Vision-based Wildfire Smoke Detector

yoolim99@stanford.edu | link to Google Drive

Introduction

Background: Wildfire has been deadly and destructive at unprecedented levels over the past 2 years in California. To combat this, there are over 300 cameras installed for wildfire— some repurposed from monitoring wildlife, others new — and this number is expected to grow within the next few years [1]. And as cameras become more universal and strategically placed to monitor wildfires in areas of risk, we can leverage computer vision algorithm to detect and localize smoke, alerting those at watch quickly and at scale.

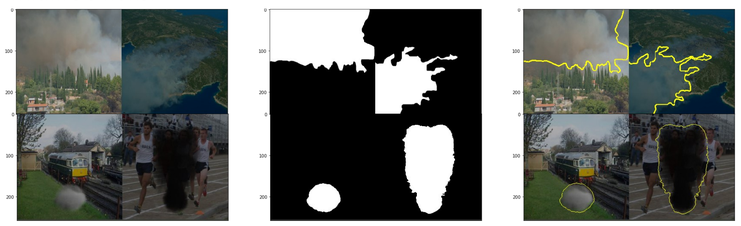
Objective: The primary objective of the project is to be able to isolate the boundaries of smoke for more accurate detection and guidelines. The scope is only for smokes since they are often early indicators for wildfires. For this we can use the U-net architecture trained against smoke boundaries for early detection and faster human verification.

Smoke Dataset

To train a U-net, we require two types of training data: an image with smoke, and another representation of the same image with 'masks', or segmentations. We rely on external datasets that provide either a hand-segmented smoke dataset from Wildlife Observers and Smoke Recognition (WOSR) [2] or synthetic smoke overlaid on top of other images [3].

WOSR data: Our data from WOSR consists of 113 images, each 1280x960 pixels, RGB format, with three classes: *smoke* zone in white, *maybe-smoke* zone in gray, and *no-smoke* zone in black. However, majority of segmentation data labeled as maybe-smoke zones in gray clearly did not have smoke. The segmentation data was thus reformatted by changing the gray values of *maybe-smoke*. This is also consistent with the synthetic data available.

Synthetic data: The synthetic images consists of 70,000 images, each 256x256 pixels, RGB format, with two classes: *smoke* zone in white and *no-smoke* zone in black. These are simulated smoke patterns that has been overlaid in random background images. These create synthetic datasets with the appropriate segmentation labels. Despite the abundance in the number of images, we only randomly sampled 500 of these due to the possibility of overweighing the patterns of synthetic smoke and not wildfire smoke.

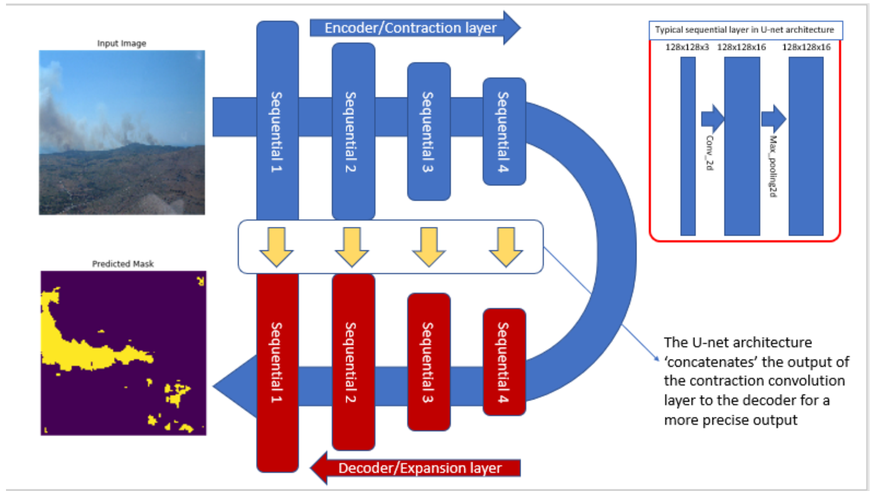


Left, original image with top two being WOSR images and bottom two synthetic images. **Middle**, ground truth of smoke segmentation. **Right**, preprocessed images and segmentation overlaid for verification.

Model Architecture

Given our objective of being able to detect and identify smoke, we choose a U-net architecture because it has a unique sequence of layers consisting of encoders and decoders that enables us to represent what was learned from the original features. A traditional CNN layer is mainly consisted of 4 operations: a convolution step that 'convolves' the features by performing an elementwise multiplication with a kernel matrix, a ReLU step that forces the function to be nonlinear, a pooling step to downsample and reduce dimensionality, and an output layer. This architecture is only able to learn the features of smoke given an image, but does not have an ability to represent back what it has learned.

A U-net overcomes this limit with two main architectures. [4] The first consists of a contracting network that acts as a encoder similar to a traditional CNN, but it increases the number of features by times 2 while decreasing the resolution of the image. The second half consists of creating a series of expansion networks where features from the contraction path is mapped back to the original pixel. This allows the features learned from smoke images from the first half of the architecture to accurately map back to the second half of expansion layers.



We standardize each image into (128, 128) pixels, and a total of 613 pictures and masks are split using a stratified sample across the WOSR and synthetic data into 500 training, 113 validation images.

Because the training size is small, we rely on a pretrained MobileNetV2 [5] model as part of the Tensorflow package as the encoder to reduce the complexity cost and model size of the network while being an effective feature extractor. Other training parameters are as follows:

Epochs: 100

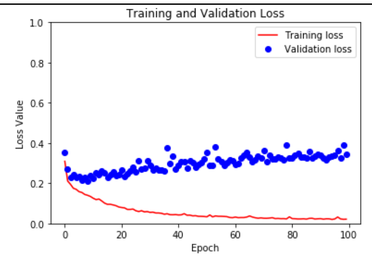
Loss: sparse categorical crossentropy, which is useful in multiclassification with initial three labels in mind (smoke, maybe-smoke, no-smoke)

Optimizer: Adam-optimizer

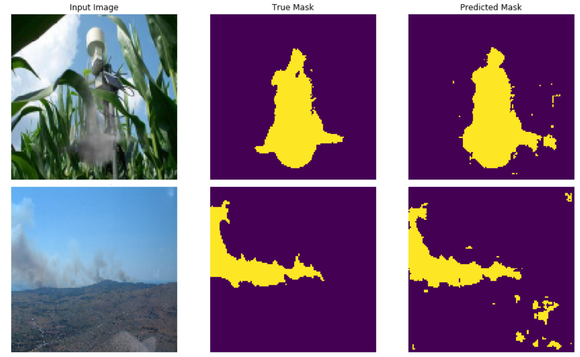
Metric: Accuracy, or (TP + TN) / (TP + TN + FP + FN)

Results and Future Steps

	Train	Validation
Accuracy	0.9965	0.9079
Loss	0.0091	0.4264



We can see from the validation loss that our model overfits with more epochs of training. The best results are yielded on the third or final epoch, though from the mask results from the initial model we do see some level of accuracy acceptable to human eye.



Left, original image with top being synthetic image and bottom WOSR image. **Middle**, ground truth of smoke mask/segmentation. **Right**, model results from the smoke-segmentation U-net model.

Although not shown here, the model was drastically improved with the introduction of synthetic datasets. For future steps we can add more data to the model to see if it increases the accuracy of the model. We can also try tuning the hyperparameters, and also explore other semantic segmentation model architectures.

References

[1] Wildfire Camera Networks Spread Across California, Celina Tebor - https://www.sandiegouniontribune.com/news/environment/story/2019-10-24/wildfire-camera-networks-spread-across-california

[2] Welcome to the Wildfire Observers and Smoke Recognition Homepage. (n.d.). Retrieved October 20, 2019, from http://wildfire.fesb.hr/

[3] Z. Xu, J. Xu, "Automatic Fire Smoke Detection Based on Image Visual Features," Proceedings of the 2007 International Conference on Computational Intelligence and Security Workshops, 316-319, 2007.

[4] O. Ronneberger, P. Fischer, T. Brox, "Medical Image Computing and Computer-Assisted Intervention (MICCAI), Springer, LNCS, Vol.9351: 234-241, 2015

[5] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, L. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," The IEEE Conference on Computer Vision and Pattern Recognition, 4510-4520, 2018.