# Fourier Domain Pruning of MobileNet-V2 with Application to Video Based Wildfire Detection

Hongyi Pan, Diaa Badawi, Ahmet Enis Cetin
hpan21@uic.edu

ELECTRICAL AND COMPUTER ENGINEERING
COLLEGE OF ENGINEERING
UIC

## MobileNet-V2 for Wildfire Detection

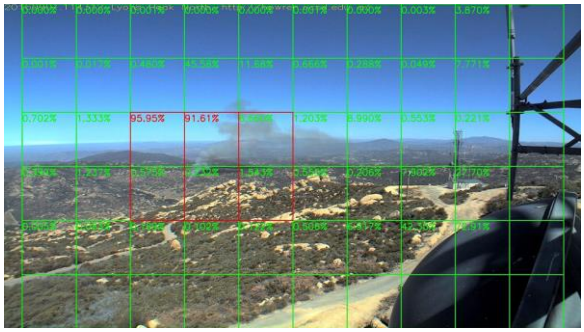
- MobileNet-V2: A mobile CNN model published in 2018.
- We totally have about 10K training images:
- 3K wildfire images.
- All images are resized into 224x224.
- We will use NVIDIA Jetson Nano in our wildfire

(a) (b) (c) (d) (e)

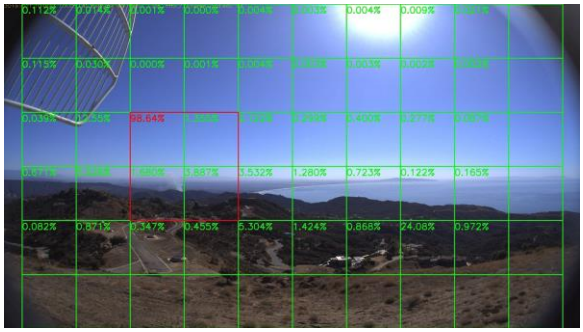
Training Dataset Samples

## Block-Based Image Frame Analysis

- Problem: In real time application, input frames are in 1080P or higher resolution, but the input of the network is 224x224. We may miss very small smoke regions, if we just down-sample the frames.
- We divide a frame into many small tiles.
- We have overlapping tiles.



## Neural Network Performance: HPWREN



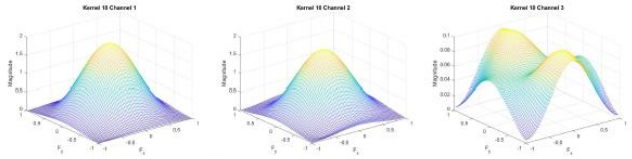
Detection Result

| Videos Name | Fire Starts | First Detected | | |
|---|---|---|---|---|
| | | Fourier Domain Pruning | Time Domain Pruning | No Pruning |
| Lyons Fire | 156 | 164 | 168 | 164 |
| Holy Fire East View | 721 | 732 | 738 | 732 |
| Holy Fire South View | 715 | 725 | 725 | 724 |
| Skylinefire | 684 | 690 | 690 | 690 |
| Palisades Fire | 636 | 639 | 640 | 639 |
| Palomar Mountain Fire | 262 | 277 | 279 | 275 |
| Banner Fire | 15 | 17 | 20 | 17 |
| Highway Fire | 4 | 6 | 6 | 6 |
| DeLuz Fire | 37 | 48 | 51 | 48 |

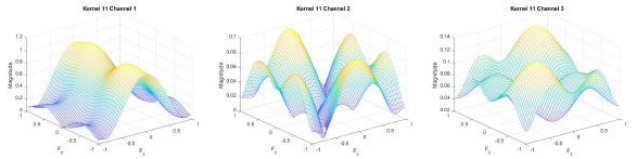
▪HPWREN dataset: https://hpwren.ucsd.edu/

## Fourier Transform Based Neural Network Pruning

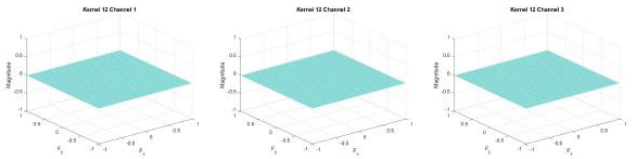
- We extract 3x3 convolution kernel weights and calculate their Fourier Transforms (FT).
- We remove similar filters according to their FT magnitudes.
- We notice that some kernels have nearly 0-response, so we remove them from the graph.
- We also eliminate those kernels with low energy at the 1x1 convolutional layers and the dense layer.
- We perform 64x64 FT and check their cosine similarity:
- Filter similarity $= \cos(\theta) = \frac{<\mathbf{X},\mathbf{Y}>}{||\mathbf{X}||\cdot||\mathbf{Y}||}$ where $\mathbf{X}$ and $\mathbf{Y}$ are the Fourier transform magnitudes of the two filters in vector form, respectively.
- We treat the kernels with similarity larger than 0.99925 as a pair of similar kernels and store only one of them. This value is chosen based on no-fire video test experiment.
- Reason to use FT: A kernel with its shifted kernel may have a small cosine similarity in time-domain but their FT magnitudes will produce a $\cos(\theta) = 1$ or $-1$.
- In this way, 22.91% kernels are removed, and 7.04% time saving is achieved by pruning these layers.



(a) Kernel 10 Channel 1 (b) Kernel 10 Channel 2 (c) Kernel 10 Channel 3



(d) Kernel 11 Channel 1 (e) Kernel 11 Channel 2 (f) Kernel 11 Channel 3



(g) Kernel 12 Channel 1 (h) Kernel 12 Channel 2 (i) Kernel 12 Channel 3

No-Fire Video Test

| Threshold | | None[a] | | 0.99925[a] | | 0.99900 | | 0.99500 | |
|---|---|---|---|---|---|---|---|---|---|
| Slimming Rate | | 0.00% | | 22.91% | | 24.90% | | 47.27% | |
| Videos Name | Frames | Num | Rate (%) | Num | Rate (%) | Num | Rate (%) | Num | Rate (%) |
| wilson-w-mobo-c[b] | 10080 | 2 | 0.0198 | 2 | 0.0198 | 5 | 0.0496 | 63 | 0.6250 |
| wilson-s-mobo-c | 10074 | 2 | 0.0199 | 2 | 0.0199 | 2 | 0.0199 | 69 | 0.6849 |
| wilson-n-mobo-c[b] | 10024 | 3 | 0.0299 | 3 | 0.0299 | 4 | 0.0399 | 71 | 0.7083 |
| wilson-e-mobo-c[c] | 10028 | 43 | 0.4288 | 43 | 0.4288 | 43 | 0.4288 | 104 | 1.0371 |
| vo-w-mobo-c | 10009 | 5 | 0.0500 | 5 | 0.0500 | 5 | 0.0500 | 64 | 0.6394 |
| 69bravo-e-mobo-c | 1432 | 1 | 0.0698 | 1 | 0.0698 | 1 | 0.0698 | 11 | 0.7682 |
| 69bravo-e-mobo-c | 1432 | 0 | 0.0000 | 0 | 0.0000 | 0 | 0.0000 | 9 | 0.6285 |
| syp-e-mobo-c | 1421 | 3 | 0.2111 | 3 | 0.2111 | 3 | 0.2111 | 13 | 0.9148 |
| sp-n-mobo-c | 1252 | 2 | 0.1597 | 2 | 0.1597 | 2 | 0.1597 | 12 | 0.9585 |
| sp-w-mobo-c[b] | 1282 | 1 | 0.0780 | 1 | 0.0780 | 2 | 0.1560 | 8 | 0.6240 |
| sp-s-mobo-c | 1272 | 2 | 0.1572 | 2 | 0.1572 | 2 | 0.1572 | 8 | 0.6289 |
| sp-e-mobo-c | 1278 | 2 | 0.1565 | 2 | 0.1565 | 2 | 0.1565 | 10 | 0.7825 |

[a]We get same false-alarm result before and after pruning and slimming in threshold of 0.99925.
[b]With lower the slimming threshold (0.99900), the false-alarm rate increases on these videos.
[c]There is an unexpected long light