

PAPER • OPEN ACCESS

Satellite Segmentation with Pre-trained CNN Models

To cite this article: Yefei Huang *et al* 2022 *J. Phys.: Conf. Ser.* **2171** 012003

View the [article online](#) for updates and enhancements.

You may also like

- [Automatic crack recognition for concrete bridges using a fully convolutional neural network and naive Bayes data fusion based on a visual detection system](#)
Gang Li, Qiangwei Liu, Shanmeng Zhao et al.
- [A segmentation method combining probability map and boundary based on multiple fully convolutional networks and repetitive training](#)
Wenshe Yin, Yangsheng Hu, Sanli Yi et al.
- [Free Core Nutation and Its Relation to the Spin-over Mode](#)
Jérémy Rekier

Satellite Segmentation with Pre-trained CNN Models

Yefei Huang¹, Tianlai Xu^{1*}, Zexu Zhang¹, Hutao Cui¹, Yu Su¹

¹Harbin Institute of Technology, Harbin, Heilongjiang, China

*email: tianlaixu@hit.edu.cn

Abstract. In a generic satellite relative pose estimation pipeline, finding sufficient features in objects is quite essential to build the correct matching relationship and then solve the relative movement. However, for low-earth-orbit (LEO) satellites, since the earth background contains much more texture than objects, an object segmentation process is necessary to provide a prior range for feature extraction. In this work, we address this task with the pre-trained Deeplabv3 and fully convolutional network (FCN). Unlike the fine-tuning or transfer learning processes in other researches, we obtain probabilistic maps from the high-dimensional output of the above-mentioned CNN models and achieve a rough satellite extraction. Our method makes Deeplabv3 and FCN models work in a totally unfamiliar LEO scene and still achieves 0.2927 and 0.2122 in average intersection over union (IoU) respectively.

1. Introduction

Estimating the relative pose between the observer and satellite is essential for many astronautic tasks, such as rendezvous and docking, formation flying, and space debris removal. Because of the low energy consumption, slight mass, and relatively high working frequency, camera units are the mainstream to acquire necessary information. According to the multiple view geometry, the relative movement can be computed by extracting and matching feature points in images. However, too many points are needed in these point-based methods, and this makes them difficult to be implemented in astronautic fields. Especially in low-earth-orbit (LEO) satellites, the movement of the background may also be extracted in the aforementioned process. Therefore, compared with feature decreasing methods [1-2], the target segmentation is a feasible and necessary solution to provide a prior location for the satellite relative pose estimation and decrease the required point amount.

With the development of convolutional neural networks (CNNs), learning-based methods show great potential in the field of image segmentation. By using high-performance graphics processing units, the classification of each pixel in images can be rapidly obtained. Recently, learning-based methods are also implemented in astronautic tasks, such as feature point extraction [3] and pose estimation [4]. Though a large number of synthesis images are generated in these works, there are few semantically annotated datasets in astronautic fields. Moreover, it is still uncertain whether these synthetic images can reflect the real aerospace scene.

Therefore, we choose pre-trained semantic segmentation models rather than construct a new model. These models are trained on numerous real images and achieve high performance in generic scenes. For the totally unfamiliar LEO scene, in which the earth texture and the satellite are not trained, we find the object can still be captured by these models. This phenomenon occurs in both fully convolutional network (FCN) [5] and Deeplab [6] models according to their output probabilistic maps. In this work, we achieve a rough satellite segmentation through this object extraction ability. Since our method does not include an additional training process, it has a higher generalization ability compared with other fine-tuning methods.



The rest of this paper is organized as follows. In Section 2, we show our proposed method in detail and how to apply it to a pre-trained semantic segmentation model. In Section 3, the extraction results are evaluated in synthetic images. The conclusion is presented in Section 4.

2. Satellite segmentation method

In this section, we first introduce the input and output process of a semantic segmentation model. Then we show the object capturing ability of FCN and Deeplabv3 [7] for a synthetic LEO image. At last, we describe our whole method in detail with an example.

2.1. Semantic segmentation process

A generic semantic segmentation process is recorded as follows. Firstly, the image is normalized by the mean and variance of the training dataset. Then the input $3 \times H \times W$ image is transformed into a $(N+1) \times H \times W$ result through the deep learning process of down-sampling and up-sampling. H , W , N are the image height, width, and the classification number of the training dataset, respectively. The additional dimension of the output in classification is used to cope with the pixels without any label, namely the background. As shown in Figure 1, each pixel corresponds with a $(N+1)$ dimensional output, the classification of this pixel is the index of the largest value.

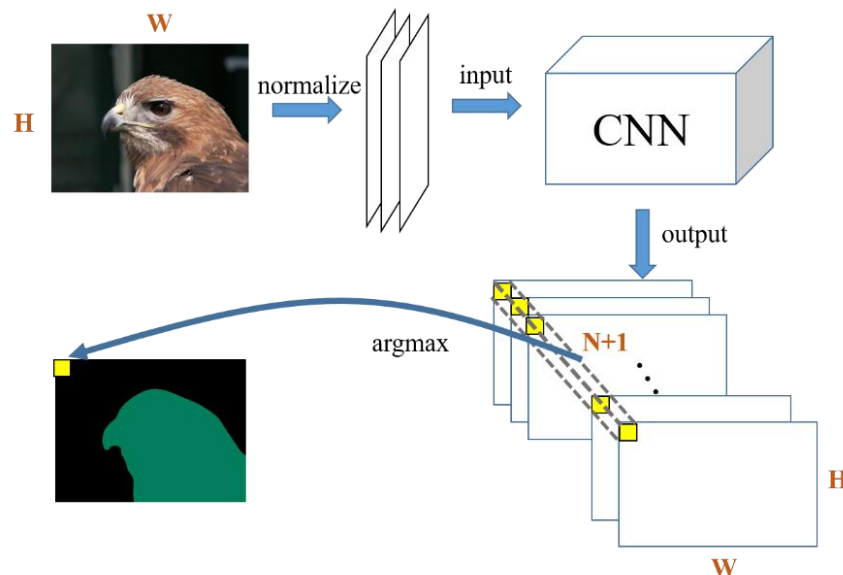


Fig. 1 Image segmentation process

Take the famous FCN and Deeplabv3 as examples, they are trained on a subset of COCO train2017, and they both take 20 classifications that are presented in the Pascal VOC dataset. Therefore, since the scene in aerospace is not trained, the segmentation for an LEO satellite image leads to a failure, and all pixels are classified as the background.

2.2. Satellite capturing by probabilistic map

Since the output of each pixel is a 21-dimension vector, it can be converted into a form of probabilities through the softmax function. We denote the i -th largest probability of all pixels as P_i ($i = 1, \dots, N+1$), and their corresponding indices map is dubbed as Ind_i . It is worth mentioning that P_0 and P_{Ind1} are completely equivalent since all pixels are classified as the background by the original CNN output. Then $(1 - P_0)$ is the probabilistic map that the pixels are not the background. The probabilistic maps are computed by Deeplabv3 and FCN with the ResNet50 or ResNet101 [8] backbone. In Figure 2, it clearly shows that the satellite object is captured even though the scene is unfamiliar. The satellite area contains an obvious larger probability than its neighbors.

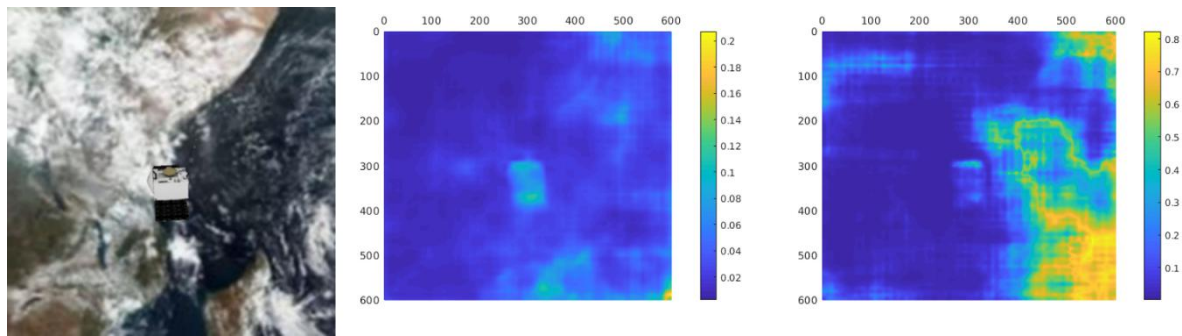


Fig. 2 $(1 - P_0)$ of Deeplabv3 and FCN. The left is an unfamiliar LEO scene, the center, and the right images are the $(1 - P_0)$ probabilistic maps of Deeplabv3 + ResNet101 and FCN + ResNet101, respectively

In addition, as shown in Figure 3, the satellite is also extracted in the second-largest classification probabilistic map, which is dubbed as P_{Ind2} . According to our experiments, even in P_{Ind3} and P_{Ind4} , the satellite can be captured. This means that, though without additional training, these pre-trained CNN models have recognized the pixels in this region belong to a whole object, but they do not know the corresponding classification. It is worth mentioning that, besides the object region, image border pixels also usually lead to high values. This phenomenon is more clear in results computed by FCN than Deeplabv3.

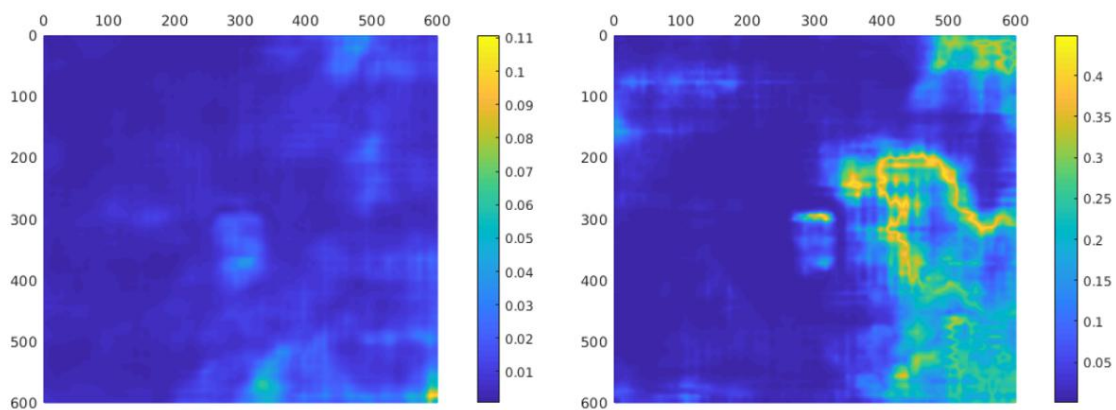


Fig. 3 P_{Ind2} of Deeplabv3 and FCN

2.3. Segmentation with a pre-trained model

For the probabilistic map of $(1 - P_0)$, we take the median value as its threshold. The pixels higher than 2 times the threshold are marked. The same operation is utilized for the map of P_{Ind1} . If a pixel is marked in both maps, it is segmented in the original image. The whole segmentation process of our method is shown in Figure 4. In the computation of the median value, the region within 50 pixels of the image edge is not counted. We take the opening operation in image morphology to filter too small regions.

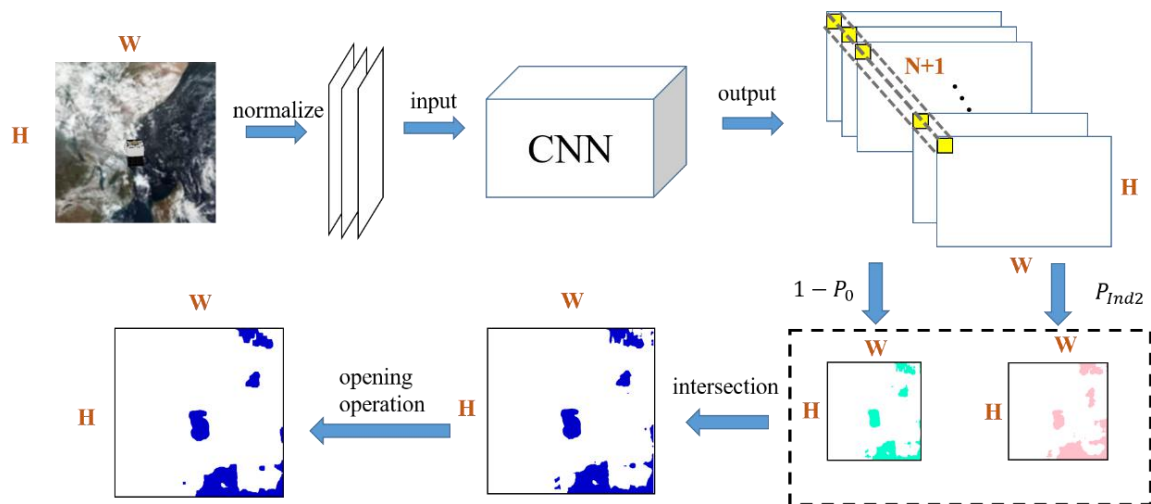


Fig. 4 The proposed segmentation method

The result is a rough segmentation compared with fine-tuning methods. However, since there is no training process in our method, this result comes from a totally unfamiliar scene, our method relies on the object capturing ability of pre-trained CNN models and has a better generalization than fine-tuning methods.

3. Experiments

3.1. LEO image synthesis

The satellite model is the OSIRIS-REx model from NASA [9], the earth and cloud map comes from Zoom Earth [10]. First, we randomly select a square region with random width and length from 50 to 200 pixels. Then this region is enlarged to a 600×600 pixels background image. The random region choice and magnification simulate the observations in different heights.

Besides, we use Pyrender to obtain the object images with different poses. The object area is covered in the background image. We also add white noise and a Gaussian filter to make the synthetic image blurred and disturbed. Some examples are shown in Figure 5.

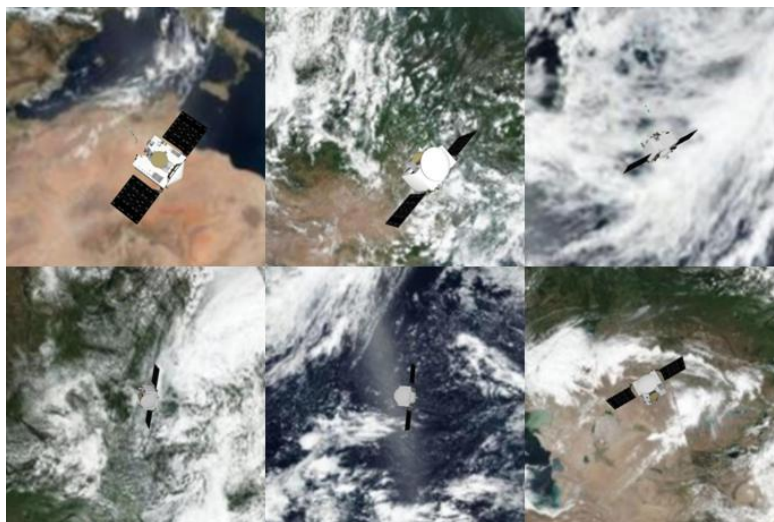


Fig. 5 Synthetic image examples

3.2. Method evaluation

We generate 500 images with the aforementioned method. For FCN and Deeplabv3, we choose ResNet50 and ResNet101 as the backbones in each model, these pre-trained models are provided by Pytorch. We compute the average intersection over union (IoU) in Table 1. The edge constraint IoU is

the result if the remarked pixels near the image edge are additionally removed. In our experiment, we find most of the outliers come from these regions.

Tab. 1 Segmentation in LEO images

Model	IoU (all pixels)	IoU (edge constraint)
FCN + ResNet50	0.0770	0.1253
FCN + ResNet101	0.1094	0.2122
Deeplabv3 + ResNet50	0.1387	0.2226
Deeplabv3 + ResNet101	0.1749	0.2927

4. Conclusion

In this work, we propose a satellite segmentation method with pre-trained CNN models. Since the LEO images are not trained by the off-the-shelf models, our method has better generalization ability compared with the fine-tuning methods. The segmentation results are rough, but they can provide prior information about the satellite location. Our method can be used in the pre-processing in aerospace tasks such as satellite tracking, and it can effectively decrease the feature points extracted from the background. However, the classification output of the pre-trained model is not utilized. A good solution is to divide the segmentation result into different regions, and a region matching algorithm can be applied according to its classification results.

References

- [1] Zhang X, Jiang Z, Zhang H and Wei Q 2018 Vision-based pose estimation for textureless space objects by contour points matching *IEEE Transactions on Aerospace and Electronic Systems* **54** pp 2342-2355
- [2] Huang Y, Zhang Z, Cui H and Zhang L 2021 A low-dimensional binary-based descriptor for unknown satellite relative pose estimation *Acta Astronautica* **181** pp 427-438
- [3] Rathinam A and Gao Y 2020 On-orbit relative navigation near a known target using monocular vision and convolutional neural networks for pose estimation *International Symposium on Artificial Intelligence, Robotics and Automation in Space (iSAIRAS), Virtual Conference (Pasadena, CA)* pp 1-6
- [4] Proença P F and Gao Y 2020 Deep learning for spacecraft pose estimation from photorealistic rendering *IEEE International Conference on Robotics and Automation (ICRA)* (IEEE) pp 6007-6013
- [5] Long J, Shelhamer E and Darrell T 2015 Fully convolutional networks for semantic segmentation *Proceedings of the IEEE conference on computer vision and pattern recognition* pp 3431-3440
- [6] Chen L C, Papandreou G, Kokkinos I, Murphy K and Yuille A L 2017 Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs *IEEE Transactions on Pattern Analysis and Machine Intelligence* **40** pp 834-848
- [7] Chen L C, Papandreou G, Schroff F and Adam H 2017 Rethinking atrous convolution for semantic image segmentation arXiv preprint arXiv:1706.05587
- [8] He K, Zhang X, Ren S and Sun J 2016 Deep residual learning for image recognition *Proceedings of the IEEE conference on computer vision and pattern recognition* pp 770-778
- [9] OSIRIS-REx 3D Model 2020 <https://solarsystem.nasa.gov/resources/2360/osiris-rex-3d-model>
- [10] Zoom Earth 2021 <https://zoom.earth>