DEEP NEURAL NETWORK APPROACH FOR AUTOMATIC FOG DETECTION USING TRAFFIC CAMERA IMAGES

Giuliano Andrea Pagani, Jan Willem Noteboom and Wiel Wauben

R&D Observations and Data Technology

Royal Netherlands Meteorological Institute (KNMI)

P.O. Box 201, 3730 AE De Bilt, The Netherlands

email: Andrea.Pagani@knmi.nl, Janwillem.Noteboom@knmi.nl and Wiel.Wauben@knmi.nl

ABSTRACT

Fog is a meteorological phenomenon that has impact on the safety and capacity of transportation and logistics activities on land, water and in the air. Modern visibility sensing technology is used to detect fog for example at airports. Nevertheless, this approach is expensive and not feasible for country-wide coverage of all roads and waterways. Alternatively, cameras showing visual landmarks at known distances allow a human operator to estimate the visibility conditions of a remote location. This approach requires, however, continuous monitoring by remote observers and therefore has limitations to the number of sites that can be supervised by one operator. Nowadays, dense networks of surveillance cameras along roads and waterways offer big amounts of image data that can be used for automated fog detection. Complicating factors are the many different types of cameras used, the wide variety in sceneries, the variations due to pan-tilt and zooming and the privacy regulations that must be respected.

We have developed an automated system to detect dense fog conditions during daytime using cameras already in place for monitoring highway traffic. To recognize fog conditions successfully from many different sceneries, a deep neural networks approach has been employed. Deep neural networks have proven to be good in adapting to the changing scenery in which cameras can zoom in and out and turn suddenly, thus changing the objects in focus. In this study, the images of more than 160 traffic cameras of the road authorities spread across the Netherlands are used. Two 5-layers neural networks have been trained using a selection of these images from cameras within different distances from visibility sensors. The paper presents details of the deep neural network and results of the automatic fog detection using traffic camera images.

The results for automatic fog detection using traffic camera images are promising. Current efforts are in improving the performance of the neural network model by collecting a larger training and test set and by the inclusion of other meteorological variables. The usage of more classes of visibility is also under evaluation. Furthermore, fog detection using cameras in artificial light (lighting along highways) will be investigated and we are committed to bring the developed fog detection system closer to operational duties for automated fog monitoring and alerting along Dutch highways.

1. INTRODUCTION

Visibility is a meteorological variable that is important for the safety and capacity of road and air traffic. Visibility is defined as the greatest distance at which a black object of suitable dimensions, situated near the ground, can be seen and recognized when observed against a background (WMO, 2014). Physically, visibility (Meteorological Optical Range) is defined as the distance required to reduce the intensity of a light source to 5% of its original value. Traditionally, visibility was estimated manually by observers using visibility markers at known distances. Currently, so-called forward scatter sensors are employed in the meteorological network of KNMI to measure visibility, but these sensors are rather costly and the sample

volume is small. Therefore, the measurements are representative for only a small area, whereas visibility can vary largely on a small spatial scale. Hence a meteorological measurement network is generally not dense enough to detect all occurrences of fog.

Nowadays, cameras are widely used for various applications such as security, supervision, traffic, construction and tourism, and images are readily available on the internet. Furthermore, image processing software is available for the interpretation of these images. Visibility extracted from camera images has therefore the potential to provide useful information. KNMI operates visibility sensors at about 25 automatic weather stations throughout The Netherlands whereas Rijkswaterstaat (the Dutch road authority) has about 5000 traffic cameras along the motorways and near tunnels, bridges etc. (see Figure 1). Using existing cameras, although that are not distributed evenly around the country, is an efficient means to obtain additional information on visibility or the occurrence of fog. Combining these different sources of (big) data is of mutual interest of meteorological institutes and road authorities.

In a previous paper (Wauben and Roth, 2016) an exploration of applying image processing techniques to camera images in order to study the feasibility of deriving visibility was performed. The exploration considered fog detection as well as quantitative estimation of the visibility. Several methods were considered to derive visibility such as; (i) edge detection; (ii) contrast and (iii) dehazing techniques. In this paper we consider traffic camera images that have been provided by the Dutch road authority. They contain sceneries that differ between cameras and may even differ over time for a specific camera since cameras can tilt, pan, and zoom. This makes the processing techniques to derive visibility developed in (Wauben and Roth, 2016) not applicable and requires a new approach. A data-driven modelling approach is considered using a supervised learning by means of deep neural networks where visibility is not estimated, but visibility classes are detected.

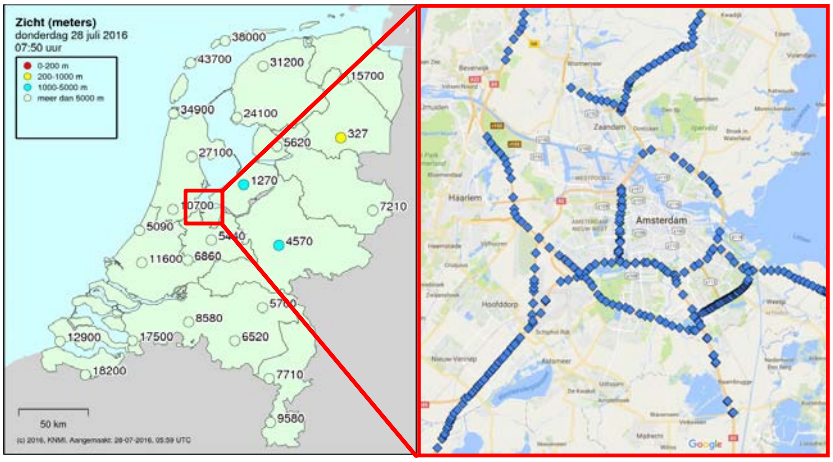

Figure 1: Overview of the automatic weather stations in The Netherlands measuring visibility (left) and the locations of traffic cameras of Rijkswaterstaat around Amsterdam.

2. DATA SETS

The data sets used in this investigation are composed of several sets of images collected by KNMI in the past years. The images are archived every ten minutes. This sampling rate is an trade-off between the dynamics of the meteorological phenomenon and the computer storage and processing constraints and costs. Four image data sets are used in this work. The first is a collection of images of the KNMI test field in De Bilt. For this collection the images are archived since 1st June 2015. The second set is composed of images from two cameras of the Cabauw atmospheric research site of KNMI and collection started on 7th October 2016. The third collection contains images of seven cameras from four civil airports, namely Eelde, Rotterdam, Schiphol, and Beek. These images are archived since 6th June 2017. The last collection of images contain images from 160 cameras for traffic monitoring along the Dutch highways. These last collection is archived since 28th June 2017. All these collections combined total to about 10 million images archived till the moment of writing this article.

While archiving the images, an important set of metadata are stored in a database to ease the selection and processing of the images. When an image is archived its timestamp, camera identifier and the phase of day (e.g., day, night, astronomical/nautical/civil dawn or dusk) are also stored in the database. In addition, the location of the camera (in longitude and latitude coordinates) are stored in the database as well as the location of the KNMI visibility sensors. Obtaining the location of a camera is not as trivial as it seems. For the cameras at KNMI locations the coordinates are available. For the cameras along the highways, the highway location marker (i.e., highway milestone marker) and the direction of travel are the only information available. The conversion from highway location marker to a longitude/latitude coordinate pair was achieved through a web resource (i.e., hmpaal.nl) or else by a careful inspection of highway pictures via google maps.

3. LABELLING THE DATA

One essential aspect when exploiting a supervised machine learning technique is to provide the data (i.e., the camera images) with true and accurate values of the variable (i.e., ground truth) that is going to be predicted by the machine learning system. The visibility measurements, in fact the so-called meteorological optical range (MOR), obtained at KNMI automatic weather stations is used as the ground truth. The values of visibility less or equal to 250m are categorized as “foggy”, and “non-foggy” otherwise. The threshold of 250m for dense fog is used by traffic management operators as a key visibility distance. The association of this categorical label (i.e., visibility class) to an image is straightforward for the images taken at KNMI and airport locations. For these cameras the 10-minute averaged MOR of the visibility sensor at the same time the image has been captured is converted to a categorical fog/non-fog value. For the images taken from cameras along the highways the following two selections have been applied: i) only cameras within 2.5km of a visibility sensor are considered and the images are labelled with the visibility class obtained from the nearest visibility sensor at the same time; ii) as (i), but using cameras within 7.5km of a visibility sensor. Selection (i) resulted in only 16 cameras along highways with 4 associated visibility sensors, while selection (ii) leaves 82 cameras with 7 associated visibility sensors. The first selection is a subset of the second one.

This labelling procedure reduces the amount of images available for training the machine learning algorithm, but allows for a fast labelling process compared to human inspection of each image. The distance between the location of the camera and the visibility sensor is a factor that affects the quality of the labelling, particularly since visibility is a phenomenon that can have large spatial variations. The two selection thresholds for assigning the labels to the images are considered in order to investigate the impact of the label quality on the performance of the algorithm both during training and while assessing of the system.

Figure 2: Example of an image of KNMI station Eelde (left) and a scenery along the N15 highway at hm 236 (right).

4. NEURAL NETWORKS FOR IMAGE CLASSIFICATION

Neural networks are a powerful class of machine learning methods that allow fitting complicated non-linear functions. Historically, these networks have been inspired by neuroscience and the functioning of the brain, i.e. many neurons evaluate a set of inputs and fire an output when a certain threshold is exceeded (Fausett, 1994). The outputs then become inputs of the following layer of neurons that perform another evaluation and a new threshold for firing is applied again. Such mechanisms go on for a set of levels or depth of the neural network and it is the essence of the feed-forward neural networks (feedback loops are not in present in the network). The goal of neural networks is to learn about the phenomenon under investigation without being explicitly providing specific rules. It is by being exposed to many (in the order of tens of thousands to millions) examples that the network will learn to approximate the output from the inputs provided.

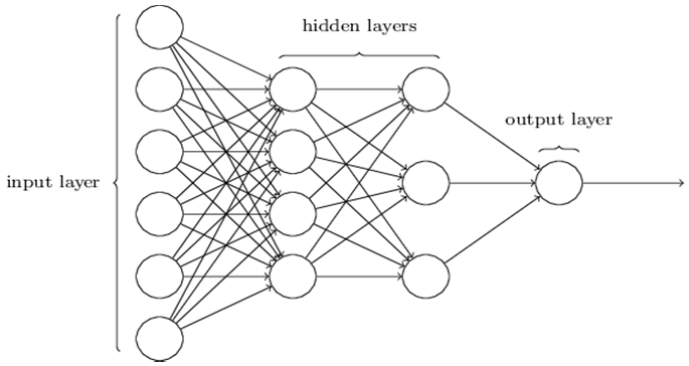

Figure 3: Illustration of a (feed-forward) neural network with two hidden layers.

The evaluation phase inside a neuron consists usually of summing the inputs by properly weighting them. In the learning phase, the algorithm tries to find the right weights that are best suited in approximating the desired output. The values assigned to the weights start from an initial random guess and they are updated iteratively in order to minimize a loss function using some form of gradient descent (Shalev-Shwartz and Ben-David, 2014).

Image processing and recognition of objects in images has recently received a boost by the abundance of images in digital form as well as by the use of deep neural networks (CireşAn et al., 2012). Deep neural networks do not have a uniquely established definition, but the concept usually refers to many layers or at least more than one hidden layer in the network. The number of hidden layers refers to the number of nodes touched by a path traversing the acyclic directed graph representing a feed-forward neural network between an input and the output node (see Figure 3).

4.1. Why neural networks?

The previous work (Wauben and Roth, 2016) used other machine learning techniques in order to estimate the visibility for the KNMI test field in De Bilt and the automatic weather station in Twenthe. For that purpose we used decision trees and a set of features such as mean edges and contrast in the image. That method provided good results for that particular circumstances: scenery is fixed and the objects in the image are constant through time. Thus training two models for both De Bilt and Twenthe was the solution. The problem has drastically changed when the goal has become to detect fog from images of surveillance cameras along the highways. First, the cameras have different sceneries among each other, thus calling for training several individual models if the decision trees approach was to be followed. Second, and most important, almost all cameras along the highways have the possibility to change the scenery by panning, tilting, and zooming. These operations are essential for the traffic monitoring personnel to assure the safety of the traffic and facilitates immediate (remote) inspection in case of accidents. The high dynamicity in the scenery called for a more powerful method that can better generalize the image properties of dense fog. Given the good results reported in the literature by the use of deep neural networks in image classification and their ability to solve more complex classes of problems (Krizhevsky et al., 2012), we have decided to apply such techniques for the detection of fog from images obtained from various cameras.

4.2. Image pre-processing and feature extraction

Before applying a neural network the images require pre-processing. This aspect is key in order to make them suitable and transform them into a set of features to be used as the predictors for the phenomenon one wants to predict. In this case two operations have been performed in the pre-processing phase:

1. Dimension reduction of the image to a 28-by-28 pixel image. This operation has three goals, the first is to homogenize the pictures that come from different cameras that have different resolutions into a single common format; the second goal is to reduce the amount of variables (i.e., pixels) that have to be processed, thus reducing computation resources and time needed; the third goal is to reduce overfitting by having a reduced variable space.

2. The image is blurred to avoid the presence of specific pixels representing signs, objects or written information on the image (e.g., camera location) that could be learnt by the network, thus compromising the generalizability of the knowledge learnt by the network.

Two examples of the effect of dimension reduction and blurring of images are provided in Figure 4. Once these two pre-processing steps are applied, the red, green, and blue (RGB) channels containing the intensity of each pixel in the image are extracted and used as features for the machine learning problem. Therefore, the input of the image to the neural network is constituted by a vector of 28x28x3=2352 variables.

Figure 4: Same images of Figure 2, but after dimension reduction and blurring.

5. MODEL FITTING AND SETUP

Once the data have been pre-processed as described in Section 4.2, the neural network model can be fitted. The fitting was performed individually for the 2.5 and 7.5km selections as described in Section 3.

Both data sets have been split in a training, a validation, and a test set. The split is random and the training set contains 60% of the data (i.e., 60% of the total dense fog cases and 60% of the total non-fog cases), while the validation and test sets contain evenly the remaining data (i.e., each 20% of the original data). In order to train the neural network the R programming environment has been used. In particular the library H2O (www.h2o.ai) has been used which is a specialized package for modelling machine learning problems. H2O has been used since it is a modern, up-to-date library that gets updated frequently and contains cutting edge machine learning techniques. Furthermore, H2O allows for high performance by facilitating parallel and distributed computation, and in-memory computations.

Fitting a neural network is not as trivial as it might seem since there are many parameters that need to be tuned in order to obtain a good performance of the network for the classification task. The identification of the parameters (e.g., number of layers, amount of nodes per layer, type of activation function, training with or without balanced classes, see Montovan et al. (1998)), known as hyper-parameter space, is a critical task that requires time due to re-fitting of the network at any new configuration of the parameters. The H2O library has a solution for effective searching in the hyper-parameter space that is based on a grid search on all the possible combinations of the hyper-parameters and fitting models for each combination. A random exploration (i.e., a random subset of all the possible

hyper-parameters combinations is used) of the hyper-parameters space using various initial configuration has been used, selecting the models that provided the best F1-score (Powers, 2011). Balancing of classes in training is applied to tune the results using even fog and non-fog occurrences, since the model should be tuned to recognize the fog events. The validation set has been used to evaluate the performance of various models produced with different configurations of hyper-parameters. The fitted models and their main parameters are presented in Table 1.

Table 1: Fitted deep neural networks for the two data sets.

Case	Number of layers	Number of nodes in hidden layers	Activation function in hidden layers	F1 score training subset*
2.5km data set	7 (Input, 5 hidden layers, output)	75, 75, 50, 50, 10	Rectifier	0.986
7.5km data set	7 (Input, 5 hidden layers, output)	50, 50, 50, 25, 10	Rectifier	0.981

6. RESULTS

The results of each model are evaluated against the respective test set, which is an unseen data set. The results are reported by using a confusion matrix that is a 2 by 2 contingency table containing the number of cases correctly or incorrectly classified as fog or non-fog. Columns indicate the ground truth and rows the prediction of the neural network model. Figure 5 contains a representation of how to interpret the confusion matrix and the metrics used to assess the performance. In the matrix, TRUE means a positive case of dense fog, that is a visibility equal or lower than 250m, while FALSE means a negative case, thus a visibility higher than 250m or non-fog.

Figure 6 (left) shows the results of the model built on training cameras within 2.5km from a visibility sensor. Overall the accuracy (fraction of cases correctly predicted) is very high (99.6% accurate detection by the neural network). However, since the data set is unbalanced towards non-fog, this indicator is misleading. Therefore it is more informative to look at the indicators precision and recall. Precision (or 1 – false alarm rate, 1 – FAR) is the fraction of predicted fog cases correctly predicted by the model. Recall (or probability of detection, POD) is the fraction of dense fog cases correctly predicted. F1 score is the harmonic mean of precision and recall.

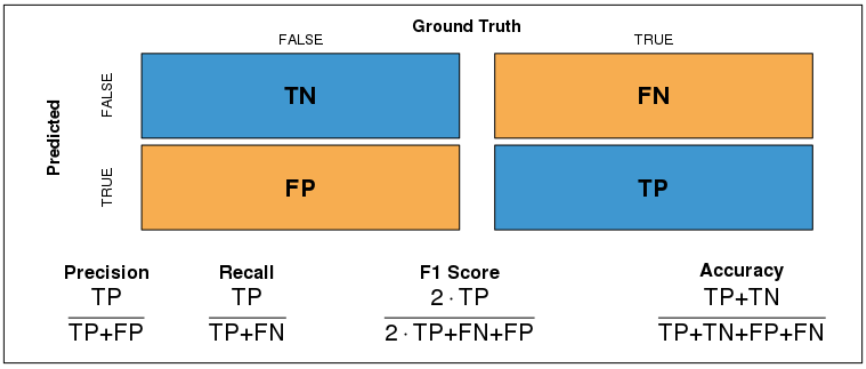

Figure 5: Confusion matrix elements and metrics for performance assessment.

* F1 score computed on a balanced subset from the training set of 10000 images per class.

The model built on training cameras within 7.5km from a visibility sensor has also been assessed on the respective test set. The results are shown in Figure 6 (right). The overall accuracy is similar to the 2.5km case, but precision, recall and F1 score are much lower. This is probably caused by the ground truth labels that are generally less accurate when considering cameras further away from the measurement locations.

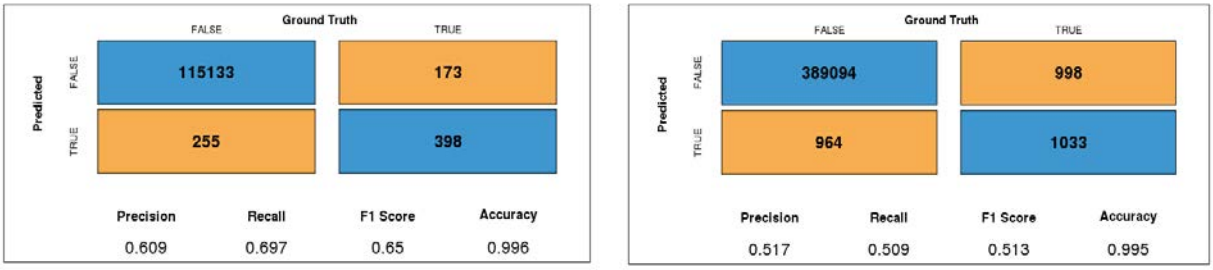

Figure 6: Confusion matrix for 2.5km-trained model (left) and 7.5km-trained model (right) on their respective test set.

In addition to assessing the performance of the models on their respective test set, the models have also been applied to a joint test set. This joint test set is built from merging the test sets of the 2.5km case and the 7.5km case, but removing any image that is present in the training or validation set of the 2.5km case or the 7.5km case. The creation of such a set is due to guarantee that the images are unseen by both the models. The results on this joint test set when using the 2.5km- and the 7.5km-trained model are shown in Figure 7. The 2.5km-trained model performs much worse for F1 score compared to the 7.5km-trained model. This is caused by the cameras (and therefore scenery) in the 7.5km set that are unknown to the 2.5km-trained model. This indicates issues of generalization (i.e., overfitting) of the model and limitations of applying the model trained on one domain to a different one.

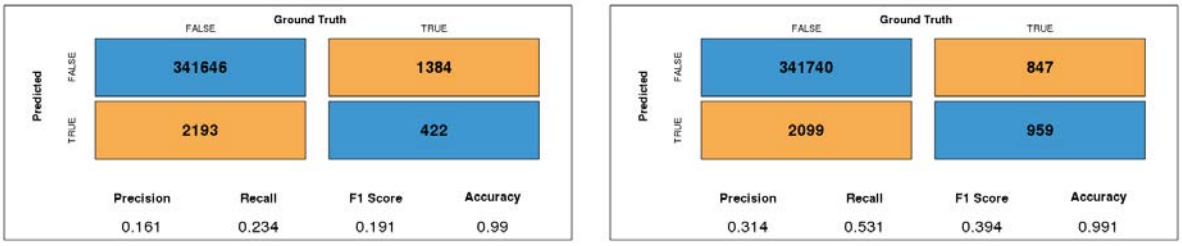

Figure 7: Results on the joint test set with 2.5km-trained model (left) and 7.5km-trained model (right).

7. ANALYSIS OF RESULTS

The combination of the training, validation and test set of the 2.5km case is considered here to investigate the FP (false positive) and FN (false negative) cases in more detail (see Figure 8). There are 1296 FP cases where the model predicts dense fog and the sensor reports no dense fog. Of these cases, 707 (55%) the sensor reports fog (MOR<1000m), so the visibility is indeed reduced. However, in 235 cases (18%) the sensor does not even report haze (MOR<5000m).

Analysis of the FP cases showed that they occur mainly isolated in time (603 or 47%) and the longest duration is 24 consecutive 10-minute intervals. The latter occurred for camera A15-HM207 on March 16, 2018 from 09:30 to 13:20 UT. During this period a visibility reduction is clearly visible on images, but the visibility sensor, at 2.3 km distance, reports visibilities above 1 km. The sensor shows a visibility reduction of 4 km to 1.1 km and back to 4 km. A similar trend can be seen in the images. The number of isolated cases in time is less for FN (270, 32%) and TP (302, 15%) cases. The longest duration for FN and TP are 15 and 46 consecutive 10-minute intervals, respectively. The longest duration for FN occurred for camera A4-HM59 on August 16, 2017 from 04:30 to 6:50 UT. The images

show a clear visibility reduction. It is unclear why the model did not predict dense fog. Possibly the camera location printed on the image affected the results. Also, the scenery on the images at the start of the FN event was different from the generally observed scenery that is shown again on images near the end of the FN event.

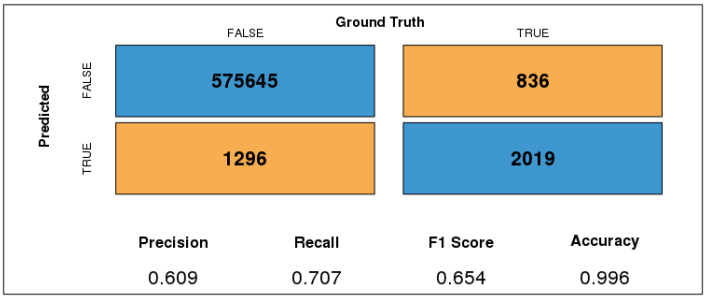

Figure 8: Results on the combination of the training, validation and test set of the 2.5km-trained model.

The scores for individual cameras show spatial differences, also for nearby cameras using the same visibility sensor as ground truth. The higher FP rates for the cameras at Cabauw are probably related to scenery (meadows with few objects and large fraction of grass with more sky), whereas the higher FP for camera A15-HM207 (camera with longest FP event) is not present at nearby camera A15-HM206. The largest TP numbers occur on sites (Cabauw and Eelde) that are prone to fog events. The FP events occur mainly isolated in space (914 or 71%) and at most 4 FP occur at the same time. FN occurs less often spatially isolated (305 or 37%) with at most 7 events at the same time. For TP the numbers are 522 (26%) for spatially isolated with at most 9 events at the same time.

Table 2: Results of various post processing steps on the combination of the training, validation and test set of the 2.5km-trained model.

Post processing	Precision	Recall	F1 score	Accuracy	% omitted	Precision*	% fog
none	60.9%	70.7%	65.4%	0.9963	0.00%		
change	70.2%	74.7%	72.4%	0.9975	0.26%	22.5%	11.7%
change F --> T	70.2%	69.3%	69.7%	0.9972	0.11%	21.5%	4.8%
change T --> F	60.9%	76.0%	67.6%	0.9967	0.15%	23.3%	6.9%
difference with nearest	74.8%	77.6%	76.2%	0.9982	0.37%	31.2%	23.7%
change OR nearest	79.7%	80.6%	80.2%	0.9986	0.54%	28.4%	31.0%
change AND nearest	67.4%	72.7%	69.9%	0.9971	0.09%	23.4%	4.3%

The characteristics of FP, FN and TP given above can be used to post process the predicted model values in order to improve the scores. The characteristics used for this purpose are the isolation in time and space of the faulty cases. In order to make post processing usable for real-time application only the current predicted model values and the prediction 10 minutes before are considered. If there is a change in predicted value the prediction is considered INCONSISTENT, otherwise the original model prediction is kept. The scores after this change with respect to previous post processing and omitting the INCONSISTENT cases are given in Table 2. The scores increase, but 0.26% of the images are classified as INCONSISTENT, containing 12% of the true dense fog cases. The precision* of the omitted INCONSISTENT cases is 23% (i.e. 23% of INCONSITENT cases are true dense fog cases), so less than for the predicted dense fog of the original data and for the post processed data. Note that specific changes (TRUE into FALSE or FALSE into TRUE) can also been considered, but allowing both changes gives the optimal results since Precision and Recall improve. Change TRUE into FALSE improves the Recall and FALSE into TRUE improves the Precision score while hardly affecting the other score. The

F1 score reaches intermediate values and the true dense fog cases that become INCONSISTENT are about 6% for both changes. Another post processing of the predicted model values is obtained by comparison with the predicted model value of the nearest camera that uses the same visibility sensor as the ground truth. Here the predicted value is set to INCONSITENT when the camera pair gives different predictions. The scores after this difference with pair post processing gives better results than change with respect to previous, but the number of images set to INCONSISTENT increases, containing 24% of the true dense fog cases. Finally, the two post processing methods of the predicted model values are combined. When it is required that either criterion (change of prediction with respect to previous or difference with nearest) needs to be met in order to obtain INCONSISTENT the scores are the best, but 0.5% of the images are set to INCONSISTENT, containing 31% of the true dense fog cases. Requiring that both criteria need to be met in order to obtain INCONSISTENT gives optimum results while only 0.09% of the images are set to INCONSISTENT, containing 4% of the true dense fog cases.

Clearly the scores can be improved by post processing the predicted model values, but there is a trade-off with the number of cases that will be made INCONSISTENT. The INCONSISTENT cases can be made available to the users as third possible output of the prediction with lower quality. These cases can be treated as either fog or non-fog depending on the users wish to obtain higher TP or smaller FP values.

8. CONCLUSIONS AND OUTLOOK

In this paper a deep neural networks method to predict the presence of dense fog from daytime camera images has been implemented and evaluated. The results are promising especially when using a data set whose labels are acquired from a visibility sensor close to the location of the camera. Overall, accuracy is very high, but false positives and false negatives have an impact on the precision and recall scores. The faulty predictions are caused by several factors such as scenery, contamination and precipitation on the camera lens, and training with ground truth labels not fully representative for the camera locations. The scores for precision and recall are 0.61 and 0.70 for the 2.5km-trained model and 0.52 and 0.51 for the 7.5km-trained model. The difference points to the reduced quality of the ground truth when the distance between camera and visibility sensor increases, but also the larger fraction of images with varying scenery. Differences between the 2.5 and 7.5km-trained models indicate issues of generalization. Post processing can be used to optimize the results, but this introduces INCONSISTENT cases with lower quality.

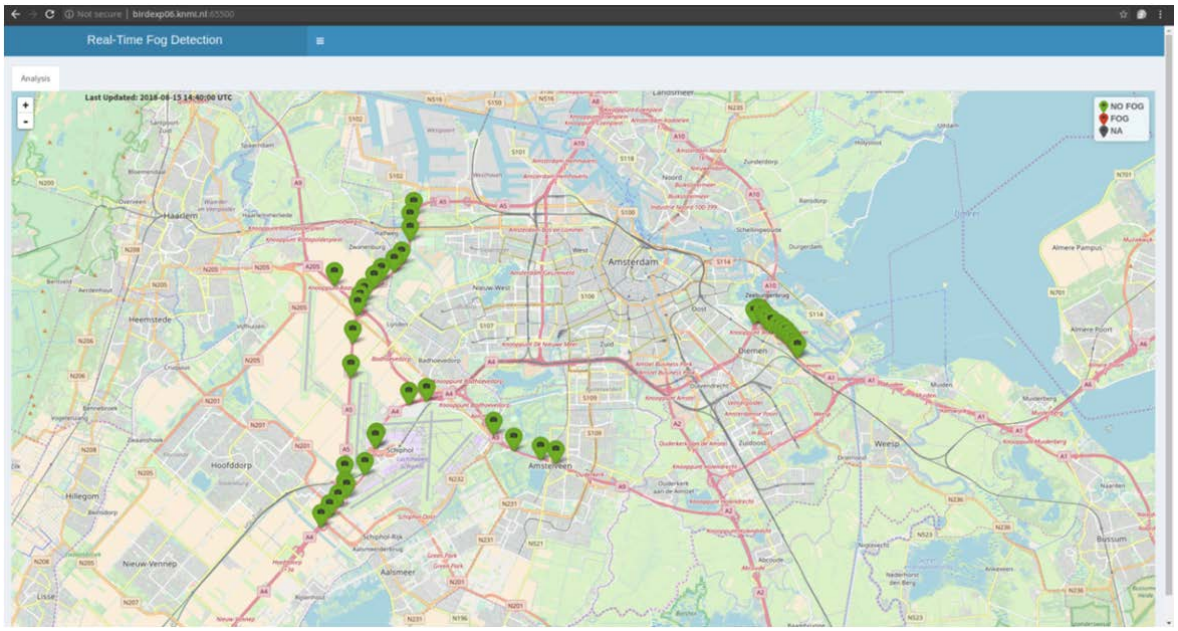

Figure 9: Geographical map showing the predicted values of the cameras (2.5km case). Selecting a camera shows the corresponding image.

It must be emphasize that the results presented in this paper are work in progress. The neural network will further improve by additional training using more labelled images with a higher variety. For the future we will look into fog detection during non-daylight conditions by relying on artificial lighting. In addition, the amount of visibility classes will be increased to better support the weather room and traffic control centres. A near real-time visualisation of the fog detection based on deep neural networks is under construction (see a preview in Figure 9) so that users can give feedback on the results and possibly it will allow users to provide input that can be used in further re-training of the model.

ACKNOWLEDGEMENTS

It is a pleasure to thank Rijkswaterstaat and especially Johan van Eyk for facilitating the research presented in this paper.

REFERENCES

CireşAn, D., U. Meier, J. Masci, and J. Schmidhuber: Multi-column deep neural network for traffic sign classification, Neural networks 32, 333-338, 2012.

Fausett, L.V.: Fundamentals of neural networks: architectures, algorithms, and applications (Vol. 3), Englewood Cliffs: Prentice-Hall, 1994.

Krizhevsky, A., I. Sutskever, and G.E. Hinton: ImageNet Classification with Deep Convolutional Neural Networks. Advances in Neural Information Processing 25, 2012.

Montavon, G., G.B. Orr, and K.R. Müller: Neural Networks: Tricks of the Trade, Springer, 1998.

Powers, D.M.: Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation, Bioinfo Publications, 2011.

Shalev-Shwartz, S. and S. Ben-David: Understanding machine learning: From theory to algorithms, Cambridge University Press, 2014.

Wauben W., and M. Roth: Exploration of fog detection and visibility estimation from camera images, WMO Technical Conference on Meteorological and Environmental Instruments and Methods of Observation (CIMO TECO 2016), 27-30 September 2016, Madrid, Spain.

WMO: Guide to Meteorological Instruments and Methods of Observation, World Meteorological Organization No. 8, Geneva, Switzerland, 2014 Edition (https://www.wmo.int/pages/prog/www/IMOP/CIMO-Guide.html).