Pixelwise Image Segmentation with Convolutional Neural Networks for Detection of Resident Space Objects

Douglas Woodward, Celeste Manughian-Peter,
Timothy Smith, Elizabeth Davison
The Aerospace Corporation

ABSTRACT

We contribute a novel convolutional neural network based approach for the detection of Resident Space Objects in Geosynchronous orbit from ground-based, rate-tracked, electro-optical telescope imagery through image segmentation. We compare this approach with alternative deep learning approaches and demonstrate state of the art performance at lower pixel error thresholds with smaller model footprint. We are able to achieve state of the art results at lower pixel error thresholds with only 2 million parameters. We report improved F1 scores at the 1, 2, and 4 pixel thresholds. These scores are 0.80, 0.935, and 0.964 respectively — improvements from the 0.78, 0.906, and 0.956 previously reported. We additionally contribute pixelwise masks for the SatNet dataset, and code for generation of masks from bounding boxes to the Unified Data Library and community. This technique allows for pixel-wise labeling of an image, expanding possibilities for follow-on work in the field such as automated Resident Space Object breakup and collision detection, debris tracking, and closely spaced object detection.

1. INTRODUCTION

As space becomes increasingly congested across all orbits, Space Domain Awareness (SDA), requires ever more accurate detection, cataloging, tracking, predicting, and contextualizing of all space objects and activity. A fundamental challenge in SDA is source extraction: identifying which pixels in an image contain Resident Space Objects (RSOs). For RSOs in Geosynchronous Earth Orbit (GEO), ground based, rate-tracked, electro-optical imagery is the primary data used for source extraction as radar techniques become less effective at higher orbits[1]. Though more effective for GEO, optical sensor data faces challenges including atmospheric noise, stray light, and high apparent visual magnitude of targets[2][3][4]. Manual source extraction methods are rendered overly time consuming and costly to scale with the rapid proliferation of RSOs[1], leading to a need for efficient, accurate automated alternatives. Current automated source extraction approaches include intensity thresholding, peak detection, gaussian processes, connected components, Random Sample Consensus(RANSAC), frame stacking, manually tuned convolutional filters, wavelet filtering, and Bayesian methods. Most recently, convolutional neural networks (CNNs) have been applied as a method for object detection[2]. Though their performance has been shown to be state of the art, object detection CNNs require large numbers of parameters and do not provide pixelwise labels.

Space Domain Awareness is defined by the US Space Force as consisting of five pillars[1]:

1. Detect, Track, and Identify
2. Characterization
3. Tactical Warning and Attack Assessment
4. Data Integration and Exploitation
5. Spacecraft Protection and Resiliency

This work falls under the fourth pillar, Data Integration and Exploitation.

Nearly every week, new spacecraft take up residency around Earth in a variety of orbits including Low Earth Orbit (LEO) and Geosynchronous Earth Orbit (GEO). As of 2015, there were 1,000 active satellites, as of this work that number has more than doubled to over 2,000. Additionally, there are over 7,000 defunct objects in orbit, along with an

order of magnitude more pieces of space debris[1]. Unsurprisingly, tracking, cataloging, and detecting these objects is of great interest for managing an increasingly congested space.

One approach to Space Domain Awareness (SDA) Resident Space Object (RSO) detection is to employ ground-based sensors. Typically, this is a radar-based approach. Radar is preferred up to LEO for greater information capture than electro-optical telescopes. Beyond LEO, the effectiveness of radar sharply decreases, and optical methods are preferred[1]. An optical approach requires the resolution of apparently small objects moving very fast, very far away from the observing telescope, a non-trivial task. In this work, we focus on rate-tracked GEO electro-optical imagery. It is desirable to automate the process of detecting and locating RSOs within an observation frame to allow analysts to focus on more intensive work, and to this end, we approach this as a computer vision task.

In the past decade, computer vision has been dominated by the rise of convolutional neural networks. We follow this trend by framing the RSO detection approach as a pixelwise binary classification problem and pose the hypothesis that deep learning image segmentation models are highly effective for the task of GEO RSO source extraction.

2. DATA

As is the case for most all machine learning applications, the availability, quality, and standardization of a benchmark dataset is the foundation of any results and progress. Fortunately, we are not the first to consider this problem.

2.1 SatNet Dataset

This work uses the SatNet dataset (v.1.2.0.0) curated and collated by [2]. The SatNet dataset contains 104,100 512x512 single channel 16-bit images. The data is sourced from electro-optical ground based telescopes. Each image is a capture of a GEO target in rate-track mode. Annotations provided include a set of bounding boxes for each image. These annotations were created by a trained analyst with the assistance of software. See [2] for more details.

2.2 Data Preprocessing

In many related works, preprocessing techniques such as a median filter and more advanced methods are employed[5][3][4], all of which could provide future improvements to the work at hand, which for now uses very minimal preprocessing. Here, we eschew most techniques, and allow the model for the most part to learn optimal image processing steps. Per frame normalization is the only preprocessing applied to the dataset in which each frame is linearly scaled to mean 0 and variance 1. However, this is not to say that preprocessing is not needed in such applications, rather that the SatNet dataset already encompasses some preprocessing and correction such as to make it amenable to minimal preprocessing. In fact, a direction of future work might be to implement improved data preprocessing with this segmentation model.

2.3 Mask Generation

In order to facilitate the shift from object detection to image segmentation, pixelwise masks were generated for the SatNet dataset. As the process of labeling over 100,000 images pixel by pixel presents a highly labor-intensive task, an automated process was developed. The process is as in Alg. 1 for each image .

The algorithm at the core of this is Otsu's Method[6]. Otsu's thresholding is especially effective when an image has a binomial distribution of intensity values. Fortunately, the single channel 16-bit images that comprise SatNet have a predominantly binomial distribution in regions around RSOs. By searching for the threshold that minimizes the intra-class variance, Otsu's method automatically returns the optimal threshold for a given image.

For the vast majority of RSOs in the SatNet dataset, Otsu's binarization[6] produces excellent results. However, there are failure modes, often when the background is highly saturated, the RSO is dim and fails to stand out from the background, or when a streak, or streaks of a background star cross the region. When this occurs, Otsu's binarization produces a threshold that results in a binarized image region of moistly noise. Such failure modes occur on 3% of the dataset. To account for this, a fallback method is triggered when the mean of the resulting binary region exceeds 24% of the total pixels in the region. When this occurs, the region is reset to zeros, the brightest pixel selected from the bounding box region, and that pixel is set as the center of a 3x3 grid of positive values in the resulting mask. An alternate fallback case is to locate this 3x3 grid at the center of the bounding box, however, by selecting the brightest pixel as the center, the small errors present in the centering of the bounding boxes can be alleviated.

This augmented dataset is contributed back to the SatNet dataset in the Unified Data Library for use by future collaborators along with the mask generation code.

Algorithm 1: Mask Generation

Input: NxNx1 Image
Input: Bounding Box Coordinates
Output: Pixelwise Mask

M = 0s size NxN
for *bounding box in image* **do**
    Initialize region mask, R with 0s
    $R = GaussianBlur(R)$
    $R = OtsuBinarization(R)$
    `/* Fallback Condition */`
    if $R_{mean} > T$ **then**
        $R = 0s$
        $C = ArgMax(R)$
        $R[C-1:C+1; C-1:C+1] = 1$
    end
    $M[R_{coords}] = R$
end
return M

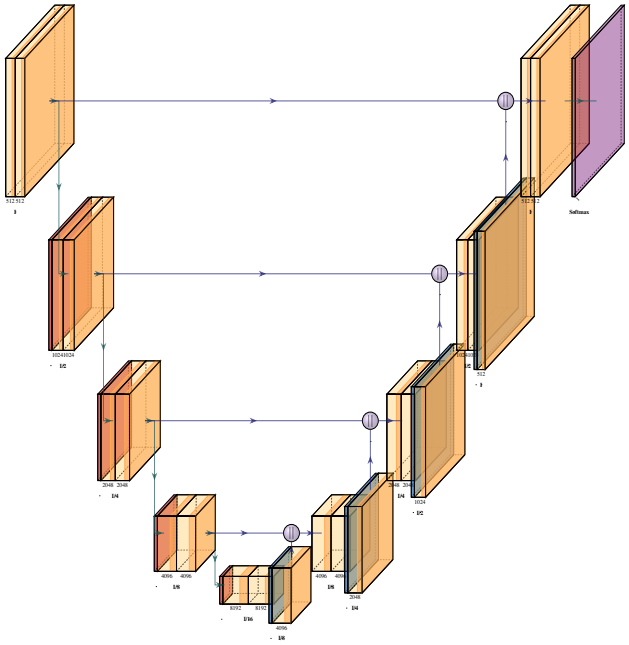
3. RELATED WORKS

The work in [2] provides the SatNet dataset employed here, as well as a first foray in the use of convolutional networks in this domain from an object detection perspective. SExtractor: Software for source extraction[7] provides a baseline comparison from a non-deep learning method. Xue et. al in "Dim Small Target Detection Based on Convolutional Neural Network in Star Image"[3] employ a similar approach to what is outlined here but tackling the problem of star detection in an image as opposed to RSO detection. The RSO detection is a more complex problem as it is necessary to distinguish between background stars, foreground objects, and the RSOs of interest. [3] also uses a standard cross entropy loss with a computed weight map for handling dim targets, whereas this work does not require weight maps and employs a focal loss. In "Detection of Artificial Satellites in Images Acquired in Track Rate Mode"[8] provides an excellent discussion of the qualities of images from Rate Tracked electro-optical imagery as well as an in-depth discussion of the difficulties of star, RSO, and background separation and detection. Many observations in their work could provide greater performance if applied in conjunction with deep learning-based methods. Their work is integrated in the Canadian Space Situational Awareness arrays. Flewelling and Sease in Computer Vision Techniques Applied to Space Object Detect, Track, ID, and Characterize[9] discuss the effectiveness of traditional computer vision techniques such as the Harris Corner methods and Phase Congruency. Another set of related works encompasses using multiple observation frames in sequence for detection. Among these methods are LINE[5], RANSACing Optical Image Sequences for GEO and near-GEO Objects[10], and the Image Stacking Method[5]. These methods rely on either detecting primitive objects in frames and then linking them to tracks or processing all frames in a sequence concurrently. This work can be integrated with either of these approaches as an effective primitive detector, or as a direct sequence processor as Fletcher et al do with a YOLOv3[11] based model.

4. MODEL

For this work, a fully convolutional neural network is employed, with the input being the image, and the output being an equivalent to input sized binary mask. Specifically, the model is a UNet[12] variant. A UNet derives its name from the fact that the model consists of N layers that consist of standard convolutions on the down(encoder) side of the model, before a bottom layer, and then up the decoder side of the U. With each encoder block having a skip connection across the U to the corresponding decoder block where the output of the decoder and encoder blocks are concatenated. By passing forward the features from each level of the U, the model is able pass each level of feature extraction deeper into the model. Additionally, the concatenation aids in enforcing equal quality of the convolutions and transpose convolutions.

Fig. 1: Model



The UNet architecture is most well-known for its' prevalence in the medical imaging field, often used for segmentation of tumors[13][14]. Despite the apparent distance in the application of medical imagery and RSO imagery, the underlying data for both share several features. Firstly, both can be expressed as single channel images, with the distinction between the object of interest in higher intensity and the low intensity background being desired at pixel level accuracy. Both types of images are prone to background noise, in the RSO imagery this comes in the form of background stars, light

Each down convolution block consists of two pairs of two-dimensional convolution with a 3×3 kernel plus batch normalization and a ReLu activation. Then a Max Pooling layer and a dropout layer. The number of filters doubles with each descending layer, up to $16\times$ the first layer filter count at the base. The standard number of filters in the model is 16×2^{layerDepth}. The base layer does not have the max pooling and dropout. On the decoder side of the U, each block up consists of a Transpose Convolution layer, a concatenation with the corresponding encoder block, dropout, and then a $2\times$ Convolution, Batch Norm, ReLu block. The final layer is a Convolution layer with a 1x1 kernel and a sigmoid activation to produce a $512 \times 512 \times 1$ output with values between $[0,1]$. The full model architecture employed here can be seen in Figure 1.

Motivated by the imbalanced nature of true positives to the total pixels in the image, focal loss[15] was selected for the training loss function. Focal loss seeks to handle this foreground background extreme imbalance by down-weighting easy examples and focusing training on harder examples. This is done by adding a modulating parameter to cross entropy loss. It can be defined as

$$\mathbf{FL}(p_t) = -\alpha_t(1-p_t)^{\gamma}\log(p_t) \tag{1}$$

See [15] for more detail on Equation 1.

In order to evaluate the model against prior work the objectwise F1 score must be computed. Computing the objectwise F1 score from this softmax mask must be converted to a list of object detections. From the model output, an optimal value between [0,1] is selected. We determine this by a search over a validation set. In empirical results, this threshold typically falls between 0.5 and 0.65 dependent on the specific criteria being optimized for. Smarter approaches to thresholding were briefly pursued, including an attempt to develop an auxiliary network to select a threshold automatically for each image, but results were not promising.

Following the binary threshold, individual object are identified using the OpenCV2 findContours method[16]. The centroids of those objects are found and used as the reference point for the objectwise F1 score. There exists the

possibility to improve this second step of the model and this could be an avenue for expansion on this work. Effectively the model outputs both a raw softmax prediction mask and a binarized object detection mask. Both are of interest - the object detection mask as a harder prediction on the input, for comparison with prior results, and as a top level data point. The raw softmax prediction provides some insight into the decision process of the model and allows for easier adjustment of the thresholding step. The softmax mask can also be used a soft prediction mask, indicating where the model was more or less confident in predictions. It should be noted that softmax outputs are not probabilities without proper calibration and should not be interpreted as such[17]. Should it be desired, the second part of the detection pipeline, the thresholding and contour detection, can be replaced with other methods that may provide improved results. This flexibility also allows for the UNet based model here to be used as a first step or part of a more sophisticated pipeline.

5. RESULTS & CONTRIBUTIONS

Table 1: Object Detection Performance

1px	F_1^*	Precision at F_1^*	Recall at F_1^*
SExtractor	0.599	0.673	0.540
YOLOv3(Fletcher et al)	0.780	0.773	**0.788**
UNet (Ours)	**0.800**	**0.816**	0.784
2px	F_1^*	Precision at F_1^*	Recall at F_1^*
SExtractor	0.818	0.901	0.750
YOLOv3(Fletcher et al)	0.906	0.911	0.901
UNet (Ours)	**0.935**	**0.954**	**0.917**
4px	F_1^*	Precision at F_1^*	Recall at F_1^*
SExtractor	0.843	0.924	0.776
YOLOv3(Fletcher et al)	0.956	0.960	**0.953**
UNet (Ours)	**0.964**	**0.983**	0.946
8px	F_1^*	Precision at F_1^*	Recall at F_1^*
SExtractor	0.843	0.924	0.776
YOLOv3(Fletcher et al)	**0.971**	0.973	**0.969**
UNet (Ours)	0.966	**0.985**	0.947

We achieve state of the art performance on the SatNet dataset using our image segmentation convolutional neural network model as seen in 1. In addition to improved performance, our model is approximately 3% of the size of prior state of the art models, and can be trained in under 24 hours on SatNet dataset. The model also exhibits a notable increase in performance at smaller pixel error thresholds compared to prior results. This last improvement is likely a result of the pixelwise emphasis of the segmentation approach and is a very promising improvement for future applications including breakup detection and tracking. With each increase to the F1 object wise score at the single pixel threshold, we approach the hypothetical maximum information exploitation from the data. As laid out in the pillars of SDA, this ability to extract maximum information is a key for future improvements in the RSO detection domain.

This work demonstrates that the deep learning based computer vision techniques have great potential in the Space Domain Awareness field. Of particular excitement is the pixelwise nature of out model. As discussed, this can be very benefitial for further applications of the technique. We encourage the community to pursue such research further and find these to be promising results.

REFERENCES

[1] M. R. Ackermann, R. Kiziah, P. C. Zimmer, J. McGraw, and D. Cox, "A systematic examination of ground-based and space based approaches to optical detection and tracking of satellites," in *31st Space Symposium*, 2015.

[2] J. Fletcher, I. McQuaid, P. Thomas, J. Sanders, and G. Martin, "Feature-based satellite detection using convolutional neural networks," in *Proceedings of the Advanced Maui Optical and Space Surveillance Technologies Conference*, 2019.

[3] D. Xue, J. Sun, Y. Hu, Y. Zheng, Y. Zhu, and Y. Zhang, "Dim small target detection based on convolutinal neural network in star image," *Multimedia Tools and Applications*, vol. 79, no. 7, pp. 4681–4698, 2020.

[4] H. N. Do, T.-J. Chin, N. Moretti, M. K. Jah, and M. Tetlow, "Robust foreground segmentation and image registration for optical detection of geo objects," *Advances in Space Research*, vol. 64, no. 3, pp. 733–746, 2019.

[5] T. Yanagisawa, H. Kurosaki, H. Banno, Y. Kitazawa, M. Uetsuhara, and T. Hanada, "Comparison between four detection algorithms for geo objects," in *Proceedings of the Advanced Maui Optical and Space Surveillance Technologies Conference*, vol. 1114, 2012, p. 9197.

[6] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE transactions on systems, man, and cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.

[7] E. Bertin and S. Arnouts, "Sextractor: Software for source extraction," *Astronomy and astrophysics supplement series*, vol. 117, no. 2, pp. 393–404, 1996.

[8] M. P. Lévesque, "Detection of artificial satellites in images acquired in track rate mode." in *Proc. AMOS-Tech. Conf., Wailea, Maui, Hawaii, 13–16 September 2011 E*, vol. 66, 2011.

[9] B. Flewelling and B. Sease, "Computer vision techniques applied to space object detect, track, id, and characterize," AIR FORCE RESEARCH LAB KIRTLAND AFB NM, Tech. Rep., 2014.

[10] R. Šára, M. Matoušek, and V. Franc, "Ransacing optical image sequences for geo and near-geo objects," in *Proceedings of the Advanced Maui Optical and Space Surveillance Technologies Conference*, 2013.

[11] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *arXiv preprint arXiv:1804.02767*, 2018.

[12] O. Ronneberger, P. Fischer, and 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.

[13] N. Heller, F. Isensee, K. H. Maier-Hein, X. Hou, C. Xie, F. Li, Y. Nan, G. Mu, Z. Lin, M. Han *et al.*, "The state of the art in kidney and kidney tumor segmentation in contrast-enhanced ct imaging: Results of the kits19 challenge," *Medical Image Analysis*, vol. 67, p. 101821, 2021.

[14] F. Isensee and K. H. Maier-Hein, "An attempt at beating the 3d u-net," *arXiv preprint arXiv:1908.02182*, 2019.

[15] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2980–2988.

[16] S. Suzuki *et al.*, "Topological structural analysis of digitized binary images by border following," *Computer vision, graphics, and image processing*, vol. 30, no. 1, pp. 32–46, 1985.

[17] C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger, "On calibration of modern neural networks," in *International Conference on Machine Learning*. PMLR, 2017, pp. 1321–1330.