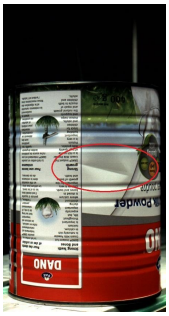
ICT for Brain, Body & Behavior (Eagle Vision): Deep learning for visual verification

Participants: Guillaume Dupont, Jan van Gemert (academic leader), Simon Haafs (industry representative), Miriam Huijser, Osman Semih Kayhan, Yancong Lin, Steve Nowee, Remco Veltkamp, Napolean Yeshwanth

Context and problem

Industrial problem. The workshop case involved automatically detecting small damages in the form of dents and scratches on images of baby food cans, while not rejecting cans where the accidental scene recording causes highlights, shading, shadows, etc. The company involved (Eagle Vision) currently has a system that is based on manually defined image processing operations. Their question is how much a machine learning approach could alleviate the manual operations. With machine learning the 'programming' is done by example. The system is given a training set of examples of the input and the expected output. The task is to automatically predict the output on unseen images. Using machine learning would circumvent the difficult trial-and-error of manual image processing operations by automatically learning them.

Scratch



Dent

Nothing wrong

Academic perspective. The state of the art of image recognition has significantly increased recently with the rise of deep learning. Current deep learning systems[1] give excellent results on classifying thousands of classes of objects, such as man-made objects such as cars, airplanes, chairs but also including fine-grained classes such as 200 different sub-species of dogs. Similar good results are obtained for detecting where an object is (object localization[2]), and for assigning a label to each pixel (semantic segmentation[3]). The academic viewpoint in this case is that it is not interesting to detect the object itself --we already know the object that will be present-- yet we want to visually verify the quality of the object. The difficulty is that there are a huge amount of appearance differences between all possible scratches/dents, which in turn can occur at various locations in the image. There

1 Huang G, Liu Z, Weinberger KQ, van der Maaten L. Densely connected convolutional networks. In computer vision and pattern recognition 2017 Jul 1 (Vol. 1, No. 2, p. 3).

2 Lin TY, Goyal P, Girshick R, He K, Dollar P. Focal Loss for Dense Object Detection. In International Conference on Computer Vision 2017 (pp. 2980-2988).

3 Shelhamer E, Long J, Darrell T. Fully Convolutional Networks for Semantic Segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2017 Apr 1;39(4):640-51.

is, however, not much training data available for all possible damages, since damages do not occur frequently. These settings make the case interesting from an academic perspective.

Approach

We split our team of 6 people into three groups of two persons each. The three groups worked on the following topics:

A. Supervised learning. The standard machine learning approach is learning from labeled examples. Team A explores the effect of the limited amount of example damages
B. Synthetic generation of negative examples. Deep learning thrives on huge amounts of data. When such data is not available, current methods [REF] aim to add synthetic data by generating new images. Team B explores this direction for generating scratches/dents to be used as negative examples in supervised learning.
C. Anomaly detection. Given the huge appearance and location variation of all possible dents/scratch damages, a method that does not need damaged examples and can only uses the undamaged can images is valuable. Team C explored such an approach.

Results

Results team A: Supervised learning. The dataset contained 144,000 good images, and 2,360 images of bad cans. The source came from 6 cameras, so they first annotated the exact image of the damaged can (1 out of 6). A standard deep net (VGG16) is trained on a balanced subset of the dataset. The network got 65.54% accuracy, with a False Acceptance of 10.31% and False Rejection of 24.15%. These inital results are not good enough in an industrial setting. More images of damaged cans would improve results.

Results team B: Synthetic generation of damages. The team investigated generating synthetic scratches and synthetic dents. Results are shown below.

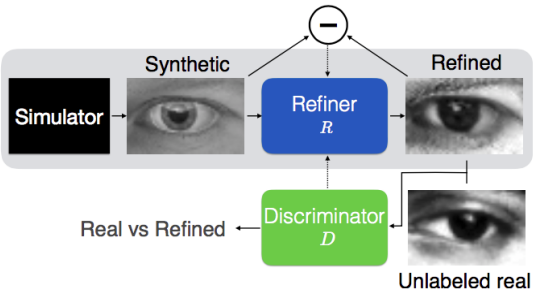
Synthetic scratches

Synthetic dents

Note that these synthetic images deviate from realistic scratches and dents. Recent work[4] has offered to learn a "refiner" network, that aims to make synthetic images similar to realistic images by fooling a "discriminator" such that synthetic and real images cannot be discriminated. See figure below (left). Results for cans (figure below, right) proved inconclusive. Especially the time to train

4 Shrivastava A, Pfister T, Tuzel O, Susskind J, Wang W, Webb R. Learning from simulated and unsupervised images through adversarial training. In Computer Vision and Pattern Recognition (CVPR) 2017 Jul 1 (Vol. 3, No. 4, p. 6).

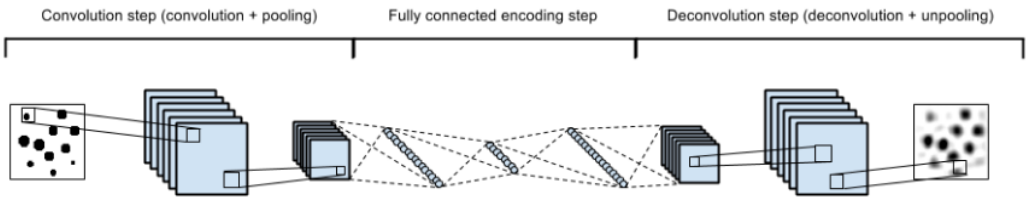
these networks is in the order of several days, which proved infeasible in one week. Also, the small number of negative images made training this difficult.



Shrivastava et al. 2016

Refined can images

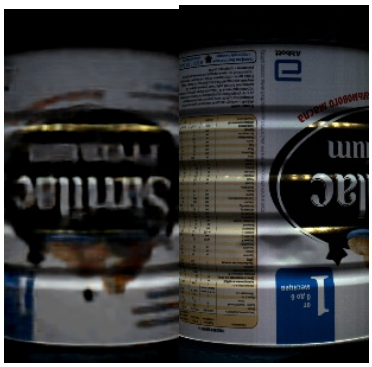
Results team C: Anomaly detection. The team started with an autoencoder, which can be seen as a non-linear variant of PCA for dimension reduction. It takes an input images, forces it through a bottleneck to reduce the dimensionality, and then reconstructs the image again.



An autoencoder aims to encode and reconstruct the input image

The anomaly detection is trained on positive images only. The idea is that it cannot well reconstruct images of damaged cans. Then we can measure the reconstruction error, and the hypothesis is that the reconstruction error of damaged images is higher than images it was trained on. Below are some reconstructions by the autoencoder.

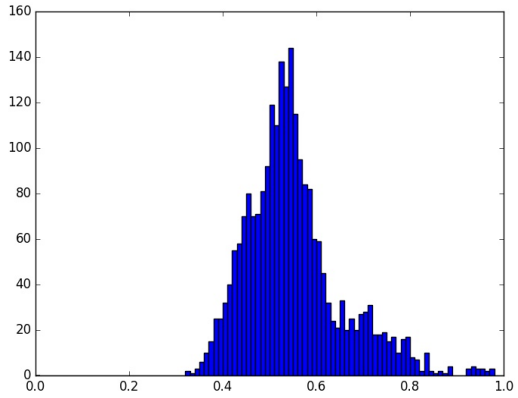
Undamaged reconstruction



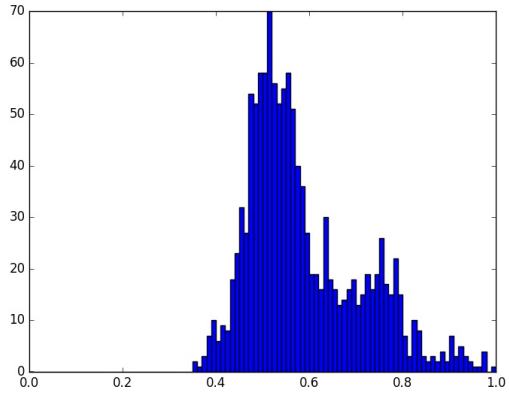
Dent reconstruction

Scratch reconstruction

The reconstruction errors, however, do not differ that much between the “good” cans and the “bad” cans. A histogram of the reconstruction MSE (Mean Square Error) is shown below.



Good cans MSE



Bad cans MSE

Future work

We have several future research directions to investigate. This may be done by a student in the future, possibly in combination with the company. We will definitely investigate closer collaboration between the university and Eagle Vision (an appointment has already been made).