



Semantic Segmentation of Radio-Astronomical Images

Carmelo Pino^{1(✉)}, Renato Sortino^{1,2}, Eva Sciacca¹, Simone Riggi¹,
and Concetto Spampinato²

¹ INAF, Catania, Italy

{carmelo.pino,eva.sciacca,simone.riggi}@inaf.it,
renato.sortino@phd.unict.it

² PeRCeiVe Lab, University of Catania, Catania, Italy
concetto.spampinato@unict.it

Abstract. In the context of next-generation radio-astronomical visual surveys, automated object detection and segmentation are necessary tasks to support astrophysics research from observations. Indeed, identifying manually astronomical sources (e.g., galaxies) from the daunting amount of acquired images is largely unfeasible, greatly limiting the huge potential of big data in the field. As a consequence, the astrophysics research has directed its attention, with increasing interest given the recent success in AI, to learning-based computer vision methods.

Several automated visual source extractors have been proposed, but they mainly pose the source identification as an object detection. While this may reduce the time needed for visual inspection, it presents an evident shortcoming in case of objects consisting of multiple, spatial distant, parts (e.g., the same galaxy appearing as a set of isolated objects). This specific limitation can be overcome through semantic segmentation. Consequently, in this paper we evaluate the performance of multiple semantic segmentation models for pixelwise dense prediction in astrophysical images with the objective to identify and segment *galaxies*, *sidelobe*, and *compact sources*. Performance analysis is carried out on a dataset consisting of over 9,000 images and shows how state-of-the-art segmentation models yield accurate results, thus providing a baseline for future works. We also employ the output segmentation maps for object detection and results are better than those obtained with Mask-RCNN based detectors that are largely used in the field.

Keywords: Tiramisu · Object detection · Galaxy visual inspection

1 Introduction

The recent technological advancement has led to an exponential growth in data availability, which in turn, has brought out the pressing need for computational

C. Pino and R. Sortino—Equal contribution.

© Springer Nature Switzerland AG 2021

Y. Hernández Heredia et al. (Eds.): IWAIIPR 2021, LNCS 13055, pp. 393–403, 2021.

https://doi.org/10.1007/978-3-030-89691-1_38

tools and innovative knowledge extraction methods to make sense of the collected data. Astronomy and astrophysics, among the others, are fields that in the last decades have produced an impressive amount of data coming from sky observation and surveys. This trend is bound not to change, but even more data is expected to be collected. For example, the Evolutionary Map of the Universe (EMU) [16] planned with the ASKAP system [3] will survey 70% of the sky, leading to an unprecedented quantity of data.

Typically, astronomy and astrophysics visual data may be of different modalities (e.g., radio-interferometric images, infrared images, etc.), and the main required task for supporting surveys is *source finding*, i.e. identifying and extracting astronomical sources like compact or point-like source, galaxies and sidelobes. However, beside cumbersome, this task is far from being trivial (both for humans and for computational methods) because of strong artifacts due to physical limitations of the acquisition process, especially in cases of extended sources or diffuse emissions. This requires an extensive manual pre- and post-processing phase that, however, is error-prone and time-consuming, other than almost infeasible on a volume of data such as the one predicted with systems like ASKAP.

Thus, there is an unmet need for automated and reliable computational methods for source detection. Indeed, several automated astronomical source detectors have been proposed, such as CAESAR [18], to address this need, yet these are based on classic computer vision methods requiring ad-hoc and complicated calibration and tuning steps. Standard learning-based techniques, e.g., shallow neural networks [18], have been adopted to overcome the limitations of computer vision methods. Despite the initial encouraging results, these methods tend to fail with extended and faint objects. At the moment, few source finders [19,20] are providing dedicated algorithms for extended sources but their performance is still inferior to what is achieved for compact sources.

With the resurgence of artificial intelligence, due to deep learning architectures, object detection methods based on convolutional neural networks have been proposed for galaxy classification [23,25], supernova remnant detection [2] and celestial object detection [4,6,8,11,26]. Nevertheless, even these deep learning-based object detectors are not able to detect accurately specific astronomical sources, especially galaxies that usually appear as composed by several fragments (see Fig. 1), thus limiting the effectiveness of the existing solutions. Motivated by the failures of the existing object detectors, in this paper we face the source identification problem from a different perspective, i.e., pixel-wise dense prediction for segmenting astronomical sources (Fig. 1 shows the advantage of semantic segmentation models over object detectors in case of galaxy detection). More in detail, we pose the source localization problem as a semantic segmentation task and propose a first, to our knowledge, benchmark analysis of state of the art approaches on astronomical images. Beside evaluating the performance in terms of segmentation accuracy, providing a first baseline for future works, we also leverage the segmentation masks to perform source detection obtaining better performance than Mask-R CNN [7], which is the most employed detector in prior works. These obtained results thus highlight that

employing semantic segmentation models is an interesting research direction in the astronomical image analysis field, as they allow scientists not only to detect automatically objects/sources but also to study morphological information about these sky objects.

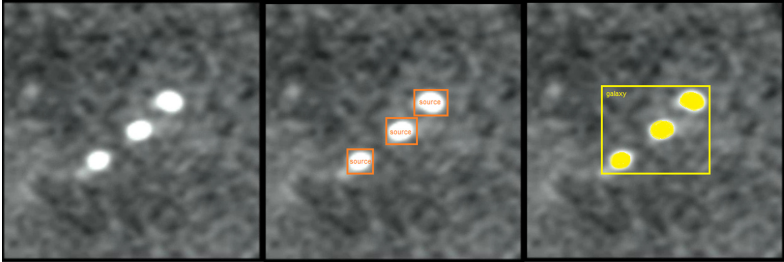


Fig. 1. Typical example of object detector failure. Usually astronomical images, especially small crops, contain one galaxy consisting of multiple non-connected parts [Left]. In this example, MaskR-CNN detects three single objects as sources [Center], while instead there is only one galaxy. A semantic segmentation method, as the one tested in this paper, identifies correctly the three objects as part of only one galaxy [Right].

2 Related Work

Automatic source detection in astronomical images has been developed mainly along two directions: either using classic computer vision techniques or deep learning methods. There exist several works on source finding based on classic computer vision techniques, such as [5], that applies Latent Dirichlet allocation to image pixels in order to segment them as source or background and [18], which performs source segmentation using the k-means algorithm based on pixels spatial and intensity proximity measure. Such works are mainly limited by the impossibility of generalizing well on unseen data. For this reason, recent works have been increasingly focused on deep learning models for automated source detection.

ConvoSource [14] uses a minimal configuration of a CNN, composed by three convolutional layers, one dropout layer and a dense layer to generate a binary map containing sources. Such an approach lacks the ability to distinguish among classes as it performs only binary classification. DeepSource [24] uses a CNN architecture composed by 5 layers with ReLU activation, residual connection and batch normalization, to first increase the signal-to-noise ratio of the input image and then apply a post-processing technique to identify the predicted source. In this case, the CNN is not used to directly perform object detection, but only to enhance image quality. The described methods use basic implementations of CNNs and do not allow for learning high-level features, which could be a problem in the case of more complex sources or faint objects. An improvement on

this architecture is made with the employment of state-of-the-art object detection methods that make use of RPN (region proposal network) backbones, to yield more accurate results. CLARAN [25] performs domain adaptation on the Faster R-CNN architecture [17], replacing the RoI Pooling layer with differentiable affine transformations and fine-tuning the model from weights pre-trained on the ImageNet Dataset [22]. Astro R-CNN [1] applies the evolution of Faster R-CNN model, Mask R-CNN [7], to perform object detection on a simulated dataset. Mask Galaxy [4] uses Mask R-CNN as well to adapt it to the astronomical domain by performing transfer learning from weights learned on COCO dataset [13] using only one class. Thus, the state of art contains several works employing object detection in astronomical images, but, to the best of our knowledge, no study yet exists that applies semantic segmentation to the source finding task. Hence, the main contribution of this work is to explore the application of such approach to the source finding task as to provide a proper baseline for future works.

3 Semantic Segmentation

This section briefly describes the semantic segmentation models applied to astronomical images. Existing semantic segmentation methods typically use an encoder-decoder architecture based on U-Net [21]. The base U-Net model in the years has been improved through combining segmentation maps created at different scales [12], or devising new loss functions [28] or through deep supervision [27] or through residual and squeeze excitation modules [15]. One significant change in the U-Net architecture was introduced in Tiramisu [10] that employs a sequence of DenseNet [9] blocks, rather than standard convolutional blocks. The *Tiramisu* network consists of a downsampling path for feature extraction and an upsampling path for output generation, with skip connections. Its architecture is shown in Fig. 2.

The input to the model consists of an image resized to 132×132 (in our case) and pre-processed by applying z-scale transform to adjust the contrast. Each image is passed to a convolutional layer to expand the feature dimensions. The resulting feature maps obtained from the first block, traverse a downsampling path consisting of five sequences of dense blocks, and transition-down layers. The transition-down layers are implemented to employ max-pooling in order to reduce feature map size. After the transition-down step, the encoded representation of the input image is obtained. The following upsampling path is symmetric to the downsampling one. Finally, a convolutional layer outputs a 2-channel segmentation map, respectively encoding the log-likelihoods of object and non-object pixels.

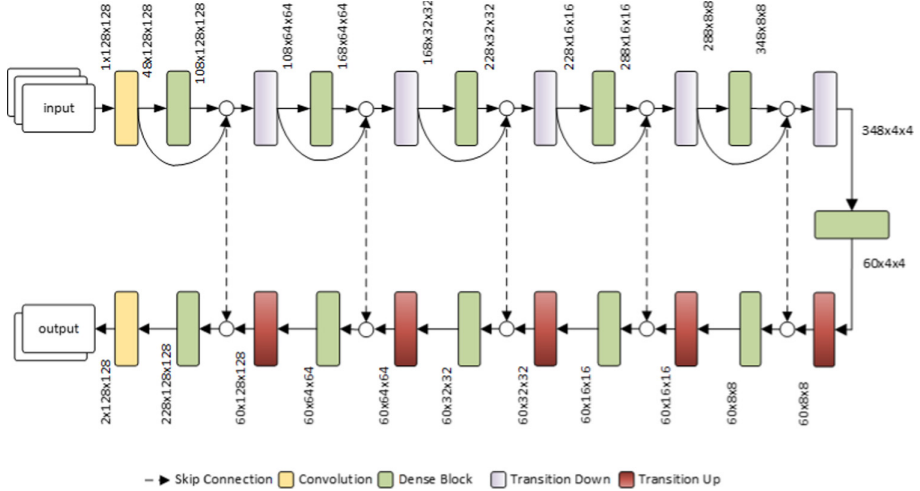


Fig. 2. The proposed Tiramisu segmentation architecture, consisting of a downsampling path and an upsampling path, interconnected by the bottleneck layer.

4 Experiments

4.1 Dataset

Performance analysis is carried on dataset containing 9,192 grayscale image cutouts extracted from different radio-astronomical survey maps taken with the Australian Telescope Compact Array (ATCA), the Australian Square Kilometre Array Pathfinder (ASKAP) and the Very Large Array (VLA). Each image has size 132×132 and may contain multiple objects of the following three classes (examples of them are in Fig. 3):

- **Source** (*19,000 samples*): Compact or point-like radio sources, with unknown astrophysical classification, having rounded and single-component morphology.
- **Sidelobe** (*1,280 samples*): A class of imaging artefacts, introduced by the map making process, often mimicking real radio sources and mostly appearing as elongated or ring-like regions around bright compact sources.
- **Galaxy** (*3,202 samples*): Extended multi-component radio galaxies, often comprising two or more disjoint regions (or islands), typically aligned along the radio structure axis and symmetrical around a center or core region.

The images are stored in FITS file format, although, for being fed to the model, they are converted into PNG format. Before conversion, each crop is normalized using a Z-Scale value of 0.3, in order to enhance the contrast. Each image in the dataset comes with a color-coded segmentation mask (see Fig. 4), which serves as ground truth during training. The whole dataset contains 23,481 different objects that are split for training, validation and test as shown in Table 1.

Table 1. Object splits. The whole dataset consists of 9,192 images containing about 23,000 objects.

Class	Train	Validation	Test
Source	13,300	3,800	1,900
Sidelobe	896	256	128
Galaxy	2,241	640	320
Total	16,437	4,696	2,348

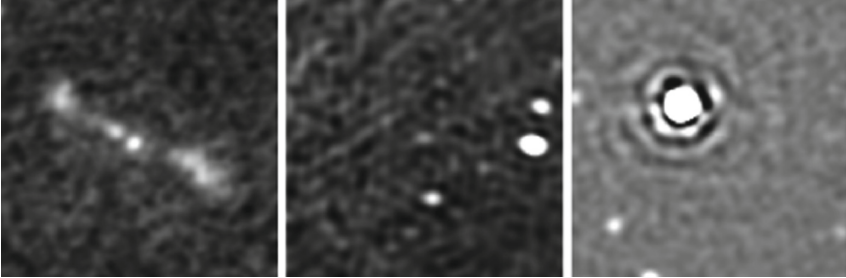


Fig. 3. Examples of (left) galaxies, (center) sources and (right) sidelobes.

4.2 Architecture and Training Details

We test multiple segmentation models on our dataset, namely, a standard encoder-decoder model, Tiramisu and U-Net. The latter has been tested in two variations: baseline and with deep supervision. The baseline version is the one reported in [21], which includes skip connections. Deep supervision consists in computing the distance between the deeper stages of the decoder and the down-sampled ground truth mask and add these distances to the final loss, so to guide the decoder to give a meaningful output even in the deeper layers. Input size is set to 132×132 , training is carried out for 100 epochs using negative log likelihood as a loss function. Initial learning rate is set to 0.0001, weight decay to 0.0001 with RMSProp as optimizer. Given a strong imbalance among classes, the loss is weighed by a different factor for each class, which results in a different update in the gradients during backpropagation, according to the class of the ground truth. For each class, the factor is computed as

$$w_j = S/(C * S_j) \quad (1)$$

where w_j is the weight for the j -th class, S stands for the total number of samples in the dataset, C is the number of classes and S_j is the number of samples for the j -th class.

This way, the classes with a smaller number of samples will have a higher loss, which pushes the model to better learn such underrepresented classes, counterbalancing the bias. Code is written in Pytorch and experiments executed on a NVIDIA GPU RTX 3090 (24 GB memory).

4.3 Results

For performance evaluation, commonly employed metrics for semantic segmentation and object detection are used. Accuracy, precision, recall and F1 score are computed according to their definition, by using true positives, true negatives, false positives and false negatives. More in detail:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F_1 = \frac{2 * Precision * Recall}{Precision + Recall} = \frac{2 * TP}{2 * TP + FP + FN}$$

For semantic segmentation and object detection, TP, TN, FP, FN are computed in different ways:

- **Semantic Segmentation:** For each class i , with $i = 1 \dots N$ (number of classes) a binary mask is generated, where values are ones if they correspond to pixels predicted class i , zeros otherwise. True positives and true negatives correspond to correctly predicted pixels (respectively for the correct class or for the background). False positives correspond to pixels not belonging to class i , predicted as class i . False negatives correspond to pixels with zero prediction where the ground truth is class i .
- **Object Detection:** To allow comparison with object detection models, the binary segmentation mask is converted into a sparse matrix where each connected component (i.e. an object) is identified separately from the others. Then, each object O_i is compared with the corresponding ground truth GT_i using the Intersection over Union (IoU) metric and defining a threshold α .

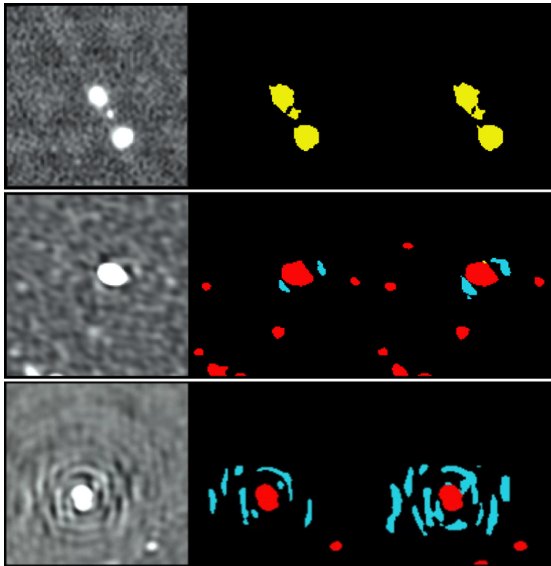
$$IoU = \frac{O_i \cap GT_i}{O_i \cup GT_i}$$

True positives are objects of class i with $IoU > \alpha$. False positives occur when the predicted object is not in the correct position with respect to its ground truth (i.e. $IoU < \alpha$). False negatives mean no prediction for the ground truth object GT_i . In this case, there are no true negatives, so the accuracy is not computed.

Table 2 reports semantic segmentation accuracy indicating how the Tiramisu models is the best performing one. All models yield good performance, especially for source and galaxy classification. Sidelobe segmentation performance is in generally lower because of both the limited representativeness in the dataset and their morphological structure. Indeed, sidelobes show a huge appearance variability as they are generated by distortions. This explains the lower number

Table 2. Comparison between Tiramisu and U-Net variations. *DS* stands for deep supervision.

Model	Class	Accuracy	Recall	Precision	F ₁
Encoder-Decoder	Sidelobe	0.45	0.38	0.26	0.33
	Source	0.65	0.65	0.72	0.68
	Galaxy	0.74	0.70	0.84	0.76
U-Net	Sidelobe	0.57	0.5	0.33	0.41
	Source	0.74	0.70	0.79	0.76
	Galaxy	0.83	0.76	0.86	0.81
↳+DS	Sidelobe	0.56	0.52	0.35	0.42
	Source	0.76	0.74	0.79	0.76
	Galaxy	0.87	0.88	0.93	0.90
Tiramisu	Sidelobe	0.85	0.85	0.50	0.63
	Source	0.90	0.90	0.82	0.86
	Galaxy	0.97	0.97	0.90	0.93

**Fig. 4.** Output segmentation maps. (left) input image, (middle) ground truth mask, (right) prediction mask. Yellow pixels belong to galaxies, blue ones to sidelobes and red ones to sources. First two rows show success cases, while the last row shows some failures on sidelobe segmentation.

of sidelobe samples in our dataset w.r.t. the other two classes: annotators often mislabel or miss often them. Among all the U-Net variants, the one employing deep supervision outperforms the others, while it underperforms the Tiramisu

Table 3. Object detection results of Tiramisu and MaskR-CNN.

Model	Class	F_1	Precision	Recall
Tiramisu	Sidelobe	0.52	0.46	0.59
	Source	0.83	0.81	0.84
	Galaxy	0.90	0.91	0.90
MaskR-CNN	Sidelobe	0.30	0.47	0.23
	Source	0.74	0.66	0.86
	Galaxy	0.77	0.66	0.93

model. Examples of good and wrong segmentations are given in Fig. 4. The failures (last row of Fig. 4) mainly pertain identification of sidelobes due to the reasons highlighted earlier.

Table 3 shows the object detection results, computed using a *IoU* threshold value of 0.5 and compared to those obtained by MaskR-CNN. Here we observe how Tiramisu model outperforms (in terms of F_1 measure, MaskR-CNN one, especially on the precision metrics for galaxy class, thus substantiating our original claim on a major effectiveness of semantic segmentation models over object detectors for that class. Similar to semantic segmentation task, lowest performance is achieved on sidelobes.

5 Conclusion

Both detection and segmentation of astronomical objects in radio images are of key importance for extracting useful information to support astrophysics research. In this work we provide a different perspective to the current object detection approach employed for source identification, i.e., performing semantic segmentation followed by a downstream localization method. To this end, we carried out a benchmark analysis of state-of-the-art semantic segmentation methods to define a baseline for future works. Beside this, we show that using semantic segmentation leads to better detection performance than MaskR-CNN, especially for galaxies. As in terms of segmentation performance, Tiramisu yields an average F_1 score of about 0.93 for galaxies, 0.86 for sources and 0.63 for sidelobes. The reduced performance on sidelobes mainly lies in the low quality of the annotations in the employed dataset. Indeed, the massive presence of sidelobes in astronomical images and their huge variability in appearance make rather complex to annotate all instances. This opens two possible research directions: (a) enhancing the quality of annotated datasets beside increasing the number of classes and instances per class; (b) investigating unsupervised and semi-supervised methods to reduce the annotation burden while keeping the same level of accuracy.

Acknowledgements. The research leading to these results has received funding from the European Commissions Horizon 2020 RIA programme under the grant agreement No. 863448 (NEANIAS) and from the INAF PRIN TEC programme (CIRASA).

References

1. Burke, C.J., et al.: Deblending and classifying astronomical sources with mask R-CNN deep learning. *Monthly Not. R. Astron. Soc.* **490**(3), 3952–3965 (2019)
2. Cunningham, F., et al.: Ensembl 2019. *Nucleic Acids Res.* **47**(D1), D745–D751 (2019)
3. DeBoer, D.R., et al.: Australian SKA pathfinder: a high-dynamic range wide-field of view survey telescope. *Proc. IEEE* **97**(8), 1507–1521 (2009)
4. Farias, H., et al.: Mask galaxy: morphological segmentation of galaxies. *Astron. Comput.* **33**, 100420 (2020)
5. Friedlander, A.E.A.: Latent Dirichlet allocation for image segmentation and source finding in radio astronomy images. In: *ACM International Conference Proceeding Series* (2012)
6. González, R.E., Muñoz, R.P., Hernández, C.A.: Galaxy detection and identification using deep learning and data augmentation. *Astron. Comput.* **25**, 103–109 (2018)
7. He, K., et al.: Mask R-CNN. In: *ICCV* (2017)
8. Hou, Y.C., et al.: Identification and extraction of solar radio spikes based on deep learning. *Solar Phys.* **295**(10), 1–11 (2020)
9. Huang, G., et al.: Densely connected convolutional networks. In: *CVPR* (2017). <https://doi.org/10.1109/CVPR.2017.243>
10. Jégou, S., et al.: The one hundred layers tiramisu: fully convolutional densenets for semantic segmentation. In: *CVPRW* (2017)
11. Jia, P., et al.: Detection and classification of astronomical targets with deep neural networks in wide field small aperture telescopes. *arXiv* 159(5) (2020)
12. Kayalibay, B., et al.: CNN-based segmentation of medical imaging data. *arXiv preprint arXiv:1701.03056* (2017)
13. Lin, T.Y., et al.: Microsoft COCO: common objects in context. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) *ECCV 2014*. LNCS, vol. 8693, pp. 740–755. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10602-1_48
14. Lukic, V., Gasperin, F.D., Brüggem, M.: ConvoSource: radio-astronomical source-finding with convolutional neural networks (2019)
15. Murabito, F., et al.: Deep recurrent-convolutional model for automated segmentation of craniomaxillofacial CT scans. In: *2020 25th International Conference on Pattern Recognition (ICPR)* (2021)
16. Norris, R.P., et al.: EMU: evolutionary map of the universe. *Publ. Astron. Soc. Aust.* **28**(3), 215–248 (2011)
17. Ren, S., et al.: Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **39**, 1137–1149 (2015)
18. Riggi, S., et al.: Caesar source finder: recent developments and testing. *Publ. Astron. Soc. Aust.* **36** (2019)
19. Riggi, S., et al.: Automated detection of extended sources in radio maps: progress from the SCORPIO survey. *Monthly Not. R. Astron. Soc.* **460**(2), 1486–1499 (2016)
20. Robotham, A., et al.: ProFound: source extraction and application to modern survey data. *Monthly Not. R. Astron. Soc.* **476**(3), 3137–3159 (2018)
21. Ronneberger, O., et al.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W., Frangi, A. (eds.) *MICCAI 2015*. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28

22. Russakovsky, O., et al.: ImageNet large scale visual recognition challenge. *Int. J. Comput. Vis. (IJCV)* **115**(3), 211–252 (2015)
23. Shimwell, T., et al.: The LOFAR two-metre sky survey-II. First data release. *Astron. Astrophys.* **622**, A1 (2019)
24. Vafaei-Sadr, A., et al.: DeepSource: point source detection using deep learning. *Monthly Not. R. Astron. Soc.* **484**(2), 2793–2806 (2019)
25. Wu, C., et al.: Radio galaxy Zoo:CLARAN - a deep learning classifier for radio morphologies. *Monthly Not. R. Astron. Soc.* **482**(1), 1211–1230 (2018)
26. Xie, Z., Ji, C., Wang, H.: Single and multiwavelength detection of coronal dimming and coronal wave using faster R-CNN. *Adv. Astron.* **2019** (2019)
27. Zhou, Z., et al.: Unet++: a nested U-Net architecture for medical image segmentation. In: Stoyanov, D., et al. (eds.) *DLMIA 2018, ML-CDS 2018. LNCS*, vol. 11045, pp. 3–11. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00889-5_1
28. Zhu, Q., et al.: Deeply-supervised CNN for prostate segmentation. In: 2017 International Joint Conference on Neural Networks (IJCNN). IEEE (2017)