

NEW ACTIVE LEARNING APPROACH FOR SEABED SEGMENTATION

Lionel Pibre^{+†} Jérôme Pasquet^{+†} Vincent Douzal[†] Claire Noël^{*}

⁺AMIS, Université de Montpellier 3, Montpellier, France

[†]TETIS - Inrae, AgroParisTech, Cirad, CNRS, Univ. Montpellier, Montpellier, France

^{*}Semantic TS, Sanary-sur-Mer, France

ABSTRACT

Many fields are interested in mapping and monitoring the sea floor natural structure and biology, for environmental surveys, including identifying macro waste or detection of submerged artifacts such as cars, tyres, wrecks, and even military applications, e.g., to mine warfare whose detection depends heavily on the seabed structure. In this paper, we propose a new active learning method to improve seabed segmentation by deep learning. We perform segmentation of the sea floor using two data sources, sonar, and bathymetry. We train a network to fuse these two modalities and segment each sea floor pixel into nine fine ecological classes, then into three gross class sets, alive/not alive, and, in two different ways, whether mines can be hunted for or not. Once this training is done, a second stage involving a new active learning method based on network uncertainties greatly improves the performance.

Index Terms— Deep learning, Active learning, Sea floor segmentation, Sea floor biology, Multimodal fusion

1. INTRODUCTION

Semantic segmentation, i.e., the classification of an image at the pixel level, is an extremely important task for many fields. This task is often performed using deep neural networks. For example, in remote sensing this type of approach is used for land cover mapping [1], or for land-cover change detection monitoring [2]. These same techniques can be applied to seabed segmentation [3].

However, even though deep neural networks are very effective, this task remains extremely complex. Current state-of-the-art methods typically require specific architectures and fine-tuning to achieve high performance but still imperfect results. Moreover, the switch on “real” data sets remains tedious due to various factors. For example, the training and test data may be mismatched due to different acquisition conditions, or due to different sensor calibration. Human intervention may then be required.

One possible way to solve these problems is to add a human in the loop to work in synergy with a learning algorithm. The role of the human in this loop will then be to train, refine or adapt the learning algorithm. This method is called interactive learning [4, 5]. However, including a human in the learning loop is unwieldy with deep neural networks because they require a long time to converge.

A more appropriate strategy is *active learning* [6, 7]. The basic idea is that the most informative data to improve the model are the yet unlabeled cases for which its predictions are worst. These cases are closely related to Shannon’s definition of information as expectation. [8]. Thus the model points to cases to be added to the learning

base, either by using the uncertainty of the responses [9], or an indication of the representativeness of the resulting dataset [7, 10].

This paper has the following structure: Section 2 presents the related work, and Section 3 details our seabed segmentation approach. Section 4 describes the method used to perform active learning, and Quantitative and qualitative results including a discussion are given in Section 6. Last, Section 7 concludes and mentions future work.

2. RELATED WORK

An experimenter in automated learning obviously builds a database reflecting “as much as possible” that part of the world aimed at representing. This is of course open-ended and cannot be controlled as in classical design of experiments [11, 12], although principles found in stratified sampling are at least intuitively applied for ‘representativeness’ [13]. Note however that the pursued aim is determining: for a poll, sampling should reflect the size of the classes; whereas to train a classifier, it should balance the classes. On a freezed database, basic model identification is termed passive learning. On complex datasets, the experimenter cannot completely anticipate which data are more informative. In (inter)active learning a human oracle keeps interacting in the process at various stages.

An active learning algorithm includes an acquisition function, based either on intrinsic uncertainty of prediction (low contrast in the model itself) [14] or on an extrinsic diversity measure (contrast within the dataset, expressed by adjunct estimators) [15], which are used to request additional labels on examples. Both aim at pointing what data is most relevant, *in the terms of the model* or additional measures, not the eye of the experimenter.

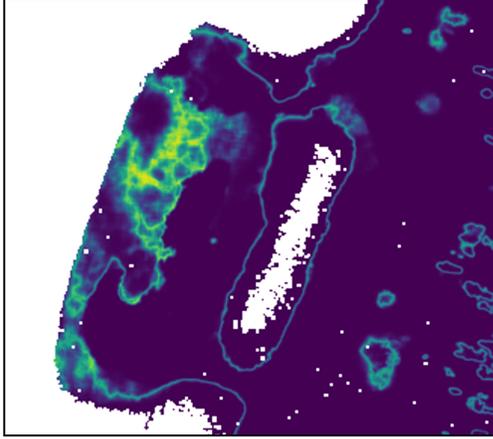
Uncertainty-based methods can rely on simple criteria such as entropy [16]. When dealing with an ensemble of models, the authors of [17] estimate the prediction confidence of the model using the disagreement between the models in the ensemble.

Independently of uncertainty-based acquisition functions, methods to maximize information content when building the training base [18] (or as mentioned, the contrast between samples), their diversity or representativeness can be utilized to actively boost learning by reducing the training set, by doing stratified sampling in the candidate population. To this end, some authors use pre-clustering [19, 20] to sample from each cluster and build learning batches. Others use a representativeness measure built from a radial basis function on observed frequencies, combined with an informativeness measure built from a best versus second best strategy [21].

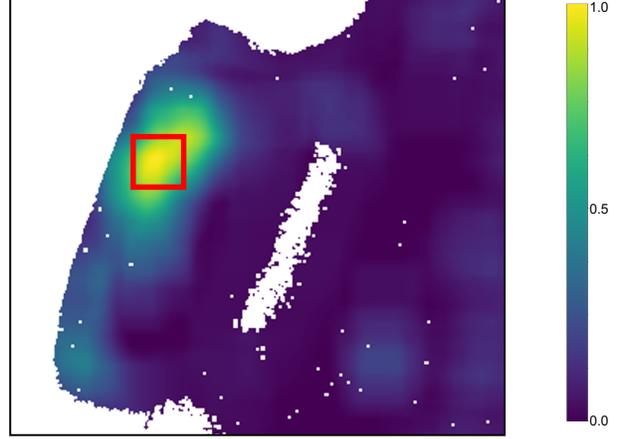
Over the past decade, active learning has been deeply explored in the field of remote sensing to train algorithms for animal detection [22], image classification [23], image segmentation [24] and recently for change detection [25].

Our approach utilizes the intrinsic uncertainty of the network

This paper and the research behind it would not have been possible without the work of Semantic TS which provided us with high quality data.



(a) Display of $1 - P(c_{max})$, uncertainty of network prediction.



(b) Result of applying a smoothing kernel to wipe out boundaries.

Fig. 1: Choosing patches of uncertain prediction in test images for fine tuning the network by active learning. Figure 1a maps the uncertainty of network predictions $1 - P(c_{max})$ on a test image. Two different features stand out: the boundaries between classes as fine contrasted lines, and large contiguous areas. Only the latter are informative of needing prediction improvement, because decisions on boundaries should be divided. These boundaries are canceled by a smoothing kernel in Figure 1b, keeping clear, large zones of high uncertainty to focus for improvement by active learning.

expressed by its softmax output to spot the particular areas needing improved prediction.

3. SEABED SEGMENTATION

For the fusion of the the two modalities, sonar and bathymetry, we used a cross fusion [26]. An illustration of this fusion is shown in Figure 2. Taking the i -th pixel as an example, the fusion representation is then

$$\mathbf{a}_{1,i}^{(l)} = f_{\mathbf{w}_1^{(l)}, b_1^{(l)}}(\mathbf{a}_{1,i}^{(p)}) + f_{\mathbf{w}_1^{(l)}, b_1^{(l)}}(\mathbf{a}_{2,i}^{(p)}), \quad (1)$$

$$\mathbf{a}_{2,i}^{(l)} = f_{\mathbf{w}_2^{(l)}, b_2^{(l)}}(\mathbf{a}_{2,i}^{(p)}) + f_{\mathbf{w}_2^{(l)}, b_2^{(l)}}(\mathbf{a}_{1,i}^{(p)}), \quad (2)$$

$$\mathbf{v}_i = \begin{bmatrix} f_{\mathbf{w}_0^{(l+1)}, b_0^{(l+1)}}(\mathbf{a}_{1,i}^{(l)}) & f_{\mathbf{w}_0^{(l+1)}, b_0^{(l+1)}}(\mathbf{a}_{2,i}^{(l)}) \\ f_{\mathbf{w}_1^{(l)}, b_1^{(l)}}(\mathbf{a}_{1,i}^{(p)}) & f_{\mathbf{w}_2^{(l)}, b_2^{(l)}}(\mathbf{a}_{1,i}^{(p)}) \\ f_{\mathbf{w}_1^{(l)}, b_1^{(l)}}(\mathbf{a}_{2,i}^{(p)}) & f_{\mathbf{w}_2^{(l)}, b_2^{(l)}}(\mathbf{a}_{2,i}^{(p)}) \end{bmatrix} \quad (3)$$

where the three components (each row of the matrix) of \mathbf{v}_i in Eq. 3 share the same to-be-learned parameters. In other words, they can be also seen as three “new” different samples for the input of the next layer to enforce a more compact fusion.

Before the cross fusion, each modality goes through 3 blocks of two convolutions with a ReLU [27] activation followed by a batch normalization [28] and an averaging pooling.

Once this fusion was done, we used a U-Net residual network [29, 30] architecture. The encoder and decoder are each composed of 50 convolutional layers. At the end of our decoder, we have a branch composed of eight convolutional layers per task. This network is trained in a multi-task way, i.e. all our tasks are learned at the same time.

4. ACTIVE LEARNING METHOD

We were allowed to reuse data acquisitions made for different authorities, with classifications meant for their specific purposes. The

classes, although reworked to be made homogeneous and apparently consistent, proved inappropriate from the point of view of automated learning. No less than three data labeling revisions were necessary through interactions back and forth between the sonar and sea floor specialists and the deep learning team. This was effectively a necessary *interactive learning*, also building a common ground, a shared context for mutual understanding of the two teams, and decisive for the results. During the process, the inconsistent output of the network signaled labeling inconsistencies to be revised.

There remained unavoidable gaps. For instance, Cymodocea often form thin, branched colonies which cannot be hand-labeled at the pixel level, only by a reasonable convex envelope. In spite of this gross, and partly erroneous indication, the network learnt to finely outline Cymodocea, resulting in a humanly-judged better result than what was given as “ground truth”.

As seen in Table 1, the classes are highly unbalanced, which implied an appropriate stratified sampling as a first active step. Furthermore, as can be seen in Figure 1, we have missing data. Indeed, all the white areas of the image correspond to pixels for which we have no sonar or bathymetry information, or which have no annotation. Once an appropriate network architecture was found, with satisfactory performance, we took an additional active step, which again seeks to compensate for unbalanced classes:

- We select areas from the test images where the network gives the predicted class with a low score, and pick there a small area, say 100×100 pixels.
- We dispatch these additional samples in the learning batches, and fine tune the network (of course, removing these patches from the test scores).

The predicted class is that of the highest value at the softmax output of the network, numerically behaving like a probability, $P(c_{max})$. When $P(c_{max})$ is low, the response is uncertain. Thus the image resulting from $1 - P(c_{max})$, shown in Figure 1a, expresses uncertainty. The boundary lines, where the network switches between classes are visible, as they should. We wipe them out by

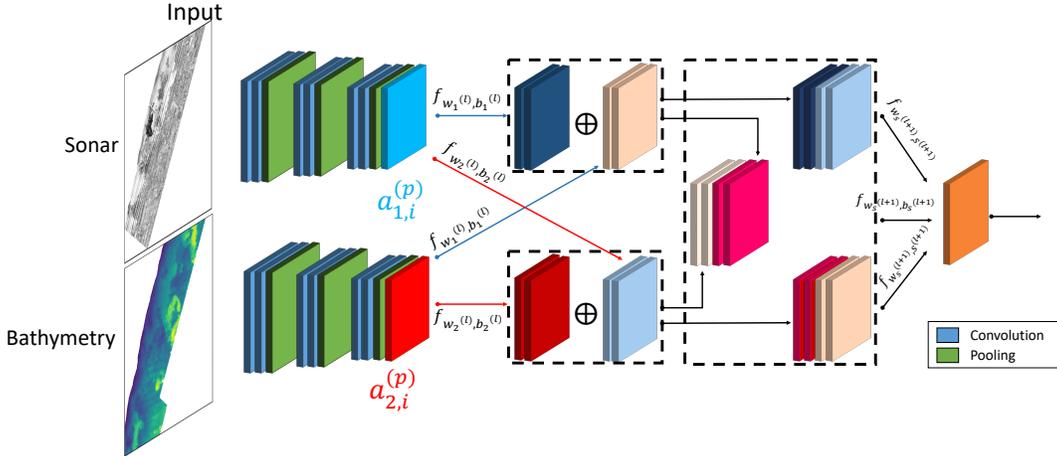


Fig. 2: Illustration of the cross fusion [26] used in our architecture.

applying a smoothing kernel, keeping only substantial areas of uncertainty, in Figure 1b. To ensure that we select an area that will allow the network to improve, the smoothing kernel ignores pixels without information (white pixels). Thus, we make sure to avoid taking an area of 100 by 100 pixels containing lots of pixels without information that will not be used to refine the network. On this image, we select the 100×100 sample of highest value for additional active learning. Of course we could choose to select more than one zone, and vary their sizes. We complete the fine tuning batch with random samples from the training database.

5. EXPERIMENTAL PROTOCOL

5.1. Data

Our database is composed of 32 pairs of sonar and bathymetry images. For each pair, we have four different annotations, detailed in Table 1. The finest, “REAL” has 9 classes: Cymodocea, Posidonia, Mat (a decimeter-thick slab formed by dead Posidonia rhizomes and sediment), Sediment, Rippled sediment, Gravel, Rock, Riprap, and Anthropic (such as pipes and other pieces of work). The 3 others gather these classes for specific purposes, into two or three classes: “LiNI” (Living/Non-living), and for mine warfare “HuNh” (Huntable/Non huntable) or “HoAnCo” (Homogeneous/Anisotropic/Complex). Many pixels are unlabeled, or lack either sonar or bathymetry. They are set to NO-VALUE and masked out for processing.

The images are of variable size, the smallest is 584×572 pixels, and the largest is 9238×15851 pixels, with an average of 2606×3239 .

5.2. Learning

Training settings. Training batches are composed of 10 patches of size 512×512 , sampled from the database by stratification to compensate for unbalanced classes (see Section 4). We use the Adam [31] optimizer with a learning rate of $1e-4$. Training is stopped after 200 epochs of 200 batches each.

Active learning settings. The batches for fine tuning keep 9 patches as above from the learning base and one patch selected for low prediction certainty on a test image, as described in section 4. In this 512×512 patch, only a 100×100 central square is annotated and used for learning. Of course these central squares are masked out

Classes	digram	# Pixels
REAL:		
Cymodocea	Cy	11 226 658
Posidonia	Po	128 337 503
Mat	Ma	2 118 059
Sediment	Se	145 111 715
Rippled sediment	Rs	3 565 197
Gravel	Gr	1 540 421
Rock	Ro	5 056 930
Riprap	Ri	2 941 621
Anthropic	Ah	40 645
Living	Li	Cy Po Ma 141 682 220
Non-living	Nl	Se Rs Gr 158 256 529
		Ro Ri Ah
Homogeneous	Ho	Se 145 111 715
Anisotropic	An	Rs 3 565 197
Complex	Co	Cy Po Ma Gr 151 261 837
		Ro Ri Ah
Huntable	Hu	Se Rs Gr 159 903 570
Non-huntable	Nh	Cy Po Ma 140 035 179
		Ro Ri Ah
NO-VALUE	Nv	361 517 967

Table 1: Number of pixels for each class in our database.

of the images to establish the test results. We use the Adam [31] optimizer with a learning rate of $1e-5$. Training is stopped after 5 epochs, thus for each test image, the network checks the area on which it is least confident five times.

6. EVALUATION OF OUR APPROACH

We used ensembles of three networks, on six cross-validation sets, with results based on accuracy, balancing True Prediction (TP) and False Prediction (FP): $Accuracy = TP / (TP + FP)$.

Table 2 shows our baseline results on the four tasks. To evaluate our ensembles, we removed from the test images the 100×100 squares that will be used to perform our active learning. Table 3 shows the results obtained after our active learning method.

As expected, Table 2 shows lowest accuracy values for the REAL line, the most demanding task. The three other, simpler, classification tasks reach naturally higher scores, probably because they somewhat present more balanced classes. The other main effect is the dis-

Table 2: Baseline results obtained with our ensembles of networks on our six cross-validations.

Task Exp	CV1	CV2	CV3	CV4	CV5	CV6	Average
LiNI	86.17	94.46	93.06	90.62	93.13	96.49	92.32
HuNh	87.97	92.56	93.31	86.64	94.10	96.58	91.86
HoAnCo	88.20	92.51	93.36	86.94	94.22	96.57	91.97
REAL	75.24	91.47	90.33	83.87	86.34	95.82	87.18

Table 3: Results obtained after active learning with our network ensembles on our six cross-validations, using a single additional 100×100 annotated area picked from each test image.

Task Exp	CV1	CV2	CV3	CV4	CV5	CV6	Average
LiNI	94.80	97.17	95.92	95.39	96.67	97.34	96.22
HuNh	95.57	96.84	96.36	93.72	97.26	97.37	96.19
HoAnCo	95.64	96.84	96.35	93.69	97.30	97.37	96.20
REAL	90.65	96.46	94.85	93.69	94.92	97.06	94.44

crepancy between cross-validations, with CV1 and to a lesser extent CV4 displaying lower accuracy scores. We attribute this to the compound effect of highly unbalanced classes, and highly unbalanced distribution of the classes among images in the database as well as the lack of a more comprehensive data set. All these reasons can lead to a mismatch between training and test data.

These effects are spectacularly compensated by picking just a few areas of low prediction for active fine tuning, as shown in Table 3: the first three task accuracies shoot up by 4%, leaping from about 92% to 96%, and remarkably homogeneous results which tend to confirm that learning suffers from a small image database with unbalanced representation of sea floor environments.

The greatest improvement is for the full classification task, on the REAL line, with an average rising by 7%, from 87% to 94%, and again, perhaps more significantly, considerably leveling off the differences between cross-validations.

We also investigated the effects of our approach on accuracy by class. These results are presented in Table 4. As we can see, our approach allows a gain on all the classes, with as expected a higher gain on the less represented classes. Also, when a class is represented in the selected area of 100×100 pixels, its gain is more significant, as for example the Mat class (Ma) gains 47%. However even the under-represented classes also show a significant gain, for example, the Anthropoc class (Ah) gains 28%. We can therefore

Table 4: Comparison of accuracy per class between baseline and active learning results.

Classes	Baseline Acc %	Active learning Acc %	%Pixel in training	% Label in active learning area	Gain
Cy	58.56	76.51	3.74	7.54	17.95
Po	92.88	93.24	42.79	24.87	0.36
Ma	7.17	54.31	0.71	15.63	47.15
Se	89.72	93.52	48.38	31.37	3.80
Rs	32.23	71.89	1.19	4.86	39.66
Gr	4.24	26.19	0.51	6.71	21.95
Ro	24.66	59.39	1.69	6.83	34.74
Ri	63.31	78.81	0.98	2.19	15.50
Ah	10.54	38.69	0.01	0.00	28.15
AVG	42.59	65.84			23.25

deduce that this allows a more efficient network generalization.

Finally, comparing the distribution of classes in the training and selected active learning areas (Table 4), we notice that active learning helps reduce even more the class imbalance.

Since the diversity of samples presented for training seemed at stake, we experimented fine tuning by picking two 50×50 (non-overlapping) areas instead of one 100×100 . Table 5 compares the results of the two strategies. Using two smaller samples for active learning improves the three coarse classification tasks, but degrades the accuracy on the full, REAL one.

Table 5: Comparison of fine tuning by active learning with one area of 100×100 pixels, or two areas of 50×50 pixels.

Exp Task	LiNI	HuNh	HoAnCo	REAL
1 area	3.89	4.32	4.23	7.26
2 areas	3.96	4.45	4.38	7.05

The observed effect seem to indicate that losing 3/4 of additional information on the highest uncertain zone is harmful to the REAL classification, not compensated by having another sample somewhere else on the image. One substantial peak zone of uncertainty is more effective for the REAL task, and appears necessary to tweak the network in the direction of improvement. On the other hand, for the three coarser classification tasks, pointing one decisive zone proves less effective. This seems consistent with the REAL task being the most significantly improved in Table 3. A fine analysis would be needed to show whether the peak uncertainty area is actually not informative regarding the corresponding coarse classes. At any rate, the observed differences show that “maximizing information”, diversity or representativeness is not an easy concept to grasp; it is highly dependent on the actual task, and it is therefore appropriate to use the model itself—here, the network—to indicate where the needed information seems to rest.

7. CONCLUSION

This paper presented a new active learning approach to significantly improve performance. The main idea is to fine tune our networks using a very small area of the test images. Areas where the network is the least confident in its prediction are candidates to be the most informative for fine tuning, and prove to greatly improve the results.

We performed two experiments. In the first one, we used a single area of 100×100 pixels in each test image. In the second, we extracted two areas of 50×50 pixels. Whenever possible, these two areas should not overlap. Our approach significantly improves the performance.

Indeed, by using only a very small area of 100×100 pixels in the test image (less than 0.12% when taking a medium size image), we have a gain of more than 7% on the REAL task. Furthermore, even though we relied on the output of the REAL task to select our uncertainty areas, we improved all other tasks by 4%.

8. ACKNOWLEDGEMENT

This work is supported by a DGA-RAPID project under grant [202906108].

9. REFERENCES

- [1] Krishna Karra, Caitlin Kontgis, Zoe Statman-Weil, Joseph C Mazzariello, Mark Mathis, and Steven P Brumby, "Global land use/land cover with sentinel 2 and deep learning," in *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*. IEEE, 2021, pp. 4704–4707.
- [2] Weiwei Yang, Haifeng Song, Lei Du, Songsong Dai, and Yingying Xu, "A change detection method for remote sensing images based on coupled dictionary and deep learning," *Computational Intelligence and Neuroscience*, vol. 2022, 2022.
- [3] Maryam Rahnemounfar and Dugan Dobbs, "Semantic segmentation of underwater sonar imagery with deep learning," in *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*, 2019, pp. 9455–9458.
- [4] Michael Schroder, Hubert Rehauer, Klaus Seidel, and Mihai Datcu, "Interactive learning and probabilistic retrieval in remote sensing image archives," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 38, no. 5, pp. 2288–2298, 2000.
- [5] Bertrand Le Saux and Martial Sanfourche, "Rapid semantic mapping: Learn environment classifiers on the fly," in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2013, pp. 3725–3730.
- [6] Burr Settles, "Active learning literature survey," 2009.
- [7] Marin Ferecatu and Nozha Boujemaa, "Interactive remote-sensing image retrieval using active relevance feedback," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 4, pp. 818–826, 2007.
- [8] Claude Elwood Shannon, "A mathematical theory of communication," *The Bell system technical journal*, vol. 27, no. 3, pp. 379–423, 1948.
- [9] Yarin Gal, Riashat Islam, and Zoubin Ghahramani, "Deep bayesian active learning with image data," in *International Conference on Machine Learning*. PMLR, 2017, pp. 1183–1192.
- [10] Ozan Sener and Silvio Savarese, "Active learning for convolutional neural networks: A core-set approach," *arXiv preprint arXiv:1708.00489*, 2017.
- [11] D.C. Montgomery, *Design and Analysis of Experiments*, John Wiley & Sons, Incorporated, 9th edition, 2017.
- [12] Herbert Robbins, "Some aspects of the sequential design of experiments," *Bull. Amer. Math. Soc.*, vol. 58, pp. 527–535, 1952.
- [13] Carl-Erik Särndal, Bengt Swensson, and Jan Wretman, *Model assisted survey sampling*, Springer series in statistics. Springer, 2003.
- [14] David D Lewis and William A Gale, "A sequential algorithm for training text classifiers," in *SIGIR'94*. Springer, 1994, pp. 3–12.
- [15] Klaus Brinker, "Incorporating diversity in active learning with support vector machines," in *Proceedings of the 20th international conference on machine learning (ICML-03)*, 2003, pp. 59–66.
- [16] Seokhwan Kim, Yu Song, Kyungduk Kim, Jeong-Won Cha, and Gary Geunbae Lee, "Mmr-based active machine learning for bio named entity recognition," in *Proceedings of the human language technology conference of the NAACL, Companion Volume: Short Papers*, 2006, pp. 69–72.
- [17] Lars Kai Hansen and Peter Salamon, "Neural network ensembles," *IEEE transactions on pattern analysis and machine intelligence*, vol. 12, no. 10, pp. 993–1001, 1990.
- [18] Y. LeCun, L. Bottou, G. Orr, and K. Muller, "Efficient backprop," in *Neural Networks: Tricks of the trade*, G. Orr and Muller K., Eds. 1998, Springer.
- [19] Hieu T Nguyen and Arnold Smeulders, "Active learning using pre-clustering," in *Proceedings of the twenty-first international conference on Machine learning*, 2004, p. 79.
- [20] Begüm Demir, Claudio Persello, and Lorenzo Bruzzone, "Batch-mode active-learning methods for the interactive classification of remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 3, pp. 1014–1031, 2010.
- [21] Bo Du, Zengmao Wang, Lefei Zhang, Liangpei Zhang, Wei Liu, Jialie Shen, and Dacheng Tao, "Exploring representativeness and informativeness for active learning," *IEEE transactions on cybernetics*, vol. 47, no. 1, pp. 14–26, 2015.
- [22] Benjamin Kellenberger, Diego Marcos, Sylvain Lobry, and Devis Tuia, "Half a percent of labels is enough: Efficient animal detection in uav imagery using deep cnns and active learning," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 12, pp. 9524–9533, 2019.
- [23] Lorenzo Bruzzone and Claudio Persello, "Active learning for classification of remote sensing images," in *2009 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2009, vol. 3, pp. III–693.
- [24] Jieliang Guo, Xiong Zhou, Jun Li, Antonio Plaza, and Saurabh Prasad, "Superpixel-based active learning and online feature importance learning for hyperspectral image analysis," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 1, pp. 347–359, 2016.
- [25] Vít Ruzicka, Stefano D'Arconco, Jan Dirk Wegner, and Konrad Schindler, "Deep active learning in remote sensing for data efficient change detection," in *MACLEAN@ PKDD/ECML*, 2020.
- [26] Danfeng Hong, Lianru Gao, Naoto Yokoya, Jing Yao, Jocelyn Chanussot, Qian Du, and Bing Zhang, "More diverse means better: Multimodal deep learning meets remote sensing imagery classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 5, pp. 4340–4354, 2021.
- [27] Vinod Nair and Geoffrey E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proc. International Conference on Machine Learning*, 2010, pp. 807–814.
- [28] Sergey Ioffe and Christian Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *International conference on machine learning*, 2015, pp. 448–456.
- [29] Yoseob Han, Jae Jun Yoo, and J. C. Ye, "Deep residual learning for compressed sensing CT reconstruction via persistent homology analysis," *ArXiv*, 2016.
- [30] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE.
- [31] Diederik P. Kingma and Jimmy Ba, "Adam: A method for stochastic optimization," in *3rd International Conference on Learning Representations, ICLR, Yoshua Bengio and Yann LeCun, Eds.*, 2015.