FireNet: The case for simple and accurate fire detection in infrared linescan images

Vasileios Konstantakos
Institute of Informatics and Telecommunications
National Centre for Scientific Research "Demokritos"
vkonstantakos@iit.demokritos.gr

1 Introduction

Active fire detection from airborne imagery is of critical importance to the management of environmental conservation policies, supporting decision-making and law enforcement. This fact becomes especially evident during bushfire season, when accurate and quick information about the location and rate of spread of active fires is required. To collect this information, aircraft carrying infrared cameras fly over and record the intensity and location of fires. The resulting images are known as an 'infrared linescans' and are currently considered one of the best sources of information about fire intensity and location. After acquiring these images, the fire boundaries in each linescan image are manually labeled by hand-drawing polygons around the edges of the fire using geospatial software. However, during times of intense and rapid fire activity, this process can create a bottleneck in delivering timely information to operational firefighting teams. For this reason, we were motivated to propose and implement a system that performs this mapping task automatically, allowing for a more effective allocation of human resources.

This problem is formally known as image or semantic segmentation. Specifically, semantic segmentation can be defined as the process of partitioning the image into meaningful parts, and classifying each part at the pixel level into one of the pre-defined classes. This is a well-established field with many applications in the domain of fire detection. Various techniques have been proposed over the years, usually based on pixel or region-level comparisons involving sensor-specific thresholds and neighborhood statistics [1–4]. Moreover, the current success of deep learning in computer vision tasks has motivated researchers to explore such techniques for fire mapping and semantic segmentation.

In this report, we address the problem of active fire detection using a simple, accurate, and easily extensible approach. We showcase a quick and easy to follow workflow, which includes data exploration, pre-processing, model construction, and post-processing methods, involving both traditional and current deep learning techniques. The final model (FireNet) is only trained on 23 images and obtains an F_1 score equal to 71.08% in the public leaderboard.

2 Methods

This section describes the general process we followed from data exploration and preparation to model construction and refinement. It also illustrates the evolution of our approach and the decisions we made at each point. In general, the approach consists of three main pillars: (i) Image pre-processing using traditional segmentation methods [5, 6]; (ii) Segmentation model construction and training, using various architectures [7]; and (iii) Refinement of the resulting predictions with post-processing [8].

2.1 Data pre-processing

To begin with, we loaded and prepared the given dataset into a suitable format for the following steps of the process. After the initial data exploration, it became evident that we needed to clean up the noise in the images and make their signal clearer. This was especially important in order to extract meaningful information and make the subsequent model training easier and more robust. To accomplish that, we explored some traditional approaches in image segmentation, such as thresholding, edge detection and clustering. We briefly present these methods and their results in the rest of this section.

Threshold segmentation is one of the most commonly used and simplest segmentation techniques in region-based segmentation algorithms [9]. It automatically determines the optimal threshold according to a certain criterion, and uses this cut-off to achieve clustering. This threshold is determined based on the properties of the whole image or the neighboring pixels alone, defining a global and local thresholding method, respectively. We followed both manual, global thresholding approaches ("Manual thresholding"), and local, adaptive ones ("Adaptive thresholding", "Yen's method") [10]. Another effective segmentation approach uses edge detection. Edge detection algorithms use the different characteristics (e.g., local brightness, color discontinuity) of the important regions to achieve image segmentation. This discontinuity can often be detected using derivative operations and differential operators, such as the first-order Sobel and second-order Laplacian operator that we implemented [11]. Finally, we explored image segmentation based on clustering. These algorithms focus on the similarity between things as the criterion of class division, so that the same kind of samples are as similar as possible, and different ones are as dissimilar as possible [12]. The representative methods of this category, which we explored, are the Simple Linear Iterative Clustering (SLIC) method [13] and the Quick shift algorithm [14]. The results of this exploration can be seen in Figure 1.

2.2 Development of FireNet

Based on the results of the image pre-processing strategy and our further experimentation, we used the Laplacian of Gaussian (LoG) operator to process the images before feeding them into our model for training. The choice of LoG gave vastly superior results compared to the remaining operators. We also resized the input images into a shape of (768, 768).

In addition, given the recent success of deep learning on image segmentation, we explored various neural network architectures that are currently used by researchers in the field. These models are U-Net [15], LinkNet [16], Feature Pyramid Networks (FPN) [17], and Pyramid Scene Parsing Network (PSPNet) [18]. The overall architecture of each model can be seen in Figure 2. All models used the VGG16 [19] architecture for the encoder part, without loading the 'imagenet' weights.

We also selected these models since they have demonstrated good performance in situations with restricted training data. For instance, U-Nets have the ability to learn in environments of low to medium quantities of training data and can be very easily adapted to the desired application. Regarding the amount of training data, only 23 images were used; thus, data augmentation also played an essential role. In order to maintain a reasonable number of images and to avoid overfitting, we applied real-time data augmentation techniques to our training set. We used some standard augmentation techniques like horizontal/vertical flip, rotation, height/width shift, and brightness

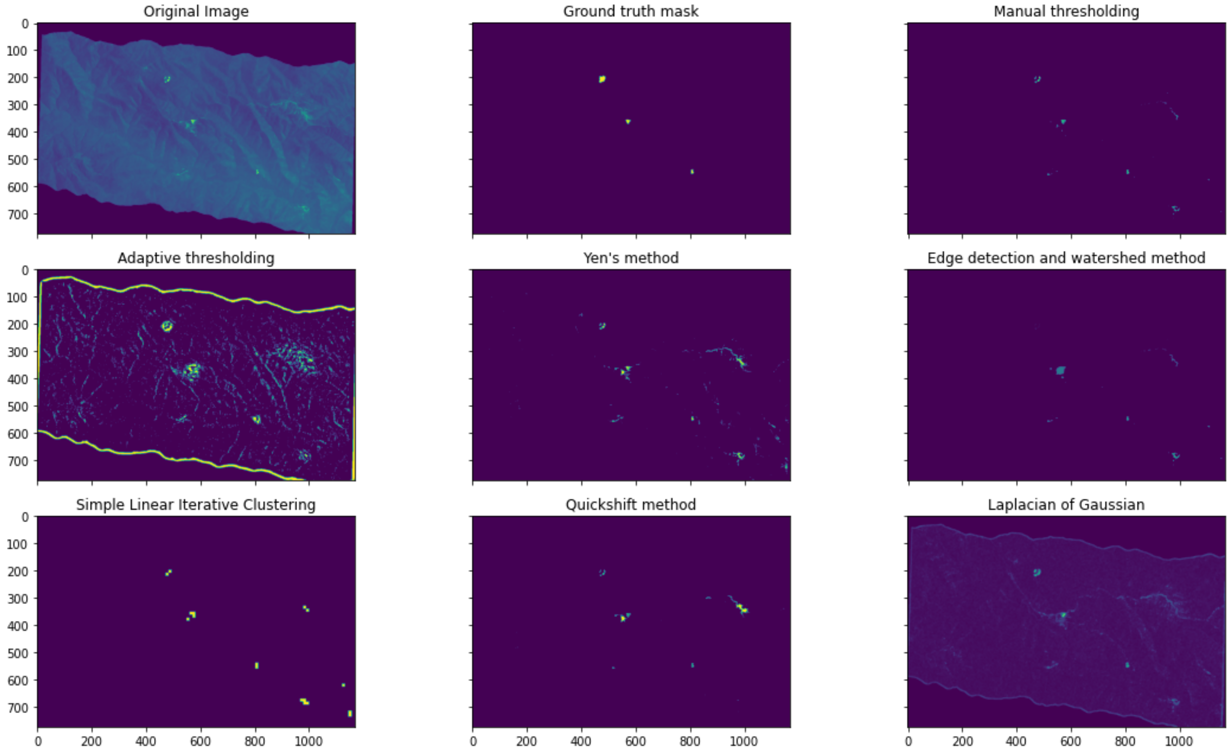

Figure 1: Image pre-processing using traditional segmentation approaches.

shift.

After evaluating the previous architectures on an 80/20 train-test split scheme, we selected the Feature Pyramid Network architecture as the model of our choice. We also compared the performance of each model trained under various loss functions (Binary Cross-Entropy, Jaccard Loss, Dice Loss, Tversky Loss, Focal Loss) [20] and optimizers (Adam, RMSProp, Stochastic Gradient Descent) [21]. The comparisons were done using the F_1 score as the evaluation metric. The results from the top performing models are shown in Table 1. Based on those, the final model was a FPN architecture trained under Jaccard loss, using the Adam optimizer. The chosen model was trained for 50 epochs with a batch size of 2, and the best model according to the validation F_1 score was kept. This final model (FireNet) was used to make predictions on the independent test set.

Table 1: Comparison of architectures and loss functions on a train-test split scheme.

Model	Training F_1 score	Validation F_1 score
FPN - Jaccard	**70.4%**	**63.4%**
FPN - Tversky	51.0%	50.2%
FPN - BCE-Dice	64.7%	60.7%
U-net Light - Jaccard	59.8%	55.6%

FPN: Feature Pyramid Network, BCE: Binary Cross-Entropy, U-net Light: Custom U-net architecture.

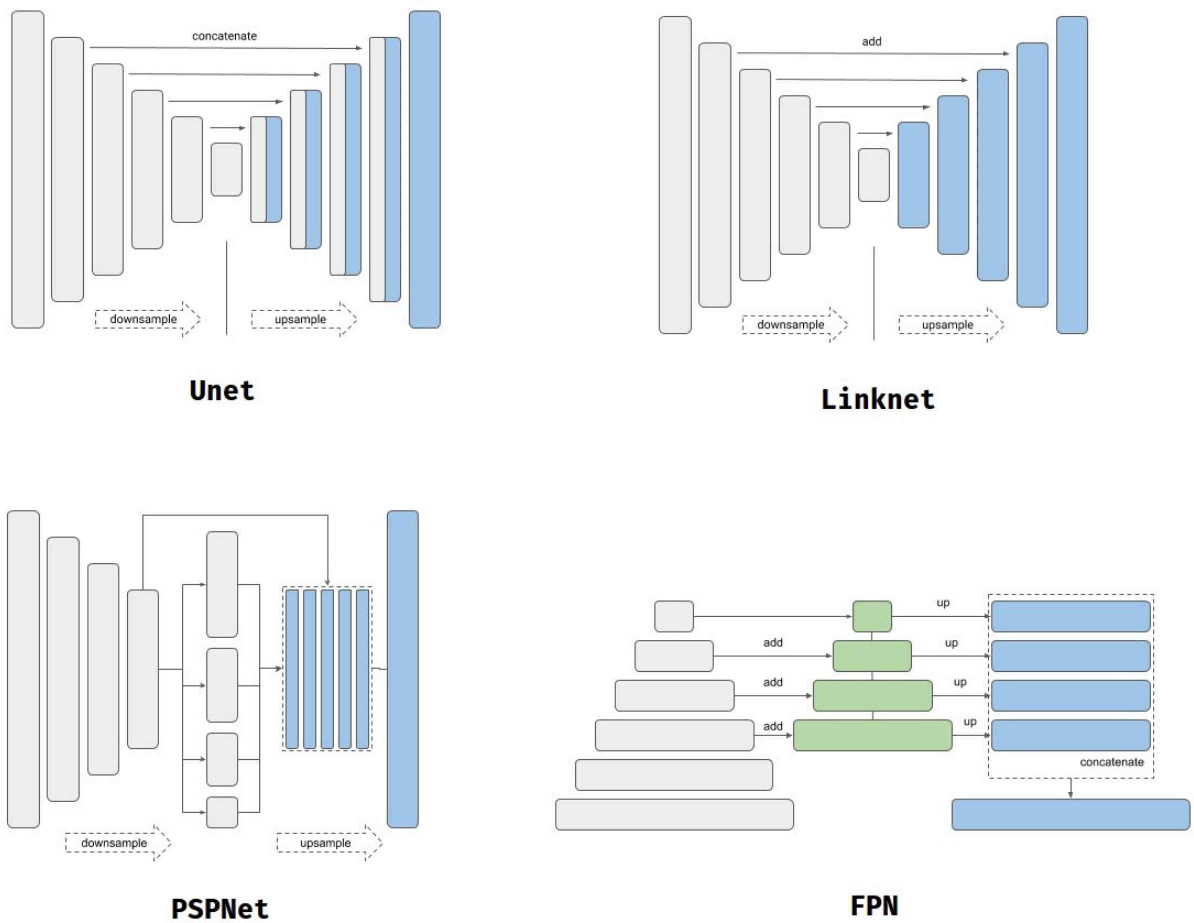

Figure 2: Architecture of various segmentation models.

2.3 Post-processing methods

Post-processing methods refine a found segmentation and remove obvious errors. For example, the morphological operations opening and closing can remove noise. The opening operation is a dilation followed by an erosion, which removes tiny segments. The closing operation is an erosion followed by a dilation. This removes tiny gaps in otherwise filled regions. They have been used previously for biomedical image segmentation [8]. Another way of refinement of the found segmentation is by adjusting the segmentation to match close edges [22, 23].

In our case, we implemented the opening and closing morphological operations to process the predicted segmentation mask, and we further refined the final predictions using a simple denoising strategy. We also tried fully connected conditional random fields as the final post-processing step in order to make our predictions smoother and found a negligible increase in the final validation metric; thus, they were not used in the final submission.

3 Discussion

Given the simplified approach that we followed, the proposed system has some benefits and drawbacks. First of all, the approach we described is unique, since it combines traditional and current deep learning segmentation techniques into one workflow. The presented pipeline can then be used to train the given model on larger datasets and extend its use case for many applications outside Australia. In addition, since the model only needs an input image to make a prediction, it can be used for mapping fires in any case around the world. That is, it may sacrifice some performance by not using location-specific features, but it acquires the additional benefit of being generalizable. Finally, due to the post-processing methods that were implemented, the model can sometimes overestimate the spread of the fire activity.

We conclude with some additional insights from our approach. The choice of the filtering strategy (i.e., Laplacian of Gaussian) and the use of post-processing to expand and denoise our predictions were crucial breakthroughs in our case. However, a lot of time was spent on getting familiar with the problem and the respective domain. Having a medical background and being familiar with biomedical image segmentation definitely helped too. Regarding the inference time, FireNet took an average of 32 ± 0.8 seconds to make the final submission.

4 Conclusion

In this report, we analyzed the provided dataset and explained our approach for effective image segmentation and fire detection in airborne images. We experimented extensively and found simple and adequate solutions for the problems we faced in the analysis of the dataset. To accomplish that, we incorporated various techniques, such as traditional segmentation approaches for pre-processing, deep learning segmentation models, and post-processing methods. We also kept the complexity of the given solution to the lowest possible level, while maintaining a good performance. Finally, the proposed approach can be easily extended to handle more training data and further improve its performance in the future.

References

[1] W. Schroeder, P. Oliva, L. Giglio, B. Quayle, E. Lorenz, and F. Morelli, "Active fire detection using landsat-8/oli data," *Remote sensing of environment*, vol. 185, pp. 210–220, 2016.

[2] S. W. Murphy, C. R. de Souza Filho, R. Wright, G. Sabatino, and R. C. Pabon, "Hotmap: Global hot target detection at moderate spatial resolution," *Remote sensing of environment*, vol. 177, pp. 78–88, 2016.

[3] S. S. Kumar and D. P. Roy, "Global operational land imager landsat-8 reflectance-based active fire detection algorithm," *International Journal of Digital Earth*, vol. 11, no. 2, pp. 154–178, 2018.

[4] P. Barmpoutis, P. Papaioannou, K. Dimitropoulos, and N. Grammalidis, "A review on early forest fire detection systems using optical remote sensing," *Sensors*, vol. 20, no. 22, p. 6442, 2020.

[5] M. Thoma, "A survey of semantic segmentation," *arXiv preprint arXiv:1602.06541*, 2016.

[6] S. Yuheng and Y. Hao, "Image segmentation algorithms overview," *arXiv preprint arXiv:1707.02051*, 2017.

[7] X. Liu, Z. Deng, and Y. Yang, "Recent progress in semantic image segmentation," *Artificial Intelligence Review*, vol. 52, no. 2, pp. 1089–1106, 2019.

[8] C. W. Chen, J. Luo, and K. J. Parker, "Image segmentation via adaptive k-mean clustering and knowledge-based morphological operations with biomedical applications," *IEEE transactions on image processing*, vol. 7, no. 12, pp. 1673–1683, 1998.

[9] R. M. Haralick and L. G. Shapiro, "Image segmentation techniques," *Computer vision, graphics, and image processing*, vol. 29, no. 1, pp. 100–132, 1985.

[10] R. C. Gonzalez, R. E. Woods, *et al.*, "Digital image processing," 2002.

[11] J. F. Haddon, "Generalised threshold selection for edge detection," *Pattern Recognition*, vol. 21, no. 3, pp. 195–203, 1988.

[12] Z. Wu and R. Leahy, "An optimal graph theoretic approach to data clustering: Theory and its application to image segmentation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 15, no. 11, pp. 1101–1113, 1993.

[13] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "Slic superpixels compared to state-of-the-art superpixel methods," *IEEE transactions on pattern analysis and machine intelligence*, vol. 34, no. 11, pp. 2274–2282, 2012.

[14] A. Vedaldi and S. Soatto, "Quick shift and kernel methods for mode seeking," in *European conference on computer vision*, pp. 705–718, Springer, 2008.

[15] 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*, pp. 234–241, Springer, 2015.

[16] A. Chaurasia and E. Culurciello, "Linknet: Exploiting encoder representations for efficient semantic segmentation," in *2017 IEEE Visual Communications and Image Processing (VCIP)*, pp. 1–4, IEEE, 2017.

[17] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2117–2125, 2017.

[18] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2881–2890, 2017.

[19] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[20] S. Jadon, "A survey of loss functions for semantic segmentation," in *2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*, pp. 1–7, IEEE, 2020.

[21] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[22] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik, "From contours to regions: An empirical evaluation," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2294–2301, IEEE, 2009.

[23] T. Brox, L. Bourdev, S. Maji, and J. Malik, "Object segmentation by alignment of poselet activations to image contours," in *CVPR 2011*, pp. 2225–2232, IEEE, 2011.