Cross Entropy Overlap Distance

Michele Fraccaroli, Alice Bizzarri, Paolo Casellati, Evelina Lamma
DE - Department of Engineering, University of Ferrara
michele.fraccaroli@unife.it, alice.bizzarri@unife.it, paolo.casellati@edu.unife.it,
evelina.lamma@unife.it

Abstract

Nowadays, deep learning is a key technology for many applications such as anomaly detection. The role of Machine Learning (ML) in this field relies on the ability of a machine to inspect images to determine the presence or not of anomalies.

Frequently, in Industry 4.0 w.r.t. the anomaly detection task, the images that compose a dataset are not optimal, contain edges or areas, not of interest. Thus, this study aims to identify a systematic way to train a neural network able to focus only on the area of interest. The study is based on the definition of a loss to be applied in the training phase of the network that, through the use of masks, gives higher weight to the anomalies identified within the area of interest.

The idea is to add an *Overlap Coefficient* to the standard cross-entropy. In this way, the more the identified anomaly is outside the area of interest greater is the loss. We call the resulting loss *Cross Entropy Overlap Distance* (CEOD).

The advantage of adding the masks in the training phase of the network is that it is forced to learn and recognize defects only in the area circumscribed by the mask itself. The added benefit is that during inference, these masks will no longer be needed. Therefore, there is no difference, in terms of execution times, between a normal Convolutional Neural Network (CNN) and a network trained with this loss.

In some applications, either the masks are determined at run-time through a trained segmentation network, as we have done for instance in the "Machine learning for visual inspection and quality control" project, funded by the MISE Competence Center Bi-REX.

1 Introduction

The detection of anomalies in industrial image data is of the highest importance for many tasks in the field of computer vision. In the task of surface analysis, very often, the images acquired in an industrial environment contain some sections that are not part of the surface to be inspected. Just think of images of products running on conveyor rollers or connected to other components not subject to inspection or simply when images are taken of the edge of the product which inevitably incorporates part of the background. In many cases, if we know the shape of a product to be inspected, we can simply use some traditional image processing techniques to remove the useless parts from the images. But, in other cases, we cannot know the exact shape of our product or where the background appears in the image.

Then, this work proposes and tests a new approach to identify a systematic way to train a CNN able to focus only on the area of interest. To do that, we identified the polygon that circumscribes the most important pixels for classification according to CNN. After that, let's calculate how much this polygon overlaps with the mask that is provided, for each image, during the training phase. The calculated overlap value is added to the loss of the network, to force the network to recognize the most important pixels, only within the area marked by the masks.

2 Cross Entropy Overlap Distance

The idea is to get an *overlap value* between objects to minimize during the training of the CNN. As an overlap value, we use the *Overlap Coefficient* (aka Szymkiewicz - Simpson coefficient) defined by Equation 1:

$$overlap_c(A_i, A_j) = \frac{|A_i \cap A_j|}{min(|A_i, A_j|)} \tag{1}$$

Where A_i and A_j are the area of two objects. In our case, the two objects are a mask that delimited the surface to be inspected and a polygon extracted from the hottest pixels in the heat map produced by the network.

Then, at the end of each forward pass of the network's training, the heat map is calculated by the *GradCam* algorithm [Selvaraju *et al.*, 2017]. After this, we can extract the pixels that were found to be the most important for classification (*hottest pixels*, see Figure 1). Then, we try to calculate how much these *hottest pixels* are contained inside the mask.

We start with the initial situation displayed in Figure 2.

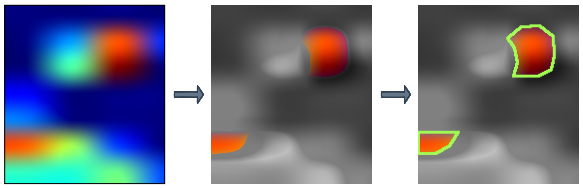

Figure 1: Creation of the polygons from hottest pixels.

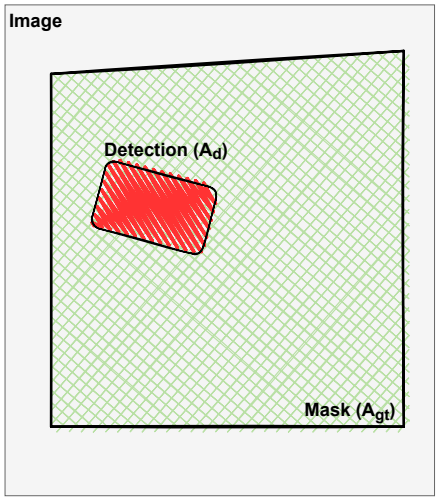

Figure 2: Stylized sample image with A_d and A_{gt} as area of the network detection and mask respectively.

Then, the *Overlap Coefficient* equation became:

$$overlap_c(A_d, A_{gt}) = \frac{|A_d \cap A_{gt}|}{min(|A_d, A_{gt}|)} \tag{2}$$

Where A_d is the area of the detection of the network obtained through *GradCam* and A_{gt} is the segmentation mask (or area of the ground truth) obtained with a manual segmentation or using a *segmentation neural network* [He *et al.*, 2018] [Chen *et al.*, 2017] [Ronneberger *et al.*, 2015]. In this case, if A_d is a subset of A_{gt} or the converse, *Overlap coefficient* is 1. If we want to add this term to the loss function of the neural network, we need to obtain the complementary of Overlap coefficient, obtaining a new value called *Overlap Distance* expressed in Equation 3:

$$overlap_d(A_d, A_{gt}) = 1 - \frac{|A_d \cap A_{gt}|}{min(|A_d, A_{gt}|)} \tag{3}$$

In this way, when A_d is subset of A_{gt}, *Overlap Distance* is 0, giving no contribution to the loss.

Now we need to add this new term to the *Cross-Entropy* loss [Mannor *et al.*, 2005]. Here, for simplicity, we use the binary cross-entropy loss defined by the equation:

$$bce = -\frac{1}{N}\sum_{i=1}^{N} y_i \log(p(y_i)) + (1 - y_i) \log(1 - p(y_i)) \tag{4}$$

Where N is the number of examples, y_i and $p(y_i)$ are the label and the result of the network for the i-*th* example respectively. Obtaining the *Cross Entropy Overlap Distance* (CEOD):

Algorithm 1 CEOD

Input: $x, y, mask$

$predict \leftarrow net(x, y)$
$heatmap \leftarrow GradCam(predict)$
$overlap\_d \leftarrow overlap\_d(mask, heatmap)$
$loss \leftarrow binary\_crossentropy + overlap\_d$
$\rightarrow$ *back propagation*

$$CEOD = bce + overlap_d(A_d, A_{gt}) \tag{5}$$

$$CEOD = -\frac{1}{N}\sum_{i=1}^{N} y_i \log(p(y_i)) + (1 - y_i) \log(1 - p(y_i)) + \omega(1 - \frac{|A_d^i \cap A_{gt}^i|}{min(|A_d^i, A_{gt}^i|)}) \tag{6}$$

The term ω in Equation 6 is a new hyper-parameter to be set which represent the degree of incision of the new term in the overall loss.

Algorithm 1 shows the process behind CEOD loss.

3 Challenges & Prospectives

One of the biggest challenges of this application is trying to force a typical CNN to focus its attention only on the important part of an image. The problem arises from two facts: the first is that the important part of an image is not always of the correct shape to be cropped and given in input to a neural network. The second is because, in the task of anomaly detection, we do not always know a priori the object (defect) that we want to identify (localize or segment). Teaching a CNN where to look, would allow us to avoid using large (and slow) neural networks for semantic segmentation or object detection, but at the same time, exploit a result that could be obtained with those networks.

This work is still under development and the final system will be compared with the state-of-the-art CNNs in the task of surface inspection on both industrial and benchmark datasets.

References

[Chen *et al.*, 2017] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, e Alan L Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, 40(4):834–848, 2017.

[He *et al.*, 2018] Kaiming He, Georgia Gkioxari, Piotr Dollár, e Ross Girshick. Mask r-cnn, 2018.

[Mannor *et al.*, 2005] Shie Mannor, Dori Peleg, e Reuven Rubinstein. The cross entropy method for classification. In *Proceedings of the 22nd international conference on Machine learning*, pages 561–568, 2005.

[Ronneberger *et al.*, 2015] Olaf Ronneberger, Philipp Fischer, e Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[Selvaraju *et al.*, 2017] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, e Dhruv Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision*, pages 618–626, 2017.