Modeling and interpreting road geometry from a driver's perspective using variational autoencoders

Fan Wang[1], Yuren Chen*[1,2], Jasper S. Wijnands[3], Jingqiu Guo[1]

[1]Key Laboratory of Road and Traffic Engineering of the Ministry of Education，Tongji University, Shanghai,China

[2]Shanghai Institute of Intelligent Science and Technology, Tongji University, Shanghai,China

[3]Transport, Health and Urban Design Research Lab, Melbourne School of Design, The University of Melbourne,VIC, Australia

Abstract: *Quantitative description of perspective geometries is a challenging task due to the complexities of geometric shapes. In this paper, we address this gap by proposing a new methodology based on variational autoencoders (VAE) to derive low-dimensional and exploitable parameters of the perspective road geometry. Firstly, road perspective images were generated based on different alignment scenarios. Then, a VAE was built to create a regularized and exploitable latent space from the data. The latent space is a compressed representation of perspective geometry, from which six latent parameters were derived. Without prior expert knowledge, four of the latent parameters were found to represent distinctive attributes of the geometry, such as visual curvature, slope, sight distance and curve direction. The latent parameters provided quantitative measurements of how the design scheme looks like in perspective view. It was found that a road with low accident rate has low values for codes 4 and 5, high values for code 3, and low variance for codes 3 and 6. The trained VAE model also ensured accurate generation of the perspective images by decoding the latent parameters.*

*To whom correspondence should be addressed. E-mail: *lmatey@ceit.es.*

Overall, this research advances the understanding of road design by considering the driver's perception.

Keywords: geometry, driver view, variational autoencoder, road design

1 INTRODUCTION

Road geometry affects a driver's perception of the driving environment. A predictable and recognizable road layout supports driver expectations about the road type they are driving on and the safe driving behaviour that is expected of them (Stelling-Konczak et al., 2010; Janssen et al., 2006). In China, 68% of crashes with serious injuries between 2012 and 2016 were in mountainous areas, where the roads have complicated geometry and topographical conditions (Yu et al., 2018). Crash statistics on more than 2000 crash reports from California and Maryland showed that improper visual search prior to left turns causes a great proportion of accidents in younger and less experienced drivers (McKnight & McKnight, 2003). Driver-oriented infrastructure design has been found to reduce both occurrence and consequences of human error (Zheng et al., 2017; Mackie et al., 2013). For safety reasons, the visual perception of road infrastructure needs to be designed to meet the expectation that drivers have about the upcoming features (Čičković, 2016). The self-explaining road concept focuses on promoting safe behavior simply through design, where road infrastructure is designed in line with road users' expectations (Walker et al., 2013; Theeuwes & Godthelp, 1995).

The geometry of road layouts are closely related to driving behaviors, including speed, lane departure and crashes (Letty et al., 2009; Eustace et al., 2016; Banihashemi, 2016; Gargoum & El-Basyouny, 2016). For example, unexpected changes in road alignment lead to abrupt changes in driving speed (Yu et al., 2018; Donnell et al., 2009). The overlapping of horizontal and vertical curves provides more curvature information than the traditional highway design plan (Fildes & Triggs, 1982). Hassan et al. (2002) pointed out that overlapping crest curves make the horizontal curvature appear sharper, while overlapping sag curves make the horizontal curvature appear less sharp. In order to ensure appropriate visual representation of combined alignments, several qualitative design guidelines are available, such as the American Association of State Highway and Transportation Officials (AASHTO, 2018) and the Transportation Association of Canada (TAC, 2017). Geometric consistency is enhanced by uniformity and continuity in design (AASHTO, 2018; TAC, 2017).

Traditionally, engineers have utilized effective parameters to describe the road in 3D, such as horizontal curvature, vertical slope, etc., which are practical. These design parameters have also been studied to build relationships with road safety, where safety criteria are determined to guide road design. Driver perspective view is essential in road design. However, it is currently not possible to objectively examine the quality of a road from a driver's perspective, as the geometries are more irregular and complex. To fill this gap, it is important to determine how to describe the perspective road geometry objectively. In the paper, we propose using a machine leaning algorithm, variational autoencoder (VAE), to learn regularized latent codes from data. The codes are low-dimensional representations of perspective geometry. Meanwhile, the VAE model is able to generate perspective images from the latent codes.

2 LITERATURE REVIEW

2.1 Current design practice and limitations

Traditional geometric design practice involves independent design of horizontal and vertical alignments, where alignments are described by parameters such as horizontal curvature, vertical slope and vertical curve radius. Transportation authorities such as the American Association of State Highway and Transportation Officials (AASHTO) and the Transportation Association of Canada (TAC) provide general design guidelines that enhance the 3D representation of the combined alignments (AASHTO, 2018; TAC, 2017). One limitation is there is no specific regulation for quantitative coordination of combined alignments. Designers depend on their professional skills to visually interpret the roadway from the driver's point of view. There is no guarantee that the combined geometry from traditional design is a satisfactory and safe design (Gibreel et al., 1999). Poor superposition of horizontal and vertical alignments can jeopardize driver's information perception or cause hazardous driving behavior, e.g. sharp horizontal curvature with upward steep slope could lead to limited sight distance and abrupt speed changes (Wang et al., 2018).

In the past decades, the rapid development of road design software has enabled efficient generation of virtual road scenes from a driver's perspective (Autodesk, 2013; Eliseev et al., 2017). Real-time display of 3D geometry which matches traditional 2D design profiles is now a mature technology. Currently, many design agencies are using computer visualization to check the quality of combined alignments. It is also a convenient and intuitive way to show design schemes to others. However, the safety assessment for driver view geometry is mainly based on expert's experience. There is limited research quantifying the quality of the perspective geometry. For example, 3D sight distance is a substitute index to judge the safety of design from the driver's view (Jung et al., 2018; Ma et al., 2018). However, computation of this index is complex and information such as curve direction and sharpness is not taken into account, essential elements influencing driving performance. Since drivers derive their driving styles mainly from their perception of the road, it is important to consider the combined geometry from the driver's point of view in early stages of the design process.

2.2 Driver view geometry studies

It is important to extract features of the perspective alignments in order to quantitatively analyze their effects on driving behavior. Several models have been put forward to describe the geometry, including the straight line-parabola model (Jung & Kelber, 2005), the modified hyperbola model (Jung & Kelber, 2005), the cyclotron line model (Kang et al., 2010) and the Catmull-Rom spline model (Loose et al., 2009). These models fit the alignment using complex equations. It is unclear how the equation parameters are related to road design.

Some of the key papers were published by Yu et al (2016, 2018), where the drivers' visual lane was fitted with Catmull-Rom spline. The length and curvature of the visual curve were extracted as shape parameters. With these new geometric parameters, accident-prone locations on two-lane mountain highways were analyzed and safety reliability for vehicles turning right from urban major roads onto

minor ones were evaluated. These two shape parameters for visual curves are easy to understand. However, they are based on the researcher's decisions (cf. observer bias) and only represent limited attributes of the driver view geometry. Moreover, reconstructing the perspective view using the length and curvature parameters is not possible. Wang et al. (2019) studied middle-aged drivers' subjective categorization for combined alignment and found that drivers distinguish the road categories based on visual characteristics of alignment. Experiments revealed that drivers divide the visual mountainous alignment into three distinct and non-overlapping subjective categories. However, no quantitative descriptions of the visual geometry in each category were presented.

2.3 Variational autoencoder

Properly optimized feature extraction is the key to effective model construction (Guyon & Elisseeff, 2006). A common approach for unsupervised feature extraction is through autoencoders (AE), first introduced by Hinton et al. (2013). AE is a neural network architecture composed of an encoder and a decoder, creating a bottleneck in the middle. As data passes through the neural network, the process of encoding reduces the dimensions of input data, while decoding reconstructs the data from the compressed representation. The output of the encoder is also referred to as the latent space, containing a reduced number of new features extracted from the original data. Instead of using labels, input data themselves guide how to train the neural network, by minimizing the difference between the input and reconstructed data (reconstruction loss). Therefore, AE is an unsupervised machine learning algorithm.

VAE is an extension of AE where the latent space is trained to be regularized (Doersch, 2016). The term variational refers to the variational inference method in statistics, which is closely related to regularization. The output of VAE's encoder is enforced to follow a Gaussian distribution during training, by which way the latent space is regularized. A regularized latent space means it represents separate attributes that interact nonlinearly in the original data (Tran et al., 2017; Kempinska & Murcio, 2019; Krajewski et al., 2018). Therefore, this compressed feature vector is generally more interpretable. For example, the latent space of VAE trained on MNIST dataset can represent independent attributes of the characters, such as content, slope and width of the writing (Doersch, 2016). Moreover, the decoder is trained to take random points from latent distributions instead of direct single values to reconstruct the data. The decoder works in a similar way as a generator in Generative Adversarial Networks (GANs) (Radford et al., 2016).

Advanced methodologies based on machine learning enable understanding the inherent structure of complex and large-scale data. Convolutional neural networks have been used to solve complex computer vision tasks, such as feature extraction, image generation, object detection and classification (Krizhevsky et al., 2012; Rafiei & Adeli 2017a; Dai et al., 2019; Ma et al., 2019; Wijnands et al., 2019; Antoniades et al., 2018; Molina-Cabello et al., 2018; Hua et al., 2019; Zhao, Adeli et al., 2019; Zhao, Honnoral et al., 2019). Machine learning is increasingly used in civil engineering research (Rafiei & Adeli, 2016, 2017b, 2018).

It is difficult to fully model the driver view geometry based on traditional methods. To fill this gap, the paper proposes a new approach to model driver view geometry based on VAE. Instead of fitting

the alignments as curves, the neural network is designed to find a regularized set of latent parameters for the perspective road images directly. The main contributions of this paper are as follows:

- A new method to objectively analyze road geometry from a driver's perspective is developed.

- Complex perspective geometry is mapped in low-dimensional and meaningful codes, providing quantitative measurements of the road geometry.

- The new approach can generate realistic geometry, given specific latent codes.

3 METHODOLOGY

3.1 Perspective view mapping

Perspective mapping is a linear mapping, which generates realistic transformations in line with how the human eyes process images (Hearn & Baker, 1997). It enables transforming a point in the 3D coordinate system to a point in the 3D camera's (driver eye) coordinate system, and then projecting it to a point in the 2D image coordinate system. Let $P_c = [X_c\ Y_c\ Z_c]$ be a point on the road, $P_r = [X_r\ Y_r\ Z_r]$ be a point representing the camera, $P_c' = [u\ v\ 1]$ the same point of P_c in the projected 2D plane , M_δ the rotation matrix determined by the orientation ($\theta_{3\times3}$) of the camera and $K_{c,f}$ the intrinsic parameters matrix of the camera. Their combination maps a 3D point P_c into the perspective view P_c' as described in Eq. (1).

$$
\begin{aligned}
P_c' &= K_{c,f} \cdot M_\delta \cdot (P_r - P_c) \\
&= \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\theta_x) & \sin(\theta_x) \\ 0 & -\sin(\theta_x) & \cos(\theta_X) \end{bmatrix} \times \\
&\begin{bmatrix} \cos(\theta_y) & 0 & -\sin(\theta_y) \\ 0 & 1 & 0 \\ \sin(\theta_y) & 0 & \cos(\theta_y) \end{bmatrix} \begin{bmatrix} \cos(\theta_z) & \sin(\theta_z) & 0 \\ -\sin(\theta_z) & \cos(\theta_z) & 0 \\ 0 & 0 & 1 \end{bmatrix} \times \\
&\left(\begin{bmatrix} X_r \\ Y_r \\ Z_r \end{bmatrix} - \begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} \right)
\end{aligned}
\quad (1)
$$

To ensure the validity and generality of the perspective view, unified model assumptions that match driver vision were used. In the model, eye level was 1.08m above surface, positioned in the center of the driving lane. The aspect ratio was set as 4:3 (width:height) and the focal distance was 50mm with 40° horizontal and 27°vertical angle of aperture (Kühn & Jha, 2006). An illustration of perspective mapping is shown in Figure 1.
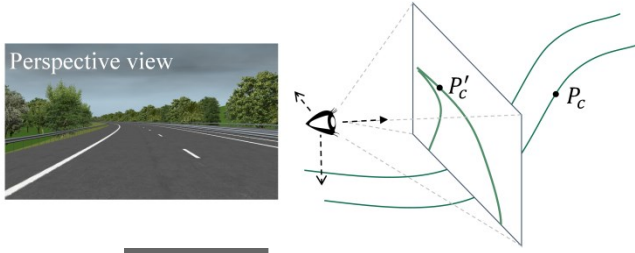
Figure 1 Perspective mapping
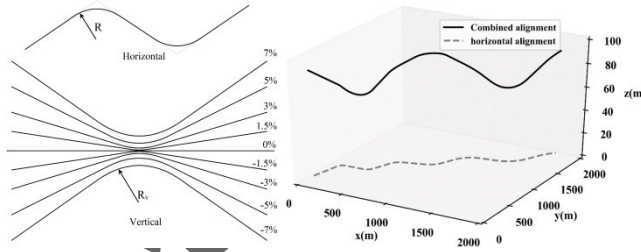
3.2 Road geometry data

In this study, the road geometry is produced according to the design specification (AASHTO, 2018). Therefore, the geometry used is not from a real road but still practical in engineering practices. Different geometries are taken into account. The minimum curve radius included is 60m, while the maximum curve radius is 1500m. 21 different kinds of curve radii are considered including tangent. Nine kinds of vertical slopes are included, ranging from -7% to 7%. A total of 100.92km road was designed with a cross-sectional width a of 3.5m. Geometric summaries are provided in Table 1. As shown in Figure 2, one horizontal curve is combined with nine different vertical slopes to produce different combined alignments

In order to map the geometry into perspective images, at the first step, 3D points were identified on the roads based on the design parameters. The points were selected with 10m intervals along the horizontal alignment. The hypothetical camera (driver eye) was set above each point with line of sight pointing to the direction of the route tangent (i.e., assumed driving direction). The 3D points in front the camera were transformed into perspective space using the mapping method described in section 3.1. The perspective points within vanishing points were kept. Then, the points were connected by straight segments to show continuous geometries. Each location corresponded to a distinctive perspective geometry. Therefore, sequences of realistic driver view images were obtained. To expand the number of samples, images were produced along both directions of the road and flipped horizontally (Shorten & Khoshgoftaar, 2019). The final dataset contains 30,000 samples of perspective images.

As the focus is on the road geometric shape, grayscale images were produced to reduce memory requirements. It was found that views of road geometry close to the driver were very similar across all images, while the geometries in the distance showed more variation. However, the perspective mapping has the effect that more distant objects appear smaller (Hearn & Baker, 1997). The original size of the images was 400×300 pixels. To focus on the variations in geometry, the central square part was cropped, resulting in an image size of 256×256. Each pixel is one of 256 different shades of gray with 0 representing black and 255 showing white. To train the neural network, pixel values were normalized into values between 0 and 1. Data processing procedures were implemented using python-3.7.3 (Python Software Foundation, 2019).

Modeling and interpreting road geometry from a driver's perspective using variational autoencoders

Table 1 Summary of road geometries

Parameters		
Horizontal radius (m)	Min	60
	Max	1500
Vertical slope	Min	-7%
	Max	7%
Length (km)		100.92
Cross-sectional width (m)		3.5

Figure 2 Road geometry combination

3.3 Variational autoencoder

The perspective geometries formed the inputs to the encoder function of a VAE (Figure 3). To minimize the loss between the original inputs and reconstructed ones, weights in all network layers are updated during training. The latent space of the VAE is enforced to follow Gaussian distributions. In practice, the encoder is trained to return the mean and the covariance matrix that describe these distributions. Each latent value is sampled from the distribution and fed into the decoder. The regularization is imposed by penalizing the difference between the latent distribution and a standard Gaussian distribution, measured by Kullback-Leibler (KL) divergence (Kullback, 1978). Both local and global regularization are ensured through the mean and variance of the distribution, independently. Generally, enforcing high regularization penalties leads to a better organized latent space but higher reconstruction errors. However, the tradeoff between KL divergence and the reconstruction error can be adjusted. An extension of VAE, $\beta-VAE$, incorporates a weight factor β in the loss function. A higher value of β emphasizes the KL divergence, leading to a more regularized latent space.

The network architecture of the VAE used is presented in

Table 2. In the encoder, 2D convolution layers encoded the images by reducing dimensionality and increasing feature depth, followed by a fully connected layer. As the output dimensions of the encoder are much smaller than the input dimensions, the encoder output forms a condensed representation of the input image. Specifically, the outputs of the encoder are multiple latent distributions, each described by a mean and log variance. These latent distributions capture the variations of extracted features in the input imagery. Latent codes were randomly sampled from these distributions. In the decoder, a fully connected layer and 2D transposed convolution layers reconstructed the images to their original size. Leaky ReLU activation functions were used in all layers, except the final layer which used the sigmoid function (Xu et al., 2015). Implicit paddings were used on both sides of the input for each convolution or transposed convolution layer.

The imagery used in this study has a size of 256×256 pixels with pixel values in the range [0, 1]. The reconstruction loss is calculated as the sum of squared errors between pixels in the original image and the reconstructed image. The customized loss function for the VAE is formulated as

$$\mathcal{L} = \sum_{i=1}^{256} \sum_{j=1}^{256} X_{i,j} + \beta D_{KL}(Q \| P) \quad (2)$$

Where $X_{i,j}$ is the squared error for pixel (i, j), Q is the latent distribution, P is the standard Gaussian distribution and $D_{KL}(Q \| P)$ denotes the KL divergence loss.

Table 2 Neural network architecture for VAE (see Figure 3 for an abstract graphical representation)

Layer	Encoder	Decoder
Hidden layer	Conv2d-(N8, K3, S2, P1), BN2(N8), LReLU	FC-(N1024), BN1(N1024), Leaky ReLU
	Conv2d-(N8, K3, S2), BN2(N8), Leaky ReLU	Reshape-(BS, 128, 4, 4)
	Conv2d-(N16, K3, S2), BN2(N16), Leaky ReLU	Deconv2d-(N128, K4, S2), BN2(128), Leaky ReLU
	Conv2d-(N32, K3, S2), BN2(N32), Leaky ReLU	Deconv2d-(N128, K4, S2), BN2(128), Leaky ReLU
	Conv2d-(N64, K3, S2), BN2(N64), Leaky ReLU	Deconv2d-(N128, K4, S2), BN2(128), Leaky ReLU
	Conv2d-(N128, K3, S2), BN2(N128)	Deconv2d-(N64, K4, S2), BN2(64), Leaky ReLU
	Leaky ReLU	Deconv2d-(N64, K4, S2),BN2(64), Leaky ReLU
	Reshape-(BS, -1)	Deconv2d-(N64, K4, S2), BN2(64), Leaky ReLU
Output layer	FC-(N (latent dimension)x2), BN1((latent dimension) x2), Leaky ReLU	Deconv2d-(N1, K3, S1), Sigmoid

Note: N-number of channels/neurons, BS-batch size, K-kernel size, S-stride, FC- fully connected layer, Conv2d-2D convolution layer, Deconv2d-2D Transposed convolution layer, BN2-2D batch normalization, BN1-1D batch normalization.
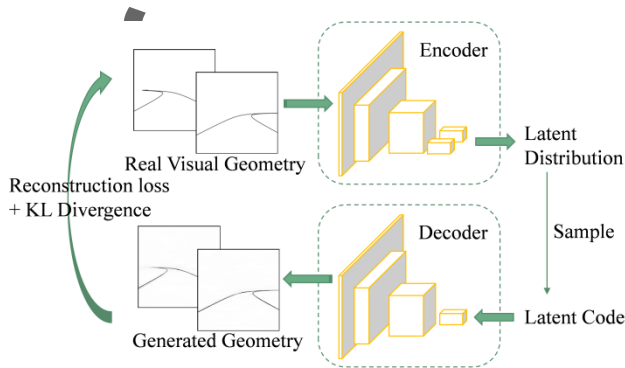
Figure 3 VAE architecture

The latent code is sampled using the following equation

$$c = \mu + \sigma \cdot e \ \ (3)$$

Where c is the latent code, μ and σ are the mean and standard deviation of the latent distribution and e is a random value from a normal distribution.

To avoid overfitting, L2 regularization was added into all weight matrixes with decay rate equal to 0.01. Batch normalizations were implemented to improve the stability of the neural network. The driver view images were randomly split into training and testing datasets with a ratio of 4:1. The VAE was trained with various latent dimensions for comparison. Further, several experiments were performed to set hyperparameter $\square$ of the VAE. With $\beta = 4$, a good balance between regularity in the latent space and reconstruction performance was achieved. We used Pytorch 1.2.0 to train the network on four P100 Nvidia GPUs with batch size of 32. Adam optimizer (Kingma & Ba, 2015) was used and the learning rate was 2×10^{-4}. The models were trained for 300 epochs, before which all models converged.

4 RESULTS

4.1 Network performance

The network performances for VAE models with different latent spaces are displayed in Figure 4. Testing losses are approximately equal to the training losses. The calibrated neural network showed good robustness, meaning it can be applied to new visual geometries not used for training. It is evident that latent dimension four is a turning point for the overall loss and reconstruction loss. As the latent dimension increases, the KL divergence increases and reconstruction loss decreases. The model loss with four or more latent codes is much smaller than that with three or fewer latent codes. Finally, the VAE model with six latent dimensions was chosen since it has the smallest overall loss. The average training and testing losses are 166.23 and 165.35, respectively. Reconstruction loss constitutes the largest part of these losses (i.e., 165.53 and 164.65).

The model's ability to accurately reconstruct driver view geometry is essential. Examples of the original images and their reconstructions are presented in Figure 5. The reconstruction losses were also compared for images with different numbers of visual curves, such as tangent, one curve and two curves. These categories were determined by human judgement. In each category, 80 images, containing different vertical slopes, were selected and average reconstruction losses were calculated. The tangent images have the lowest average reconstruction loss (87.63), while images showing two curves have the highest average reconstruction loss (212.51). Images with one curve have an average reconstruction loss of 163.42.
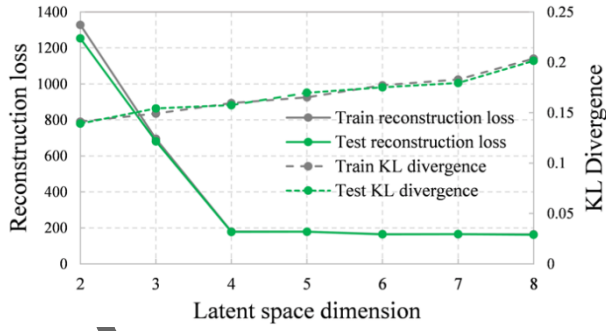
Figure 4 Network performance.

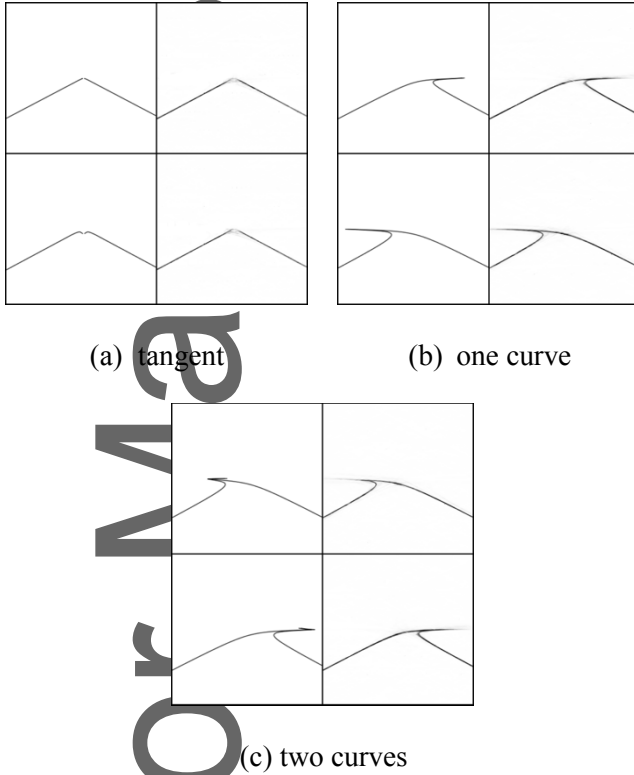(a) tangent   (b) one curve

(c) two curves

Figure 5 Visual geometries and the reconstructed
images in difference categories: a) tangent, b) one
curve, c) two curves (left: original, right:
reconstruction)

The values of latent codes are mainly between -2 and 2. As mentioned above, latent codes were random samples from the latent distributions. In feature extraction, it is more convenient to use a single value for each code than to utilize the distribution. Therefore, to map a complex image into latent codes, we randomly sampled 10 sets of latent codes from the latent distribution and chose the set with the smallest reconstruction loss as the latent codes. It should be noted that it is necessary to

use the latent distribution rather than direct latent code during training to maintain both the regularity of the latent space and the generative ability of the decoder.

4.2 Regularization of latent space

A regularized representation enables separation of attributes that interact nonlinearly in the original, high-level data (Tran et al., 2017). Regularity in latent space can also be visually presented, since changes in the latent codes result in variations in the decoded data. One specification for visually judging the regularity is: the higher the proportion of latent codes that can lead to distinct changes in the visual image, the better the regularity. When gradually changing the value of a latent parameter while keeping the other latent codes constant, a smooth variation in one (and only one) attribute is expected due to the nature of continuous latent representation.

Figure 6 shows the regularized representation of latent codes. A perspective image was encoded to six latent codes. One code at a time was replaced with a value ranging from -2 to 2, which were then decoded to perspective images. Most of the attributes can be expressed by four of the latent codes. Latent codes 1 and 2 do not have obvious impact on decoded perspective images. This is consistent with Figure 4, showing that the VAE models with four or more latent dimensions have similar loss values.

It is shown that variations on latent code 3 lead to different visual curvatures. When gradually decreasing code 3 from 2 to -2, the visual curvature increases. Similarly, latent code 4 captures variations in visual slope. In particular, the vertical slope increases when latent code 4 increases, resulting in reduced sight distance. Variations in latent code 5 also result in different visual curvature and sight distance, where the visual curvature increases as the latent code progresses from -2 to 2. It is found that latent code 6 only influences curve direction, where negative values correspond to left turns, positive values represent right turns and zero value associates with tangent.

Continuous variations in visual geometry are apparent with continuous changes in latent codes. Further, each code is correlated with a distinct attribute of road geometry. Therefore, these latent parameters are regularized representations of the driver view geometry. Note that these visual interpretations of curvature and slope are not necessarily the 2D horizontal curvature and vertical slope. They are based on the visual effect of combined horizontal and vertical alignments, reflecting what the geometry looks like from a driver's perspective.

The distributions of the latent codes can be used to create a wide variety of latent parameter combinations. Corresponding driver view geometries can be generated with the decoder, including geometries not present in the training dataset. This is especially useful for data augmentation.
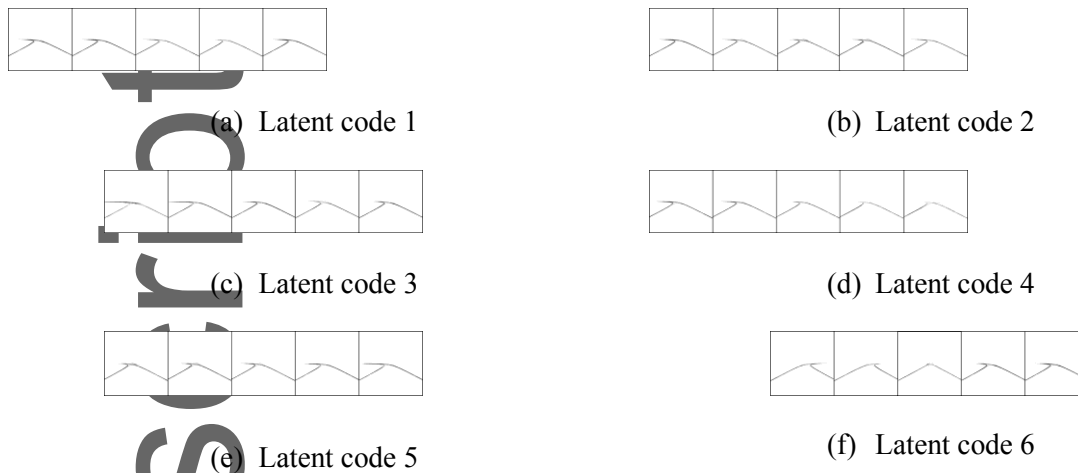
(a) Latent code 1

(b) Latent code 2

(c) Latent code 3

(d) Latent code 4

(e) Latent code 5

(f) Latent code 6

Figure 6 Regularized representation of latent codes. An original image was encoded into latent codes to form the basis, where codes 1 to 6 are [-0.738, -0.805, 0.701, 0.941, -0.729, 0.517]. For each set of five images, only one latent code is changed, and others are kept consistent with the original code vector. The varying latent code is assigned a value of -2, -1, 0, 1, 2 from left to right. For example, in (a) code 1, the first image corresponds to latent codes [-2.000, -0.805, 0.701, 0.941, -0.729, 0.517], while the second corresponds to [-1.000, -0.805, 0.701, 0.941, -0.729, 0.517]

4.3 Relations with traditional design parameters

The practical meanings of latent parameters are further explained by comparing them to traditional 2D parameters. The plots of latent codes against different horizontal curves and vertical slopes are shown in Figures 7 and 8, respectively. In Figure 7, the perspective images are taken at 100m before the midpoints of horizontal curves, then the images are encoded into latent codes by the VAE model. There are observed relations between horizontal curve radius and latent codes 3, 5 and 6. As the absolute value of curve radius decreases, latent code 3 decreases and latent code 5 increases. When the road curves left, shown as a negative radius, latent code 6 is positive. In contrast, right turn curves lead to negative values for latent code 6. The absolute value of latent code 6 for all visual curves is consistently larger than 0.5. There is no distinct relationship between horizontal curve radius and latent codes 1, 2 and 4. In Figure 8, the perspective images are taken on vertical slopes with consistent horizontal curve radius of 600m. Latent codes 3, 4 and 5 show obvious relations to vertical slope, while latent codes 1, 2 and 6 have no clear
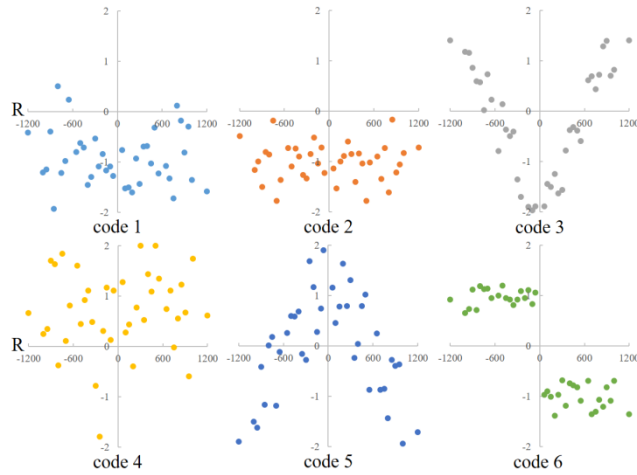
Modeling and interpreting road geometry from a driver's perspective using variational autoencoders

Figure 7 Latent code plots for different horizontal
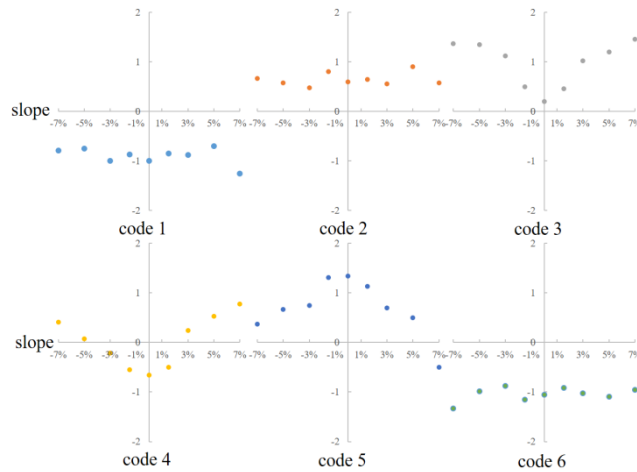curve radiuses with vertical slope equals zero

Figure 8 Latent code plots for different vertical slopes
with horizontal radius equals 600m

association with vertical slope. As the absolute value of vertical slope increases, latent codes 3 and 4
increase, while latent code 5 decreases.

The variations of latent codes against different horizontal and vertical design parameters are
consistent with the variations of perspective images against different latent codes. For example, when
the absolute value of the left turn curve radius increases, the value of latent code 3 increases (Figure
7). A large horizontal curve radius leads to smaller visual curvature in perspective images. In Figure
6, increasing values of latent code 3 result in a more moderate visual curvature. Hence, the latent
codes capture attributes of the perspective geometries that are related to traditional design parameters.

Plots of the latent codes as well as the curvature and altitude of a 5km mountainous highway are
shown in Figure 9. This road contains many curves, variations in slope and combinations of horizontal

and vertical curves. The latent codes represent parameters calculated in one direction from abscissa 0 to abscissa 5000.

The existence of visual curves is correctly reflected in latent code 6, where a large negative value indicates a right turn and a large positive value represents a left turn. The traditional curvature plot shows that 38% of the road consists of circular curves and 27% are transition curves. However, 91% of the absolute values of latent code 6 are larger than 0.5, where the latent code reveals more curvature information than the design parameter does. This is reasonable since the latent codes cover a section of road geometry and the horizontal curve is recognizable by drivers before a curve actually starts. Moreover, the altitude plot shows the slopes are stable before abscissa 2000, while latent codes 3, 4 and 5 show dramatic variations, especially when there are changes in horizontal curvature. 61% of latent code 3 values are negative and 54% of latent code 5 values are positive, meaning over half of the alignment show sharp visual curvatures. Horizontal curvature with slope could lead to limited sight distance (Wang et al., 2018), which is captured in these latent codes. The visual information predicted by latent codes is consistent with previous research, stating that superposition of horizontal and vertical alignments provides more curvature information (Fildes & Triggs, 1982).

A comparison of these latent codes between accident sites and non-accident sites was performed using a 40km two-lane divided mountainous highway in Zhejiang, China. 39 dangerous, crash-prone roads were determined in two driving directions according to historical crash records from 2011 to 2013. The mean length of crash-prone roads was 432m (SD=102.23m). In addition, 39 normal road sections, non-crash-prone roads, were randomly selected (length: M=432m, SD=102.23m). The mean and standard deviation of latent codes on each road were calculated. A t-test showed that the mean of latent code 3 on crash-prone roads are significantly lower than on normal roads ($p<0.01$). Further, the means of latent codes 4 and 5 are significantly higher than on normal roads ($p<0.05$). This indicates that crash-prone roads have a sharper visual curvature, steeper slope and shorter sight distance than normal roads. Furthermore, the standard deviation of latent codes 3 and 6 are significantly higher on crash-prone roads than on normal roads ($p<0.05$). Hence, crash-prone roads show more variation in perspective geometry. Meanwhile, the mean of horizontal curvature on crash-prone roads show significantly higher values than on normal roads ($p<0.05$). However, the standard deviation of horizontal curvature, and the mean and standard deviation of vertical slope show no significant difference ($p>0.05$). In summary, good perspective road quality is characterized by low values for codes 4 and 5, high values for code 3, and low variance for codes 3 and 6.

These analyses show that the latent parameters provide new means for road design assessment by considering a whole section of road from the driver's perspective. Also, the direct overlapping effects of horizontal and vertical alignments are captured in the parameters.
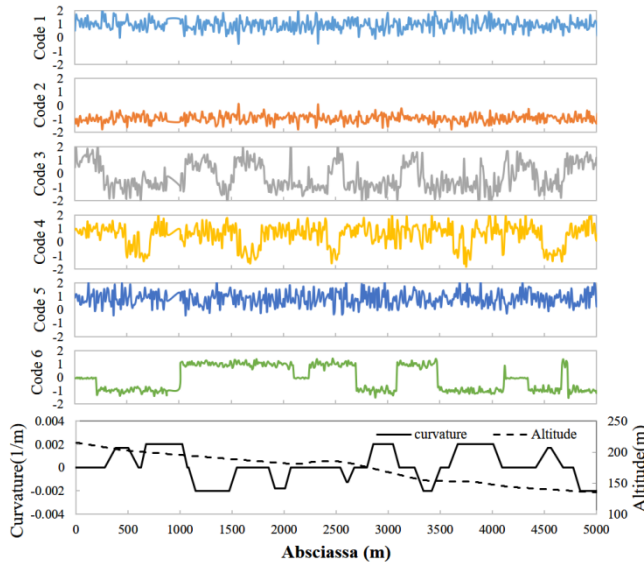
Modeling and interpreting road geometry from a driver's perspective using variational autoencoders

Figure 9 Latent parameters, curvature and altitude of a
5km mountainous highway in Zhejiang, China

5 DISCUSSION

By changing the dimension of the latent space, architecture of encoder and decoder, weight of KL
divergence and diversity of training dataset, various experiments were performed. Although the
results vary, slightly different settings of these hyper-parameters still lead to a disentangled,
exploitable latent space. Finally, the model with the best interpretation of neural network outputs was
present.

As shown in Figure 10, the road design parameters are used to produce perspective images. The
alignment used can be a design plan or an existing road. For existing road, numbers of studies have
been done on extracting design parameters using technologies such as GPS and Lidar (Holgado-Barco
et al., 2015; Ai et al., 2015; Rasdorf et al., 2012). Once the road geometry is obtained, perspective
road images can be produced using the methodology proposed. The VAE compressed the geometric
information stored in perspective images into low-dimensional space. No prior expert knowledge was
imposed when training the model. However, the latent parameters were found to contain visually
interpretable meanings. The compressed latent space purely represents the perspective geometric
information by six parameters. Our study has shown that such parameters are meaningful, especially
since no quantitative assessment for perspective geometry was previously available. The regularized
latent parameters correspond to a distinct visual characteristic, e.g. visual curvature, indication of

curve direction and sight distance. Engineers can obtain an objective assessment of the design scheme from a driver's perspective through these latent parameters in the early stages of design.
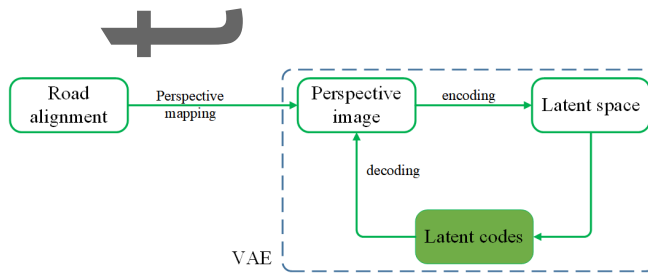

Figure 10 Flow chart for producing latent codes and generating new perspective image.

In addition to feature extraction through the encoder, the VAE also enables generating realistic imagery using the decoder. With the enhanced regularity in the latent space, a gradual change in each latent code will lead to a change in the corresponding geometric attribute, shown in Figure 6. Therefore, controlled generation of visual geometry is possible. Changing the primary design scheme by varying specific latent codes could lead to a series of design schemes for comparison, assisting with choosing the optimum design. The generation process is very quick with the trained VAE model. The change of road environment from complex to monotonous has a negative impact on distance estimation and reaction time (Liu & Wu, 2009; Zhao & Rong, 2012). Indications are that such bias can be applied to road safety assessment by identifying the occurrence of complex design.

The training dataset was created based on a driver eye height of 1.08m, which is the appropriate eye height for passenger vehicle drivers (AASHTO, 2018). For different road users, there are various eye heights, leading to different shapes of the perspective images at the same site. In road design practice, different types of road user, such as car drivers and truck drivers, are usually considered independently. Therefore, a distinct VAE model could be used for each type of driver, based on a dataset with specific eye height.

Studies on risk-compensation behaviors indicated that a good design condition, such as large curvature and small slope, is not always a safe design, since such design is associated with higher speeds and subsequently higher safety risks. In contrast, poor road conditions can lead to higher driving cautions and decreased speeds, thereby resulting in lower crash rates (Mannering, 2009; Labi, 2011, 2016). Overall, a satisfactory geometry is to deliver correct road information to guide safe driving behaviour. Driving behavior is continuously influenced by the perception of road geometry. The quality of perspective road can be assessed by the values and variances of latent parameters.

The training dataset was created based on a driver eye height of 1.08m, which is the appropriate eye height for passenger vehicle drivers (AASHTO, 2018). The eye height will vary for different road users, leading to different shapes of the perspective images at the same site. In road design practice, different types of road users, such as car and truck drivers, are usually considered independently.

Therefore, a distinct VAE model could be used for each type of driver, based on a dataset with specific eye height.

There are a few limitations in this research. Firstly, cross-sections and road environments are not considered. This could be explored in future research by adding more attributes to the images. Secondly, a challenge related to machine learning algorithms is to explain their results (Došilović et al., 2018). In this research, the obtained latent codes represent geometric features with physical meanings related to road design. However, more work is needed to interpret them from a road safety perspective. Finally, the generation quality of complex visual geometry (e.g., multiple curves) could be further improved.

6 CONCLUSIONS

Interpretation and modeling of road geometry from a driver's perspective were the main objectives of this study. Starting from traditional 2D design schemes, the alignment was mapped into perspective geometries. Then, a VAE model, an unsupervised machine learning algorithm, was built to extract low-dimensional features of the geometry as well as generate realistic samples with controlled attributes. A complete translation from traditional design parameters to latent parameters was obtained.

This paper utilizes machine learning algorithms to quantitatively model perspective road geometry. VAE was applied to directly extract latent features for perspective images. The derived six latent parameters are well regularized and exploitable, as they correlate to a combination of geometric attributes. The latent parameters also showed relations with road crashes.

The latent parameters provide quantitative measurements of the perspective view of a design scheme, in line with driver's visual perception. Indications are that they are valuable parameters to assist with road safety assessments. Overall, our research is useful for transportation engineers aiming to improve road design in the perspective of road users. In future research, in-depth analysis of the correlation between crashes, driving speed and the latent parameters could advance our understanding of safe road design.

ACKNOWLEDGMENTS

AUTHOR CONTRIBUTION

The authors confirm contribution to the paper as follows: study conception and design: Fan Wang, Yuren Chen; data collection: Fan Wang; analysis and interpretation of results: Fan Wang; manuscript preparation: Fan Wang, Yuren Chen, Jasper S. Wijnands, Jingqiu Guo. All authors reviewed the results and approved the final version of the manuscript.

CONFLICT OF INTEREST

The authors declare no potential conflict of interests.

REFERENCES

AASHTO. (2018). Policy on Geometric Design of Highways and Streets. Washington DC: American Association of State Highway and Transportation Officials.

Ai, C., & Tsai, Y. (2015). Automatic horizontal curve identification and measurement method using GPS data. *Journal of Transportation Engineering*, 141(2), 1–30

Antoniades, A., Spyrou, L., Martin-Lopez, D., Valentin, A., Alarcon, G., Sanei, S., & Took, C. C. (2018). Deep Neural Architectures for Mapping Scalp to Intracranial EEG. *International Journal of Neural Systems, 28*(8), 1–15.

Autodesk. (2013). Autodesk Introduces Breakthrough Technology for Infrastructure. *Business Wire (English)*. Autodesk Inc.

Banihashemi, M. (2016). Effect of horizontal curves on urban arterial crashes. *Accident Analysis and Prevention, 95,* 20–26.

Chen, X., Duan, Y., Houthooft, R., Schulman, J., Sutskever, I., & Abbeel, P. (2016). InfoGAN: Interpretable representation learning by information maximizing generative adversarial nets. *Advances in Neural Information Processing Systems*, 2172–2180.

Čičković, M. (2016). Influence of Human Behaviour on Geometric Road Design. In *Transportation Research Procedia,* 4363–4373.

Dai, X., Duan, Y., Hu, J., Liu, S., Hu, C., He, Y., … Meng, J. (2019). Near infrared nighttime road pedestrians recognition based on convolutional neural network. *Infrared Physics and Technology*, 97, 25–32.

Doersch, C. (2016). Tutorial on Variational Autoencoders. *ArXiv:1606.05908*.

Donnell, E. T., Himes, S. C., Mahoney, K. M., & Porter, R. J. (2009). Understanding Speed Concepts: Key Definitions and Case Study Examples. *Transportation Research Record: Journal of the Transportation Research Board, 2120,* 3–11.

Došilović, F. K., Brčić, M., & Hlupić, N. (2018). Explainable artificial intelligence: A survey. In *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 1–6.

Eliseev, M., Tomchinskaya, T., Lipenkov, A., & Blinov, A. (2017). Using 3D-modeling Technologies to Increase Road Safety. *Transportation Research Procedia*, *20*, 171–179.

Eustace, D., Almutairi, O., & Hovey, P. W. (2016). Modeling factors contributing to injury and fatality of run-off-road crashes in Ohio. *Advances in Transportation Studies*, *40*(B), 53–68.

Fildes, B. N., & Triggs, T. J. (1982). Effects of road curve geometry and approach distance on judgements of curve exit angle. In *Proceedings - Conference of the Australian Road Research Board* (Vol. 10, pp. 135–144).

Gargoum, S. A., & El-Basyouny, K. (2016). Exploring the association between speed and safety: A path analysis approach. *Accident Analysis and Prevention, 93,* 32–40.

Gibreel, G. M., Easa, S. M., Hassan, Y., & El-Dimeery, I. A. (1999). State of the art of highway geometric design consistency. *Journal of Transportation Engineering*, *125*, 305–313.

Guyon, I., & Elisseeff, A. (2006). An introduction to feature extraction. *Studies in Fuzziness and Soft Computing*, 1–25.

Hassan, Y., Sayed, T., & Bidulka, S. (2002). Influence of vertical alignment on horizontal curve perception. Phase II: Modeling perceived radius. *Transportation Research Record: Transportation Research Board,* 24–34.

Hearn, D., & Baker, M. (1997). *Computer Graphics, C Version*. Englewood Cliffs, Prentice Hall.

Hua, C., Wang, H., Lu, S., Liu, C., & Khalid, S. M. (2019). A Novel Method of Building Functional Brain Network Using Deep Learning Algorithm with Application in Proficiency Detection. *International Journal of Neural Systems, 29*(1), 1–17.

Janssen, W. H., Ridder, S. De, Brouwer, R. F. T., Thomson, R., Fagerlind, H., Lanner, G., … Figaredo, A. (2006). Roadside infrastructure for safer European roads: D02 Summary of driver behaviour and driver interactions with roadside infrastructure.

Jung, C. R., & Kelber, C. R. (2005). An improved linear-parabolic model for lane following and curve detection. In *Brazilian Symposium of Computer Graphics and Image Processing*, 2005, 131–138.

Jung, J., Olsen, M. J., Hurwitz, D. S., Kashani, A. G., & Buker, K. (2018). 3D virtual intersection sight distance analysis using lidar data. *Transportation Research Part C: Emerging Technologies, 86,* 563–579.

Kang, M. W., Yang, N., Schonfeld, P., & Jha, M. (2010). Bilevel highway route optimization. *Transportation Research Record: Transportation Research Board*, 107–117.

Kempinska, K., & Murcio, R. (2019). Modelling urban networks using Variational Autoencoders. *Applied Network Science*, *4*, Article number: 114.

Kingma, D. P., & Ba, J. L. (2015). Adam: A method for stochastic gradient descent. *ICLR: International Conference on Learning Representations*.

Krajewski, R., Moers, T., Nerger, D., & Eckstein, L. (2018). Data-Driven Maneuver Modeling using Generative Adversarial Networks and Variational Autoencoders for Safety Validation of Highly Automated Vehicles. *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, 2383–2390.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 1–9.

Kühn, W., & M. K. Jha. (2006). Using Visualization for the Design Process of Rural Roads. In *5th International Visualization in Transportation Symposium and Workshop*. Denver, Colorado: 5th International Visualization in Transportation Symposium and Workshop.

Kullback, S. (1978). *Information Theory and Statistics*. Dover Publications.

Labi, S. (2011). Efficacies of roadway safety improvements across functional subclasses of rural two-lane highways. *Journal of safety research*, 42(4), 231-239.

Labi, S. (2016). The tri-zone concept: Exploring the relationships between the intensity of safety features and safety performance. *International conference on transportation & development, Houston, TX, June*, 26-29.

Letty, A., Nicole, V. N., Fred C. M., W., Ingrid, V., & Robert, L. (2009). Safe Speeds and Credible Speed Limits (SaCredSpeed): New Vision for Decision Making on Speed Management. In *Proceedings of the 88rd Annual Meeting of the Transportation Research Board*. Washington DC.

Liu, Y. C., & Wu, T. J. (2009). Fatigued driver's driving behavior and cognitive task performance: Effects of road environments and road environment changes. *Safety Science*, *47*(8), 1083–1089.

Loose, H., Franke, U., & Stiller, C. (2009). Kalman particle filter for lane recognition on rural roads. In *IEEE Intelligent Vehicles Symposium, Proceedings*.

Ma, Y., Zheng, Y., Cheng, J., Guo, L., & Zhang, Y. (2018). A Novel Real-time Visualization Algorithm for Computing Three-dimensional Sight Distance. In *Transportation Research Board Annual Meeting*. Transportation Research Board.

Ma, Y., Zheng, Y., Cheng, J., Zhang, Y., & Han, W. (2019). A convolutional neural network method to improve efficiency and visualization in modeling driver's visual field on roads using MLS data. *Transportation Research Part C: Emerging Technologies*, *106* , 317-344.

Mackie, H. W., Charlton, S. G., Baas, P. H., & Villasenor, P. C. (2013). Road user behaviour changes following a self-explaining roads intervention. Accident Analysis and Prevention, 50, 742–750.

Mannering, F. (2009). An empirical analysis of driver perceptions of the relationship between speed limits and safety. *Transportation Research Part F: Traffic Psychology and Behaviour*, 12(2), 99–106.

McKnight, A. B., & McKnight, A. S. (2003). Young novice drivers: careless or clueless? *Accident Analysis and Prevention*, *35*(6), 921–925.

Molina-Cabello, M. A., Luque-Baena, R. M., López-Rubio, E., & Thurnhofer-Hemsi, K. (2018). Vehicle type detection by ensembles of convolutional neural networks operating on super resolved images. *Integrated Computer-Aided Engineering, 25*(4), 321–333.

Puente, I., Akinci, B., GonzÃąlez-Jorge, H., DÃŋaz-Vilar iÃśo, L., & Arias, P. (2016). A semi-automated method for extracting vertical clearance and cross sections in tunnels using mobile LiDAR data. *Tunnelling and underground space technology*, 59, 48–54.

Python Software Foundation. (2019). *Python release 3.7.3*. Retrieved 2020-05-20, from https://www.python .org/downloads/release/python-373/

Radford, A., Metz, L., & Chintala, S. (2016). Unsuper vised representation learning with deep convolutional generative adversarial networks. 4th Inter national Conference on Learning Representations, *ICLR 2016 - Conference Track Proceedings*, 1-16.

Rafiei, M. H., & Adeli, H. (2017). A novel machine learning-based algorithm to detect damage in high-rise building structures. *Structural Design of Tall and Special Buildings*, *26*, 18.

Rafiei, M. H., & Adeli, H. (2018). A novel unsupervised deep learning model for global and local health condition assessment of structures. *Engineering Structures*, *156*(1), 598–607.

Rafiei, M. H., & Adeli, H. (2016). A novel machine learning model for estimation of sale prices of real estate units. *Journal of Construction Engineering and Management, 144*, 1–12.

Rafiei, M. H., & Adeli, H. (2017). A New Neural Dynamic Classification Algorithm. *IEEE Transactions on Neural Networks and Learning Systems*, *28*(12), 3074–3083.

Rasdor f, W., Findley, D. J., Zegeer, C. V., Sundstrom, C. A., & Hummer, J. E. (2012). Evaluation of GIS applications for horizontal curve data collection. *Journal of Computing in Civil Engineering*, 26 (2), 191–203.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (2013). Learning Internal Representations by Error Propagation. *Readings in Cognitive Science: A Perspective from Psychology and Artificial Intelligence*.

Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on Image Data Augmentation for Deep Learning. *Journal of Big Data*.

Stelling-Konczak, A., Aarts, L., Duivenvoorden, K., & Goldenbeld, C. (2010). Supporting drivers in forming correct expectations about transitions between rural road categories. *Accident Analysis and Prevention, 43*(1), 101–111.

TAC. (2017). *Geometric Design Guide for Canadian Roads*. Ottawa, Ontario: The Transportation Association of Canada.

Theeuwes, J., & Godthelp, H. (1995). Self-explaining roads. *Safety Science*, *19*(2–3), 217–225.

Tran, L., Yin, X., & Liu, X. (2017). Disentangled representation learning GAN for pose-invariant face recognition. In *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*.

Walker, G. H., Stanton, N. A., & Chowdhury, I. (2013). Self Explaining Roads and situation awareness. *Safety Science*, *56*, 18–28.

Wang, F., Chen, Y., Guo, J., Yu, C., Stevenson, M., & Zhao, H. (2019). Middle-aged drivers' subjective categorization for combined alignments on mountainous freeways and their speed choices. *Accident Analysis & Prevention*, *127*, 80–86.

Wang, X., Guo, Q., & Tarko, A. P. (2018). Analyzing Speed on Mountainous Freeways with Longitudinal Data: A Driving Simulator Study. In *Transportation Research Board 97th Annual Meeting*. Washington DC, United States: Transportation Research Board 97th Annual Meeting.

WHO. (2018). *Global status report on road safety*. Geneva: World Health Organization

Modeling and interpreting road geometry from a driver's perspective using variational autoencoders

Wijnands, J. S., Nice, K. A., Thompson, J., Zhao, H., & Stevenson, M. (2019). Streetscape augmentation using generative adversarial networks: Insights related to health and wellbeing. *Sustainable Cities and Society*, *49*, 1–12.

Xu, B., Wang, N., Chen, T., & Li, M. (2015). Empirical evaluation of rectified activations in convolutional network. *arXiv:1505.00853,* 1–5.

Yu, B., Chen, Y., Bao, S., & Xu, D. (2018). Quantifying drivers' visual perception to analyze accident-prone locations on two-lane mountain highways. *Accident Analysis and Prevention, 119*, 122–130.

Yu, B., Chen, Y., Wang, R., & Dong, Y. (2016). Safety reliability evaluation when vehicles turn right from urban major roads onto minor ones based on driver's visual perception. *Accident Analysis and Prevention, 95,* 487–494.

Zhao, Q., Adeli, E., Honnorat, N., Leng, T., & Pohl, K. M. (2019). Variational autoencoder for regression: Application to brain aging analysis. International Conference on Medical Image Computing and Computer- Assisted Intervention, 823–831.

Zhao, Q., Honnorat, N., Adeli, E., & Pohl, K. M. (2019). Truncated gaussian-mixture variational AutoEncoder. arXiv:1902.03717.

Zhao, X., & Rong, J. (2012). The Relationship between Driver Fatigue and Monotonous Road Environment. *Computational Intelligence for Traffic and Mobility, Atlantis Computational Intelligence Systems, 8*, 19–36.

Zheng, Z., Du, Z., Yan, Q., Xiang, Q., & Chen, G. (2017). The impact of rhythm-based visual reference system in long highway tunnels. *Safety Science*, *95*, 75–82.

Author/s:
Wang, F;Chen, Y;Wijnands, JS;Guo, J

Title:
Modeling and interpreting road geometry from a driver's perspective using variational autoencoders

Date:
2020-10

Citation:
Wang, F., Chen, Y., Wijnands, J. S. & Guo, J. (2020). Modeling and interpreting road geometry from a driver's perspective using variational autoencoders. COMPUTER-AIDED CIVIL AND INFRASTRUCTURE ENGINEERING, 35 (10), pp.1148-1159. https://doi.org/10.1111/mice.12594.

Persistent Link:
http://hdl.handle.net/11343/276006