Computer Vision Aided Fire Localization for Wildfire Monitoring

Laura Spector

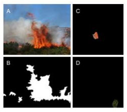
Department of Genetics, Stanford University

Stanford
CS230 Final Project

Introduction

- Every year since 2000, an average of 72,400 wildfires burned an average of 7.0 million acres in the U.S.
- Scarcity of wildfire image and video databases make training with smaller data sets necessary.
- Previous studies have used pixel-level fire identification for the purpose of non-temporal fire localization with deep neural networks, either end-to-end or for classification of image patches or superpixels pre-processed from fire images.
- I present a transfer learning approach that improves our ability to specifically localize wildfires.
  - Input: isolated superpixels that are pre-segmented from images of wildfires.
  - Method: fine tune a reduced Inception v1 CNN to classify wildfire superpixels as 'fire' or 'non-fire'. Positively predicted ('fire') superpixels for each image are then reconstructed to localize the predicted fire within the original image space.

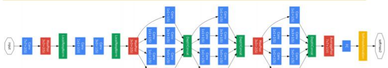
Data and Features



- Corsican Fire Database (595 images paired with binary ground truth images)
- Split 7:2:1 into train/val/test (41,594/11,899/6,000 superpixels)
- See Workflow below for pre-segmentation and labeling method

Examples of original images and input. A) original image, B) ground truth image, C) fire superpixel (input), D) non-fire superpixel (input).

Model

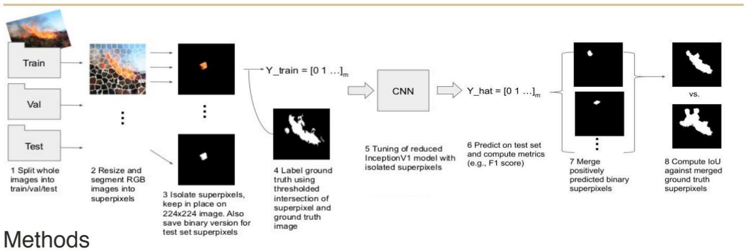


InceptionV1-OnFire [1]: a reduced Inception v1 model with three inception modules and softmax output activation. First two inception modules were frozen for tuning and fully connected layer was replaced.

Loss Function

$$L = -\sum_i y_i log\hat{y}_i + (1-y_i)log(1-\hat{y}_i)$$

Workflow



Methods

Superpixel Segmentation

Simple linear iterative clustering (SLIC) was used to segment images into perceptually meaningful regions that are similar in color and texture. It adapts K-means clustering to reduced spatial dimensions for computational efficiency. The distance measure D combines color proximity (d_c, in CIELAB color space) and spatial proximity (d_s) and normalizes each by its respective maximum within a cluster (N_s and N_d). [2]

$$d_c = \sqrt{(l_j-l_i)^2+(a_j-a_i)^2+(b_j-b_i)^2}$$

$$d_s = \sqrt{(x_j-x_i)^2+(y_j-y_i)^2}$$

$$D' = \sqrt{\left(\frac{d_c}{N_c}\right)^2+\left(\frac{d_s}{N_s}\right)^2}$$

Metrics

$$\text{precision} = \frac{\text{tp}}{\text{tp}+\text{fp}} \qquad \text{recall} = \frac{\text{tp}}{\text{tp}+\text{fn}}$$

$$F_1 = 2\cdot\frac{\text{precision}\cdot\text{recall}}{\text{precision}+\text{recall}}$$

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

Class saliency
[3]

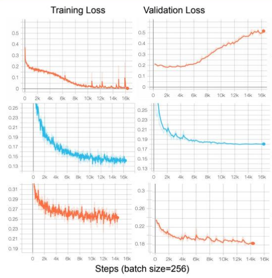
We want to rank the pixels of an image I_0 based on their influence on the class score $S_c(I)$, which can be approximated by

$$S_c(I) \approx w^T I + b \qquad w = \left.\frac{\partial S_c}{\partial I}\right|_{I_0}$$

The map M is generated at each pixel (i,j) by taking the maximum magnitude of w across all color channels

$$M_{ij} = \max_c |w_{h(i,j,c)}|$$

Model training and validation



Adam

Momentum with learning rate decay

Adam with data augmentation

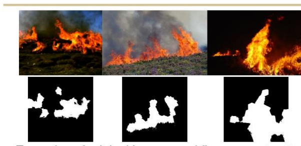
Model	Accuracy	F1	IoU
Original pre-trained	0.90	0.70	0.49
Adam	0.93	0.84	0.71
Adam w/data aug.	0.93	0.83	0.69
Momentum w/α decay	0.93	0.84	0.70

Batch sizes tested: 64 and 256

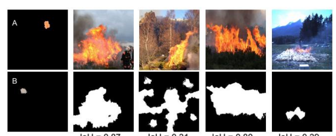
Learning rate:

- Adam, alpha = 0.001 (above) and alpha = 0.01
- Momentum with learning rate decay $\alpha = \alpha \cdot x^{(\text{total steps}/100)}$
  - alpha = 0.01
  - x=0.97 (above) and x=0.993

Results

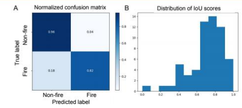


Examples of original images and fires reconstructed from classified superpixels (following image resizing).

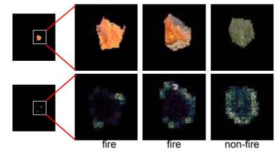


Error analysis. A) correctly classified, B) incorrectly classified. Resized original images and reconstructed classified superpixels.

Results

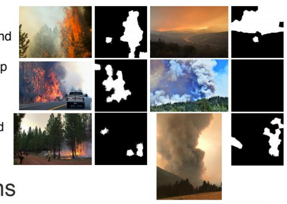


A) Representative normalized confusion matrix for isolated superpixel classification. B) Histogram showing distribution of IoU scores for reconstructed wildfires.



Input images (top) and saliency maps showing magnitude of the saliency (bottom).

Examples from recent California, Colorado, and Oregon fires including the Woosley Fire, Camp Fire, and Carr Fire. Original images (left); resized and reconstructed predicted superpixels (right).



Conclusions

- This network outperformed a pre-trained fire network on this dataset, achieving an F1 score of 0.84 and mean IoU score of 0.71 compared to 0.70 and 0.49, respectively, suggesting that wildfires have unique features compared to urban fires.
- The network successfully identifies fire while excluding smoke, but misclassifies orange 'fire-glow'. It is affected by lack of context.
- Training initially resulted in overfitting to the training set, which was mitigated by data augmentation or early stopping.
- Saliency maps suggest that the most salient regions for predicting the output class are pixels that are close to but not exactly the color of fire. Bright orange pixels have smaller gradients.
- Suggested future directions include increasing other regularization methods such as dropout, testing a larger hyperparameter space, unfreezing additional inception modules during training, and training on remote sensing images.

References/Acknowledgements

Many thanks to the CS230 TAs for their assistance and support!

[1] Dunnings and Breckon, 2018

[2] Achanta et al., 2012.

[3] Szegedy et al., 2015