



# AI IN VIDEO RECOGNITION

ASSESSING VIDEO  
FOOTAGE WITH A  
MACHINE LEARNING  
ALGORITHM



# AI IN VIDEO RECOGNITION

## ASSESSING VIDEO FOOTAGE WITH A MACHINE LEARNING ALGORITHM

**MANY COMPANIES CREATE VIDEO FOOTAGE FOR APPLICATIONS LIKE INSPECTION, QUALITY CONTROL, SURVEILLANCE OR PROCESS MANAGEMENT. A SHARED FEATURE OF ALL SITUATIONS WHERE VIDEO MONITORING IS APPLIED IS THAT IT INVOLVES A LOT OF MANPOWER TO ASSESS THE FOOTAGE. THIS IS WHERE ARTIFICIAL INTELLIGENCE (AI) CAN HELP.**

It's almost impossible to do without AI anymore, and its applications are widespread: from self-driving cars and apps that can assess whether a birthmark will lead to skin cancer to the automatic detection of cows going into heat based on drone-mounted camera footage. All this is made possible by artificial intelligence, especially machine learning and deep learning. In this whitepaper we outline the uses of machine learning for the automatic recognition and classification of video footage. For which situations in video recognition can AI be applied? What are the benefits? And what does the business case look like?

### VARIOUS USES OF VIDEO RECOGNITION

This spring many birds nests in the Netherlands were outfitted with camera's, so that bird enthusiasts were able to follow the breeding process. Yet most of the day nothing much happens on these feeds. Wouldn't it be useful if an algorithm could automatically detect whenever something special happened and would then select that footage into a highlights video at the end of the day? A similar AI application can be used to assess video footage of surveillance camera's.

Surveillance personnel then no longer has to monitor all the different camera feeds, but only gets to see real-time footage whenever something is happening. The assessment of that footage is still done by people, but they no longer have to monitor footage where nothing is going on.

Another AI category is automatic recognition and inventory of objects, like products in a packaging line. In most meat factories different meats are packaged in succession, since the packaging line itself is unable to identify the types of meat. The meat cutter selects the meat in crates, causing the meat to lie for too long and lose its juices. It would be very useful indeed if a camera would be able to tell the difference between a round steak and a pork rib and could manage the follow-up procedures automatically. Then serial production could be transformed into parallel processing.

Other examples of video recognition for the inventory of objects are counting the number of containers on a transshipment site, monitoring the growth of agricultural crops or discerning between crops and weeds.

A last category in video recognition applications is inspection. Water authority experts still physically inspect dikes to assess whether they are sturdy enough. When the water level suddenly rises they work overtime to inspect all critical dikes. Wouldn't it increase our safety if a drone was able to shoot close range video footage of a dike and a trained algorithm would notify dike inspectors of potential risks?

# 'USING TRAINED AI MODELS, IT IS POSSIBLE TO CLASSIFY FOOTAGE AUTOMATICALLY. THIS IS DONE BY A MODEL TRAINED IN IMAGE RECOGNITION'

## AI FOR VIDEO RECOGNITION: HOW DOES IT WORK?

Using trained AI models, it is possible to classify footage automatically. This is done by a model trained in image recognition. One condition to properly train the model is the availability of sufficient image data that also has a clear relation to the goal for which the AI is used. A surveillance camera that needs to distinguish burglars from employees and passers-by will need sufficient images of all three these situations. In reality the number of images showing burglars will be quite low, making it all but impossible to distinguish between an employee and a burglar. However, it is possible to distinguish images of people entering the premises from images of people merely passing by.

Aside from having sufficient data with a clear relation to the goal, it is also important to monitor the learning process. One feature of AI is that the model itself learns from all new input. This can only happen if the new data also has a clear relation to the goal, so that the algorithm understands what it needs to learn. If the surveillance camera's register small animals such as cats or rabbits entering the premises, it can be trained to recognise them as such. However, this learning process needs to be monitored so the algorithm will not mistake a man crawling underneath a barbed wired fence for a dog.

The reason this is so crucial is that AI often fails at things that we as human beings are good at: recognising context and reacting accordingly. AI only recognises things it is specifically trained for, not things it isn't trained for. For example, take the AI for recognising breast cancer from a mammograph: these algorithms are better and faster at diagnosing the beginning stages of cancer than a human radiologist, but they are unable to recognise other diseases. That makes them well equipped for breast cancer screening, but to help women visiting the clinic with general health issues a broader view is needed.

## USE CASE: ROAD INSPECTIONS

We like to clarify these conditions and the underlying business case by means of a practical use case: the inspection of asphalt by construction company BAM. This inspection is done by means of camera footage made by a recording car. All the footage is viewed by experts. When they see damage, they draw a frame around it and annotate it.

In cooperation with OrangeNXT and ICT Group BAM has investigated whether damage recognition can also be executed by a self-learning algorithm. For this purpose a model trained in image recognition was used. The model was trained with a set of 2500 images of eight different types of asphalt damage. This is how the model learns to recognise damages and discern the different types of damage.

During the learning process the team discovered that some of the damages were framed too broadly, which made it difficult for the algorithm to learn. For instance, the algorithm did not properly recognise animal remains on the asphalt, because large areas of clean asphalt were included in the annotation. The annotation framework is now being narrowed so there is less noise for the algorithm to process, which leads to more accurate predictions. Furthermore, the algorithm struggles with damages that are also difficult for people to recognise, such as asphalt raveling. This makes sense, since images of raveling that haven't been annotated as such have also ended up in the training set. Sometimes the difference between a healthy and a damaged road is very subtle. For these kinds of damage the algorithm is presented with a larger training set. The ultimate goal is to replace classification by human experts by a qualification of AI detected damages.

The model is structured in such a way as to prevent false negatives, meaning that no footage containing damages is missed. This is why the first version contains comparatively more false positives: footage the algorithms has doubts about, but that through further assessment by a human expert appears to be undamaged. This input is used to fine-tune the algorithm, so it will be able to better predict the degree of asphalt damage. The first algorithm that BAM used was able to assess 80 percent of the footage as undamaged. That meant the inspectors only had to review the remaining 20 percent of the footage. The extra saved time was used to further train the algorithm and monitor the learning process.

In the long run this ratio will grow towards 99 percent automatic detection and 1 percent human intervention. For a long time human intervention will still be necessary, because footage may always contain unusual images that the algorithm has never seen before and is unable to explain. Because these exceptions occur so rarely, the algorithm will necessarily have too little training data to recognise these situations of its own accord. That's why applying AI continues to remain a combination of man and machine.

### ONE DATA SET, VARIOUS TRAINING GOALS

In some cases a dataset can be used for different purposes. In the BAM case the footage of the asphalt is an adaption of footage shot with a Horus car camera, a device similar to what Google Streetview uses. This means the original footage contains much more information than only asphalt damage. For example, they also show the location and state of street signs. Maintenance of these signs can be scheduled based on that information. It's also possible to review the orderliness of traffic situations: are the signs positioned clearly and in locations that make sense? The rise of self-driving vehicles increases the importance of orderly traffic situations. Using AI this can be evaluated much faster.

The example of traffic signs also shows that the needed training set varies according to the application. Each country has a limited amount of traffic signs. The algorithm has to learn them only once in order to recognise them in the future. Still, the AI will not be able to correctly discern a person wearing a t-shirt with an image of a traffic sign, so human intervention will always be needed to assess all situations. But the training set can be much smaller than in an application where the algorithm has to recognise different types of asphalt damage.

## BENEFITS OF AI USE

THERE ARE MANY DIFFERENT BENEFITS OF USING AI:

### 1. HIGHER EFFICIENCY

*People no longer have to look at all the camera footage, but only at the footage that shows probable deviations. This enables them to do more work in less time;*

### 2. MORE INTERESTING WORK

*Work becomes more interesting to the experts reviewing the footage. They only have to assess the more complicated cases;*

### 3. HIGHER QUALITY

*Unlike people, algorithms do not get tired and don't lose their concentration. An AI model always delivers a predictable output. The quality of that output depends on how well the algorithm is trained. In the beginning an algorithm will perform sub-optimally. The challenge is to train the model in such a way that the number of false negatives is reduced to 0;*

#### 4. FASTER DECISIONS

*Because people are able to do more work in less time, they can speed up the decision making process for camera footage that requires an intervention. This is a big advantage in emergency situations;*

#### 5. CONTINUITY

*An AI model doesn't fall sick, doesn't go on holiday and can work 24x7;*

#### 6. SCALEABLE

*An AI model can be easily duplicated on other virtual machines to increase the processing speed. This makes it possible to process an enormous amount of video footage simultaneously;*

#### 7. LOWER COSTS

*The aforementioned benefits - especially the efficiency benefit - lead to much lower costs;*

#### 8. NEW BUSINESS MODELS

*The combination of lower costs and higher quality creates opportunities for new business models. For example, inspections can be conducted more frequently, so that errors can be corrected earlier. It's also possible to detect damages much earlier and to accurately monitor the damage process, opening up the possibility for a different maintenance schedule;*

#### 9. HUMAN CAPACITY

*Human capacity is no longer a limiting factor.*

### BUSINESS CASES JUST AROUND THE CORNER

The number of possible AI applications in image recognition is large. Yet real-world implementation is low. This is caused by the notion that the technology is still in its infancy. Nothing could be more wrong. Neural networks were already deployed in the 1980s. The first image recognition applications were developed in the 1990s. In the current millennium image recognition is widely used by the police in criminal investigations, such as number plate recognition and automatic face recognition. In the medical world AI is widely used as well, for instance in finding cancer cells in an MRI. In short, the technology is mature, which has lowered the costs of selflearning models.

The real costs are determined by the quality of the training data. The greater the dataset and the more accurate the annotations, the faster the algorithm can be trained. However, if training data are badly labeled, the quality has to be boosted first. As we said earlier, an algorithm learns much faster if it immediately understands what it should recognise in an image or video.

If you would like to know what your business case may look like, feel free to talk to us. We are able to evaluate the quality of your data and offer an estimate of the costs needed for improvement.

### FOR MORE INFORMATION PLEASE CONTACT US THROUGH:

#### BAS LAMME

BUSINESS DEVELOPMENT MANAGER

+31 (0)6 25 73 33 27

bas.lamme@orangenxt.com

#### FRANK THOMSON

DATA SCIENTIST

+31 (0)6 28 60 09 89

frank.thomson@orangenxt.com



# TURBO-BOOST YOUR DIGITAL TRANSFORMATION



## ORANGENXT

PEOPLE • DEVICES • DATA

John F. Kennedylaan 2 • 5612 AB Eindhoven • P.O. Box: Kopenhagen 9 • 2993 LL Barendrecht  
+31 (0)88 908 2000 • [info@orangenxt.com](mailto:info@orangenxt.com)

[WWW.ORANGENXT.COM](http://WWW.ORANGENXT.COM)

an **ICT|GROUP** company