Crash Rate Estimation by Aerial Image Analysis

Nils Kornfeld, Leonhard Lücken, Andreas Leich, Peter Wagner, Hagen Saul, Ragna Hoffmann

German Aerospace Center, Institute of Transportation Systems, Rutherfordstrasse 2, 12489 Berlin, Germany, Telephone: +49 30 67055-380, Telefax: +49 30 67055-291

Abstract - Estimating road safety is a major concern of a large body of theoretical research as well as for practitioners all over the world. Most related studies rely heavily on structured data as tables concerning the road geometry, infrastructural items, traffic volumes, etc., which are not always available.

A more and more universally available source of data, which has rarely been used in conjunction with road safety research are aerial or satellite images. These images potentially contain a wealth of information relevant to the prediction of road safety if they could be thoroughly analyzed in great numbers.

Coincident with the widespread availability of satellite and aerial images, machine learning algorithms for image processing and automatic object detection and classification are maturing. This allows the automated processing of huge amounts of image data by artificial neural networks (ANNs) or related machine learning systems – an area in which convolutional neural networks have shown a significant improvement over conventional methods.

In the submitted work initial results on the application of machine learning on aerial images are presented. The goal is to determine an estimation of road safety levels. ANNs were trained to predict crash frequencies for road intersections relying merely on aerial images of the intersections. The used data consists of police recorded crashes in the city of Berlin and aerial images provided by the Berlin Senate Department for Urban Development.

The performance of the ANN suggests that the line of research is worth further pursuit. For instance, the trained ANN was able to predict the presence of crashes on intersections in a Berlin district excluded from the training process with an accuracy of approximately 74%.

NOTATION

L_{VAE} Loss of the variational autoencoder

L_{VRM} Loss of the variational regression model

D_{KL} Kullback-Leibler divergence

INTRODUCTION

Crash frequencies are an immensely important factor for practitioners in road safety work. For example, the standard process in German road safety assessment is the following:

First, crash frequencies within the road network under consideration are reviewed and crash hotspots identified based on the average annual number of crashes over the past three years. As a second step, countermeasures are devised and implemented for a manually selected set of hotspots.

Finally, the effectiveness of the countermeasures is evaluated using the crash frequencies during following three years.

This process is taken out by the local traffic accident commission, which is constituted by representatives from the local police, the local traffic authority and the municipal administration.

For the city authorities one of the main options to influence the road safety is the manipulation of infrastructural conditions, especially if a particular location is to be addressed. Therefore, the identification of infrastructural deficiencies is of highest relevance as it directly points out possible improvements of road safety. Infrastructural factors, which have been ascribed an important role for road safety, include road construction sites, lane markings, routing layout at the intersection, speed limits and obstructions to the line of sight of road users. For these factors, aerial photographs may contain important hints that might be exploited for identifying potentially dangerous locations in the road network in an automated manner.

This paper deals with the application of modern deep learning approaches for the analysis of crash frequency relevant features in aerial images. The following research questions are addressed:

- Methods for estimating the number of annual crashes at a given patch of an aerial image of the road network (the regression problem)
- Methods for classifying given patches of aerial images into traffic safety classes, e.g. "**dangerous** hot spot", "junction with **moderate** crash frequencies" and "**average** crash frequencies" (the classification problem)
- Methods for understanding what's under the hood of the deep learning algorithms. Thus, looking for an answer to the question "Why does the deep learning algorithm think that this might be a dangerous location?"
- Exploring options that could help practitioners in their everyday road safety work. Possible Examples:
    - "If the deep learning algorithm thinks that this particular image feature is typical for dangerous locations, maybe countermeasures for this part of the road should be considered".
    - "What would the deep learning algorithm say if I add a road marking here?"
    - "What road junction in a region of interest would the deep learning algorithm consider as the most dangerous one?"

The remainder of this paper is organized as follows. First, related work in the field of crash frequency estimation and the specific field of ANNs and Variational Autoencoders is presented. Second, the database used for training and performance testing of the ANNs is described. In the third section the proposed machine learning method and its specific implementation introduced and experimental results are presented in the fourth section. The last chapter gives a discussion of results and an outlook to future work.

RELATED WORK

A common approach for crash rate estimation for practitioners is the use of safety performance functions (SPF) [8]. The approach of SPFs is as follows: a uniform distribution of the crash rate over the road network is assumed (base condition). This crash rate is normalized by the average annual daily traffic (AADT). The average annual crash rate (crashes per km and year) for a country can be estimated from the average annual traffic volume, the size of the road network and the number of

road traffic accidents per year. Crash hotspots are seen as deviations from the mean. It is assumed that the amplitude of deviations from the mean can be calculated using so called crash modification factors (CMF). Crash modification factors are for example lane width, shoulder width and height, curvature, speed limits and others. This final crash rate estimate is calculated by multiplying all CMFs with the base condition crash rate. The SPF method gives course estimates and some work in literature is dedicated to the problem of calibration issues [9]. The SPF approach, as well can be used to get better estimates improve the estimation of crash numbers from conflict rates in traffic [7].

While the computation of local crash rates based on SPF is very cost efficient, the estimation based on conflict rates is comparatively costly.

Najjar et al.[3] used artificial neural networks (ANNs) to Map Road Safety from Satellite Imagery and Open Data. They trained a convolutional neural network, based on the AlexNet-Architecture by Krizhevsky et al.[4] on Satellite Imagery, collected from Google Maps and traffic-accident reports collected and published by the NYPD.[3]

Chen et al.[5] created road safety maps by using a combination of traffic accident and GPS data, collected by 1.6 million anonymous users in Japan. The proposed architecture is based on denoising autoencoders, which apply a lossy compression of the information on the input data in each step. In a final block, the safety level estimation is computed by one final dense layer.[5] In contrast to VAEs, no latent variable representation is learned and there are no additional constraints on the distribution of the intermediate representations. The output loss consists only of the reconstruction loss.

Najjar et al. and Chen et al. use a classification approach, assigning one of three safety levels on the input data.[1, 2] The method proposed in this paper estimates crash rates, by applying a regression, given the training data. The estimated crash rates can then easily be processed further, to discriminate the given data into an arbitrary number of safety levels.

Variational Autoencoders (VAEs) were first developed bei Kingma et al. [1] and Rezende et al.[2] independently. The created neural networks can be used to estimate the lower bound for directed graphical models with continuous latent variables. VAEs can be used for maximum likelihood (ML) or maximum a posteriori (MAP) inference on input data. Because the loss-function is a linear combination of the reconstruction loss of the autoencoder and the Kullback-Leibler divergence of the distribution of the latent variables from the normal Gaussian-distribution, the latent variables are forced into a dense arrangement around the point of origin of the latent space.[1]

DATA

For the analysis in this paper, a data set was compiled, combining

1. images covering the road network of the city of Berlin and
2. crash data for the accidents that happened during a given time period.

The proposed approach was tested on aerial images provided the Berlin Senate Department for Urban Development [10] and road crash data for the year 2016 provided by the Berlin Police

Department [12]. The aim was to estimate of the annual crash frequencies at road intersections specifically.

To enable an ANN to predict crash rates within a particular area, square patches divided into focal and contextual regions [see Figure *1*] were used. The focal region corresponds to the center of the patch, while the contextual region is the remainder of the patch encircling the focus. The task of the ANN was the prediction of two safety relevant parameters in the focal region given the information of the whole patch consisting of context and focus. The first parameter was the annual number of crashes and the second parameter was a classification task for the patch (potential crash site, if the number of accidents is > 0, safe site otherwise).

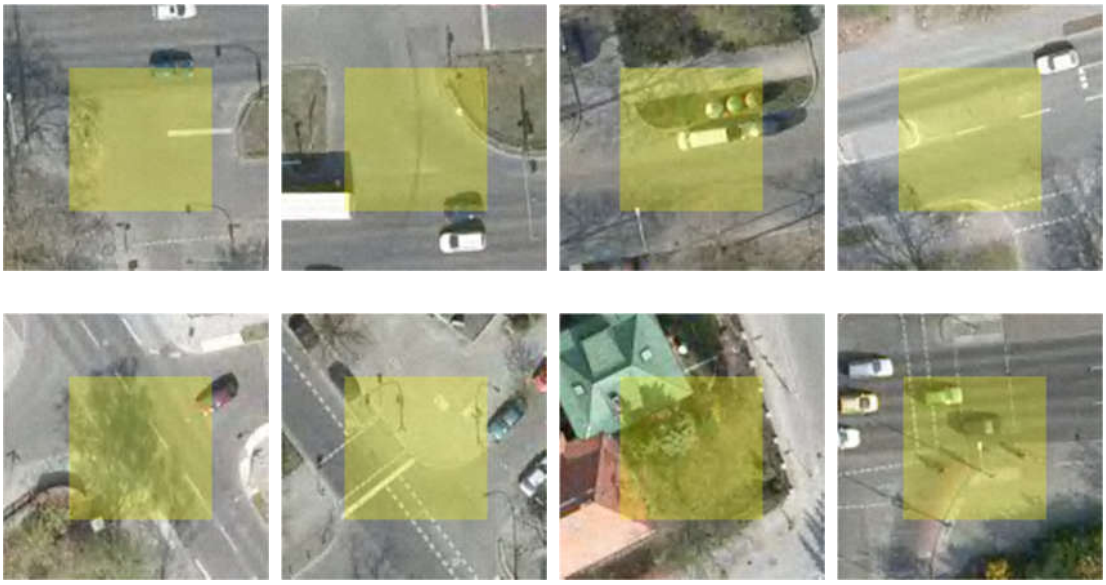

Figure 1: Example image patches close to intersections in Berlin. Focal regions are shaded. Source: Geoportal Berlin / Digitale farbige Orthophotos 2016; extended by own elements.

To generate training and test sets, the area of interest was defined as the conjunction of all intersection neighborhoods. Only junctions involving at least one main road were considered, to avoid a “zero crashes bias” in the data. The road geometry data used for locating the intersections was provided by the Berlin Geodata portal [11]. The training and test sets were constructed iteratively by first adding patches centered in the location of crashes from the year 2016, which were not yet contained in the focal region of any patch added previously. As a next step, patches with random center points in the neighborhood of junctions were included, which were not covered by focal regions of hitherto added patches. This was repeated until a 1:1-proportion of patches with and without a crash location inside their foci was established [see Figure *2*].

Figure 2: Intersection covered by patches (grey shading)centered in crash locations (red marker) and random locations (green markers).Legend: "*Kreuzungsumgebungen*"=Region of Interest around the junction; "*Kontextbereiche*"=contextual regions; "*Fokusbereiche*"=focal regions; "*Zufallspunkte"*=random locations; "*Unfallorte*"=crash locations; "*Hauptnetz*"=main road network. Source: Geoportal Berlin/Digitale farbige Orthophotos 2016; extended by own elements.

A LATENT VARIABLE MODEL FOR CRASH RATE ESTIMATION

The method presented in this paper uses only the information, which can be inferred from aerial images of a potential crash site to estimate the annual crash rate on this site.

The proposed method is based on the following simplifying assumptions:

- There is a latent, unobserved variable z, which is an instantiation of a random process having the prior distribution $p_\theta(z)$.
- There is an observed image $x \in X$, that can be seen as a sample drawn from the conditional distribution $p_\theta(x|z)$, where θ denotes the model parameters of the observation model.
- There is a recognition model $q_\varphi(z|x)$, as an approximation of $p_\theta(z|x)$, as introduced by Kingma et al. in [1].

As shown in [1] the model parameters can be learned with a variational autoencoder.

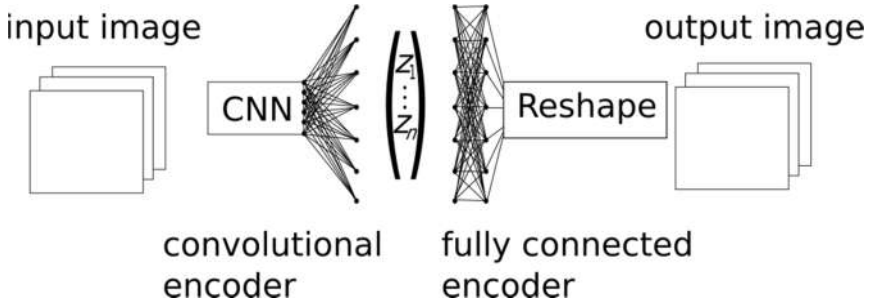

Figure 3. Structure of the Variational Autoencoder (VAE)

To use this model for crash rate estimation, the VAE is trained on aerial images. The target function is the linear combination of the reconstruction loss L_x of the images and the Kullback-Leibler divergence $D_{KL}(q_\varphi(z|x)||p_\theta(z+\epsilon))$ between the recognition model $q_\varphi(z|x)$ and the prior distribution $p_\theta(z+\epsilon)$, where ϵ is randomly sampled noise (see [1] for the derivation of the estimation algorithm). In contrast to the original derivation, here the factors λ, ν of the terms of the linear combination are introduced to be able to weight the two individual loss functions against each other. This leads to the following combined loss function L_{VAE}:

$$L_{VAE} = \lambda D_{KL}(q_\varphi(z|x)||p_\theta(z+\epsilon)) + \nu \sum(x - x^*) \quad [1]$$

The output image is denoted by $x^* \in X^*$. For simplicity the prior distribution $p_\theta(z+\epsilon)$ is assumed to resemble the standard normal distribution $N(0,1)$. The architecture of the specific VAE in this paper is shown in Figure 3. The input images are fed into a convolutional encoder network. The encoder network compresses the information of the input images into an n-dimensional vector, which approximates the latent variable z corresponding to the input image x, maximizing the likelihood $p_\theta(x|z)$. As the input to the encoder a new latent variable z^*, which is close to z is created, by sampling the random variable ϵ:

$$z^* = z + \epsilon \quad [2]$$

So the final loss function can be written as:

$$L_{VAE} = \lambda D_{KL}(q_\varphi(z|x)||N(0,1)) + \nu \sum((x|y) - (x^*|y^*)) \quad [3]$$

When the network shown in Figure 3 is trained with the aerial images, the latent variables are grouped densely around the point of origin resembling a Gaussian distribution in n-dimensional space. This leads to good generalization capabilities of the proposed VAE method.

To leverage the generalization of the VAE, the weights of the encoder part of the network are saved after training. To infer crash rates, the second part of the VAE gets modified for obtaining a variational regression model (VRM), in the following way: The encoder is replaced by a fully connected regressor network, which calculates a single floating point value, as an estimate of the annual crash rate c at the road site shown in the aerial image x.

To train the VRM, in an initial warmup phase, the weights of the encoder are kept the way they are initialized by the already trained encoder of the VAE. Only the weights of the fully connected regression subnetwork are trained for eight epochs on the input data, to keep the already densely

packed distribution in the latent space. So during this warmup phase, the loss function consists only of the mean of the squared errors of one batch of training images with batch size m:

$$L_{Reg} = \frac{1}{m}\sum(c-t)^2 \quad [4]$$

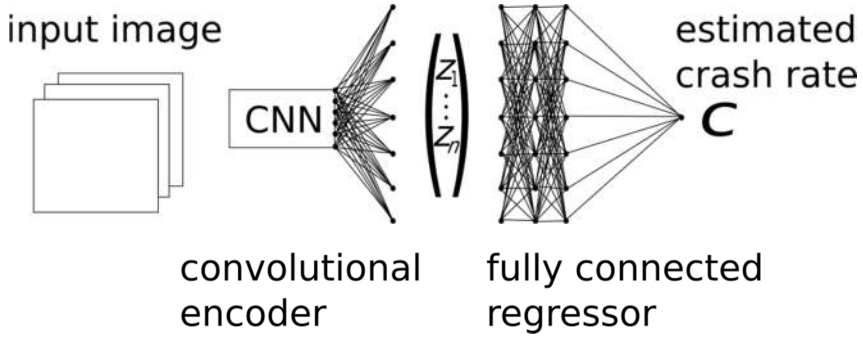

Figure 4. Structure of the Variational Regression Model (VRM)

To further optimize the features of the latent vector z for the regression task, in a second training phase, the whole network is trained, this time including the CNN. The CNN is responsible for feature extraction from the image data. To train the whole network, a new combined loss function is needed:

$$L_{VRM} = \lambda D_{KL}(q_\varphi(z|x)||N(0,1)) + \nu\frac{1}{m}\sum(c-t)^2 \quad [5]$$

To increase the ability to generalize to data not used during training, it is recommended to increase λ in this scenario. After this final training phase, the VRM can be used for crash rate estimation from aerial image data.

The trained networks can now be used for a variety of tasks. The VAE can be used to determine visual cues, which lead the network to believe that a specific road site is especially dangerous. By sampling data from the recognition model $q_\varphi(z|x)$, new artificial data can be created, s. t. the network synthetizes images of safer or more dangerous road sites. The VRM can be used for road safety classification tasks, like in [3] and [5]. The regression output can be fed into a logistic regression algorithm, which now does not need to classify a road site by its aerial image, but by a single scalar value, which corresponds to the road safety at that site.

EVALUATION

In an initial stage the VAE was trained for 400 epochs on a training set, consisting of 37436 aerial images of road sites in Berlin. In Figure 5, ten example training images from the dataset are shown.

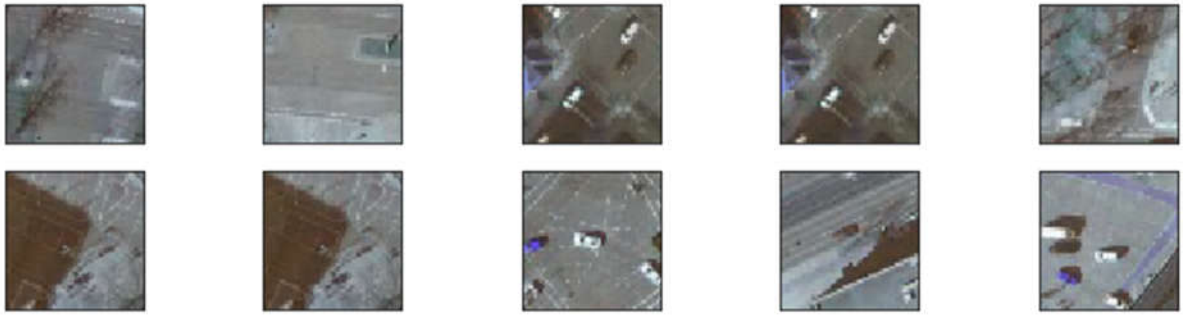

Figure 5. Example images from the training dataset

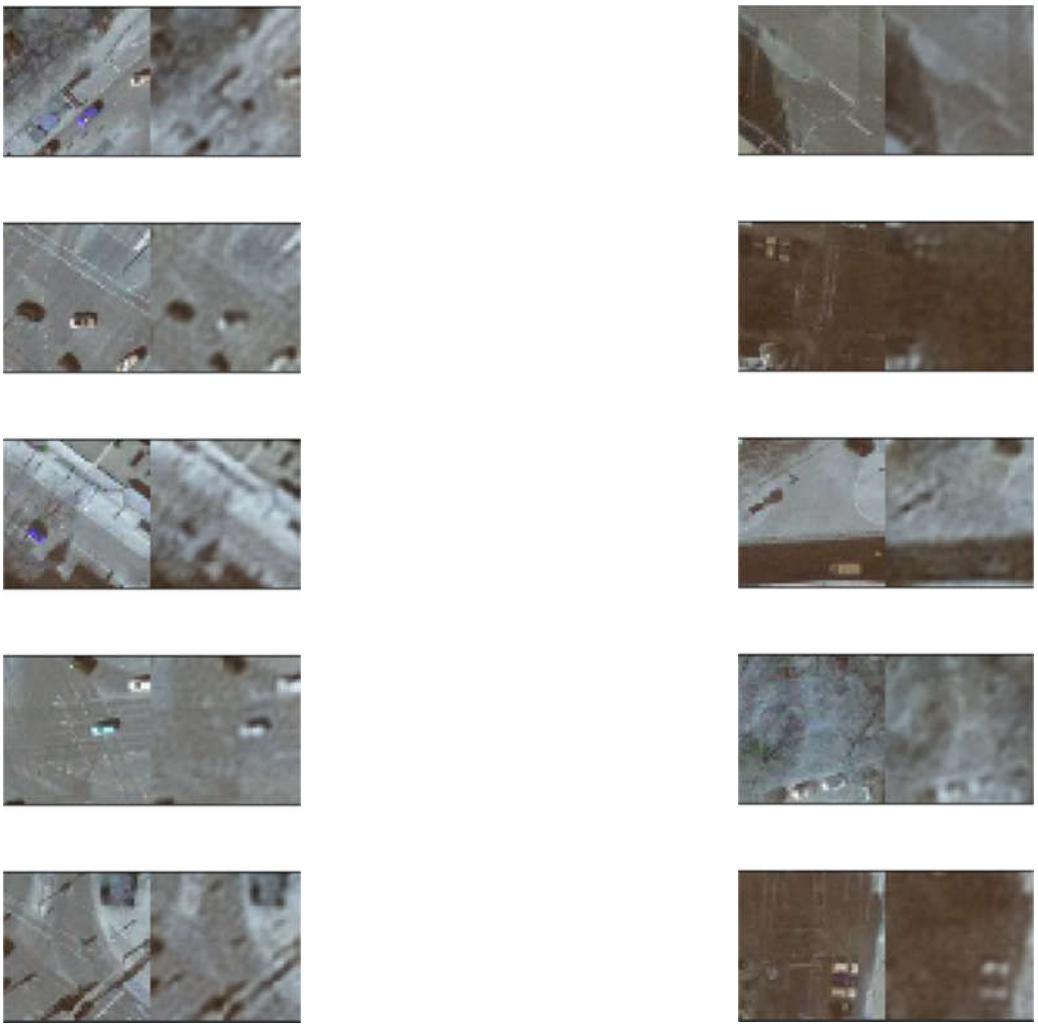

Figure 6. Examples of reconstructed images (left: training set, right: test set)

To accelerate the training process, the ADADELTA[6] optimizer was used. To illustrate the performance of the autoencoder, example results are shown in Figure 6.

The left column of the image pairs shows the input to the network, on the right, the reconstructed output is visualized. The left image pairs are examples from the training dataset, the image pairs on the right are from the test set, to illustrate the generalization capabilities of the network.

To show the estimated likelihood of the crash rate $p(c|x)$, the computed conditional distribution is compared to the actual distribution of the labeled data in Table 1. As can be seen in this table, the

VMR estimator tends to underestimate on test data, thus to predict lower crash rates on unfamiliar data.

To evaluate the performance of the VRM, it is sensible to take a look at the mean absolute error after training. On the training data, the mean regression error is 0.08. On the unfamiliar test data, the mean regression error is 1.91.

	crash rate (ground truth)	**crash rate estimation result**
mean train	1.45	1.47
std train	6.67	6.75
mean test	1.40	0.90
std test	4.23	1.60

Table 1. Mean and standard deviations of the estimated and ground truth distributions

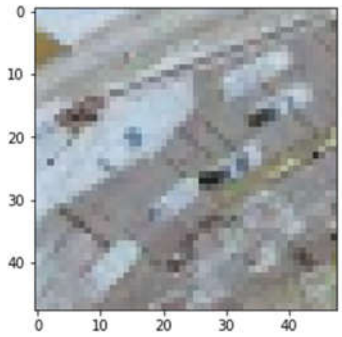

Figure 7. Image created from randomly sampled latent variable z

After the crash rates of the images were estimated, they can be visualized by the initially constructed VAE. Because of the dense representation of the data in the latent space, even artificial images can be created. An example of one decoded image from a randomly sampled latent variable is shown in Figure 7.

To classify the road sites into safe and unsafe sites, the data was divided into two labels. Road sites are considered safe, when there were no reported accidents within the year, the data was collected. All other road sites are considered unsafe. In this classification scenario, a simple fully connected neural network with one hidden layer, with four neurons was used to compute a logistic regression on the regression output of the VRM. In this simple classification test scenario, the proposed algorithm achieves an accuracy of 74.2 %.

RESULTS

A VAE can be trained on aerial image data to create a latent variable model on the images of road sites. The trained weights of the VAE can be used to initialize the encoder subnetwork of a VRM, which is then used to estimate crash rates, based on aerial images, as well as the coding in the latent space. Because of the constraints on the distribution in the latent space, induced by the combined loss functions, the trained machine learning algorithm generalizes well on unknown test data.

An advantage of the dense representation of the prior distribution is that artificial road sites can be sampled from the distribution in the latent space.

In a regression task, the trained network achieves a mean regression error of 1.91 at a standard deviation of the approximated distribution of 4.23. Used for classification, the algorithm achieved an accuracy of 74.2 %.

DISCUSSION

Latent variable models are a feasible way for crash rate estimation from image data. They provide possibilities for assisting authorities in planning scenarios for higher road safety levels. To tackle the problem that the crash rate of a road site is not only dependent on visual cues, that can be determined from aerial images, additional features, e. g. traffic density can be added to the training and inference processes, to learn common latent variables for these input features.

Another way to increase the performance of the proposed method can be the use of higher resolution images, like the ones used in [3], from Google Maps. For increasing generalization capabilities, data from different regions or cities should be added to the training dataset.

The proposed method appears to be a promising approach in using collected data in well generalizing machine learning models to assess road safety. It can be used as part of a software tool that could help practitioners to assess possible variants of proposed changes to the infrastructure. This requires rendering proposed changes into the aerial image.

Moreover, in principle, it can help practitioners by proposing changes to existing sites, because images of safer roads, which differ only slightly from the existing road, can be synthetized by a gradient descent on the latent space. While this is subject to future work, an example of a synthesized image is given in Figure 7.

REFERENCES

1 D P Kingma & M Welling, Auto-Encoding Variational Bayes, Amsterdam, 2014

2 D J Rezende & S Mohamed & D Wierstra, Stochastic Backpropagation and Approximate Inference in Deep Generative Models, London, 2014

3 AAAI-17 Proceedings - A Najjar & S Kaneko& Y Miyanaga, Combining Satellite Imagery and Open Data to Map Road Safety, Hokkaido, 2017

4 A Krizhevsky & I Sutskever & G E Hinton, ImageNet Classification with Deep Convolutional Neural Networks, Toronto, 2012

5 AAAI-16 Proceedings - Q Chen & X Song & H Yamada & R Shibasaki, Learning Deep Representation from Big and Heterogeneous Data for Traffic Accident Inference, Tokyo, 2016

6 M D Zeiler, ADADELTA: An Adaptive Learning Rate Method, New York, 2012

7 E Sacchi & T Sayed: Bayesian estimation of conflict-based safety performance functions, Journal of Transportation Safety & Security, 8:3, 266-279, DOI: 10.1080/19439962.2015.1030807, 2016

8 Highway Safety Manual, Volume 2. AASHTO, Washington, D.C., 2010.

9 B Brimley, M Saito & G. Schultz: Calibration of Highway Safety Manual safety performance function: development of new models for rural two-lane two-way highways. Transportation Research Record: Journal of the Transportation Research Board, (2279), 82-89. 2012

10 Geoportal Berlin / Digitale farbige Orthophotos 2016, http://fbinter.stadt-berlin.de/fb/index.jsp

11 Geoportal Berlin / Übergeordnetes Straßennetz Bestand, http://fbinter.stadt-berlin.de/fb/index.jsp

12 Crash database of the Berlin Police Department.Polizei Berlin, PPr St II 4 (2016)