Classifying Wildfires in Satellite Imagery

Connor Couture
Department of Computer Science
University of California, Los Angeles
connorcouture@g.ucla.edu

Shirley Qi
Department of Computer Science
University of California, Los Angeles
shirleyqi@g.ucla.edu

Joseph Wong
Department of Computer Science
University of California, Los Angeles
josephwong@g.ucla.edu

Abstract

As the impacts of climate change begin to be felt, along with certain forestry policies, wildfires have become a regular occurrence all over the world. Research has shown that climate change will likely increase fire risks in Pacific Northwest forests [Halofsky et al., 2020]. At the same time, forestry practices employed in the West Coast of the United States arguably make extreme fire risks worse as prior indigenous forestry practices which included more regular small-scale deliberate burns reduced the risk of extreme wildfires [Hessburg et al., 2021]. With large-scale wildfires more common, it is important to have an efficient and accurate way to identify the areas burned by wildfires. With satellite data widely available in today's world, there is now enough data to train a model to identify these areas. Using the Satellite Burned Area Dataset, we trained several models to identify areas from satellite images which have burn scars from recent wildfires. The UNet architecture (implemented in PyTorch) is one of the top models for medical image segmentation, so this was one of the models we decided to experiment on for this dataset. Some other models we explored on this dataset include a pre-trained version of Meta's Segment Anything model, and FastAI's UNet ResNet34 model implementation.

1 Introduction

1.1 Dataset

The dataset that we are using is called the Satellite Burned Area Dataset [Colomba et al., 2022]. The dataset consists of images from 73 various areas in Europe. The dataset include multiple satellite images of a given area taken on different dates (generally a month or two apart), as well as cloud coverage images to indicate any areas of the image covered by clouds. The pre- and post-burn dates are stored in a CSV file, which contains lots of metadata for each region, including the name of its folder, latitude, longitude, height, width, pre-burn date, post-burn date, and many other features which are not relevant to this project. Most images are provided in both PNG and TIFF formats: the TIFF images are higher resolution and contain 12 channels, while the PNG files are lower resolution and only contain the standard 3 channels. A few regions only contain images in TIFF format. Each different area has an associated mask, which was created using the Copernicus API. Since the images are large, we split the images into 512 x 512 pixel tiles. Finally, pixels in the mask are encoded as 0 if it is not a burned pixel and 1 if it is a burned pixel.

(a) Example of post-fire image from dataset.

(b) Associated mask from image in Figure 1a

Figure 1: Example input and target from Satellite Dataset.

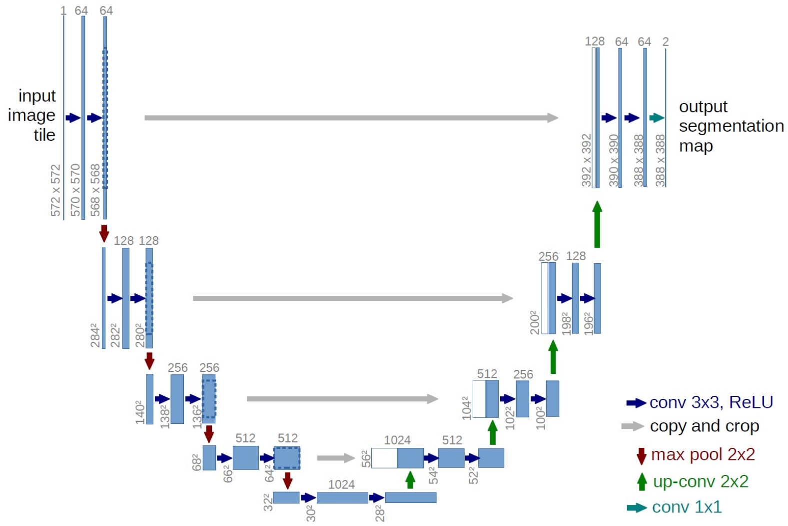


Figure 2: Example of UNet architecture [Ronneberger et al., 2015].

1.2 UNet Model

We are using a UNet Model [Ronneberger et al., 2015] to help classify whether an image has wildfire burns or not. The UNet model consists of three parts: a contracting path, an expanding path, and skip connections. The contracting path contains encoder layers that capture contextual information and reduce the spatial resolution of the input, while the expansive path contains decoder layers that decode the encoded data and use the information from the contracting path via skip connections to generate a segmentation map.

The contracting path in UNet is responsible for identifying the relevant features in the input image. The encoder layers perform convolutional operations that reduce the spatial resolution of the feature maps while increasing their depth, thereby capturing increasingly abstract representations of the input. This contracting path is similar to the feedforward layers in other convolutional neural networks. On the other hand, the expansive path works on decoding the encoded data and locating the features while maintaining the spatial resolution of the input. The decoder layers in the expansive path upsample the feature maps, while also performing convolutional operations. The skip connections from the contracting path help to preserve the spatial information lost in the contracting path, which helps the decoder layers to locate the features more accurately.

2 Methodology

2.1 UNet Model

The UNet model was trained on the Satellite Burned Area Dataset, which was split into 512 x 512 pixel tiles. Specifically, we only used the post-burn data to train the model - this was identified by the date of the image in comparison with the post-burn dates provided in the CSV file. Including the pre-burn data would have resulted in further data imbalance, which could influence the model to predict that an area is unburned due to the number of unburned pixels it was trained with. Since burned tiles also contain unburned regions, adding additional unburned training data is not necessary: in the entire dataset, around 91.9% of the areas in the images already consists of unburned tiles; adding pre-burn data would only increase the amount of unburned training data being fed into the model.

In training the model, we used the PNG images due to the restrictions from Google Colab: training on the TIFF images would result in our instance crashing due to running out of GPU. As a result, 9 of the 73 areas in the dataset were skipped in this training due to not having PNG versions of the satellite images for their regions.

The UNet model was trained over 30 epochs with a batch size of 4. This small batch size was used in consideration with the GPU limits from Google Colab. We split the data into training and validation data, where 80% of the data was used as training data, and 20% was used for validation. We used the dice coefficient as the loss function for training the model. The dice coefficient is computed using the following equation:

$$DSC = \frac{2 * |A \cap B|}{|A| + |B|}$$

We use the dice coefficient as the loss function since it is more sensitive to overlap between the predicted and ground truth masks. A represents the set of pixels of the predicted mask, and B represents the set of pixels of the ground truth mask.

A high dice coefficient value indicates a high level of similarity between the predicted and ground truth masks, meaning that the segmentation model or algorithm is performing well. Conversely, a low dice coefficient value indicates poor segmentation performance.

We used the IOU (Intersection over Union) metric to measure the accuracy of the model. This is generally the standard metric used for measuring image segmentation accuracy as it measures the similarity between the predicted area and the true area; however, the dice score coefficient is used in training because the IOU metric is not differentiable, while the dice score metric is. In order for gradient descent to function, the function must be differentiable. The IOU metric is given by the following equation:

$$\text{IOU} = \frac{A \cap B}{A \cup B}$$

In the equation above, A is the set of pixels which are predicted to be burned, and B is the set of pixels which are labeled as burned in the true mask. For this particular problem, since we are doing binary classification, this equation can be rewritten as:

$$\text{IOU} = \frac{TP}{TP + FP + FN}$$

TP denotes the true positives, FP denotes the false positives, and FN denotes the false negatives.

2.2 Segment Anything

The Segment Anything Model (SAM) by Meta AI is an image segmentation model inspired by NLP tasks. The model can be prompted with bounding boxes, points, or even full masks to help the model generalize to any segmentation task.

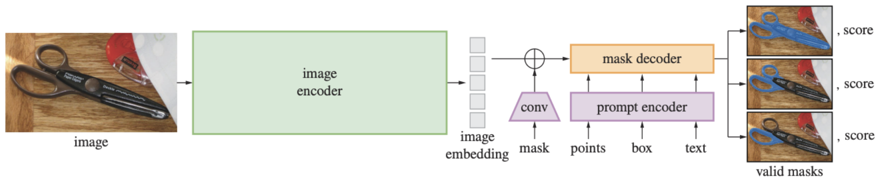


Figure 3: SAM architecture overview. The model is broken up into three parts: image encoder, prompt encoder, and mask decoder [Kirillov et al., 2023].

SAM is comprised of three parts: an image encoder, a prompt encoder, and a mask decoder. The image encoder uses an MAE pre-trained Vision Transformer that is adapted for higher resolution images. The prompt encoder works with 2 sets of prompts: sparse and dense prompts. Sparse prompts consists of bounding boxes, points, or text, where as dense prompts consits of masks. Bounding boxes and points are encoded using a positional encoder, and text is encoded using the CLIP text encoder. Masks are embedded using convolutions and summed with the image. The mask decoder takes the summed image embedding and prompt embeddings to produce a list of masks, each with their respective IOU scores [Kirillov et al., 2023].

2.3 FastAI UNet-ResNet34

The FastAI library (https://docs.fast.ai/) provides a more abstracted method of experimenting with models, while we were having trouble getting the main UNet working, we decided to run several experiments using FastAI. FastAI uses UNet for image segmentation, and we went with the UNet model with the ResNet34, this works by having the encoder of the UNet encoder-decoder architecture being the ResNet34 model, while the decoder part of the model is still the regular UNet architecture, this is supposed to lead to better results with image segmentation [Thomas, 2019]. The FastAI UNet-ResNet34 model uses cross entropy loss as the loss function. ResNet is a foundational model architecture that introduces a conception of residuals where a function with the input subtracted from it is learned, which makes it feasible to train even deeper neural networks, the authors of ResNet achieved state-of-the-art performance on a variety of image related tasks such as image classification and image segmentation at the time of publication [He et al., 2015]. ResNet34 in particular has 34-layers [He et al., 2015].

3 Model Results

3.1 UNet

The UNet model achieved an average IOU score of **0.6296** with 30 epochs, with a maximum score of 0.8753 on one of the images in the validation dataset. (With 50 epochs, the model achieved an average IOU score of 0.6077 and a similar maximum score). We also tried training the model over 100 epochs, but Google Colab had problems timing out: regardless, it is not clear that this would have improved our model's results.

3.2 Segment Anything

We downloaded the trained SAM model. The image encoder has 632M parameters, and the prompt encoder and mask decoder have 4M parameters. We ran our dataset through the mask decoder to see how well the model would segment post-fire images. In general, we saw that the model was able to generate a segment for the burned section, but it also had masks for other parts of the image. The average predicted IOU score when running SAM on the entire dataset is **0.94459**. An example can be seen in Figure 5.

3.3 FastAI UNet-ResNet34

Our first experiment was using 100 epochs trained on the entire wildfire dataset (so images that were before land was burned, and after land was burned). With this we achieved a training accuracy of 0.9736 and a validation accuracy of 0.4372 (it appears overfitting occurred here). After this

(a) Example of post-fire image from dataset.

(b) Associated mask from image in Figure 7a

(c) Predicted mask from image in Figure 7a using trained UNet

Figure 4: Example predicted mask from trained UNet.

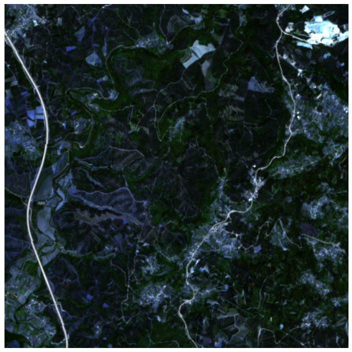
experiment, we conducted further experiments with the post burn dataset we used with our UNet and SAM models. Using the post burn dataset, we achieved better results, as expected. With 30 epochs, we achieved a training accuracy of 0.8714 and a validation accuracy of **0.8809**. With 50 epochs, we achieved a training accuracy of 0.9636 and a validation accuracy of 0.8601, it seems after 30 epochs, the model started overfitting the data.

4 Conclusions

All of the models we trained showed good performance given our resource limitations. While not developed specifically for the task of segmenting wildfire burn scars from satellite imagery, the UNet architecture was able to generalize quite well when given correctly processed training data and masks. The SAM model was able to identify the burned tiles without any form of training, but had some extra mask noise generated as well. With fine-tuning we believe that SAM could identify which masks represent the burned areas. And finally, we have found that the FastAI UNet-ResNet34 architecture resulted in the best performance, revealing the strength of combining a ResNet encoder with the UNet decoder in this particular segmentation task.

5 Future Work

Given more time and computing resources, we would experiment with training the UNet model over more epochs: we only trained the model on 30 epochs due to Google Colab's time limits for GPU units. It would be interesting to see the model's performance given more epochs of training. Additionally, we would use the higher-resolution TIFF images, which have 12 input channels instead of the 3 input channels of PNG images. This would potentially allow the model to observe more patterns which could be used to more accurately identify burn scars from satellite imagery. Due to time and resource limitations, we did not consider the cloud coverage images provided in the dataset

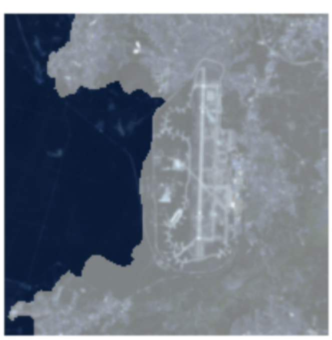


(a) Example of post-fire image from dataset.

(b) Associated mask from image in Figure 5a

(c) Predicted mask from image in Figure 5a using pre-trained SAM model

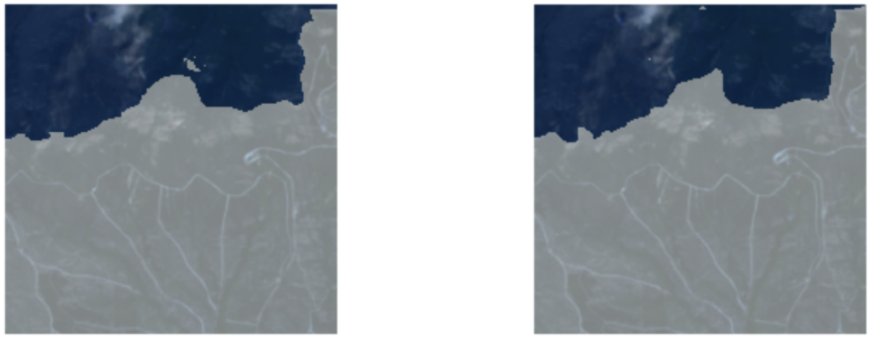
Figure 5: Generated mask from pre-trained SAM model. When compared to the original mask in Figure 5b, we can see that it has a mask generated in the same area as the original mask (green mask bottom-middle right), but also has masks generated in other areas of the image. IOU scores for all the masks are very high, and the mask that is the same as the original mask has a predicted IOU score of 0.975.



(a) Ground Truth.

(b) Prediction.

Figure 6: Example predicted masks from 30 epochs trained FastAI UNet-ResNet34.



(a) Ground Truth.　　(b) Prediction.

Figure 7: Example predicted masks from 50 epochs trained FastAI UNet-ResNet34.

in training our model. This could also potentially improve our model's accuracy, as our model would sometimes predict the shadows of the clouds as burn scars.

For SAM, we looked into fine-tuning the SAM model in order to give a single mask to the burned area and ignore other segmentations. However, given limited GPU availability and inefficiency of training with CPU, we were unable to run our fine-tuning code. This is something that we would like to explore in the future given more resources.

For the FastAI UNet-ResNet model, in the future, we could experiment with other ResNet model types (we only used the ResNet34 encoder) to see how this affects the performance.

Additionally, the original dataset labeled burned areas under five different classes of burn levels, ranging from class 0 (no burn) to class 4 (completely destroyed). Our model only predicted whether an area was burned or not for simplicity due to time limitations - in the future, we could expand our model to predict the severity of the burn in addition to whether or not the area is burned or not.

6 Code, Weights, and Dataset

We have placed the relevant Jupyter notebooks for our models, model weights, and dataset in a shared google folder. In the Jupyter notebook for our final UNet model, we also have several cells at the bottom that have a function that allows one to input an image batch from the dataset, and get (and see) the model predictions of the burned area.

References

Luca Colomba, Alessandro Farasin, Simone Monaco, Salvatore Greco, Paolo Garza, Daniele Apiletti, Elena Baralis, and Tania Cerquitelli. A dataset for burned area delineation and severity estimation from satellite imagery. In *Proceedings of the 31st ACM International Conference on Information amp; Knowledge Management*, CIKM '22, page 3893–3897, New York, NY, USA, 2022. Association for Computing Machinery. ISBN 9781450392365. doi: 10.1145/3511808.3557528. URL `https://doi.org/10.1145/3511808.3557528`.

Jessica Halofsky, David Peterson, and Brian Harvey. Changing wildfire, changing forests: the effects of climate change on fire regimes and vegetation in the pacific northwest, usa. *Fire Ecology*, 16:4, 12 2020. doi: 10.1186/s42408-019-0062-8.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015.

Paul F. Hessburg, Susan J. Prichard, R. Keala Hagmann, Nicholas A. Povak, and Frank K. Lake. Wildfire and climate change adaptation of western north american forests: a case for intentional management. *Ecological Applications*, 31(8):e02432, 2021. doi: https://doi.org/10.1002/eap.2432. URL `https://esajournals.onlinelibrary.wiley.com/doi/abs/10.1002/eap.2432`.

Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C. Berg, Wan-Yen Lo, Piotr Dollár, and Ross Girshick. Segment anything, 2023.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation, 2015.

Christopher Thomas. U-nets with resnet encoders and cross connections, 2019. URL `https://towardsdatascience.com/u-nets-with-resnet-encoders-and-cross-connections-d8ba94125a2c`.