Sustainable Inference of Remote Sensing Data by Recursive Semantic Segmentation — a Flood Extent Mapping Study

Thomas Brunschwiler[1], Tobia Claglüna[1], Michal Muszynski[1], Tobias Hölzer[1], Paolo Fraccaro[2], Maciel Zortea[3], Jonas Weiss[1]

[1] IBM Research Europe, Switzerland
[2] IBM Research Europe, United Kingdom
[3] IBM Research Brazil, Brazil

tbr@zurich.ibm.com, tcl@zurich.ibm.com, mmu@zurich.ibm.com, tobias.hoelzer@ibm.com, Paolo.Fraccaro@ibm.com, mazortea@br.ibm.com, jwe@zurich.ibm.com

Abstract

In times of climate change and large machine learning models with petabytes of training data, the demand for responsible AI methodologies is more pressing than ever. This is in-particular true for the vast amount of remote sensing data. Its value to explore and inform about earth processes is of paramount importance, especially in times of global warming. Thus, it is nearly ironic that such applications can be the cause of substantial green-house-gas emissions, through energy demands from compute and communication systems. Thus, this study aims at reducing the data transfer between data centers while maintaining near-real time insights from remote sensing data. A recursive inference approach is introduced consisting of three steps: i) Data pyramid preparation in the host data center (sequence of upscaled raster data). ii) The transfer of low-resolution images to the service data center, where a deep-learning model performs a semantic segmentation task, including an uncertainty estimation. Images of higher resolution are then requested and segmented in a recursive fashion, in areas of high uncertainty only. iii) Finally, the merging of the predictions at different resolutions is performed to result in the final pixel-wise segmentation at scale. The method is demonstrated on synthetic and real-world data for a flood mapping task. A U-Net encoder-decoder model is used for the semantic segmentation task, using Monte-Carlo dropout to result in the uncertainty map. The proof-of-concept demonstrated a 35-38% performance gain per transferred pixel compared to high-resolution image segmentation only. Further, we perform a scaling study to estimate the true potential of the recursive inference approach, indicating the potential to reduce a data transfer by up to 98%, considering four hierarchy levels in the data pyramid. With this study, we hope to have contributed a small but important step towards sustainable machine learning.

Introduction

Extracting insights from vast amounts of remote sensing data in near real-time has become feasible by machine learning techniques. In 2021 only, the total Sentinel data volume available for retrieval from the Copernicus Data Access System was 42 petabytes, with a total download volume of 81 petabytes (Copernicus 2022). As indicated by these numbers, for services at national or global scale, petabytes of data are needed to be transferred to data centers where geospatial insights are being computed and from where services are being provided. Hence, the cost of moving data to the compute infrastructure can no longer be neglected and becomes economically and environmentally taxing, due to carbon emissions from communication networks. The transfer of 1 petabyte of data between data centers results in 588 t-CO_2 emissions considering the energy mix in Japan, according to Tabata et al. (Tabata and Wang 2021). To render the insight generation more sustainable, the co-location of data and compute should be considered whenever possible. Further, data compression should be applied to reduce the data volumes. Multi-resolution image processing approaches were proposed in the medical domain, to reduce the computational cost in segmenting high-resolution radiological images with gigapixel size (Seyedhosseini and Tasdizen 2016). The authors used three sets of network weights, each trained on a different resolution of downsampled images. Others like Thandiackal et al. (Thandiackal et al. 2022) or Maksoud (Maksoud et al. 2020) considered hierarchical inference approaches for binary classification.

In this study, we are proposing a recursive inference approach reducing the overall data transfer between a data center of the remote sensing host and the entity generating the insights for a semantic segmentation task. The method is evaluated on a flood extent mapping task based on synthetic and Sentinel-2 data, in an attempt towards responsible AI.

Methodology

The recursive inference methodology can be split into three main parts: i) data pyramid preparation, ii) inference with uncertainty quantification and iii) merging of the predictions, as depicted in Figure 1.

Data Pyramid Preparation

The host of the remote sensing data performs an upscaling of the original raster data (hierarchy level k_1) by a linear scaling factor r, which is larger than 1, resulting in the raster image of hierarchy level k_2. This process is repeated until hierarchy level k_n is reached, where n is a hyperparameter which depends on the specific use-case. All these hierarchy levels form a data pyramid. The resulting storage overhead in the host data center considering $n \rightarrow \infty$ is the limes of

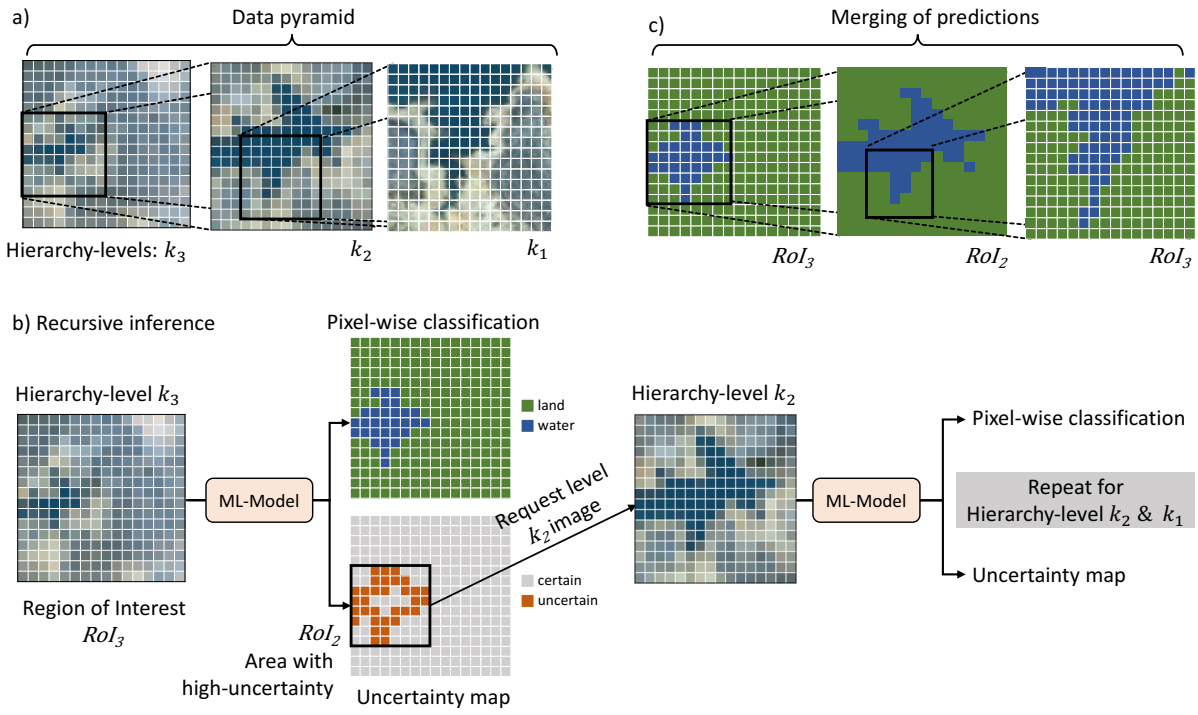

Figure 1: Depiction of the recursive inference methodology with a) data pyramid creation, b) recursive inference, and c) the merging of the predictions.

a geometric series with the common ratio between adjacent terms of $\frac{1}{r^2}$ being $\frac{1}{r^2-1}$. Considering $r = 2$ the overhead equals to 33%.

Recursive Inference

The initial semantic segmentation iteration starts with the transfer of the raster image from the highest hierarchy level to the inference data center. The pixel-wise classification is then performed on this low resolution image by a machine-learning (*ML*) model. In addition, the ML-model provides also a pixel-wise uncertainty estimate. For regions of high uncertainty, which we call regions of interest (*RoI*), the image from the next lower hierarchy-level is requested, to again perform the pixel-wise classification and uncertainty map estimation. This process is repeated recursively for areas of high uncertainty, through all the hierarchy-levels, until the last hierarchy-level is reached.

Merging of Predictions

Finally, the semantic segmentation maps of the different *RoI*'s with different resolutions are upsampled to the highest resolution (i.e. of hierarchy-level k_1) and merged to the final semantic segmentation map.

General Algorithm

Algorithm 1 depicts the formal view of the general recursive inference algorithm. The assumption is to have access to a neural network (NN) for semantic segmentation with weights trained on data from all n hierarchy levels of the data pyramid, resulting in predictions $\Omega(S_k), \forall k \in [1, n]$. The method Φ identifies RoIs with high uncertainty $\chi_k = \chi(\Omega, S_k)$ and $H(\chi_k, \chi_{k-1}) = |\chi_k - \chi_{k-1}|/|\chi_k| < \alpha$ is the stopping criteria with α as the threshold for the recursive algorithm indicating diminishing relative returns of additional high-resolution data.

Algorithm 1 Pseudo-code of recursive inference

Input: Image instance S with access to n resolutions
Output: Prediction $\Omega(S_n)$ for image instance S

$k \leftarrow n$
Request sample S_k (lowest resolution)
Semantic segmentation $\Omega(S_k)$
Define stopping threshold α
while $k > 1$ **do**
  Identify RoI by Φ
  Request S_{k-1} from RoI
  Create a prediction $\Omega(S_{k-1})$ from S_{k-1}
  Merge $\Omega(S_{k-1})$ into $\Omega(S_n)$
  if $H(\chi_k, \chi_{k-1}) < \alpha$ **then**
    break
  else
    $k \leftarrow k - 1$
  end if
end while

Experiment

The validation of the applicability of the recursive inference methodology was done based on the flood extent mapping task (i.e., semantic segmentation task). The goal is to perform a pixel-wise classification of water-bodies and landmasses on remote sensing data. All experiments in this study we carried out on two hierarchy levels (i.e. k_1 and k_2).

Flood Datasets

For the study, we considered i) a synthetic coastline dataset and ii) a real-world flood dataset.

The **synthetic dataset** mimics a coastline based on fractals, as explored by Mandelbrot (Mandelbrot 1967). In particular, 400 binary images based on the Julia set were generated for the same RoI with 150×150 and 300×300 pixel resolution, representing the hierarchy-levels k_2 and k_1, respectively. The high resolution images (k_1) represent not only the input features, but also the ground-truth. One instance of a synthetic coastline image is depicted in Figure 2d.

The **Sen1Flood11 dataset** (Bonafilia et al. 2020) which consists of Sentinel-1 and 2 imagery at 10m spatial resolution and manually labeled ground-truth from 11 flood events at different locations around the world was used for the real-world validation. The VV and VH-bands, as well as the optical bands (RGB) from Sentinel-1 and 2 were selected respectively as features for the segmentation task, as proposed by (Muszynski et al. 2022). In total, 446 scenes with 512×512 pixels are available, representing hierarchy level k_1 (see Figure 3a). Those 446 scenes were upscaled by $r = 2$ resulting in a spatial resolution of 20m and 256×256 pixels for the same RoI. It is worth mentioning that both datasets were split into 80/20 training and validation sets, to evaluate the proposed recursive inference methodology.

Semantic Segmentation Model

A U-Net with encoder-decoder architecture (Ronneberger, P.Fischer, and Brox 2015) and a ResNext-101 ($32 \times 8d$) encoder was used as the semantic segmentation model. One U-Net model was trained with images from both hierarchy levels, for each of the datasets as done in (Wang 2021). Training was done over 100 epochs with batch size 32 and a learning rate of 0.001. We also used dropout regularization with a probability of 0.05. For the applied *Tversky* loss function (Salehi, Erdogmus, and Gholipour 2017) (Abraham and Khan 2019) we chose a penalty score for *False Negatives* of $\beta = 0.9$ to compensate for the uneven distribution (90% landmass, 10% water) of the classes in the dataset. For the implementation we used the PyTorch Machine Learning Framework (Paszke et al. 2019).

Uncertainty Quantification

During inference, Monte-Carlo dropout was used to estimate the uncertainty of the NN prediction. Accordingly, the semantic segmentation of one image was repeated $T = 10$ times with a dropout of 0.05. The final prediction for a hierarchy level resulted from a pixel-wise majority vote. We applied the formulation derived by Kwon et al. (Kwon et al. 2020) to estimate the uncertainty for a one-hot encoded prediction y^*. The estimator for their variational formulation of the uncertainty is as follows:

$$\underbrace{\frac{1}{T}\sum_{t=1}^{T}\left[\text{diag}\{p(y^*|x^*,\hat{w}_t)\} - p(y^*|x^*,\hat{w}_t)^{\otimes 2}\right]}_{\text{aleatoric}} + \tag{1}$$

$$\underbrace{\frac{1}{T}\sum_{t=1}^{T}\left[p(y^*|x^*,\hat{w}_t) - \hat{p}_{\hat{\theta}}(y^*|x^*)\right]^{\otimes 2}}_{\text{epistemic}} \tag{2}$$

where $v^{\otimes 2} = vv^T$ and $\hat{p}_{\hat{\theta}}$ is the Monte-Carlo estimator.

Region of Interest Identification

A sliding window approach was used to identify regions with high uncertainty in the final prediction of a hierarchy level. A window size with half the number of pixels compared to the full image and a dilation of half the window size was chosen. For each of the resulting 16 windows we computed a 2D image entropy measure with the method proposed in (Larkin 2016). The window with the highest entropy value has to be declared as RoI_1, where an image with higher resolution from the next lower hierarchy level (k_1) will be segmented.

Merging of Predictions

For the validation experiments, the high-resolution prediction needs to be merged at the position of the RoI_1 in the low-resolution prediction, to result in the final mixed-resolution prediction. To perform this task, the low-resolution prediction is upsampled by a scaling-factor of two, considering bilinear interpolation. Subsequently, the values in the RoI_1 are being updated by the high-resolution predictions.

Evaluation Metric

To evaluate the semantic segmentation results we employ the Jaccard coefficient (*J*) also known as Intersection over Union (IoU), representing the overlap between the prediction (*A*) and the ground-truth (*B*):

$$J(A,B) = \frac{A \cap B}{A \cup B}. \tag{3}$$

$J = 0$ indicates no overlap, whereas $J = 1$ implies a complete overlap between the sets.

Further, we would like to quantify the J-gain achieved by transferring additional pixels for obtaining a prediction at hierarchy level k_j compared to one at a lower-resolution hierarchy level k_i (with $p_i < p_j$). Therefore, we define the J-gain density g as

$$g(y,\hat{y}_i,\hat{y}_j) = \frac{J(y,\hat{y}_j) - J(y,\hat{y}_i)}{p_j - p_i}p_1, \tag{4}$$

with y being the ground-truth, $\hat{y}$ a prediction and p the pixel count for a given prediction. The subscripts i and j on $\hat{y}$ and p indicate the baseline case with the low resolution and the

mixed or high-resolution case, respectively. The multiplication by p_1 results in a normalization of the pixel transfer difference w.r.t. the pixel count of the highest resolution of the data pyramid (level k_1).

Results

Synthetic Data

The feasibility of the recursive inference approach is depicted by the segmentations and uncertainty maps of the synthetic dataset (see Figure 2). The low-resolution segmentation (a) does not reveal details present in the central upper part of the ground-truth image (d). Accordingly, the uncertainty in this area is high as depicted in (b) and thus, the recursive inference algorithm requested a higher-resolution image for the mixed-segmentation in the RoI depicted with the white window.

The mean prediction scores ($\bar{J}$) and number of pixels to transfer per sample for the low, high and mixed-resolution segmentations are listed in Table 1-Synthetic. The low-resolution prediction acts as baseline to compute the J-gain density. As expected, the recursive inference with the mixed-resolution performs better than the low, but worse than the high-resolution prediction. However, the mixed-resolution prediction yields the highest performance J-gain density, due to the 2-fold lower pixel transfer than the high-resolution prediction.

Sen1Flood11

Given the proof-of-concept with the synthetic dataset, we now report on the results with real-world data from the Sen1Flood11 dataset, represented by one instance in Figure 3a). Again, the selection of the RoI window along the river bank, based on the maximum entropy algorithm, turns out to be appropriate. Accordingly, the segmentation details in the window with high-resolution (d) is reflecting the ground-truth data (b) better, than the low-resolution version in image (c).

For the real-world test case, the overall performance is lower compared to the synthetic case as depicted in Table 1-Sen1Flood11. This decline of performance can be attributed to the inherent complexity of the real-world scenario, which involves three (highly unbalanced) segmentation classes, measurement noise, capturing 11 flood events in very diverse environmental conditions. Further, the mixed-resolution segmentation performance lies again between the low and high-resolution segmentation performance. The J-gain density g is also substantially higher due to the 2-fold reduced pixel transfer in case of the mixed-resolution prediction. These results demonstrate the applicability of the methodology also for real-world applications.

Discussion

The recursive inference algorithm, which identifies and performs a higher resolution inference for a single RoI per tile is suitable to proof the concept based on the sample tiles from the synthetic and Sen1Floods11 dataset. However, it does not depict its full potential which can unfold for large scale regional segmentation in locations of extended regions of a single class, e.g. in oceans and lakes or in forests and deserts, for the water and landmass classes, respectively. High-resolution images would only have to be requested along shore-lines. The RoI would be identified by an entropy threshold for each of the sliding windows, instead of single RoI window.

Resolution	$\bar{J}$	**Pixels** p	J**-gain density** $\bar{g}$
Synthetic			
Low k_2	0.923	22'500	baseline
Mixed k_2+k_1	0.935	45'000	0.048
High k_1	0.949	90'000	0.035
Sen1Flood11			
Low k_2	0.532	65'536	baseline
Mixed k_2+k_1	0.546	131'072	0.056
High k_1	0.563	262'144	0.041

Table 1: Mean semantic segmentation performance $\left(\bar{J}\right)$ for low, mixed and high-resolution images and the resulting mean J-gain density compared to the low-resolution baseline case, for the synthetic and Sen1Flood11 dataset.

Threshold Based Recursive Inference

A data transfer mitigation estimation for the entropy threshold based recursive inference approach was done. It is a function of the uncertainty area ratio, which represents the ratio of the area in the region with an uncertainty beyond the threshold, relative to the entire region to be segmented and the hierarchy levels. Figure 4 depicts the result considering a scaling factor of $r = 2$. The case with one hierarchy level performs only one inference step, based on images with the highest resolution for the entire region. Thus, its data transfer volume is independent of the uncertainty area ratio and will be the baseline for the cases with two and more hierarchy levels. The recursive inference mitigates data volumes most effectively considering a low uncertainty area ratio and many hierarchy levels. At an uncertainty area ratio of 0.01% savings of 75% to 98% can be expected for hierarchy levels from 2 to 4 (see also Table 2). Those values are close to the asymptotic case, which follows

$$\frac{1}{(r^2)^{n-1}} \tag{5}$$

for n hierarchy levels. For those cases none of the higher-resolution images need to be requested. However, in regions with 100% uncertainty area ratio, the data transfer increases up to 33% compared to the case with a single hierarchy level, as all data from all hierarchies has to be transferred, for the entire region. This result is equivalent to the data pyramid overhead as discussed in the section Data Pyramid Preparation. The relative data volume is one for all the cases considering an uncertainty area ratio of 75%.

In the future, this theoretical scaling study will be tested on real-world data from regions with varying uncertainty area ratios, to identify the practical range of savings.

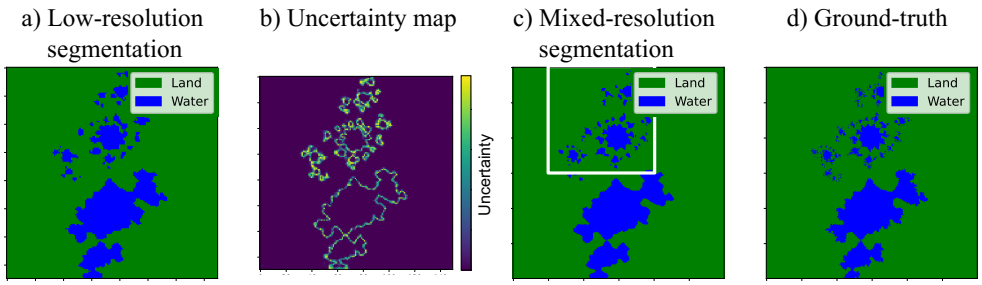

Figure 2: Segmentation results and uncertainty map of a synthetic sample: a) with low-resolution k_2, c) with mixed-resolution (mainly k_2, but with k_1 inserted into the window) segmentation and d) high-resolution ground-truth. b) is the uncertainty map from the low-resolution image segmentation.

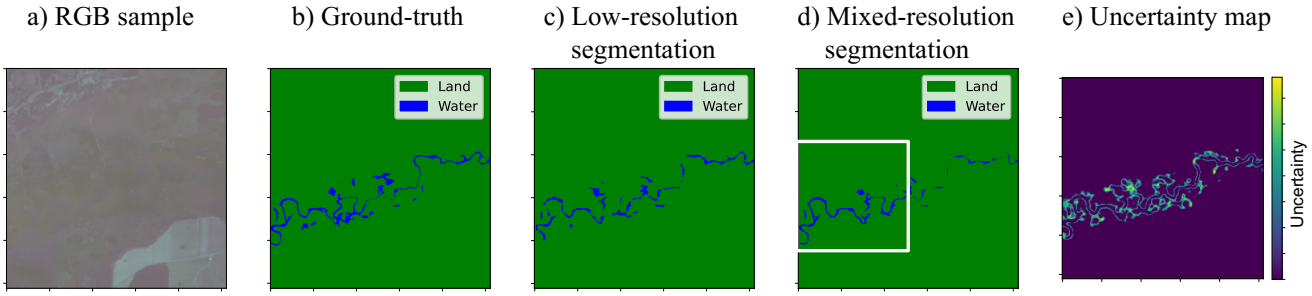

Figure 3: Segmentation results and uncertainty map of a Sen1Flood11 sample: a) RGB remote-sensing sample, b) high-resolution ground-truth, c) low-resolution k_2 and d) mixed-resolution (mainly k_2, but with k_1 inserted into the window) segmentation. e) is the uncertainty map from the low-resolution image segmentation.

Hierarchy levels	**Area ratio [%]**		**Resolution**
	0.01	100	[m]
1	1.000	1.000	10
2	0.250	1.25	20
3	0.063	1.31	40
4	0.016	1.33	80

Table 2: Relative data volume for extreme uncertainty area ratios vs. hierarchy levels for a scaling factor $r = 2$ and image resolution example considering a 10m spatial-resolution for hierarchy level 1.

Conclusion

In this study, we introduced a recursive inference approach to mitigate the data transfer between data centers of a remote sensing data host and the service provider producing insights based on those data through machine learning. The recursive inference approach is based on i) data pyramid preparation in the host data center, ii) a semantic segmentation model also predicting uncertainty maps from low-resolution data and a request of higher resolution data in areas of high uncertainty, and iii) the merging of the predictions with the different resolutions.

The proof of concept was performed applying a U-Net with Monte-Carlo dropout on synthetic and real-world remote sensing data with the aim of flood segmentation. Two hierarchy levels were considered and a maximum entropy sliding window approach was applied to identify the Region of Interest where higher resolution data is being requested for detailed segmentation. As expected, the uncertainty was high between regions of different classes, e.g. along rivers or shore lines. The approach yielded a 35-38% performance gain per pixel for the synthetic and real-world use-case, respectively.

However, the true potential of the algorithm was estimated through a theoretical scaling study which indicates a data traffic reduction of up to 98% in areas of extended water or land masses and by applying a sliding window and entropy threshold approach, to identify Regions of Interest.

The results showcase the effectiveness of the recursive inference methodology in achieving sustainable and accurate semantic segmentation of remote sensing data. The approach holds promise for mitigating the economical and environmental impacts associated with processing large-scale remote sensing data sets, opening possibilities for real-time applications in various domains.

Future work should focus on the implementation of the threshold-based recursive inference approach and demon-

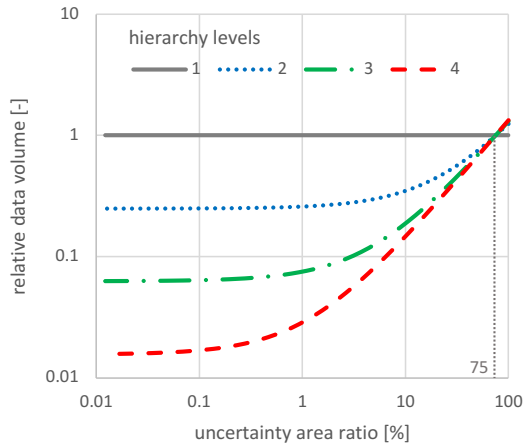

Figure 4: Relative data volume scaling vs. uncertainty area ratios and hierarchy levels with a scaling factor of $r = 2$.

strate it on real-world data. In addition, a holistic system analysis should be performed to estimate the energy and green-house-gas emission gains from the recursive inference approach, including data transfer, as well as energy spent on computation for the inference, which was not investigated in this study.

References

Abraham, N.; and Khan, N. M. 2019. A Novel Focal Tversky Loss Function With Improved Attention U-Net for Lesion Segmentation. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, 683–687.

Bonafilia, D.; Tellman, B.; Anderson, T.; and Issenberg, E. 2020. Sen1Floods11: A Georeferenced Dataset to Train and Test Deep Learning Flood Algorithms for Sentinel-1. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*.

Copernicus. 2022. Copernicus Sentinel Data Access Annual Report 2021. https://scihub.copernicus.eu/twiki/pub/SciHubWebPortal/AnnualReport2021/COPE-SERCO-RP-22-1312_-_Sentinel_Data_Access_Annual_Report_Y2021_merged_v1.0.pdf. Accessed: 2023-11-19.

Kwon, Y.; Won, J. H.; Kim, B. J.; and Paik, M. C. 2020. Uncertainty quantification using Bayesian neural networks in classification: Application to biomedical image segmentation. *Computational Statistics Data Analysis*, 142: 106816.

Larkin, K. G. 2016. Reflections on Shannon Information: In search of a natural information-entropy for images. *CoRR*, abs/1609.01117.

Maksoud, S.; Zhao, K.; Hobson, P.; Jennings, A.; and Lovell, B. C. 2020. SOS: Selective Objective Switch for Rapid Immunofluorescence Whole Slide Image Classification. *CoRR*, abs/2003.05080.

Mandelbrot, B. 1967. How Long Is the Coast of Britain? Statistical Self-Similarity and Fractional Dimension. *Science*.

Muszynski, M.; Hölzer, T.; Weiss, J.; Fraccaro, P.; Zortea, M.; and Brunschwiler, T. 2022. Flood Event Detection from Sentinel 1 and Sentinel 2 Data: Does Land Use Matter for Performance of U-Net based Flood Segmenters? In *2022 IEEE International Conference on Big Data (Big Data)*, 4860–4867.

Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; Desmaison, A.; Kopf, A.; Yang, E.; DeVito, Z.; Raison, M.; Tejani, A.; Chilamkurthy, S.; Steiner, B.; Fang, L.; Bai, J.; and Chintala, S. 2019. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In Wallach, H.; Larochelle, H.; Beygelzimer, A.; d'Alché-Buc, F.; Fox, E.; and Garnett, R., eds., *Advances in Neural Information Processing Systems 32*, 8024–8035. Curran Associates, Inc.

Ronneberger, O.; P.Fischer; and Brox, T. 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 9351 of *LNCS*, 234–241. Springer. (available on arXiv:1505.04597 [cs.CV]).

Salehi, S. S. M.; Erdogmus, D.; and Gholipour, A. 2017. Tversky loss function for image segmentation using 3D fully convolutional deep networks. *CoRR*, abs/1706.05721.

Seyedhosseini, M.; and Tasdizen, T. 2016. Semantic Image Segmentation with Contextual Hierarchical Models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(5): 951–964.

Tabata, T.; and Wang, T. Y. 2021. Life Cycle Assessment of CO2 Emissions of Online Music and Videos Streaming in Japan. *Applied Sciences*, 11(9).

Thandiackal, K.; Chen, B.; Pati, P.; Jaume, G.; Williamson, D. F.; Gabrani, M.; and Goksel, O. 2022. Differentiable Zooming for Multiple Instance Learning on Whole-Slide Images. In *European Conference on Computer Vision*, 699–715. Springer.

Wang, L. 2021. SaNet: Scale-aware neural Network for Parsing Multiple Spatial Resolution Aerial Images. *CoRR*, abs/2103.07935.