A LIGHTWEIGHT, NON-COMPLEX CNN FOR FIRE DETECTION IN SATELLITE IMAGES

Tuck Chong Lee, Chenguang Hou, Aik Song Chia, Celvin Tan, Leong Keong Kwoh, Soo Chin Liew

Centre for Remote Imaging, Sensing and Processing, National University of Singapore
10 Lower Kent Ridge Road, Blk S17 Level 2, Singapore 119076, Singapore
Email: {crsltc, crshc, crscas, crsct, crsklk, crslsc}@nus.edu.sg

KEY WORDS: Fire detection; Convolutional Neural Networks; DenseNet; Machine learning; High resolution.

ABSTRACT: In this paper, we describe our development of FireNet, a lightweight, non-complex convolutional neural network, specifically for the task of detecting fires in high resolution satellite images. This CNN is aimed at assisting our regional fire monitoring operation that has been conducted by visual scanning of high resolution images for signs of smoke plumes. The FireNet is adapted from a variety of DenseNet and trained using a dataset of high-resolution satellite images accumulated through the many years of fire monitoring operation. Overall, our FireNet has comparable performance to that of DenseNet despite having a reduced model capacity and simple construct.

1. INTRODUCTION

Land and forest fires with the associated transboundary haze pollution is a main environmental concern in the Southeast Asia region. Regional fire monitoring is normally done with hotspots derived from thermal sensors such as the Moderate Resolution Imaging Spectrometer (MODIS) on-board the Terra and Aqua satellites (Giglio et al. 2016). Such sensors typically have low resolution ranging from a few hundred meters to a kilometer or so. The Centre for Remote Imaging, Sensing and Processing (CRISP) at the National University of Singapore has been monitoring the regional fires using high resolution multispectral satellite imagery since 1998, shortly after the major fire episode in 1997/1998 (Miettinen et al. 2013, Liew et al. 2003).

Detection of fires in high resolution satellite imagery usually involves domain experts who pre-process and visually scan the displayed images for sign of smoke plumes. The procedure is laborious and time consuming. Machine learning is a good approach to automate this process of searching for smoke plumes in the satellite images. In this paper, we propose a lightweight and non-complex Convolutional Neural Network (CNN) that is developed by adapting a widely accepted CNN architecture, DenseNet (Huang et al. 2017), through ways of revising and scaling based on a set of non-rigorous design heuristics. Overall, our CNN has comparable performance to that of DenseNet despite having a reduced model capacity and simple construct.

Convolutional neural networks (CNN) are a specific type of machine learning methods commonly used for analyzing visual imagery (Albawi et al. 2017). Many large pre-trained neural networks are available. However, they are mainly trained with ground-level image datasets. The image features are inherently incompatible with those in aerial or satellite images. Thus, we build and train the base models from scratch, using satellite images and ground truth data derived from CRISP high resolution satellite image dataset accumulated during the fire monitoring operation that spans over a period of more than 20 years. As the labelled dataset for satellite images is limited in quantity compared to the ground-level images, we tweaked the architecture of a suitable CNN to overcome this constraint. A densely connected convolutional network (DenseNet) has an architecture that connects each layer to

every other layer in a feed-forward manner (Huang et al. 2017). This design gives DenseNet many advantageous properties such as the implicit deep supervision (Wang et al. 2016) and reduced number of parameters, among others, that facilitate the construction of a CNN that circumvents the aforementioned constraints. Based on the DenseNet, we constructed a lightweight and non-complex CNN (nicknamed "FireNet") specifically for the detection of fires in high resolution satellite images. The FireNet is an attempt in effective utilization of currently available shared resources to meet our operational requirements.

2. METHODOLOGY

We followed a set of simple rules to construct our model. First, a non-rigorous approach to compound scaling (Tan et al. 2019) is performed by manually adjusting the network architecture to achieve an optimal balance between the three aspects of the model, i.e. input resolution, depth, and width of the CNN, for efficient execution. With a limited labelled dataset, we next utilized implicit deep supervision to perform feature extraction and representation learning. We trained our model from scratch using our training dataset in order to effectively match the CNN to our intended application with considerations to the practicality and extensibility/flexibility of machine learning (ML) to meet our operational needs.

Our dataset is derived from high resolution satellite images used in CRISP fire monitoring operation since 1998. We generated the ground truth data from the dataset by annotating the locations of all detected fires with visible smoke plumes in the satellite images. The images were then partitioned into almost 100 thousand image chips and categorized into the fire and background (no fire) groups. About 24,000 of these image chips contain fire locations with smoke plumes. About 19,000 fire image chips together with the same number of background chips were used in the training and validation of the FireNet. A quarter of these samples were used for training and the rest for validation. A separate testing of the FireNet used about 54,000 image chips that included about 4,000 positive (fire) and 50,000 negative (background) samples. Figure 1 shows examples of the dataset used in the training, validation, and testing of our FireNet.

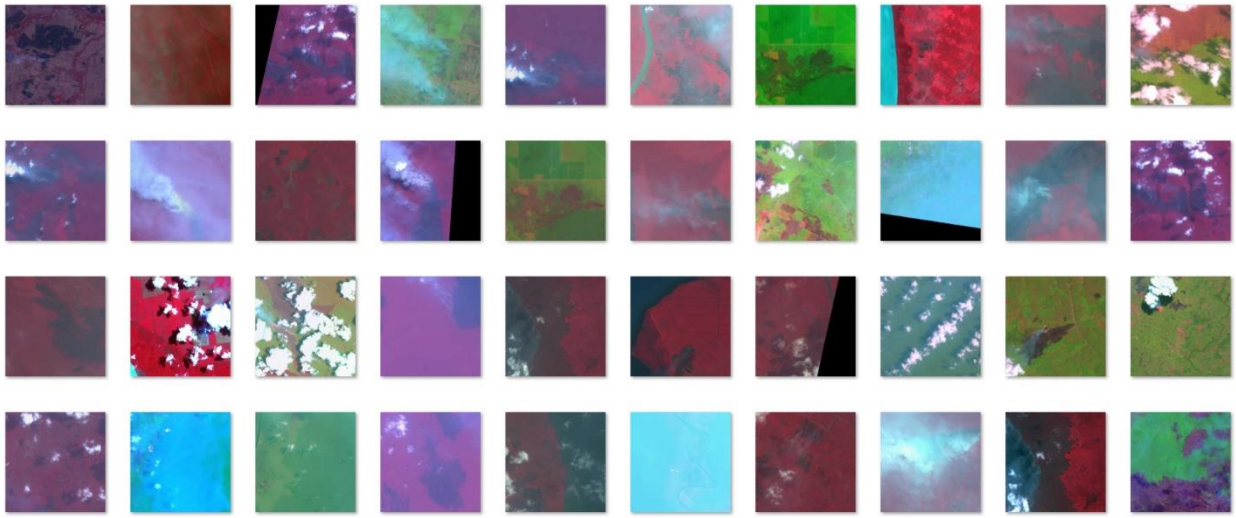

Figure 1. Examples of CRISP high-resolution satellite image dataset

Our FireNet is architecturally similar to DenseNet-121 (Huang et al. 2017), a variant of the DenseNet. Each dense block of the DenseNet is changed to a cascade of 3 × 3 (growth-rate=32)-grouped convolution layers and a shrunk block configuration of (2, 4, 8, 6), as illustrated in Table 1. The FireNet has about 1.05 million parameters and requires a memory size of 8603 kB which are more than a factor of 6 reduction compared to DenseNet-121 (Huang et al. 2017). The FireNet uses the same model

hyperparameters as the DenseNet-121 (Huang et al. 2017) without any fine-tunings and randomization (by setting the same random seed). The FireNet is expected to perform better operationally with a more optimal adjustment of the parameters. The locations fire sources in the images can also be visualized by overlaying the score-weighted class activation map (CAM) to highlight the possible positions of the fire sources (Figure 2).

Table 1. FireNet architecture

Convolution	7×7 conv, stride 2
Pooling	3×3 max pool, stride 2
Dense Block (1)	[32-grouped 3×3 conv] × 2
Transition Layer (1)	1×1 conv, 2×2 average pool, stride 2
Dense Block (2)	[32-grouped 3×3 conv] × 4
Transition Layer (2)	1×1 conv, 2×2 average pool, stride 2
Dense Block (3)	[32-grouped 3×3 conv] × 8
Transition Layer (3)	1×1 conv, 2×2 average pool, stride 2
Dense Block (4)	[32-grouped 3×3 conv] × 6
Classification Layer	7×7 global average pooling, 2D fully connected, softmax

Figure 2. Examples of visualizing fire locations by overlaying the CAM on the image

3. RESULTS

For each input image chip, the FireNet output a decimal number ranging from 0 to 1 representing the confidence level that a fire is present in the image. Fire is considered to be present in the image chip if the confidence level is greater than a predefined threshold value. In principle, an optimal threshold value can be determined according to the probability density functions of the confidence level for both the fire and no-fire samples (Liew et al. 2005). For operational purposes, we fix the threshold value to be 0.5. We used 9,657 image chips consisting of approximately equal number of fire and no-fire samples for validation of both the DenseNet-121 and FireNet. The FireNet is comparable to the DenseNet-121 in performance. As shown in Table 2, FireNet correctly classifies 93.8% of fire samples and 94.6% of no-fire samples giving an overall accuracy of 94.2% compared to 94.9% for DenseNet-121.

Two tests were carried out. The first test used 9,424 unseen image chips with an equal number of fire and no-fire samples for testing both the FireNet and DenseNet-121. Both CNNs are comparable in accuracy, with an overall accuracy of 89.6% for FireNet and 89.7% for DenseNet-121 (Table 3). In the second test, 45,809 all no-fire image chips were used for testing the false positive response of both CNNs. Our FireNet falsely classified 2,670 image chips as fires giving a false positive rate of 5.83% versus 5.54% for DenseNet-121.

Table 2. Validation results for fire detection by FireNet and DenseNet-121 with equal number of fire and no-fire samples

FireNet		
Outputs / Targets	No fire	Fire
No fire	**0.946** (TN)	0.054 (FP)
Fire	0.062 (FN)	**0.938** (TP)
Overall accuracy: 0.942		

DenseNet-121		
Outputs / Targets	No fire	Fire
No fire	**0.954**	0.046
Fire	0.060	**0.940**
Overall accuracy: 0.949		

Notes: For each CNN, "Targets" are the ground truths while "Outputs" are the results of detection. The numbers in the four cells (clockwise, from top left cell) indicate respectively, the true negative (TN), false positive (FP), true positive (TP) and false negative (FN) rates.

Table 3. Test results for fire detection by FireNet and DenseNet-121 with equal number of fire and no-fire samples

FireNet		
Outputs / Targets	No fire	Fire
No fire	**0.936**	0.064
Fire	0.145	**0.855**
Overall accuracy: 0.896		

DenseNet-121		
Outputs / Targets	No fire	Fire
No fire	**0.936**	0.064
Fire	0.143	**0.857**
Overall accuracy: 0.897		

Table 4. Test results for false fire detection by FireNet and DenseNet-121 with all no-fire samples

FireNet		
Outputs / Targets	No fire	Fire
No fire	0.9417	0.0583
Fire	0.0000	0.0000
False positive rate: 0.0583		

DenseNet-121		
Outputs / Targets	No fire	Fire
No fire	0.9446	0.0554
Fire	0.0000	0.0000
False positive rate: 0.0554		

The apparently high accuracy of the CNNs may not be indicative of their performance during the actual fire detection operation. In a more realistic scenario of fire detection, the fire and no-fire test samples do not appear in equal proportions. There are a lot more image chips without fire than those with fires. In this case, the more realistic performance metrics are the commission error (CE) and omission error (OE) defined as (Liew et al. 2005),

$$\mathrm{OE} = \mathrm{FN} \tag{1}$$

$$\mathrm{CE} = \frac{(1-f)\mathrm{FP}}{(1-f)\mathrm{FP} + f\mathrm{TP}} \tag{2}$$

where FN, FP, TP are respectively, the false negative, false positive and true positive rates as in Tables 2 and f is the fire fraction, i.e. the proportion of fire image chips presented to the CNN. The two errors are related to the more familiar producer accuracy (PA) and user accuracy (UA),

$$\mathrm{PA} = 1 - \mathrm{OE} \tag{3}$$

$$\mathrm{UA} = 1 - \mathrm{CE} \tag{4}$$

If we use the test results for FireNet shown in Table 3, the OE and CE can be expressed in terms of the fire fraction by the following equations,

$$OE = 0.145 \quad (5)$$

$$CE = \frac{0.064(1-f)}{0.064(1-f)+0.855f} \quad (6)$$

Note that the commission error is dependent on the fire fraction *f*. If the fire fraction is assumed to be 1%, i.e. $f = 0.01$, then the commission error is CE = 0.881. Hence, even though the FireNet has a high producer accuracy of 85.5%, the user accuracy is not so impressive, at 11.9%. To put these numbers in context, suppose that in a typical fire detection operation the CNN detects 100 fires, only 12 are true fires while the other 88 are false alarms. So, instead of potentially scanning about 10,000 image chips by a human operator to visually look for fires, the human operator now only needs to inspect the 100 outputs of the CNN to confirm or reject the presence of fires. The CNN would also missed about 14 fires. The fires that escape detection are likely to be small fires with vague smoke plumes. These fires would probably also be missed by a human observer.

4. CONCLUSIONS

The FireNet described in this paper is a revised and scaled down CNN adapted from a widely accepted DenseNet and constructed for the purpose of detecting fires in high resolution satellite images. It produced an acceptable accuracy comparable to DenseNet-121, no significant miscalibration nor overfitting in our domain of interest. In addition, many images in the dataset are quite diversified in appearance as the images were acquired over many different land cover types with various atmospheric conditions. There is no deliberate data cleaning performed. It has performed reasonably well compared to the DenseNet-121 despite having reduced model capacity and a simple construct. With further improvements (such as better hyperparameter tuning, data pre-processing, different training-validating-testing procedures, etc.), its performance may be improved and it may be adaptable for other detection tasks.

The FireNet has a high true positive detection rate (i.e. the producer accuracy) of about 85% and a low false positive rate of about 6%. However, these seemingly good performance should be put into the context of a realistic fire operation scenario where the number of fire samples is usually much fewer than the background null samples. With a fire fraction of approximate 1%, the user accuracy is estimated to be about 12%. The FireNet would still be useful in this case as it would reduce the need for a human operator to scan through a large volume of imagery. The role of the human would be to verify the detection outputs of the CNN which would require much less human efforts.

5. REFERENCES

1. Albawi, S., Mohammed, T. A., Al-Zawi, S., 2017. Understanding of a convolutional neural network. In IEEE International Conference on Engineering and Technology 2017, pp. 1-6.
2. Giglio, L., Schroeder, W., Justice, C. O., 2016. The collection 6 MODIS active fire detection algorithm and fire products. Remote Sensing of Environment, 178, pp. 31-41.
3. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K. Q., 2017. Densely connected convolutional networks. IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2017, pp. 2261-2269. Doi: 10.1109/CVPR.2017.243.
4. Liew, S. C., Shen, C., Low, J., Lim, A., Kwoh, L. K., 2003. Validation of MODIS fire product over Sumatra and Borneo using high resolution SPOT imagery. Proc. 24th Asian Conference on Remote Sensing, Vol. I, pp. 671-673.
5. Liew, S. C., Lim, A., Kwoh, L. K., 2005. A stochastic model for active fire detection using the thermal bands of MODIS data. IEEE Geoscience and Remote Sensing Letters, 2 (3), pp. 337-341.
6. Miettinen, J., Hyer, E., Chia, A. S., Kwoh, L. K., Liew, S. C., 2013. Detection of vegetation fires

and burnt areas by remote sensing in insular Southeast Asian conditions: current status of knowledge and future challenges. International Journal of Remote Sensing, 34 (12), pp. 4344-4366. Doi: 10.1080/01431161.2013.777489.

7. Tan, M. and Le, Q., 2019. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. Proc. 36th International Conference on Machine Learning, in PMLR, 97, pp. 6105-6114.
8. Wang, L., Lee, C. Y., Tu, Z., Lazebnik, S., 2015. Training deeper convolutional networks with deep supervision. Preprint, arXiv:1505.02496.