

Segmentation of remotely sensed images with a neuro-fuzzy inference system

Giovanna Castellano, Ciro Castiello, Andrea Montemurro, Gennaro Vessio and Gianluca Zaza

Department of Computer Science, University of Bari Aldo Moro, Italy

Abstract

The semantic segmentation of remotely sensed images is a difficult task because the images do not represent well-defined objects. To tackle this task, fuzzy logic represents a valid alternative to convolutional neural networks—especially in the presence of very limited data—, as it allows to classify these objects with a degree of uncertainty. Unfortunately, the fuzzy rules for doing this have to be defined by hand. To overcome this limitation, in this work we propose to use an adaptive neuro-fuzzy inference system (ANFIS), which automatically infers the fuzzy rules that classify the pixels of the remotely sensed images, thus realizing their semantic segmentation. The resulting fuzzy model guarantees a good level of accuracy in the classification of pixels despite the few input features and the limited number of images used for training. Moreover, unlike the classic deep learning approaches, it is also explanatory, since the classification rules produced are similar to the way of thinking of human beings.

Keywords

Remote sensing, semantic segmentation, neuro-fuzzy model, ANFIS

1. Motivations and objectives

Remote sensing images cover diverse applications in meteorology, agriculture, geology, biodiversity conservation, land use planning, education, intelligence and warfare (see, for example, [1, 2]). The images can be acquired in color in the visible spectrum (for example in RGB) or in other electromagnetic spectra (for example in the infrared). The *semantic segmentation* of this type of images is an important task to make them usable and understandable also by human beings, in the fields of application mentioned above [3].

In recent years, the dominant approach to tackle this task has been the one based on Convolutional Neural Networks (CNNs). Popular models used for semantic segmentation include U-Net [4] and SegNet [5], and many variants have been proposed in the particular remote sensing domain (e.g., [6, 7]). However, while successful, the CNN-based approach has some major drawbacks. First, huge labeled training sets are typically required to develop accurate models, and the data collection process can be very expensive and time-consuming. Second, the

WILF 2021: 13th International Workshop on Fuzzy Logic and Applications

✉ giovanna.castellano@uniba.it (G. Castellano); ciro.castiello@uniba.it (C. Castiello); a.montemurro23@studenti.uniba.it (A. Montemurro); gennaro.vessio@uniba.it (G. Vessio); gianluca.zaza@uniba.it (G. Zaza)

🆔 0000-0002-6489-8628 (G. Castellano); 0000-0002-8471-5403 (C. Castiello); 0000-0002-0883-2691 (G. Vessio); 0000-0003-3272-9739 (G. Zaza)



© 2021 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

CNN-based approach generally does not provide human-understandable explanations of the process by which a certain output is obtained. This clearly limits the end user’s confidence in the developed model and does not help to use the given prediction in decision-making processes.

On the other hand, fuzzy inference systems, as an extension of traditional expert systems, can best use small and inherently inaccurate data to generate simple, but accurate models that are much easier to interpret and explain. A fuzzy system, in fact, can explicitly represent image classification decisions in the form of IF-THEN rules, which are very easy to understand [8, 9]. For example, a fuzzy inference system for the explainable semantic segmentation of trees in satellite images has recently been proposed in [10]. The system uses the HSV color space as input, and is optimized using a Big Bang-Big Crunch evolutionary procedure. Similarly, a semantic segmentation method based on an interval type-2 fuzzy membership function for high-resolution remote sensing images has been presented in [11]. Domain knowledge is the primary source of the fuzzy rules and fuzzy set parameters on which an inference system is based. Therefore, acquiring this knowledge is laborious, error-prone and highly subjective.

One way to get around the limitations of both neural networks and fuzzy systems is to integrate them into a unified framework in which the shortcomings of one method are mitigated by the advantages of the other, and vice versa. This is already done in the literature in the form of so-called *neuro-fuzzy* models [12]. Basically, a neuro-fuzzy model is a fuzzy system in which a learning algorithm typically employed for neural networks, such as the popular backpropagation algorithm, is exploited to learn fuzzy set parameters directly from data. This paradigm is attracting growing interest in the remote sensing research area. For example, a fuzzy extreme learning machine has recently been proposed in [13] for the classification of remote sensing images based on a domain adaptation approach. Instead, a new deep fuzzy convolutional network has been presented in [14], which leverages fuzzy logic to provide more reliable segmentation results than classic CNN models.

To contribute to this research effort, in this paper we propose to use an adaptive neuro-fuzzy inference system (ANFIS) [15] for the semantic segmentation of remotely sensed images. The proposed method provides a better understanding of how the decision made by the system was reached. Furthermore, since it transforms the image-level segmentation of remotely sensed images into a pixel-level classification, the method requires a much lower labeled training set than is typically required. It is worth noting that in this work we focus on RGB images, which are much easier to collect with traditional optical cameras.

The next section presents the proposed method. Section 3 describes the experimental setup and the results obtained. Section 4 concludes the paper.

2. Neuro-fuzzy modeling

To learn fuzzy rules for pixel classification starting from available data, we trained a neuro-fuzzy model, that is a neural network that encodes a collection of IF-THEN fuzzy rules in its structure. To this end, we considered zero-order Takagi-Sugeno (TS) fuzzy models [16], whose rules have fuzzy sets in the antecedents and singleton values in the consequent parts. Formally, the TS fuzzy rules can be formalized as:

$$\text{IF } (x_1 \text{ is } A_{k1}) \text{ AND } \dots \text{ AND } (x_m \text{ is } A_{km}) \text{ THEN } (y_1 \text{ is } b_{k1}) \text{ AND } \dots \text{ AND } (y_C \text{ is } b_{kC}),$$

for $k = 1, \dots, K$, where K is the number of rules. Moreover, A_{ki} , ($i = 1, \dots, m$) are fuzzy sets defined over the input feature vector $\mathbf{x} = [x_1, \dots, x_m]$, C is the number of classes and b_{kj} is a fuzzy singleton expressing the degree of certainty of the output class y_j , ($j = 1, \dots, C$). Each fuzzy set is defined by a Gaussian membership function:

$$u_{ki} = \mu_{ki}(x_i) = \exp \left[- \left(\frac{x_i - c_{ki}}{\sigma_{ki}} \right)^2 \right], \quad (1)$$

where c_{ki} and σ_{ki} are the center and width of the Gaussian function, respectively. Gaussian functions are quite popular in the fuzzy literature: they are always continuous functions which can be defined on the basis of a reduced number of parameters. Also, their employment in a fuzzy rule base implies the functional equivalence between the fuzzy system and the radial basis function network [17], which is the inspiring idea of neuro-fuzzy modeling.

Once fuzzy rules have been defined, they are used to provide degrees of certainty for each class by means of the following fuzzy reasoning procedure:

1. Compute fuzzy membership values u_{ki} according to Eq. (1);
2. For $k = 1, \dots, K$, compute the activation level of the k -th rule using the product operator:

$$u_k(\mathbf{x}) = \prod_{i=1}^m u_{ki}; \quad (2)$$

3. Normalize the rule activation levels:

$$\mu_k(\mathbf{x}) = \frac{u_k(\mathbf{x})}{\sum_{h=1}^K u_h(\mathbf{x})}; \quad (3)$$

4. For each class, compute the degree of certainty:

$$y_j = \sum_{k=1}^K \mu_k(\mathbf{x}) \cdot b_{kj} \quad j = 1, \dots, C. \quad (4)$$

Fuzzy rules are typically defined manually by the domain expert who specifies the fuzzy terms in the antecedent and consequent of each rule. When expert knowledge is lacking (as it may be the case in an image segmentation task), the parameters of fuzzy rules can be better learned from the data by leveraging the learning capabilities of neural networks. This combination yields neuro-fuzzy networks, which are adaptive fuzzy systems capable of learning fuzzy rules from data [18]. Thus, a model learned by a neuro-fuzzy network can be represented by a set of fuzzy rules understandable to humans.

To learn the parameters of the fuzzy rules, namely the antecedent fuzzy set parameters c_{ki} and σ_{ki} , and the consequent parameters b_{kj} , we use a neuro-fuzzy architecture similar to ANFIS [15], which is a feed-forward neural network reflecting the fuzzy rule base in its parameters and topology. The architecture of the network (sketched in Fig. 1) includes four layers that realize the inference of the fuzzy rules by calculating, respectively: the membership degree to fuzzy sets; the activation level of each fuzzy rule; the normalized activation level for each fuzzy rule; the final output. Training is carried out through an optimization procedure of backpropagation-type based on gradient descent over the parameter space.

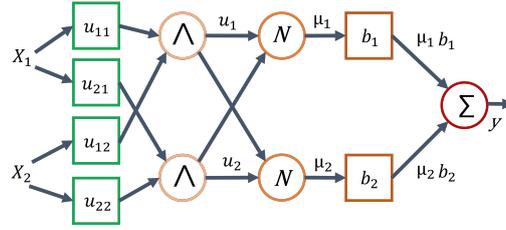


Figure 1: Simplified schema of the neuro-fuzzy network architecture.

3. Experiment

The experiment was conducted through an implementation of ANFIS in PyTorch.¹ The complexity of the network architecture may vary with respect to the scheme depicted in Fig. 1 since in our experiments we adopted different levels of granularity to partition the input ranges (as discussed in Section 3.2). As a running environment, we used Google Colaboratory, which provides powerful GPUs for free. Below, we will present the data and discuss the experimental results.

3.1. Data

To collect the data useful for our experiments, we referred to the Wuhan Dense Labeling Dataset (WHDLD) [19], which is a publicly available dataset suitable for image retrieval, classification or semantic segmentation. WHDLD includes images cropped from a larger satellite image of the Wuhan urban area: 4940 RGB images were obtained with the same pixel size (256×256) and resolution equal to 2 meters. The pixels of each image are manually labelled considering six classes: *vegetation*; *building*; *pavement*; *water*; *road*; and *bare soil*.

Since our goal is to produce a model for pixel classification, we considered a limited number of images from the WHDLD dataset. Actually, considering only 18 images, we were able to collect the information relating to 1,179,648 labelled pixels. As stated earlier, this represents a major advantage of our approach, as it is possible to train a classification model by referring to a limited amount of images (provided that they are chosen in such a way to preserve the balanced co-occurrence of different pixel classes, with reference to the distribution within the original dataset). As a reference, it can be observed that the training of the fully convolutional networks proposed in literature chosen for comparison (see the subsequent discussion in Section 3.2) has been performed on the basis of more than 250 million labelled pixels. The selected images consist of the following pixel categories: *vegetation* (26%); *building* (19%); *pavement* (18%); *water* (14%); *road* (12%); and *bare soil* (11%). The predominance of vegetation pixels reflects the composition of the original satellite image, where only a few limited portions of roads and bare soil were involved.

To tackle our pixel classification task, we built up a dataset starting from the analysis of the selected images. For each of them, a four-dimensional feature vector was considered. The

¹<https://github.com/jfpower/anfis-pytorch>

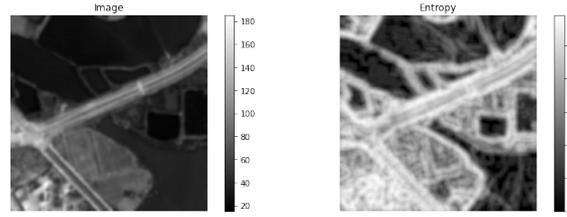


Figure 2: A sample (gray scale) image and the corresponding one obtained applying the entropy filter.

information embedded in the feature vector concerns the values of the red, green and blue channels of the pixel, and the value of the pixel when an entropy filter is applied to the image. In other words, each pixel is described in terms of its RGB components, together with an additional feature related to the local entropy information. This is evaluated by applying a filter to the gray-scale version of the image, which returns the minimum number of bits needed to encode the local gray-scale distribution. Entropy analysis is a way to classify textures in image processing applications. In fact, certain textures can be associated to specific entropy values depending on the repetition of patterns within the image. Therefore, low entropy values concern more homogeneous components than those characterized by higher entropy values. This information was included in the feature vector to provide an additional low-level feature for distinguishing image pixels. In Fig. 2, we show one of the images (on the left, in its gray scale version) chosen to build our dataset and the corresponding filtered image (on the right) obtained at the end of the entropy filter application. After feature extraction, each sample in the dataset is completed by the information concerning the pixel class.

3.2. Results and discussion

To derive the classification model from the data, some parameters must be set. A first choice concerns the number of fuzzy sets to be associated with the input features (i.e., the level of granularity to partition the ranges of input values). This is a crucial point, since more fuzzy sets generally lend themselves to a more adequate representation of the data distribution. However, higher granularity also increases the complexity of the resulting predictive model. Since 4 input features are involved, the choice of 2, 3, and 4 fuzzy sets per input leads to the realization of a model including 16, 64, 256 fuzzy rules, respectively. The additional parameters are the number of training epochs, which was set to 200, and the size of the training batch, which was set to 32.

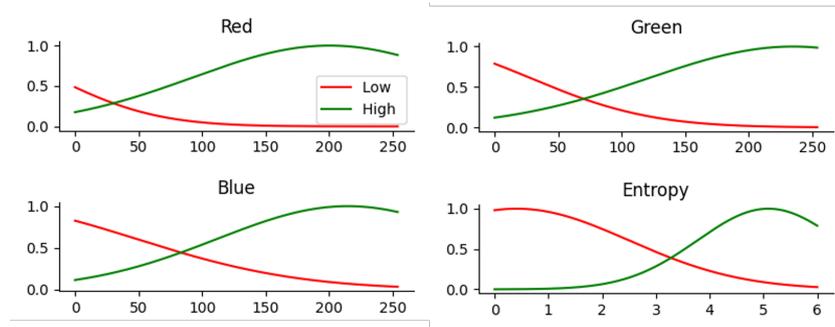
Table 1 reports the information relating to the three derived models (namely, NFC1, NFC2, and NFC3), including their composition (number of fuzzy sets per input feature, total number of rules), the training time of the corresponding neuro-fuzzy networks, and the value of their overall accuracy. The experiments were conducted considering the previously described dataset composed by 1,179,648 pixel datapoints, which was arranged in a stratified split of training and test sets equal to 70% and 30% of the samples, respectively.

As expected, the accuracy of the predictive models improves as their complexity increases. However, the aggravation in terms of training time and, most importantly, complexity exhibited by the NFC2 and NFC3 models is not rewarded by a significant boost in classification perfor-

Table 1

Details of the three neuro-fuzzy classification models derived from the data.

Model ID	# of fuzzy sets per input feature	# of rules	Training time	Overall accuracy
NFC1	2	16	3h 39m 13s	74.98%
NFC2	3	64	5h 20m 5s	76.50%
NFC3	4	256	7h 15m 51s	77.53%

**Figure 3:** Membership functions composing the rule antecedents of NFC1.**Table 2**

Performance of NFC1 evaluated in terms of precision, recall, and f1-score.

Class	Precision	Recall	F1-score
Vegetation	0.73	0.79	0.75
Building	0.75	0.83	0.79
Pavement	0.75	0.72	0.73
Water	0.93	0.87	0.90
Road	0.57	0.47	0.51
Bare Soil	0.77	0.72	0.75

mance. Therefore, we can refer in the following analysis to NFC1 which offers the benefit of being quite simple and readable for a human user. In Fig. 3 the membership functions composing the antecedents of the fuzzy rules involved in NFC1 are depicted for the sake of illustration. When applied to the test set, the neuro-fuzzy model showed the performance reported in Table 2, described in terms of precision, recall, and f1-score. To appreciate the interpretability of the resulting fuzzy model, an excerpt of rules of the NFC1 model is depicted in Fig. 4. These fuzzy rules offer an understanding of how the neuro-fuzzy system performs the image classification process. For example, it can be seen that pixels with low values in all channels are classified with high certainty as water, while pixels with high values in all three channels are classified as urban lands, i.e. building, but also road and pavement. Vegetation pixels generally exhibit a high value of the green channel, as expected.

Finally, to better assess the suitability of our proposal, we considered a different method

```

IF (red_value is LOW) and (green_value is LOW) and (blue_value is LOW) and (filtered_value is LOW)
THEN pixel class is (building with degree 0.53), (road with degree 0.33), (pavement with degree 0.35),
(vegetation with degree 0.0), (bare soil with degree 0.61), (water with degree 1.0)

IF (red_value is LOW) and (green_value is HIGH) and (blue_value is LOW) and (filtered_value is LOW)
THEN pixel class is (building with degree 0.0), (road with degree 0.24), (pavement with degree 0.43),
(vegetation with degree 1.0), (bare soil with degree 0.53), (water with degree 0.69)

IF (red_value is HIGH) and (green_value is HIGH) and (blue_value is HIGH) and (filtered_value is LOW)
THEN pixel class is (building with degree 1.0), (road with degree 0.92), (pavement with degree 0.92),
(vegetation with degree 0.54), (bare soil with degree 0.70), (water with degree 0.0)

IF (red_value is LOW) and (green_value is HIGH) and (blue_value is LOW) and (filtered_value is HIGH)
THEN pixel class is (building with degree 0.0), (road with degree 0.58), (pavement with degree 0.71),
(vegetation with degree 0.53), (bare soil with degree 1.0), (water with degree 0.38)

```

Figure 4: An excerpt of the rules produced by NFC1.

introduced in the literature for the sake of comparison. The work described in [19] consists in a multi-label approach for image retrieval: it is based on the employment of fully convolutional networks (FCNs) trained on the same WHDL D dataset adopted in our experimental session. More precisely, the cited work relies on the use of the entire image dataset (split into a training and a validation set). On the other hand, our proposal can be implemented considering only a limited subset of images: this represents a first advantage in view of the comparison. From a methodological point of view, the FCN proposed in [19] is first trained to derive a segmentation map of the images, that is preparatory for a subsequent process of multi-label image retrieval performed on the basis of a region-based similarity measure. Although our approach does not aim to fulfil the entire pipeline of an image retrieval process, we can compare our results with those appeared in literature, limited to the segmentation performance, evaluated in terms of some metrics related to the pixel classification level. Table 3 illustrates this comparison. The involved metrics are the overall accuracy, the averaged recall and the mean Jaccard index formally described in [20]: they have been chosen in order to allow a comparison with the results reported in [19]. As it can be observed, the proposed model was able to produce satisfactory results: even if it does not match the accuracy performance of the FCN-based model, it is still able to exhibit higher values of the remaining considered metrics. More precisely, Table 2 shows that the overall accuracy of the proposed model is penalized by a poor ability to recognize the pixels relative to the *road* class, probably due to the fact that from the chromatic point of view these are very similar to those of the *building* and *pavement* class, which are also often close to gray. Fig. 5 shows the pixel classification for some test images, i.e., images taken from the WHDL D collection which did not contribute to the realization of our training set. Each image is coupled with a map corresponding to the pixel classification of the image, thus providing a segmentation of the illustrated scene.

4. Conclusion

In this work, the ANFIS neuro-fuzzy system has been applied to the semantic segmentation of remotely sensed images. The results obtained are encouraging, since the proposed approach

Table 3

Comparison between our proposed model and the best performing one proposed in [19].

Model	Overall accuracy	Averaged recall	Mean Jaccard index
NFC1	0.75	0.73	0.60
FCN-based	0.81	0.68	0.56

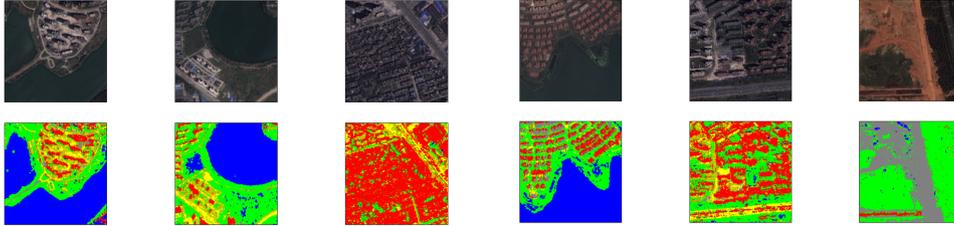


Figure 5: Test images and the corresponding pixel classification produced by NFC1.

can learn an effective model using even a very limited dataset for training. Moreover, the rules derived can provide an understandable explanation of why some areas have been segmented in some way, which is a desirable feature in different planning scenarios, where experts must justify their decisions based on the support provided by the machine. While the use of pixel-level information is beneficial, it limits the model to specific features of certain objects instead of using features that can improve generalization. As a future work, we expect to include more complex features that provide more contextual information of the pixel and do not limit the model to the color values of a given area and its local entropy. We can also improve system performance by adding new rules to those automatically derived from ANFIS, removing incorrect rules or updating existing rules. In particular, the interpretability and relevance of the learned fuzzy rules may be further investigated using user feedback. In this way, the results may contribute to show that accuracy is not necessarily the single indicator to be considered when studying the quality of an automatic classifier, and a more “sensible” learned classification rule could match expert judgment better than purely data-driven approaches.

Acknowledgment

The research is partially supported by the Italian Ministry of University and Research (MUR) under the PON ARS01_00141 “CLOSE” funding. Giovanna Castellano, Ciro Castiello, Gennaro Vessio, and Gianluca Zaza are members of the INdAM Research group GNCS.

References

- [1] R. Wang, J. A. Gamon, Remote sensing of terrestrial plant biodiversity, *Remote Sensing of Environment* 231 (2019) 111218.

- [2] M. Weiss, F. Jacob, G. Duveiller, Remote sensing for agricultural applications: A meta-review, *Remote Sensing of Environment* 236 (2020) 111402.
- [3] X. Yuan, J. Shi, L. Gu, A review of deep learning methods for semantic segmentation of remote sensing imagery, *Expert Systems with Applications* (2020) 114417.
- [4] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: *International Conference on Medical image computing and computer-assisted intervention*, Springer, 2015, pp. 234–241.
- [5] V. Badrinarayanan, A. Kendall, R. Cipolla, Segnet: A deep convolutional encoder-decoder architecture for image segmentation, *IEEE Trans. on PAMI* 39 (2017) 2481–2495.
- [6] R. Dong, X. Pan, F. Li, Denseu-net-based semantic segmentation of small objects in urban remote sensing images, *IEEE Access* 7 (2019) 65347–65356.
- [7] Y. Yi, Z. Zhang, W. Zhang, C. Zhang, W. Li, T. Zhao, Semantic segmentation of urban buildings from vhr remote sensing imagery using a deep convolutional neural network, *Remote sensing* 11 (2019) 1774.
- [8] J. M. A. Moral, C. Castiello, L. Magdalena, C. Mencar, *Explainable Fuzzy Systems: Paving the way from Interpretable Fuzzy Systems to Explainable AI Systems*, Springer, 2021.
- [9] G. Casalino, G. Castellano, C. Castiello, V. Pasquadibisceglie, G. Zaza, A fuzzy rule-based decision support system for cardiovascular risk assessment, in: *International Workshop on Fuzzy Logic and Applications*, Springer, 2018, pp. 97–108.
- [10] H. Leon-Garza, H. Hagra, A. Peña-Rios, A. Conway, G. Owusu, A big bang-big crunch type-2 fuzzy logic system for explainable semantic segmentation of trees in satellite images using hsv color space, in: *IEEE Int. Conf. on Fuzzy Systems (FUZZ-IEEE)*, 2020, pp. 1–7.
- [11] C. Wang, A. Xu, X. Li, Supervised classification high-resolution remote-sensing image based on interval type-2 fuzzy membership function, *Remote Sensing* 10 (2018) 710.
- [12] K. Shihabudheen, G. N. Pillai, Recent advances in neuro-fuzzy system: A survey, *Knowledge-Based Systems* 152 (2018) 136–162.
- [13] S. K. Meher, N. S. Kothari, Interpretable rule-based fuzzy elm and domain adaptation for remote sensing image classification, *IEEE Trans. on Geoscience and R. Sensing* (2020).
- [14] Z. Tianyu, J. Xu, Hyperspectral remote sensing image segmentation based on the fuzzy deep convolutional neural network, in: *13th International Congress on Image and Signal Processing, BioMedical Eng. and Informatics (CISP-BMEI 2020)*, IEEE, 2020, pp. 181–186.
- [15] J.-S. Jang, Anfis: adaptive-network-based fuzzy inference system, *IEEE Transactions on Systems, Man, and Cybernetics* 23 (1993) 665–685.
- [16] T. Takagi, M. Sugeno, Fuzzy identification of systems and its applications to modeling and control, *IEEE Transactions on Systems, Man, and Cybernetics* (1985) 116–132.
- [17] J.-S. Jang, C.-T. Sun, Functional equivalence between radial basis function networks and fuzzy inference systems, *IEEE Transactions on Neural Networks* 4 (1993) 156–159.
- [18] M. Brown, C. J. Harris, *Neurofuzzy adaptive modelling and control*, Prentice Hall, 1994.
- [19] Z. Shao, W. Zhou, X. Deng, M. Zhang, Q. Cheng, Multilabel Remote Sensing Image Retrieval Based on Fully Convolutional Network, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 13 (2020) 318–328.
- [20] E. Shelhamer, J. Long, T. Darrell, Fully convolutional networks for semantic segmentation, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39 (2017) 640–651.