web-based image/video annotation

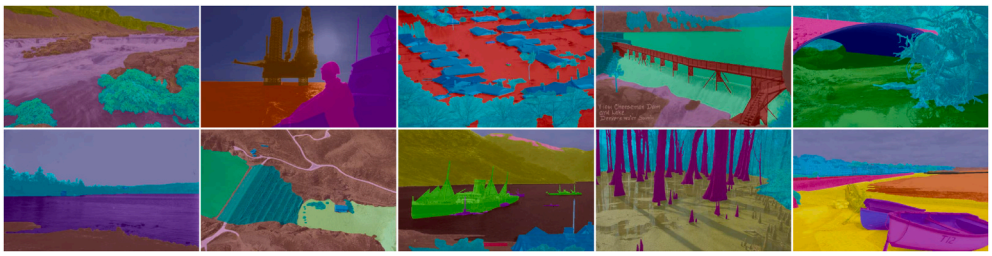

Fig. 1. ATLANTIS, ArTificiaL And Natural waTer-bodIes dataSet.

tool.

3.1. Dataset statistics

Fig. 2 shows the frequency distribution of the number of images and the percentage of pixels for waterbody labels. Such a long-tailed distribution is common for semantic segmentation datasets Lin et al. (2014); Zhou et al. (2019) even if the number of images that contain specific label are pre-controlled. Such frequency distribution for pixels would be inevitable for objects existing in real-world. Taking "water tower" as an example, despite having 219 images, the percentage of pixels are less than many other labels in the dataset. Fig. 3 shows the positive but weak correlation between number of images for each label and the corresponding pixels. In total, only 4.89% of pixels are unlabeled, and 34.17% and 60.94% of pixels belong to waterbodies (natural and man-made) and general labels, respectively. As it is evident, the main proportion of pixels belongs to general labels. This clearly shows the importance of general labels for better scene understanding Caesar et al. (2018) and accurate object classification in semantic segmentation network.

3.1.1. Spatial analysis

Following ADE20K dataset Zhou et al. (2019), a spatial analysis, known as "mode of the object segmentation" has been done on the ground truth segmentation map for each label. Specifically, considering a waterbody label L with n images in ATLANTIS, we resize their corresponding n ground-truth segmentation maps to 512 × 512 pixels. We then count the most frequent label at each pixel of the map and construct a "mode of segmentation" map for label L, as shown in Fig. 4. This map demonstrates the spatial distributions of the most frequent co-occurred labels with respect to a given waterbody label. Based on this, we

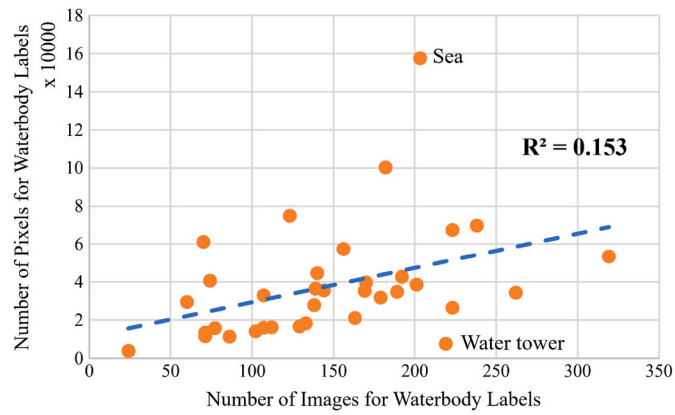

Fig. 3. Correlation between number of images for each label and the corresponding pixels.

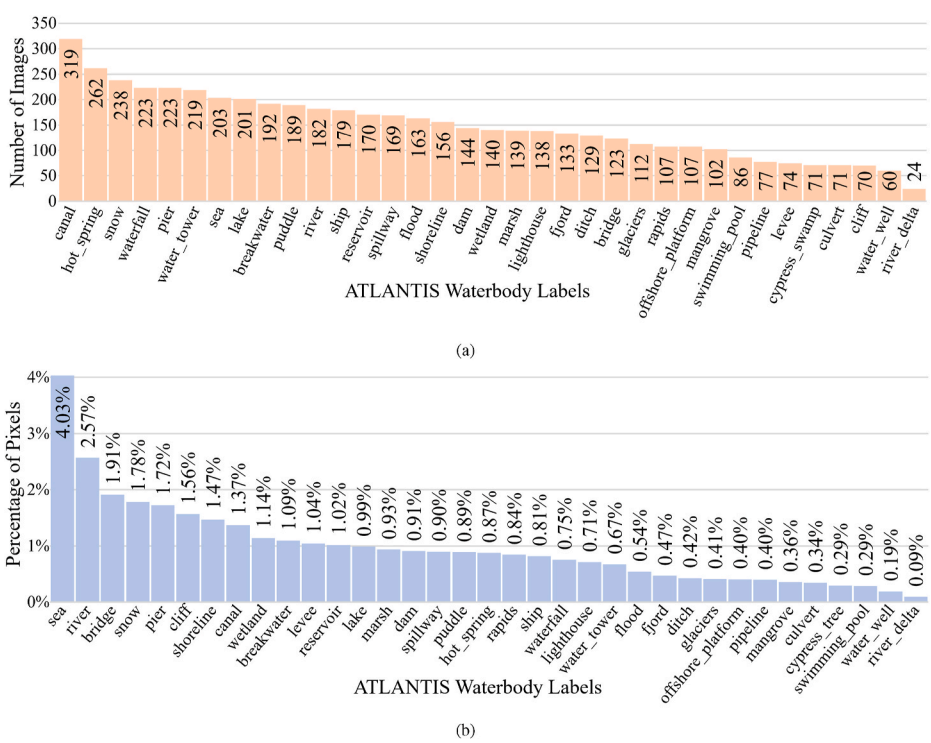

Fig. 2. (a) The frequency distribution of images for different waterbodies in ATLANTIS (b) The Percentage of pixels for waterbody labels.

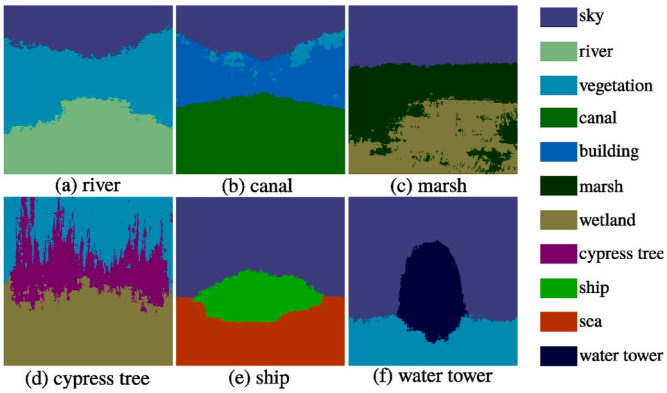

Fig. 4. Spatial analysis of four different waterbody labels.

equipped our proposed network with cross-path feature modulation to cope with the difficulties associated with the recognition of waterbody labels having visual similarities.

3.2. Image annotation

3.2.1. Annotation pipeline

ATLANTIS was annotated by six annotators having prior knowledge in the area of water resources. The goal in the annotation task is to balance speed and quality. Generally, time spent on a single image can range from 4 minutes to 25 minutes depending on the image complexity. In this project, each kind of waterbody was assigned to a specific annotator. Before annotation of a label, all the images of that label are scrutinized and discussed by a group of experts in water resources engineering. We can see that the annotation of complex flood scenes takes more time since such images are usually captured in urban areas and have more elaborated components as shown in Fig. 5.

3.2.2. Consistency analysis

While one image is annotated by one annotator for ATLANTIS, we perform additional consistency analysis across annotators and over time for an annotator. We choose 52 images from ATLANTIS, by including both images that are highly susceptible to wrong labelling and those contain objects prone to be either left unannotated or wrongly annotated. We ask three annotators to annotate them again and compare the results against the already approved ground truth in ATLANTIS. The accuracy and mIoU in terms of all 52 images (total) and the subsets of images that had been annotated by themselves before (individual) are shown in Table 2. We can see that an annotator can process the images that he/she annotated before with much better consistency.

3.3. ATLANTIS texture (ATex)

Waterbodies usually bear texture appearance and it is an interesting problem to study whether different kinds of waterbodies may show subtle differences associated with the texture features. For this purpose, we construct a new waterbody texture dataset, ATeX, by cropping patches from ATLANTIS and take the corresponding annotated waterbody label as the label of the patch. We set patch size to 32 × 32 pixels

Fig. 5. Two samples of complex flood scenes.

Table 2
Consistency analysis for three annotators.

		Ann 1	Ann 2	Ann 3
Total	Acc	84.00	79.93	79.00
	mIoU	62.29	59.33	57.34
Individual	Acc	91.11	90.44	94.69
	mIoU	72.83	76.14	75.69

and all pixels in a cropped patch must have the same waterbody label in the original image. We also ensure there is no partial overlap between any two patches. In total we collected 12,503 patches with 15 waterbody labels: Two waterbody labels "estuary" and "swamp" are added based on the nearby tree species – mangrove "estuary" and cypress for "swamp", while four waterbody labels "canal", "ditch", "reservoir" and "fjord" are omitted because of high visual similarities with other labels. Sample images of ATeX dataset are shown in Fig. 6. We split ATeX into 8,753 for training, 1,252 for validation and 2,498 for testing. In the later experiment, we train different models to evaluate their classification performance on ATeX images.

4. AQUANet

Typically, existing semantic segmentation networks are designed based on a strong backbone (e.g. ResNet He, Zhang, Ren and Sun (2016)) to extract features from images with additional feature aggregation schemes such as ASP-OC Yuan and Wang (2018) and PPM Zhao et al. (2017). Because of difficulties associated with semantic segmentation of waterbodies, we design AQUANet to segment aquatic and non-aquatic categories, separately, as shown in Fig. 7.

4.1. Network architecture

According to Fig. 7, the input image is first fed into a ResNet-101 (pretrained on ImageNet Deng, Dong, Socher, Li, Li and Fei-Fei (2009)) to extract the feature F with a size of $C \times H \times W$. Then, the feature is sent into two separate paths for further processing. The aquatic path is to segment different types of waterbodies including sea, river, waterfall, wetland and etc., while the non-aquatic path is to segment other categories such as ship and bridge. In each path, the feature F is first modulated by the low-level feature F_l extracted from the third

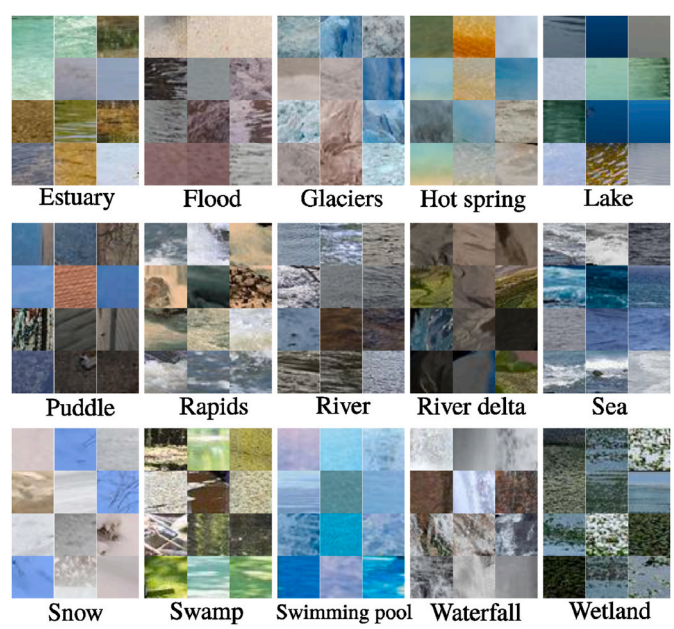

Fig. 6. Samples of ATeX texture images.

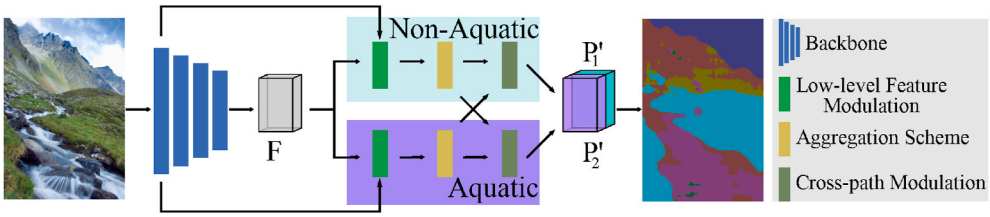

Fig. 7. The network architecture of proposed AQUANet.

convolutional layer of ResNet-101, and then passed into ASP-OC Yuan and Wang (2018) to produce the probability map. In the last step, two cross-path modulation blocks are applied to adjust the probability maps P_1 and P_2 in parallel. Finally, the resulted probability maps are concatenated and upsampled to the size of the original image.

4.2. Feature modulation

The goal of the feature modulation $\mathcal{M}$ is to adjust a feature map F_1 given feature map F_2 to represent the adjusted feature $F_1^{'}$. It can be formulated as:

$$F_1^{'} = \mathcal{M}(F_1|F_2). \tag{1}$$

To generate the modulated feature $F_1^{'}$, the parameters α and β are learned from F_2 via the feature modulation that consists of three downsampling layers, six 1×1 convolutional layers and two leakyReLU layers as shown in Fig. 8. The learned parameters α and β have the shape as the F_1. Then, the resulting feature $F_1^{'}$ is constructed as follows according to Wang et al. (2018); Park et al. (2019):

$$F_1^{'} = \alpha \cdot F_1 + \beta + F_1. \tag{2}$$

4.2.1. Low-level feature modulation

To enhance the low-level texture representation of the water-bodies, we propose to use the low-level feature F_l to modulate the feature F and the resulted feature F' is defined as:

$$F^{'} = \mathcal{M}(F|F_l). \tag{3}$$

Note that different channels of the receptive field in the third convolutional layer of ResNet-101 are used to construct the low-level feature F_l for the two paths.

4.2.2. Cross-path modulation

We also propose a cross-path modulation to aggregate the outputs of the probability maps. The adjusted probability maps resulted from this module are defined as:

$$P_1^{'} = \mathcal{M}(P_1|P_2), \tag{4}$$

$$P_2^{'} = \mathcal{M}(P_2|P_1), \tag{5}$$

where the $P_1^{'}$ and $P_2^{'}$ represent the final probability maps for the aquatic and non-aquatic objects, respectively. Note that the modulation layer is not sharing weights across the two paths.

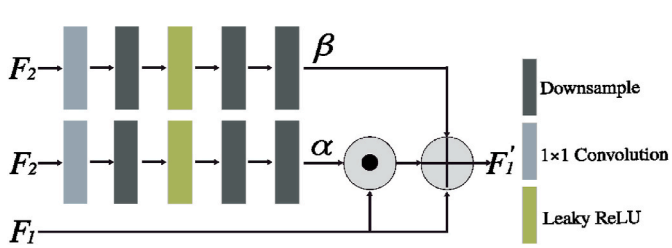

Fig. 8. An illustration of the feature modulation.

4.3. Loss function

The proposed network is trained in a fully supervised fashion constrained by the widely-used cross-entropy loss function on both final prediction P (main loss) and the intermediate feature produced by the fourth block of the ResNet-101 (auxiliary loss). Following Zhao et al. (2017), the weights of the main and the auxiliary loss are set to 1 and 0.4, respectively.

5. Experiments

To demonstrate the effectiveness of the proposed method for waterbodies semantic segmentation, we train AQUANet on the proposed ATLANTIS dataset. For performance evaluation, we take the mean of class-wise intersection over union (mIoU) and the per-pixel accuracy (acc) as the main evaluation metrics. To further evaluate the performance on the waterbodies, we calculate the mean IoU for aquatic categories (A-mIoU) and the accuracy in the aquatic region (A-acc). Aquatic categories include 17 labels showing just water content in different forms and bodies, e.g., sea, river, lake, etc.

5.1. Experimental settings

The AQUANet is implemented using PyTorch. During training, the base learning rate is set to 2.5×10^{-4} and it is decayed following the poly policy Zhao et al. (2017). The network is optimized using SGD with a momentum of 0.9 and weight decay of 0.000 1. In total, we train the network for 30 epochs, around 80K iterations with a batch size of 2. The training data are augmented with random horizontal flipping, random scaling ranging from 0.5 to 2.0 and random cropping with the size of 640×640.

5.2. Comparisons

We use several state-of-the-art networks to perform training and testing on ATLANTIS, including PSPNet Zhao et al. (2017), DeepLabv3 Chen, Papandreou, Schroff and Adam (2017b), CCNet Huang et al. (2019), EMANet Li et al. (2019a), ANNet Zhu et al. (2019), DANet Fu et al. (2019), DNLNet Yin, Yao, Cao, Li, Zhang, Lin and Hu (2020), GCNet Cao, Xu, Lin, Wei and Hu (2019), OCNet Yuan and Wang (2018), OCRNet Yuan et al. (2020). For a fair comparison, we train all the networks with the same backbone (ResNet-101) for 30 epochs. As shown in Table 3, the proposed AQUANet outperforms all these networks on waterbody image semantic segmentation. Fig. 9 shows the visualization results of some samples from ATLANTIS validation set. Considering the ground truth and comparing with other networks' outputs, the boundaries between different classes are better preserved in AQUANet output. Compared with Chen et al. (2017b), Zhao et al. (2017), Fu et al. (2019), and Yuan and Wang (2018), our method achieves better results on both the aquatic and non-aquatic regions.

5.3. Failure cases

Due to the challenges associated with segmentation of waterbody images, there are still many failure cases we found in the testing stage.

Table 3
The per-category results on the ATLANTIS test set by current state-of-the-art methods and our AQUANet. The best and the second best results are highlighted with **bold** font and underline, respectively.

Method	canal	ditch	fjord	flood	glaciers	hot spring	lake	puddle	rapids	reservoir	river	river delta	sea	snow	swimming pool	waterfall	wetland	A-acc (%)	A-mIoU	acc (%)	mIoU
PSPNet	53.8	29.0	42.9	46.5	57.2	53.9	29.7	**54.7**	38.2	29.8	28.8	65.5	63.5	49.9	47.7	48.4	47.5	66.19	46.29	72.72	40.85
DeepLabv3	52.5	27.2	52.3	43.8	58.7	42.5	31.1	54.2	46.0	32.4	27.1	51.1	61.5	46.3	**53.6**	52.8	52.9	65.83	46.25	69.21	36.23
DANet	50.5	**34.1**	37.1	37.0	51.0	61.6	23.8	51.5	42.8	30.2	31.5	63.5	60.4	50.8	43.1	55.2	54.6	62.00	45.80	74.12	39.60
CCNet	41.1	17.4	35.2	26.9	43.7	47.9	18.6	43.8	29.9	16.6	23.7	48.3	53.3	47.6	38.4	51.1	34.1	51.98	36.33	70.84	36.11
EMANet	46.1	16.6	27.1	23.0	53.8	**63.7**	17.2	43.6	42.2	17.2	21.0	**68.6**	53.5	47.3	36.1	52.1	36.2	55.88	39.13	71.93	36.43
ANNet	50.9	22.8	31.6	32.0	53.1	58.1	25.6	52.9	**48.4**	20.8	28.6	56.8	60.4	**51.1**	43.9	**57.9**	51.4	61.51	43.90	74.06	39.79
GCNet	**56.6**	19.0	44.7	34.8	46.9	36.1	**35.8**	39.4	39.9	**41.6**	32.4	67.0	62.2	46.4	42.9	50.7	59.7	**69.89**	44.48	68.64	37.73
DNLNet	54.4	26.3	48.8	36.3	**63.2**	55.3	35.5	52.3	40.4	32.1	31.3	37.1	61.7	48.3	52.4	48.7	54.6	67.72	45.80	71.95	39.97
OCNet	56.4	33.6	48.0	37.3	57.7	55.2	29.2	50.6	43.8	35.1	**35.6**	65.9	62.7	47.2	47.9	53.1	54.9	67.97	47.89	73.54	41.19
OCRNet	52.4	19.4	46.9	34.9	48.3	58.8	30.4	39.7	42.5	29.8	31.9	55.5	55.4	47.3	43.6	56.8	51.5	65.90	43.83	71.66	36.17
Ours	55.0	27.7	**53.4**	**47.0**	63.1	60.5	33.2	54.4	46.3	39.0	34.7	63.2	**64.2**	50.3	44.9	53.0	**66.1**	68.63	**50.34**	**75.18**	**42.22**

Three failure examples are shown Fig. 10, from which we observe that many aquatic classes are vulnerable to mis-classification – here sea is mis-classified to lake and river (row 1–2) and river is mis-classified to canal (row 3).

5.4. Ablation studies

We also conduct ablation studies to compare a number of different model variants of the proposed network, including the design of aquatic and non-aquatic paths, and the two feature modulations. The results are shown in Table 4. We can see that the design of two paths can improve the performance of waterbody image semantic segmentation. Moreover, both the proposed low-level feature modulation (LM) and the cross-path modulation (CM) can achieve certain performance gains in terms of acc and mIoU.

Comparing to other state-of-the-art semantic segmentation models, AQUANet provides highest mIoU and accuracy. In addition, by considering just labels which includes water content, AQUANet still works better than others (mIoU = 50.34%). In this regard, OCNet provides the second highest performance (mIoU = 47.89%) and GCNet also provides the best results for canal, lake and reservoir. These results approved AQUANet is well customized for analysis of waterbodies and water-related scenes. However, considering mIoU as a more informative metric for semantic segmentation task, all approaches could only achieve 36.11% ∼42.22% mIoU on the proposed dataset. It shows that the semantic segmentation of waterbodies and related objects is a challenging task and needs more researches.

5.5. ATeX experimental results

We further train ten well-known classification models including VGG Simonyan and Zisserman (2014), ResNet He et al. (2016), SqueezeNet Iandola, Han, Moskewicz, Ashraf, Dally and Keutzer (2016), DenseNet Huang, Liu, Van Der Maaten and Weinberger (2017), GoogLeNet Szegedy, Liu, Jia, Sermanet, Reed, Anguelov, Erhan, Vanhoucke and Rabinovich (2015), ShuffleNet v2 Ma et al. (2018), MobileNetV2 Sandler et al. (2018), ResNeXt Xie, Girshick, Dollár, Tu and He (2017), Wide ResNet Zagoruyko and Komodakis (2016) and EfficientNet Tan and Le (2019) on the proposed ATeX dataset. All models are implemented using PyTorch. Cross-entropy loss function is applied for training networks. We train all the networks with the same 30 training epochs, SGD optimizer with a momentum of 0.9 and weight decay of 0.000 1, and batch size is set to 64. For all networks, the learning rate is first set to 2.5×10^{-4}, then it is adjusted based on decaying rate of the resulted loss function during training. Table 5 shows the training time and learning rate for each models over certain 30 epochs.

Three common performance metrics including Precision, Recall and F1-score are reported to evaluate the performance of the models on ATeX. Table 5 shows weighted average (averaging the support-weighted mean per label) of these three metrics on the test set. Accordingly, EffNet-B7, EffNet-B0 and ShuffleNet V2 × 1.0 provide the best results. Considering training time, ShuffleNet V2 × 1.0 can be presented as the most efficient network.

6. Conclusion

In this paper, we introduced ATLANTIS, a large-scale dataset for semantic segmentation of waterbodies and water-related scenes, by carefully collecting images of diverse area from the internet (Flickr) and providing high-quality annotations with the help of annotators major in water resources engineering. We further provided comprehensive analysis of the characteristic of ATLANTIS and reported the performance of current state-of-the-arts by training and testing the networks on our dataset. A novel baseline network AQUANet is also proposed for waterbody image semantic segmentation and achieves the best performance on ATLANTIS. Additionally, we constructed ATLANTIS Texture

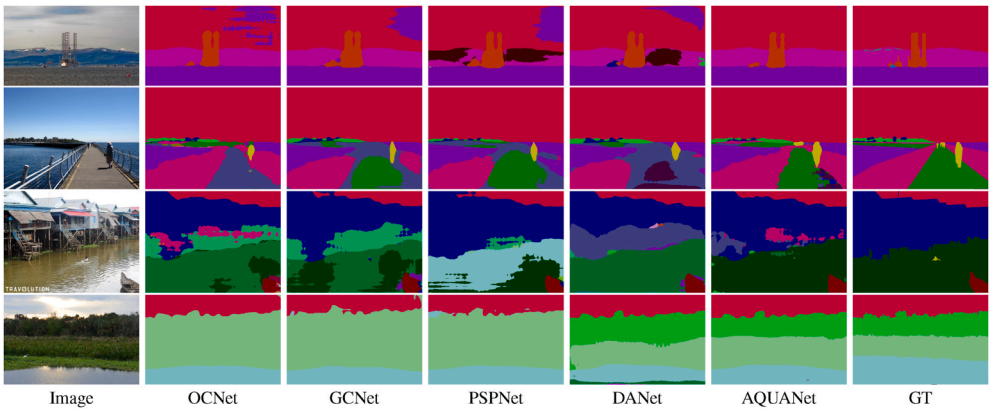

Fig. 9. Visualization comparison of AQUANet and four well-known methods on the ATLANTIS validation set.

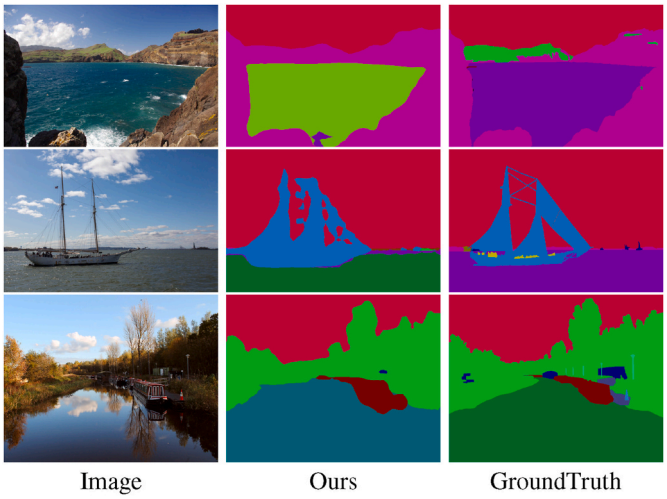

Fig. 10. Failure cases from the ATLANTIS val set.

Table 4
Ablation study of each proposed component of AQUANet on the ATLANTIS dataset.

Two Paths	LM	CM	A-acc	A-mIoU	acc	mIoU
			67.27	44.53	73.28	38.81
✓			67.73	47.89	75.29	40.28
✓	✓		68.11	47.90	75.29	40.28
✓		✓	67.21	46.63	75.81	40.57
✓	✓	✓	68.85	48.42	76.18	40.83

Table 5
The performance result on ATeX test set by well-known classification models.

Networks	Time	LR	Val	Test		
	[mm:ss]		Acc.	Prec.	Recall	F1
Wide-ResNet-50-2	06:56	2.5E-4	91	77	75	75
VGG-16	04:38	2.5E-4	90	75	72	72
SqueezeNet	00:47	7.5E-4	82	81	81	81
ShuffleNet V2 × 1.0	01:46	1.0E-2	90	90	90	90
ResNeXt-50-32 × 4*d*	03:15	2.5E-4	90	77	75	75
ResNet-18	01:28	2.5E-4	87	74	72	72
MobileNet V2	01:35	2.5E-4	88	74	72	72
GoogLeNet	01:51	5.0E-3	89	88	88	88
EffNet-B7	12:42	1.0E-2	90	91	91	91
EffNet-B0	02:38	7.5E-3	91	90	90	90
DenseNet-161	06:15	2.5E-4	91	81	79	79

(ATeX) dataset which is derived from ATLANTIS for classification and texture analysis of water. The performance of several baseline classification networks on ATeX was also evaluated and reported.

In general, digital image processing of water and water-related objects has been a complex task due to the visual challenges which are inherent in water. ATLANTIS includes images and categories beyond a specific purpose, and does not focus on a certain surrounding environment for specific purpose or limited applications. Covering such a broad and divers water and water-related categories, ATLANTIS poses significant challenges for semantic segmentation which we believe will boost new insights in both water resources engineering and computer vision communities.

Software and data availability

ATLANTIS is available through the GitHub repository [https://github.com/smhassanerfani/atlantis]. All labels, including natural waterbodies and water-related structures are listed and introduced in ATLANTIS Wiki page, which is accessible at [https://github.com/smhassanerfani/atlantis/wiki]. ADK, ATLANTIS Development Kit, is also available for researchers who are interested in developing similar image datasets [https://github.com/smhassanerfani/atlantis/tree/master/adk]. In ADK, different pipelines are provided to facilitate downloading, annotating, organizing and analyzing images. These pipelines are explained in detail, whether for contribution to this project or for personal uses.

AQUANet along with all other semantic segmentation networks in this study are discussed in detail and are available at [https://github.com/smhassanerfani/atlantis/tree/master/aquanet]. All models are implemented in Python using PyTorch machine learning library.

ATeX is also available through the GitHub repository [https://github.com/smhassanerfani/atex]. Likewise, ATeX Wiki [https://github.com/smhassanerfani/atex/wiki] documents the guidelines for the list of waterbodies considered in the dataset. Applied classification networks in this study are available in the GitHub repository as well. These networks are all implemented in Python using PyTorch machine learning library.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work is supported by Microsoft AI for Earth under USCeRA proposal number 155200-20-52426. Any opinions, findings, and

discussions expressed in this study are those of the authors and do not necessarily reflect the views of the Microsoft. University of South Carolina is acknowledged for generous allotment of computation time on the High Performance Computing (HPC) clusters.

References

Badrinarayanan, V., Handa, A., Cipolla, R., 2015. Segnet: A Deep Convolutional Encoder-Decoder Architecture for Robust Semantic Pixel-wise Labelling. arXiv preprint arXiv: 1505.07293.

Bernal, J., Tajkbaksh, N., Sánchez, F.J., Matuszewski, B.J., Chen, H., Yu, L., Angermann, Q., Romain, O., Rustad, B., Balasingham, I., et al., 2017. Comparative validation of polyp detection methods in video colonoscopy: results from the miccai 2015 endoscopic vision challenge. IEEE Trans. Med. Imag. 36, 1231–1249.

Caesar, H., Uijlings, J., Ferrari, V., 2018. Coco-stuff: thing and stuff classes in context, in: IEEE Conf. Comput. Vis. Pattern Recog. 1209–1218.

Cao, Y., Xu, J., Lin, S., Wei, F., Hu, H., 2019. Gcnet: non-local networks meet squeeze-excitation networks and beyond. In: Int. Conf. Comput. Vis. Worksh, 0–0.

Cervone, G., Sava, E., Huang, Q., Schnebele, E., Harrison, J., Waters, N., 2016. Using twitter for tasking remote-sensing data collection and damage assessment: 2013 boulder flood case study. Int. J. Rem. Sens. 37, 100–124.

Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L., 2017a. Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE Trans. Pattern Anal. Mach. Intell. 40, 834–848.

Chen, L.C., Papandreou, G., Schroff, F., Adam, H., 2017b. Rethinking Atrous Convolution for Semantic Image Segmentation arXiv preprint arXiv:1706.05587.

Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S., Schiele, B., 2016. The cityscapes dataset for semantic urban scene understanding, in: IEEE Conf. Comput. Vis. Pattern Recog. 3213–3223.

Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L., 2009. Imagenet: a large-scale hierarchical image database. In: IEEE Conf. Comput. Vis. Pattern Recog. IEEE, pp. 248–255.

Duan, L., Hu, X., 2019. Multiscale refinement network for water-body segmentation in high-resolution satellite imagery. Geosci. Rem. Sens. Lett. IEEE 17, 686–690.

Eltner, A., Bressan, P.O., Akiyama, T., Gonçalves, W.N., Marcato Junior, J., 2021. Using deep learning for automatic water stage measurements. Water Resour. Res. 57 e2020WR027608.

Eltner, A., Elias, M., Sardemann, H., Spieler, D., 2018. Automatic image-based water stage measurement for long-term observations in ungauged catchments. Water Resour. Res. 54, 10–362.

Everingham, M., Van Gool, L., Williams, C.K., Winn, J., Zisserman, A., 2010. The pascal visual object classes (voc) challenge. Int. J. Comput. Vis 88, 303–338.

Fu, J., Liu, J., Tian, H., Li, Y., Bao, Y., Fang, Z., Lu, H., 2019. Dual attention network for scene segmentation, in: IEEE Conf. Comput. Vis. Pattern Recog. 3146–3154.

Gebrehiwot, A., Hashemi-Beni, L., Thompson, G., Kordjamshidi, P., Langan, T.E., 2019. Deep convolutional neural network for flood extent mapping using unmanned aerial vehicles data. Sensors 19, 1486.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition, in: IEEE Conf. Comput. Vis. Pattern Recog. 770–778.

Hirabayashi, Y., Mahendran, R., Koirala, S., Konoshima, L., Yamazaki, D., Watanabe, S., Kim, H., Kanae, S., 2013. Global flood risk under climate change. Nat. Clim. Change 3, 816–821.

Hosseiny, H., 2021. A deep learning model for predicting river flood depth and extent. Environ. Model. Software 145, 105186.

Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q., 2017. Densely connected convolutional networks, in: IEEE Conf. Comput. Vis. Pattern Recog. 4700–4708.

Huang, X., Wang, C., Li, Z., 2018. A near real-time flood-mapping approach by integrating social media and post-event satellite imagery. Spatial Sci. 24, 113–123.

Huang, Z., Wang, X., Huang, L., Huang, C., Wei, Y., Liu, W., 2019. Ccnet: Criss-cross attention for semantic segmentation. In: Int. Conf. Comput. Vis., pp. 603–612.

Iandola, F.N., Han, S., Moskewicz, M.W., Ashraf, K., Dally, W.J., Keutzer, K., 2016. Squeezenet: Alexnet-Level Accuracy with 50x Fewer Parameters and $\lessbin $0.5 Mb Model Size arXiv preprint arXiv:1602.07360.

Jha, D., Smedsrud, P.H., Riegler, M.A., Halvorsen, P., de Lange, T., Johansen, D., Johansen, H.D., 2020. Kvasir-seg: A Segmented Polyp Dataset, in: International Conference on Multimedia Modeling. Springer, pp. 451–462.

Li, X., Zhong, Z., Wu, J., Yang, Y., Lin, Z., Liu, H., 2019a. Expectation-maximization attention networks for semantic segmentation. In: Int. Conf. Comput. Vis., pp. 9167–9176.

Li, Z., Wang, R., Zhang, W., Hu, F., Meng, L., 2019b. Multiscale features supported deeplabv3+ optimization scheme for accurate water semantic segmentation. IEEE Access 7, 155787–155804.

Lin, G., Milan, A., Shen, C., Reid, I., 2017. Refinenet: multi-path refinement networks for high-resolution semantic segmentation, in. In: IEEE Conf. Comput. Vis. Pattern Recog, pp. 1925–1934.

Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C. L., 2014. Microsoft coco: common objects in context. In: Eur. Conf. Comput. Vis., Springer, pp. 740–755.

Lo, S.W., Wu, J.H., Lin, F.P., Hsu, C.H., 2015. Visual sensing for urban flood monitoring. Sensors 15, 20006–20029.

Long, J., Shelhamer, E., Darrell, T., 2015. Fully convolutional networks for semantic segmentation, in. In: IEEE Conf. Comput. Vis. Pattern Recog, pp. 3431–3440.

Ma, N., Zhang, X., Zheng, H.T., Sun, J., 2018. Shufflenet v2: practical guidelines for efficient cnn architecture design. In: Eur. Conf. Comput. Vis., pp. 116–131.

Mottaghi, R., Chen, X., Liu, X., Cho, N.G., Lee, S.W., Fidler, S., Urtasun, R., Yuille, A., 2014. The role of context for object detection and semantic segmentation in the wild, in. In: IEEE Conf. Comput. Vis. Pattern Recog, 891–898.

Muñoz, D.F., Muñoz, P., Moftakhari, H., Moradkhani, H., 2021. From local to regional compound flood mapping with deep learning and data fusion techniques. Sci. Total Environ. 782, 146927.

Neuhold, G., Ollmann, T., Rota Bulo, S., Kontschieder, P., 2017. The mapillary vistas dataset for semantic understanding of street scenes. In: Int. Conf. Comput. Vis., pp. 4990–4999.

Noh, H., Hong, S., Han, B., 2015. Learning deconvolution network for semantic segmentation. In: Int. Conf. Comput. Vis., pp. 1520–1528.

Pally, R., Samadi, S., 2021. Application of image processing and convolutional neural networks for flood image classification and semantic segmentation. Environ. Model. Software 105285.

Park, T., Liu, M.Y., Wang, T.C., Zhu, J.Y., 2019. Semantic image synthesis with spatially-adaptive normalization, in. In: IEEE Conf. Comput. Vis. Pattern Recog., pp. 2337–2346.

Razavi, S., 2021. Deep learning, explained: fundamentals, explainability, and bridgeability to process-based modelling. Environ. Model. Software 144, 105159.

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.C., 2018. Mobilenetv2: inverted residuals and linear bottlenecks, in. In: IEEE Conf. Comput. Vis. Pattern Recog, pp. 4510–4520.

Sarp, S., Kuzlu, M., Cetin, M., Sazara, C., Guler, O., 2020. Detecting floodwater on roadways from image data using mask-r-cnn. In: 2020 International Conference on INnovations in Intelligent SysTems and Applications (INISTA). IEEE, pp. 1–6.

Sazara, C., Cetin, M., Iftekharuddin, K.M., 2019. Detecting floodwater on roadways from image data with handcrafted features and deep transfer learning. In: 2019 IEEE Intelligent Transportation Systems Conference (ITSC). IEEE, pp. 804–809.

Schnebele, E., Cervone, G., 2013. Improving remote sensing flood assessment using volunteered geographical data. Nat. Hazards Earth Syst. Sci. 13, 669–677.

Sekachev, B., Manovich, N., Zhiltsov, M., Zhavoronkov, A., Kalinin, D., Hoff, B., TOsmanov Kruchinin, D., Zankevich, A., DmitriySidnev Markelov, M., Johannes222 Chenuet, M., Melnikov, A., Kim, J., Ilouz, L., Glazov, N., Priya4607 Tehrani, R., Jeong, S., Skubriev, V., Yonekura, S., truong, Vugia, zliang7, lizhming, Truong, T., 2020. a andre telenachos opencv/cvat: v1.1.0. URL. https://doi.org/10.5281/zenodo.4009388, 10.5281/zenodo.4009388.

Simonyan, K., Zisserman, A., 2014. Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv preprint arXiv:1409.1556.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2015. Going deeper with convolutions, in: IEEE Conf. Comput. Vis. Pattern Recog. 1–9.

Tan, M., Le, Q., 2019. Efficientnet: rethinking model scaling for convolutional neural networks, in. In: Int. Conf. on Mach. Learn. PMLR, pp. 6105–6114.

Wang, X., Yu, K., Dong, C., Change Loy, C., 2018. Recovering realistic texture in image super-resolution by deep spatial feature transform, in. In: IEEE Conf. Comput. Vis. Pattern Recog., pp. 606–615.

Xie, S., Girshick, R., Dollár, P., Tu, Z., He, K., 2017. Aggregated residual transformations for deep neural networks, in. In: IEEE Conf. Comput. Vis. Pattern Recog, pp. 1492–1500.

Yin, M., Yao, Z., Cao, Y., Li, X., Zhang, Z., Lin, S., Hu, H., 2020. Disentangled non-local neural networks. In: Eur. Conf. Comput. Vis., Springer, pp. 191–207.

Yu, F., Chen, H., Wang, X., Xian, W., Chen, Y., Liu, F., Madhavan, V., Darrell, T., 2020. Bdd100k: a diverse driving dataset for heterogeneous multitask learning, in. In: IEEE Conf. Comput. Vis. Pattern Recog., pp. 2636–2645.

Yuan, Y., Chen, X., Wang, J., 2020. Object-contextual representations for semantic segmentation. In: Eur. Conf. Comput. Vis., Springer, pp. 173–190.

Yuan, Y., Wang, J., 2018. Ocnet: Object Context Network for Scene Parsing arXiv preprint arXiv:1809.00916.

Zagoruyko, S., Komodakis, N., 2016. Wide Residual Networks. arXiv preprint arXiv: 1605.07146.

Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J., 2017. Pyramid scene parsing network, in: IEEE Conf. Comput. Vis. Pattern Recog. 2881–2890.

Zhou, B., Zhao, H., Puig, X., Xiao, T., Fidler, S., Barriuso, A., Torralba, A., 2019. Semantic understanding of scenes through the ade20k dataset. Int. J. Comput. Vis. 127, 302–321.

Zhu, Z., Xu, M., Bai, S., Huang, T., Bai, X., 2019. Asymmetric non-local neural networks for semantic segmentation. In: Int. Conf. Comput. Vis., pp. 593–602.