HIGH SCHOOL EDITION
Journal of Student Research
Volume 13 Issue 1 (2024)

Satellite Imagery Data Curation Workflow for Wildfire Detection with Advanced Segmentation Modeling

Anjali Singh

Saratoga High School, USA

ABSTRACT

Across the globe, wildfires are occurring at increased frequency, significantly impacting ecosystems and human civilizations. This research paper focuses on the efficacy of two of the most advanced semantic segmentation machine learning models, specifically U-Net based on convolutional neural network and SegFormer based on Vision Transformer network for wildfire detection utilizing Moderate Resolution Imaging Spectroradiometer (MODIS) satellite imagery. The dataset is assembled using a specially built pipeline composed of 1) a workflow to obtain the wildfire candidate location and date, and 2) a subsequent step to collect satellite imagery data utilizing Google Earth Engine image collection and image download service. Experimental evaluation on this dataset shows that both models demonstrate high predictive power of fire at specific geolocations, with ViT outperforming U-Net at the edges of fire regions.

Introduction

Global warming is increasing stress on earth's ecosystems with larger and more frequent wildfires being one of the most devastating consequences. In California, 6 of the top 10 largest wildfires have occurred in the last 4 years, the most destructive “camp” wildfire occuring in 2018 [1]. According to the National Interagency Coordination Center 2022 wildland fire report, 5 year average wildfire burn area was 8 million acres over 59 thousands fire incidents in the US [2]. The 2019 Australia wildfire burnt area was estimated to be 24 million hectares [3], affecting 80 percent of Australians [4]. Due to the destructive nature and severity of wildfires, it is crucial to detect them early, and employing a computer vision approach on remote sensing data can be crucial in this regard.

Satellite imagery, which has wide reach even at remote locations, is increasingly becoming an efficient tool at all stages of wildfire management, especially if combined with recent advancements in machine learning. For example, Pham et al. analyzed the wildfire detection accuracy of five distinct machine learning models on California state, county level remote sensing and fire incident report dataset, with the best accuracy of 89% and 97% for these datasets respectively [5]. Liu et al. evaluated the efficacy of power line density, enhanced vegetation index, vegetation optical depth, and distance to the wildland-urban interface features for a machine learning based fire ignition model, yielding the area under precision-recall curve of 0.90 for populated area [6]. Ben-Haim and Nevo from google research have created real-time tracking of wildfire boundaries with a Convolutional Neural Network model employing GOES-16, GOES-18, MODIS, and VIIRS satellite imagery achieving f1-score of 0.79 on US wildfire imagery [7].

Artificial Intelligence and Machine Learning areas are undergoing exponential advancements in recent years, including image understanding, classification, and segmentation [8]. In satellite imagery each pixel represents a spectroradiometer signal for a specific geographical location at a specific date and time. A single image can represent a wide swath of the earth’s surface. Therefore, wildfire detection is modeled as semantic

segmentation, with the goal to assign semantic labels to every pixel in an image. The widely available state of the art vision segmentation models are primarily trained on everyday images like persons, animals, etc., not for wildfire. The satellite imagery band values may be very different from the typical everyday image RGB values. These could present challenges to the vision models. Therefore, it's imperative to analyze and compare performances of the models specifically for wildfire. Here, we train and analyze in detail two of the top model architectures. These two are picked because both are state of the art models in this area and have distinct architectures. The first is U-Net, which is one of the best segmentation models based on convolutional neural network (CNN) architecture [9]. The second is SegFormer, which is one of the most current state of the art vision transformer (ViT) based segmentation model architecture [10].

Dataset Curation

Satellite Imagery

MOD14A1 V6.1 imagery collection from Moderate Resolution Imaging Spectroradiometer (MODIS) sensor aboard NASA Terra satellite, is the data source for wildfire modeling [11]. The collection is maintained by the NASA EOSDIS Land Processes Distributed Active Archive Center (LP DAAC) at the USGS Earth Resources Observation and Science (EROS) Center. Here, V6.1 denotes collection version 6.1, calibration and algorithm refinements of which are outlined in NASA user guide [12].

The imagery bands relevant to this research are - MaxFRP, QA and FireMask [12]. Each pixel in the image represents 1 square km of earth surface and each band in the pixel is the NASA algorithms' composited sensor data detected over a 24 hours period.

MaxFRP is the maximum radiative power of 4-micrometer and 11-micrometer infrared waves in megawatt. QA is the quality assurance bitmap. Bit indexes 0 and 1 denote water, coast, land or missing data. Bit index 2 denotes night or day. FireMask is the fire classification. Relevant fire classifications are - 0 (no fire), 7 (low confidence fire), 8 (nominal confidence fire), and 9 (high confidence fire).

MaxFRP and QA are inputs to the model, and FireMask is the label which the models are trained to predict.

Flowchart

At Earth size scale, at any given point of time, only a very small fraction of Earth experiences wildfire events. Therefore, an efficient method is designed to create high quality dataset from satellite imagery repositories. The flowchart in Figure 1 outlines these steps. The rest of the subsections describe this process in detail.

The satellite imagery itself is obtained utilizing Google Earth Engine platform which hosts a multi-petabyte catalog of satellite imagery and geospatial datasets for scientific analysis [13].

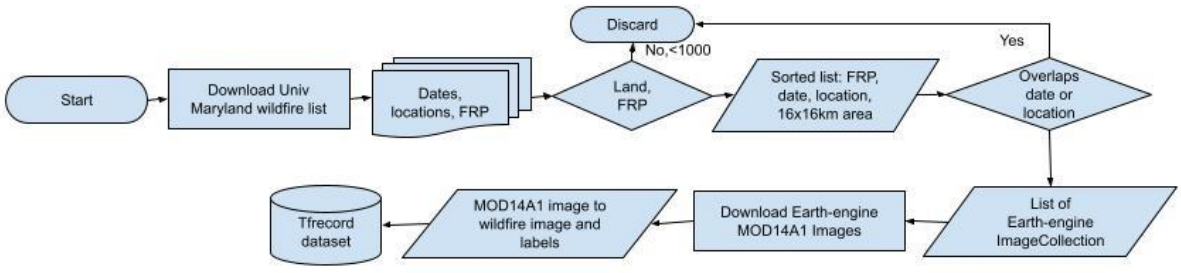

Figure 1. Modeling dataset processing flowchart. Starting from a potential wildfire list from the Univ of Maryland, it undergoes a number of steps to create a high efficacy wildfire list. These wildfire satellite images are retrieved from Google Earth Engine, transformed, and stored as tensorflow dataset.

Wildfire List

The first step is to identify time and location of the wildfire occurrences. This step is extremely important for efficiency of the data gathering process. University of Maryland maintains "Monthly Fire Location Product" in downloadable csv format at "sftp://fuoco.geog.umd.edu" [12]. For this research, all the monthly files pertaining to the year range 2019 to 2022 are chosen. These are the latest years for which complete 12 months of data are available. According to CA Fire, 6 of top 10 wildfires occurred during these 4 years [1]. This data acts as seeds for the potential wildfire time and location. For our purposes YYYYMMDD, lat, lon and FRP fields are used from the csv files. Here, lat is latitude, lon is longitude, and FRP is fire radiative power in megawatts.

In subsequent steps, a sequence of filtering logics are applied to improve efficacy of the dataset. A given monthly file could contain hundreds of thousands of lines of fire entries. For example, the 2022/07 file contains 420,871 entries. Not all these entries are equally useful or interesting. First, these are filtered to only keep the North American locations which are the chosen geo region of interest.

Next, to prioritize list entries that have higher probability of wildfires coverage, only entries with FRP value of 1000 or higher are kept. The underlying hypothesis being that the surrounding area of such a high FRP location will also have high FRP. This helps avoid cases where the corresponding satellite image only has a tiny fraction of wildfire coverage, making most of the pixels uninteresting.

The satellite image corresponding to a given entry will cover 16x16 square km of earth surface. A given image could cover many entries if the date and location overlap. Therefore, the list of entries are further pruned by removing overlapping date and location. For any overlap, the entry with highest FRP is kept, again the hypothesis being that it will improve data efficacy.

Google Earth Engine

The Earth Engine APIs are used to obtain the satellite image for each of the wildfire entries created in the previous step. The satellite dataset name is MODIS/061/MOD14A1 [13]. The longitude, latitude, radius, time, and bands are given as input to the API. The API then returns the images in numpy format with one dimension per band.

The numpy image shape is (16, 16, 3) representing 16x16 pixel image. Each pixel covers 1 square km of land. The 3 in the shape index is equivalent to RGB in a normal image format. MaxFRP band is the R, FireMask band is the G and QA band is the B.

Tensorflow Example Record

The downloaded numpy image array is used to create 2 arrays, one for input image and another for label. The input image is created by copying MaxFRP and QA band values into a (16, 16, 3) shape numpy array. Here, compared to a normal RGB image, MaxFRP is R, G is set to 0, and QA is B.

The label is created by copying FireMask band value into (16, 16, 1) shape numpy array. The FireMask value in the label represents 4 segmentation classes which the model will predict for each of the image pixels. The segmentation classes are - no fire (0 mask value), low confidence fire (7 mask value), nominal confidence fire (8 mask value) and high confidence fire (9 mask value).

The input image numpy array and the label numpy array are then each transformed into tf.train.Example features. The Example bytes are stored as TFRecord on the disk. A total of 5500 images are curated for training and validating the models. Overall, 4900 images are used for the train set and 600 images are used for the validation set.

Segmentation Modeling

Model Framework

The official keras computer vision website is chosen as the source of model artifacts. It provides high-quality, comprehensive reference model examples. The models are built according to the reference implementations available on the site. As per the instructions, U-Net is built and trained from scratch, whereas pretrained SegFormer is finetuned. ViT architecture is more complex than CNN, requiring a much larger training dataset and GPU resources. Finetuning instead of full training of SegFormer makes it a practical approach in the context of Google colab free small memory and GPU runtime resources.

U-Net Model

"U-Net: Convolutional Networks for Biomedical Image Segmentation" is based on a fully convolutional neural network, with skip connection encoder decoder architecture to make image segmentation accurate while utilizing very few training images [9]. Reference implementation for training the model end to end in keras is utilized [14]. The model has 2 million parameters.

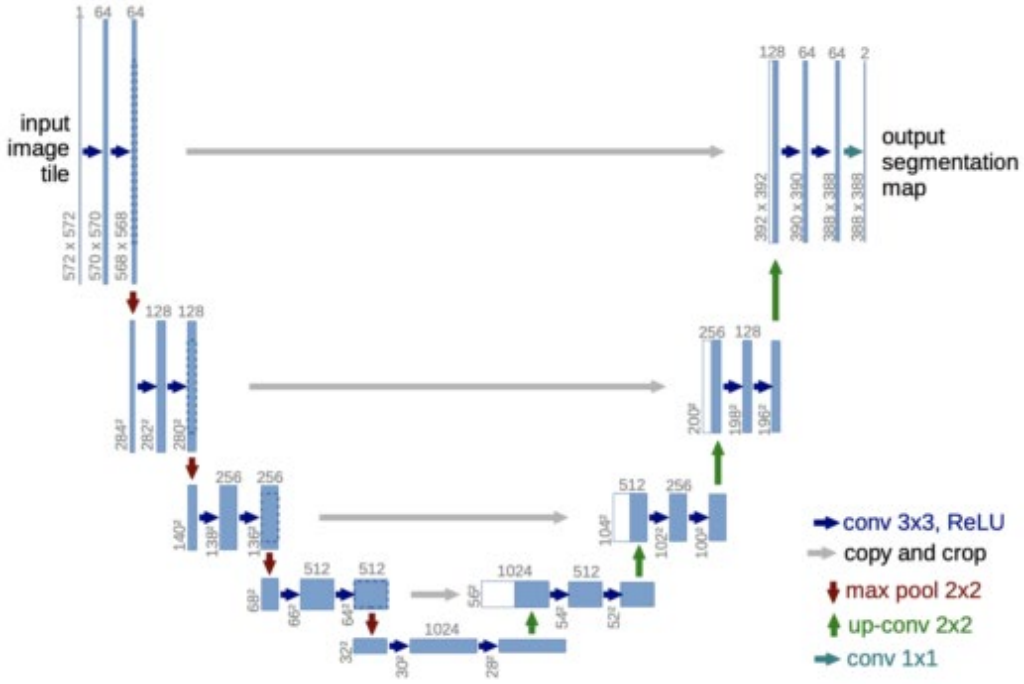

Figure 2. U-Net architecture diagram as shown in [9]. "U-net architecture (example for 32x32 pixels in the lowest resolution). Each blue box corresponds to a multi-channel feature map. The number of channels is denoted on top of the box. The x-y-size is provided at the lower left edge of the box. White boxes represent copied feature maps. The arrows denote the different operations."

SegFormer Model

"SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers" is a vision transformer (ViT) based model with changes to encoder and decoder specifically for improving segmentation performance and efficiency [10]. Pretrained model variant "nvidia/mit-b0" with 3.7 million parameters which is available at huggingface hub is finetuned with the wildfire dataset for our purposes. Reference implementation for finetuning the model in keras is utilized [15].

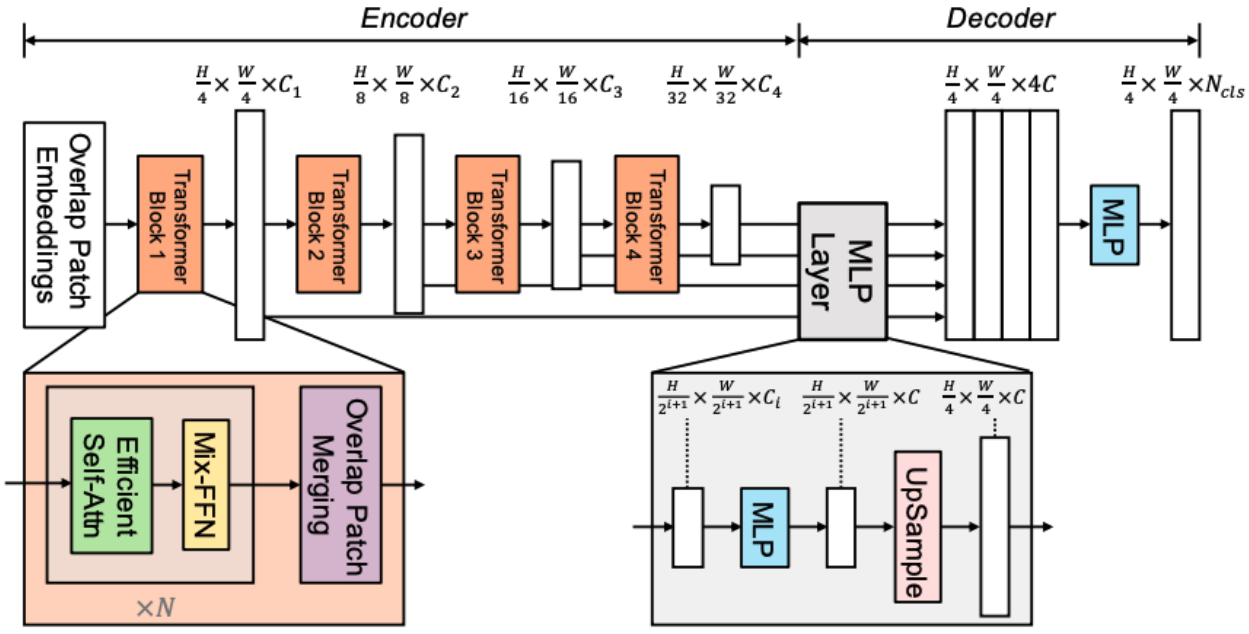

Figure 3. SegFormer architecture diagram as shown in [10]. "A hierarchical Transformer encoder to extract coarse and fine features; and a lightweight All-MLP decoder to directly fuse these multi-level features and predict the semantic segmentation mask. "FFN" indicates feed-forward network."

Model Training

Models are built and trained utilizing keras framework. U-Net is trained with sparse categorical cross entropy loss and adam optimizer. Training is stopped when validation loss plateaued, which was 75 epochs. The pretrained SegFormer has its own loss function implementation. It is finetuned utilizing adam optimizer with a really small learning rate of 6e-5. It was trained for 75 epochs, by which point validation loss plateaued.

Models were trained on free colab T4 GPU. Because of GPU resource availability limitations in the free account, models were checkpointed at the end of each epoch, and training were resumed over days as and when GPU became available in colab.

Results

Wildfire Imagery

In Figure 4 and Figure 5, I show the U-Net and SegFormer performances respectively for a 16x16 square km area of the 2023 Canada Fox Creek wildfire. The figures show fire radiative power, true labels, and predicted labels on the satellite imagery. In this specific satellite image, the fire regions are non-contiguous, which can present challenges to both models. Here, SegFormer is able to capture finer details around the fire boundaries better than U-Net.

Figure 4. U-Net qualitative result for 2023 Canada Fox Creek wildfire. Predicted accuracy around the edges of fire regions is low. Segmentation class colors are 0 - no color, 7 - purple, 8 - yellow, and 9 - red.

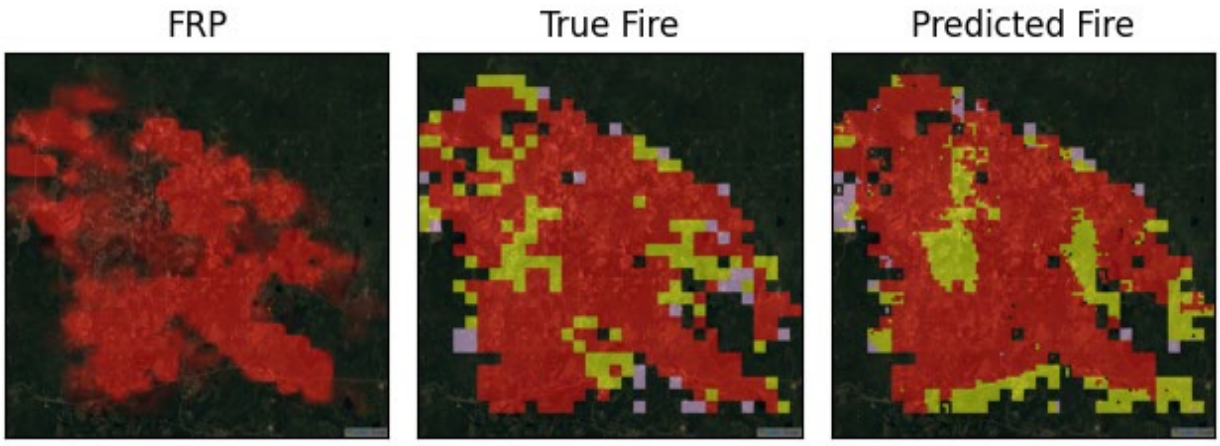

Figure 5. SegFormer qualitative result for 2023 Canada Fox Creek wildfire. Predicted accuracy around the edges of fire regions is substantially higher than U-Net. Segmentation class colors are 0 - no color, 7 - purple, 8 - yellow, and 9 - red.

Metrics

The model performances are measured on the validation dataset of 400 images. A larger number of images could not be used due to memory limitations of the free colab. The metrics are calculated using numpy, sklearn, matplot libraries.

Both U-Net and SegFormer models are able to determine no-fire class pixels with high accuracy, as shown in Table 1. The no-fire recall is at 0.99, therefore, only 1 percent of pixels are misclassified as fire regions. SegFormer outperforms U-Net on detecting edge regions of fire and differentiating nominal fire regions from high fire regions. U-Net and SegFormer f1-scores for nominal confidence regions are 0.54 vs 0.59 and high fire regions are 0.82 vs 0.84. Both U-Net and SegFomer achieve f1-score of 0.99 for detecting no fire regions.

Table 1. Comparison of U-Net and SegFormer metrics. A high precision model will help determine the right level of resources required to respond to the wildfire. A high recall model will help prevent wildfires from spreading by quick detection.

FireMask	**Precision**		**Recall**		**F1-score**	
	U-Net	**SegFormer**	**U-Net**	**SegFormer**	**U-Net**	**SegFormer**
0: **No** fire	0.99	0.99	0.99	0.99	0.99	0.99
7: **Low** confidence	0.14	0.15 (+0.01)	0.07	0.05 (-0.02)	0.10	0.07 (-0.03)
8: **Nominal** confidence	0.56	0.60 (+0.04)	0.52	0.58 (+0.06)	0.54	0.59 (+0.05)
9: **High** confidence	0.80	0.82 (+0.02)	0.83	0.86 (+0.03)	0.82	0.84 (+0.02)

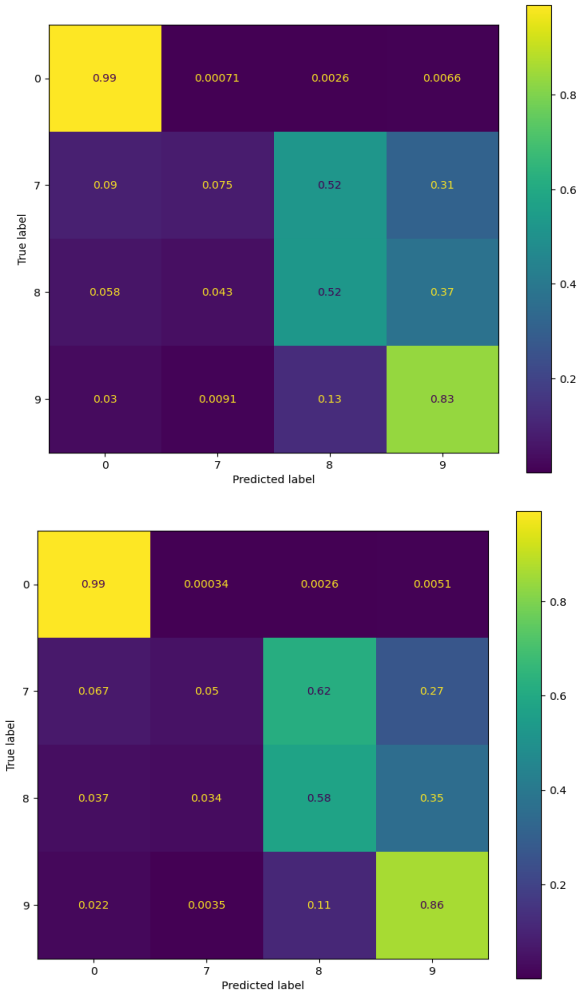

Figure 6. Confusion matrix for U-Net (left) and SegFormer (right). The fire labels are: 0 - no fire; 7 - low confidence; 8 - nominal confidence; and 9 - high confidence. The labels 0 and 9 have the highest accuracies, with SegFormer outperforming U-Net. Neither model performs well for 7 and 8.

Analysis

Segmenting within low, nominal, and high fire pixels accurately is where the models primarily differ in performances. The SegFormer f1-score is only 0.07 for the "low" class and improves to 0.84 for the "high" class. The SegFormer confusion matrix in Figure 6 shows that 62 percent of "low" are misclassified as "nominal" and 27 percent are misclassified as "high". Whereas, only 11% of "high" are misclassified as "nominal" and less than 1% are misclassified as "low". SegFormer performs better than U-Net for the "nominal" and "high" fire classes. Both models have difficulty identifying the "low" fire class.

Detecting fire class at the edges of a wildfire region is particularly difficult for U-Net. In Figure 4 "predicted fire" column, at the top and the left regions, the model misclassified all non "high" fires as "high" fires. Primary characteristics of these regions, as shown in the FRP column, is that the fire is non-contiguous. That is, fire and non fire regions are interspersed, creating many edges of fire regions. Similar misclassification can be seen at the bottom right region. As shown in Figure 5, the SegFormer model performs much better at both of these regions. The model detected fire regions comparable to the true fire regions. The "predicted" fire differs from the "true" fire at the region edges only in "high" vs "nominal". At the interspersed top left region

edges, many of the true “nominal” are predicted as “high”. At the long tail fire region at bottom right, many of the true “high” are detected as “nominal”.

Conclusion

The advanced image segmentation models, although not developed for satellite imagery signals, and these signal values are very different from RGB values in a typical photographic image, adapt very well to the wildfire detection. The state of the art ViT architecture based SegFormer outperformed U-Net, specifically at the edges of the segmentation regions with varied FRP values. The edge fire regions are important from fire management and from a human built area perspective. If applied with real time satellite imagery, the model prediction can be used to inform and respond early in those edge regions. As future work, specific deficiency of the model like “low” class accuracy can be improved in various ways by improving pixel count balance of the different classes, or by using weighted loss function. The dataset can also be improved by incorporating signals from other statellies like GOES-R. In closing, the machine learning models are a valuable resource when combined with satellite imagery.

Acknowledgments

The author would like to thank Xiao Dong, PhD of Polygence for providing the research mentorship in this paper.

References

[1] *Statistics*. (2022, October 24). California Department of Forestry and Fire Protection. Retrieved January 1, 2024, from https://www.fire.ca.gov/our-impact/statistics

[2] *National Interagency Coordination Center Wildland Fire Summary and Statistics Annual Report 2022*. (2022). National Interagency Fire Center. Retrieved January 1, 2024, from https://www.nifc.gov/sites/default/files/NICC/2-Predictive%20Services/Intelligence/Annual%20Reports/2022/annual_report.2.pdf

[3] Binskin, M. (2020, October 28). *Royal Commission into National Natural Disaster Arrangements Report*. Royal Commission into Natural Disaster Arrangements. https://www.royalcommission.gov.au/natural-disasters/report

[4] Hughes, L., Dean, A., Steffen, W., Rice, M., & Mullins, G. (2020, November 3). *Summer of crisis*. Climate Council. Retrieved December 25, 2023, from https://www.climatecouncil.org.au/resources/summer-of-crisis/

[5] Pham, K., Ward, D., & Rubio, S. (2023). California Wildfire Prediction using Machine Learning. *2022 21st IEEE International Conference on Machine Learning and Applications (ICMLA)*. https://doi.org/10.1109/ICMLA55696.2022.00086

[6] Liu, Y., Le, S., & Zou, Y. (2023). A simplified machine learning based wildfire ignition model from insurance perspective. *ICLR 2023 Workshop on Tackling Climate Change with Machine Learning*.

[7] Haim, Z. B., & Navo, O. (2023, February 3). Real-time tracking of wildfire boundaries using satellite imagery. *Google Research*. https://blog.research.google/2023/02/real-time-tracking-of-wildfire.html

[8] Thisanke, H., Deshan, C., & Chamith, K. (2023, May 5). *Semantic Segmentation using Vision Transformers: A survey*. https://doi.org/10.48550/arXiv.2305.03273

[9] Ronneberger, O., Fischer, P., & Brox, T. (2015, May 18). *U-Net: Convolutional Networks for Biomedical Image Segmentation*. https://doi.org/10.48550/arXiv.1505.04597

[10] Xie, E., Wang, W., & Yu, Z. (2021, May 31). *SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers*. https://doi.org/10.48550/arXiv.2105.15203

[11] Giglio, L., Justice, C. (2021). *MODIS/Terra Thermal Anomalies/Fire Daily L3 Global 1km SIN Grid V061* [Data set]. NASA EOSDIS Land Processes Distributed Active Archive Center. Retrieved January 1, 2024, from https://doi.org/10.5067/MODIS/MOD14A1.061

[12] Giglio, L., Schroeder, W., & Hall, J. V. (2021, May). *MODIS Collection 6 and Collection 6.1 Active Fire Product User's Guide Version 1.0*. MODIS Active Fire Products. Retrieved January 1, 2024, from https://lpdaac.usgs.gov/documents/1005/MOD14_User_Guide_V61.pdf

[13] Google Earth Engine (2023). A planetary-scale platform for Earth science data & analysis. https://developers.google.com/earth-engine/datasets/catalog/MODIS_061_MOD14A1

[14] Chollet, F. (2019, March 20). *Image segmentation with a U-Net-like architecture*. Keras. https://keras.io/examples/vision/oxford_pets_image_segmentation/

[15] Paul, S. (2023, January 25). *Semantic segmentation with SegFormer and Hugging Face Transformers*. Keras. Retrieved December 25, 2023, from https://keras.io/examples/vision/segformer/