RISK AT WORK

USING DEEP LEARNING FOR IMAGE RECOGNITION IN INSURANCE

By Richard Plat[1]

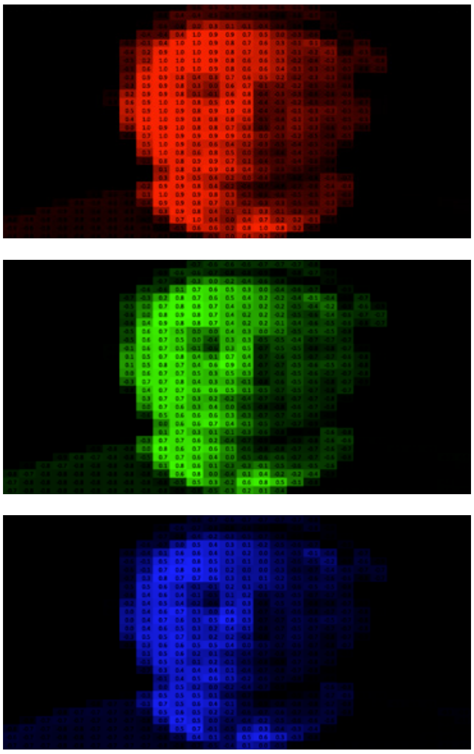
In the near future, image recognition could be integrated into the processes of insurers, for example for estimating car damage based on photos. This requires specific Machine Learning techniques, also often referred to as Deep Learning. This is discussed in this article by means of a concrete example.

1. Background

Machine Learning models are on the rise within the professional field of insurers. Most insurance companies investigate the possibilities for applying these models for, amongst others, pricing, fraud detection and reserving. However, for these purposes more traditional actuarial models (such as linear or logistic regression) can also be used.
In the near future, image recognition may also have a place in the processes of insurers, for example for estimating car damage on the basis of photos. Traditional actuarial models are not sufficient for this, and specific Machine Learning techniques are needed.

2. Structure of the data

To understand how Machine Learning models can recognize images, it is important to know the structure of the data (read: images). An image is made up of a lot of pixels. Each pixel has a specific color. This color is made up of a certain amount of red, green and blue. The amount of each of these colors in a pixel is represented by a number between 0 and 255. This is illustrated in Figure 1. The left side of the figure shows how a picture of a face is made up, on the right side it is zoomed in on the left eye.



226	178	172
229	167	133
229	143	112

190	132	127
197	134	81
192	111	54

175	117	111
182	120	64
177	105	47

Figure 1: structure image face (left), left eye zoomed (right) [2]

3. Example

Below a simple example is worked out, where the purpose of the model is to recognize whether a car has damage or not. If an insurer collects more specific information, such as the amount of the damage, the model can be further refined.

> **"Traditional actuarial models are not sufficient for this, and specific Machine Learning techniques are needed."**

3.1 Data

The available dataset[3] for 'training' (calibration) of the model contains 1840 photos of cars, half of which have damage. A total of 460 photos are available for validation of the model, evenly distributed between cars with and without damage. Examples of photos in the dataset are given in Figure 2.

Figure 2: examples from dataset without damage (left) and with damage (right)

3.2 Machine learning models applied

A multitude of Machine Learning models are available, such as different regression models, decision tree based methods and neural networks[4]. For image recognition the so-called 'convolutional neural network' (CNN)[5] is the state of the art. A CNN is a type of neural network with specific characteristics that make it suitable for image recognition. Techniques like CNNs are also often referred to as Deep Learning, because deep neural networks are required to appropriately recognize images.

An important element in a CNN is the "convolutional filter", the operation of which is shown in Figure 3.

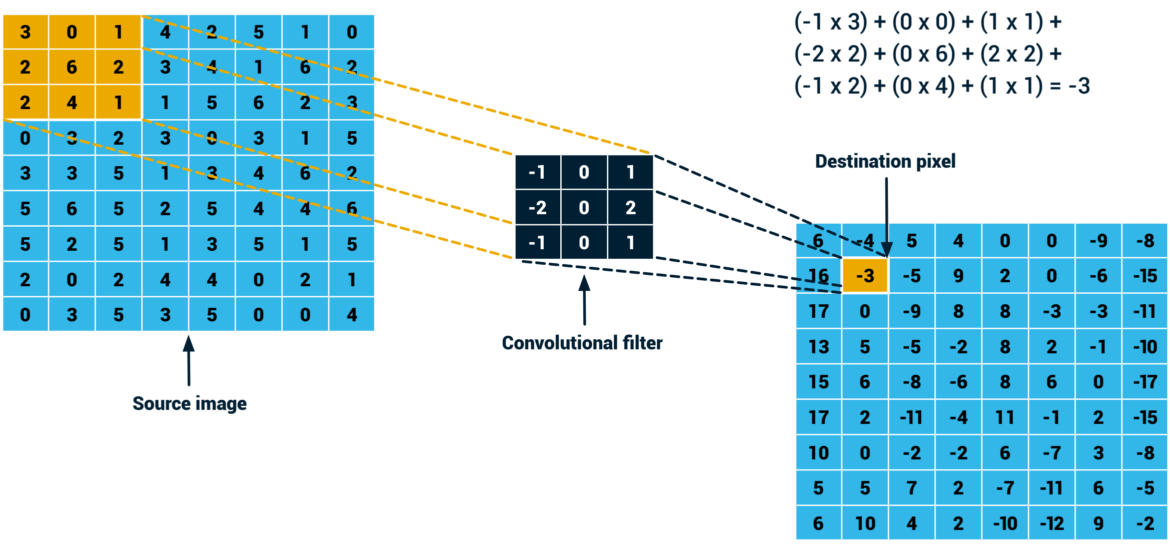


Figure 3: example working of convolutional filter

The figure shows that the convolutional filter, which in this example consists of 9 weights to be calibrated, translates a 3x3 pixel plane into a single number with simple multiplication. By shifting this filter over the entire image, a new image is ultimately created, whereby the relationship between pixels in the original image is addressed by means of the filters. In this way the model can recognize patterns in the image.

Multiple filters can be applied to an image at the same time, which are collected in one layer of the model. The model consists of different layers. In addition to the filters, the model contains layers where either a non-linear operation is applied or the largest values are collected ("pooling" layer). During the training, the CNN learns the weights associated with the filters or the other layers. The strength of the model is that through the structure with layers with filters it automatically:

- **Recognizes detailed features such as borders, color changes and shapes;**
- **Recognizes shapes based on this, followed by features such as faces, parts of a car, etc.;**
- **Finally, based on these characteristics, makes an estimate of what is on the image (with associated probabilities).**

> **"If insurers collect data sets with more images of claims and the associated costs of the claims, then these types of models could be included in the processes of insurers to be able to automatically estimate the heights of car damage."**

A CNN model has many parameters, which implies that many images are also needed to estimate the model reliably. These are generally not available from insurers for this type of purpose. A solution for this can be to take an already existing (and calibrated) CNN as a basis and only (re) calibrate the last layer based on the available data set. This is called "Transfer Learning". In the example in this article, the so-called "VGG16"[6] model is used, calibrated to the well-known ImageNet dataset (with approximately 15 million images).

3.3 Results

Five different CNN variants have been applied to the dataset, with variations in, among other things, the number of layers and the number of filters per layer. In addition, a Transfer Learning approach based on VGG16 has also been applied. The models are calibrated to the training data set, and then the accuracy of the models is measured based on the validation data set. The results are shown in Table 1.

Model	Accuracy
CNN Variant 1	77,7%
CNN Variant 2	72,3%
CNN Variant 3	69,4%
CNN Variant 4	75,7%
CNN Variant 5	78,6%
CNN / VGG16	93,0%

Table 1: accuracy different CNN models

The table shows that the accuracy of the normal CNN variants does not exceed 80%. In addition, there is probably overfitting, because millions of parameters are used. The quality of Transfer Learning based on the VGG16 model shows significantly better results. Apparently, the parameters calibrated on the basis of the general ImageNet data set are well applicable as a basis for specific data sets such as the data set in this article.

4. Conclusion

The results in this article imply that image recognition can achieve high accuracy with a relatively small data set. If insurers collect data sets with more images of claims and the associated costs of the claims, then these types of models could be included in the processes of insurers to be able to automatically estimate the heights of car damage. The models mentioned in this article can also be used for other (future) purposes where images play a role.

Dr. Richard Plat

- Partner Risk At Work.

Richard holds a Master of Science degree (1999) and PhD (2011) in Actuarial Science at the University of Amsterdam and successfully finished postdoctoral studies for actuary of the Dutch actuarial society (AAG) and investment analyst (RBA / CEFA). He has 20 years of experience in actuarial and risk consultancy for insurers, pension funds and banks. Besides that, he published several scientific and practically oriented articles and has been speaker at several seminars and workshops on different topics.

Index

1. Dr. Richard Plat AAG RBA is partner of Risk at Work.
2. Source: https://towardsdatascience.com/cutting-edge-face-recognition-is-complicated-these-spreadsheets-make-it-easier-e7864dbf0e1a
3. Source: https://github.com/halloTheCoder/car-damage-evaluation-using-deep-learning/tree/master/code
4. See Plat – 'Data Science en Machine Learning: concreet voorbeeld verzekeringsportefeuille', De Actuaris (September 2017).
5. See for example https://medium.freecodecamp.org/an-intuitive-guide-to-convolutional-neural-networks-260c2de0a050
6. http://www.robots.ox.ac.uk/~vgg/research/very_deep/

RISK
AT
WORK

Who we are

At Risk-at-Work we specialize in making models and providing solutions for quantitative risk management, trading and ALM topics. We pride ourselves on cutting through complexity by providing creative and practical solutions for hard problems. We do this by having a wealth of practical experience, state-of-the-art technical knowledge but first and foremost by understanding our clients and their problems and co-creating solutions with them that work.

What we do

Our client list covers all major Dutch financial institutions. We help our clients to get better, to go further. We help on topics where we believe we can offer unsurpassable quality. This means that we often work on topics that are on the intersection of quantitative finance, risk management and ALM. On these topics, we can help with providing insight, develop and implement models, review and improve models or just analyze understand data.

www.**riskatwork**.nl info@**riskatwork**.nl