ANOMALY DETECTION IN THERMAL IMAGES USING DEEP NEURAL NETWORKS

Cai Lile, Li Yiqun

Institute for Infocomm Research, Singapore

ABSTRACT

Infrared thermography has become an effective tool in electrical preventive maintenance program due to its high precision and the capability of performing non-contact diagnostic. Anomalies in a thermal image is typically detected by comparing the temperatures of the equipment with reference temperatures. Manual detection is time-consuming and unreliable, making it unable to meet the excessive demand for condition monitoring in industrial applications. In this paper, we propose an automatic method to detect thermal anomalies based on deep neural networks (DNNs). The DNN model is trained to learn the statistical regularities of normal thermal images, and anomalies are detected based on pixel-wise comparison between the learned reference temperatures and the actual temperatures. We test our method on a variety of electrical equipment and the experimental results demonstrated the effectiveness of the proposed method.

Index Terms— Deep convolutional neural networks, anomaly detection, infrared thermography, thermal image

1. INTRODUCTION

Infrared thermography (IRT) is a technique that detects the intensity of radiation in the infrared part of the electromagnetic spectrum and visualizes the recorded temperatures as 2-D images. It has the advantages of performing high-precision and non-contact temperature measurement. In recent years, with the availability and affordable price of commercial thermal imaging cameras, IRT has been widely employed as a non-destructive test method to detect temperature-related anomalies in building inspection, civil engineering, high/low voltage installations, mechanical installations and medical diagnosis.

Electrical equipment inspection is an important application of IRT. Anomaly detection in the thermal images of electrical equipment usually consists of two steps. First, the region of interest (ROI) or the target equipment is detected. Second, the temperature difference (also known as Delta-T value) between the ROI and a defined reference (which is typically a similar component under the same conditions or the ambient air temperature) is calculated. The severity of the problem and the recommended action can be referred to a temperature-rating table, which is usually divided into three or four different categories according to the Delta-T values. Such a table can be found in handbooks, guidelines or standards ([1]), or created based on experience [2].

To facilitate automatic anomaly detection in thermal images, previous methods typically employed segmentation or feature extraction techniques to detect and separate the ROIs from the rest of the image [3]. Various segmentation algorithms, such as thresholding, watershed transform and normalized cut, have been applied to segment hot regions in thermal images [4]. The popular SIFT algorithm, has been employed to detect similar electrical components in a thermal image [5, 6]. However, thermal images have a much lower resolution and signal-to-noise ratio than optical images, and the details in a thermal image are usually blurred, of low contrast and lacking texture. These factors make ROI detection and segmentation difficult in thermal images.

Since its record beating performance on the ImageNet 2012 challenge [7], deep convolutional neural network (CNN) has drastically altered and pushed forward the landscape of computer vision. Deep CNNs are able to discover hierarchical, distributed representations that correspond to the underlying causes of the observed data [8]. This is desirable in solving computer vision tasks where the curse of dimensionality has made algorithms based on shallow architectures fail to generalize well. Algorithms based on deep neural networks (DNNs) have obtained the state-of-the-art results on many computer vision problems, such as object detection and recognition [9, 10]. DNNs have also been widely employed for pixel-wise prediction, including semantic segmentation [11], depth map prediction [12], image super-resolution [13] and saliency prediction [14].

In this paper, we propose an automatic method to detect anomalies in thermal images based on deep CNNs. The CNN model is trained to predict a thermal image from its corresponding visible image. Specifically, the model is trained on positive (normal) samples and the prediction it produces is based on the assumption that the equipment is operating normally, otherwise the prediction error will be large. Just as the reconstruction error of an autoencoder can be utilized to indicate anomaly [15, 16], the prediction error of the CNN model can serve the same purpose in our task. In the test stage, the predicted thermal image is used as reference and the Delta-T values are calculated as the absolute difference of the predicted thermal image and the actual thermal image. Any region with Delta-T value above a pre-specified thresh-

old is detected as anomalies.

Our major contribution is a novel anomaly detection framework based on pixel-wise analysis of thermal images using DNNs. The reference temperatures used in our method are based on the statistical regularities learned from normal samples, which are reliable and do not require the thermal image contains another similar component for reference. Another contribution is a deep CNN architecture adapted for the task of thermal anomaly detection. Taking advantage of the representation learning power of the DNNs, our method is flexible and can be easily applied to detect anomalies in new categories as long as sufficient data is available for training.

2. METHOD

2.1. Model Architecture

Our model is inspired by the architectures used for saliency prediction [14, 17], where an input visible image is mapped to a saliency map. However, unlike saliency prediction that is interested in *relative* saliency values, we are interested in *absolute* temperature values. Therefore, we do not rescale the final feature map to a probability distribution to compute network loss as done in saliency prediction. Instead, we use mean squared error (MSE) as the objective function and add a batch normalization layer [18] for the network to converge.

Figure 1 illustrates the architecture of our deep CNN model. The model is built on the VGG-16 [19] that has been trained on the ImageNet dataset for object recognition. The VGG-16 consists of 5 stages of convolutional layers. We only use the first 3 stages (excluding the pooling layer following stage 3 convolution) as we found that using deeper stages did not bring significant improvement. We add a batch normalization layer to normalize the feature values output by VGG-16 to have zero mean and unit variance, which is found to be helpful for the network to converge. To facilitate pixel-wise comparison between the prediction and ground truth, we apply 2 upscale layers to offset the 2 pooling layers in the VGG-16 (pooling layer after stage 1 and stage 2 respectively) and thus the output map has the same size as the input image. To improve prediction accuracy, convolutional layers are applied after the upscaling layers, where the filters can learn to refine the coarse prediction. We use rectify activation function for all layers except for the output layer, which used linear activation to allow gradients to back-propagate.

The training process can be described mathematically as follows. Letting $\{\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^n\}$ denote the visible images of the electrical equipment under normal operation, and $\{\mathbf{y}^1, \mathbf{y}^2, \ldots, \mathbf{y}^n\}$ denote the corresponding thermal images. The model takes as input $\mathbf{x}^i$, $i = 1, \ldots, n$, and outputs the predicted thermal image $f(\mathbf{x}^i)$. The prediction error of $\mathbf{x}^i$ is measured by squared error: $\|f(\mathbf{x}^i) - \mathbf{y}^i\|^2$. The model is trained by minimizing the average prediction error:

$$J(\theta) = \frac{1}{n}\sum_{i=1}^{n}\|f(\mathbf{x}^i;\theta) - \mathbf{y}^i\|^2, \quad (1)$$

where θ denote the parameters to be learned(i.e., weights and biases of neurons). The minimization problem can be solved by stochastic gradient descent.

2.2. Anomaly Detection Pipeline

After training, the model has learned to predict the normal thermal profile for an input visible image from its visual features. If the target equipment in the input image is under normal operation, the difference between the predicted thermal image and the actual thermal image will be small and no anomaly will be detected. However, if the equipment is working anomalously, its thermal signature will be very different from the predicted one. Therefore, by using the predicted thermal image as reference and compare it with the actual thermal image, we will be able to identify where anomalies occur. We compute the Delta-T value as the absolute difference between the predicted thermal image and the ground truth thermal image. Regions with Delta-T values above a pre-specified threshold are considered as anomalies and are masked in red on the visible image to alert the user. The anomaly detection pipeline is illustrated by an example in Fig. 2.

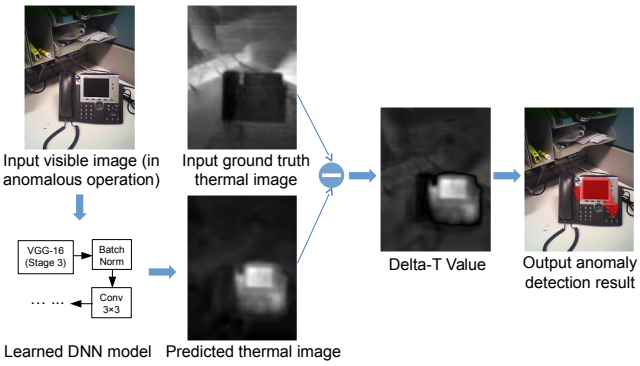

Fig. 2. The proposed anomaly detection pipeline.

3. EXPERIMENTS

3.1. Experimental Setup

Our algorithm was implemented by Python deep learning package Lasagne and run on an Nvidia GTX Titan X GPU. The infrared camera used was a FLIR One that has a thermal resolution of 160*120 and thermal sensitivity of 0.1°C. The FLIR One also has an onboard optical camera of resolution 640*480. The thermal and optical cameras can generate aligned thermal and optical images. The camera is to be used

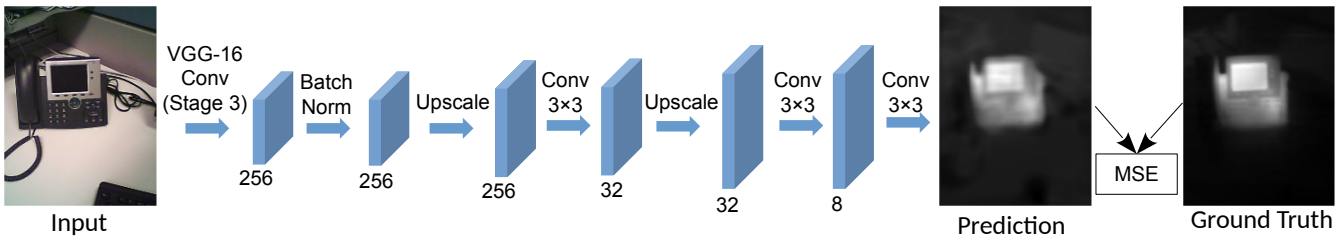

Fig. 1. Training of the deep CNN architecture.

in conjunction with an iPhone and we have developed an iOS app based on the FLIR One SDK to facilitate obtaining the thermal and visible image data. The thermal image output by the app is in unsigned int 16 bit format and the absolute temperature value (i.e., in Kelvin) of the pixel can be obtained by dividing the pixel value over 100. The original image dimension of 452*339 was resized to 224*168 for network input.

We captured images of 4 categories at various positions, including bus bar (meter), bus bar (circuit), IP phone and PC motherboard. The total number of training images is 1425, of which 20% are split for validation. Due to the limited number of training images, we randomly applied horizontal/vertical flips to 50% of images in a batch for data augmentation. The total number of test images is 142. Table 1 lists the number of training and test images for each category.

	Train	Test
PC Motherboard	331	34
Bus bar (circuit)	325	37
IP phone	349	36
Bus bar (meter)	420	35

Table 1. The number of training and test images for each category.

The DNN was trained by stochastic gradient descent with Nesterov momentum. Images are normalized to zero mean and unit variance. The pre-trained weights of VGG-16 were used to initialize the corresponding layers in the model and other layers are initialized randomly. The learning rate was set to be 0.01 and the batch size was 16. The model was trained for 100 epochs with each epoch taking approximately 31 seconds.

3.2. Experimental Results

3.2.1. Qualitative Evaluation

Figure 3 displays the prediction results for some test samples. The first row of Fig. 3 shows the input visible images, the second row are ground truth thermal images and the third row are thermal images predicted by the model. It can be seen that the model can successfully predict salient thermal features for the input visible images, with hot spots correctly predicted at various positions and shapes. It is also observed that the prediction made by the model tends to be a blurred version of the ground truth thermal image. This blurring effect has been observed in previous work [20, 21] and the explanation may be that predictive models trained with MSE would react to uncertainty by blurring. For the second sample in each category, we display the nearest neighbor of the input visible image in the training set and the corresponding ground truth thermal image to show that the model has learned to predict the thermal images for unseen samples instead of simply overfitting the training data.

3.2.2. Quantitative Evaluation

To quantitatively evaluate the prediction quality, we compute the pixel-wise absolute difference between the prediction and ground truth. Besides, for anomaly detection, we are particularly interested in maximum temperature difference, as once the maximum temperature difference between an equipment and a reference exceeds a threshold, the equipment is probably anomalous regardless of the temperature of other parts. Therefore, we also compute the maximum difference between the prediction and ground truth. The results are displayed in Table 2.

	abs-td	max-td
PC Motherboard	0.92 (0.33)	6.41 (1.10)
Bus bar (circuit)	0.57 (0.49)	3.38 (0.58)
IP phone	0.70 (0.21)	3.98 (1.08)
Bus bar (meter)	0.44 (0.32)	2.65 (0.69)

Table 2. Mean pixel-wise absolute temperature difference (abs-td) and maximum temperature difference (max-td) for positive test samples. Number in the bracket is standard deviation across images. The unit is Kelvin.

From Table 2, it can be seen that the mean pixel-wise temperature difference is very small (less than 1 K). This is probably because thermal images are dominated by black pixels and thus the DNN model can achieve small mean tempera-

Fig. 3. Model prediction for positive test samples. Top row: input visible images; middle row: ground truth thermal images; bottom row: thermal images predicted by the model. From left to right, the object category is IP phone, PC motherboard, bus bar (meter) and bus bar (circuit). Images beside the second sample in each category are the nearest neighbor of the input visible image in the training set and the corresponding ground truth thermal image.

ture difference by predicting low temperatures for most pixels. However, the maximum temperature difference is relatively large, ranging from 2.65 to 6.41 K. This may be due to the blurring effect as observed in Fig. 3, which will smooth the maximum points and cause inaccuracy in prediction.

To demonstrate how our method can be applied for anomaly detection, we simulate the anomalous condition by unplugging the power cable for IP phone and PC motherboard. According to the temperature table in [1], Delta-T values corresponding to anomalies are usually larger than 4 K. In our test, the Delta-T threshold for anomalies is set to 6 K. Regions with Delta-T values above the threshold are detected as anomalies and masked in red. Figure 4 presents the anomaly detection results on the two devices. It can be seen that when unplugging the devices, the temperatures of the original heated places (e.g., the microchips beside the fan in motherboard, the screen and control panel of the IP phone) drop significantly. As the temperature maps are normalized to [0,1] when displayed in gray scale, regions do not appear as hot spots under normal operation will pop out as hot spots under anomalous operation. By comparing the prediction and the ground truth, our method correctly detected where anomalies occur.

Fig. 4. Anomaly detection results on PC motherboard (top row) and IP phone (bottom row). From left to right, the images are: input visible images (under anomalous operation), ground truth thermal images, predicted thermal images, visualization of Delta-T values and output anomaly detection results.

4. CONCLUSIONS

In this paper, we proposed an automatic anomaly detection method based on DNNs. The model was trained to predict a thermal image from an input visible image. Experimental results revealed that the prediction made by the model was a blurred version of the ground truth, with major thermal features (e.g., hot spots) predicted correctly at various positions and shapes. The model was trained on positive samples and the prediction error was used as the Delta-T value to detect anomalies in the test stage. Anomaly detection results on electrical devices under simulated anomalous conditions demonstrated the effectiveness of the proposed method.

Several future works can be done to improve the current method. Incorporating adversarial loss [22] in training the model may reduce the blurring effect and produce higher quality prediction, as demonstrated by [21]. Besides, different equipment may need different Delta-T threshold values for anomaly detection. Knowing the category of the incoming image facilitates automatically selecting thresholds for different equipment and makes the system more intelligent. It will be interesting to study how the weights trained for a regression task can be reused for a classification task.

5. REFERENCES

[1] MTS NETA, "Maintenance testing specifications for electrical power distribution equipment and systems," *International Electrical testing Association Inc*, 2001.

[2] Infraspection Institute, "Standard for infrared inspection of electrical systems & rotating equipment," 2008.

[3] Angeliki Kylili, Paris A Fokaides, Petros Christou, and Soteris A Kalogirou, "Infrared thermography (irt) applications for building diagnostics: A review," *Applied Energy*, vol. 134, pp. 531–549, 2014.

[4] Mohd Shawal Jadin and Soib Taib, "Recent progress in diagnosing the reliability of electrical equipment by using infrared thermography," *Infrared Physics & Technology*, vol. 55, no. 4, pp. 236–245, 2012.

[5] David Wretman, *Finding regions of interest in a decision support system for analysis of infrared images*, Citeseer, 2006.

[6] Magnus Smedberg, "Thermographic decision support–detecting and classifying faults in infrared images," *Master of Science Thesis, Royal Institute of Technology, Stockholm, Sweden*, 2006.

[7] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[8] Ian Goodfellow, Yoshua Bengio, and Aaron Courville, "Deep learning," Book in preparation for MIT Press, 2016.

[9] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," *arXiv preprint arXiv:1512.03385*, 2015.

[10] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Advances in neural information processing systems*, 2015, pp. 91–99.

[11] Vijay Badrinarayanan, Ankur Handa, and Roberto Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for robust semantic pixel-wise labelling," *arXiv preprint arXiv:1505.07293*, 2015.

[12] David Eigen, Christian Puhrsch, and Rob Fergus, "Depth map prediction from a single image using a multi-scale deep network," in *Advances in neural information processing systems*, 2014, pp. 2366–2374.

[13] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang, "Learning a deep convolutional network for image super-resolution," in *European Conference on Computer Vision*. Springer, 2014, pp. 184–199.

[14] Saumya Jetley, Naila Murray, and Eleonora Vig, "End-to-end saliency mapping via probability distribution prediction," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 5753–5761.

[15] Graham J Williams, Rohan A Baxter, Hongxing He, Simon Hawkins, and Lifang Gu, "A comparative study of rnn for outlier detection in data mining.," in *ICDM*, 2002, vol. 2, pp. 709–712.

[16] Mayu Sakurada and Takehisa Yairi, "Anomaly detection using autoencoders with nonlinear dimensionality reduction," in *Proceedings of the MLSDA 2014 2nd Workshop on Machine Learning for Sensory Data Analysis*. ACM, 2014, p. 4.

[17] Xun Huang, Chengyao Shen, Xavier Boix, and Qi Zhao, "Salicon: Reducing the semantic gap in saliency prediction by adapting deep neural networks," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 262–270.

[18] Sergey Ioffe and Christian Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167*, 2015.

[19] Karen Simonyan and Andrew Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[20] Nitish Srivastava, Elman Mansimov, and Ruslan Salakhutdinov, "Unsupervised learning of video representations using lstms," *CoRR, abs/1502.04681*, vol. 2, 2015.

[21] William Lotter, Gabriel Kreiman, and David Cox, "Unsupervised learning of visual structure using predictive generative networks," *arXiv preprint arXiv:1511.06380*, 2015.

[22] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio, "Generative adversarial nets," in *Advances in Neural Information Processing Systems*, 2014, pp. 2672–2680.