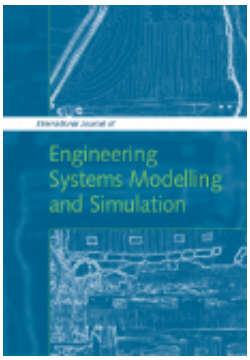



**International Journal of Engineering Systems Modelling and Simulation**

ISSN online: 1755-9766 - ISSN print: 1755-9758
https://www.inderscience.com/ijesms

# Image quality estimation based on visual perception using adversarial networks in autonomous vehicles

D. Vijendra Babu, A. Umasankar, K. Somasundaram, C.M. Velu, A. Sahaya Anselin Nisha, C. Karthikeyan

**DOI:** 10.1504/IJESMS.2022.10046831

**Article History:**
Received: 11 September 2021
Accepted: 28 January 2022
Published online: 01 December 2023

# Image quality estimation based on visual perception using adversarial networks in autonomous vehicles

## D. Vijendra Babu*

Department of Electronics and Communication Engineering,
Aarupadai Veedu Institute of Technology,
Vinayaka Mission's Research Foundation,
Deemed to be University,
Paiyanoor 603104 Tamil Nadu, India
Email: drdvijendrababu@gmail.com
*Corresponding author

## A. Umasankar

Department of Electrical and Electronics Engineering Technology,
Yanbu Industrial College,
Yanbu, Kingdom of Saudi Arabia
Email: umashankarhris1975@gmail.com

## K. Somasundaram

Institute of Information Technology,
Saveetha School of Engineering,
SIMATS, Chennai, India
Email: soms72@yahoo.com

## C.M. Velu

Saveetha School of Engineering,
Saveetha Institute of Medical and Technical Sciences,
Deemed to be University,
Chennai, Tamil Nadu 602105, India
Email: velucm.sse@saveetha.com

## A. Sahaya Anselin Nisha

Department of Electronics and Communication Engineering,
Sathyabama Institute of Science and Technology,
Deemed to be University,
Chennai 600199, Tamil Nadu, India
Email: anselinnisha.ece@sathyabama.ac.in

## C. Karthikeyan

Koneru Lakshmaiah Education Foundation,
Deemed to be University,
Vaddeswaram, Guntur,
Andhra Pradesh 522502, India
Email: ckarthik2k@gmail.com

**Abstract:** To improve autonomous cars, the dynamic systems method is re-enacted. Due to the unreality of the sensors employed in vehicles, human creation of the surrounding environment and objects is necessitated. We propose a novel efficient method for generating accurate scenario sensor data using limited LIDAR and video data from an autonomous vehicle. A new SurfelGAN network recreates realistic camera pictures to recognise the cars and moving objects in the scenario. The suggested approach uses real-world camera image data from Waymo Open Dataset to evaluate actual scenarios for autonomous vehicle movement. A new dataset allows for simultaneous analysis of two autonomous cars. This dataset is used to test and explain the proposed SurfelGAN model. GAN is the greatest technique for capturing realistic pictures. The

machine generates precise sensor data that is used to identify obstacles, cars, and other moving objects in the route of an autonomous vehicle. The autonomous car approaches the destination by recreating a surfel scene. Pictures are collected using semantic and instance segmentation masks.

**Keywords:** generative adversarial networks; GAN; visual perception; image quality assessment; IQA; autonomous vehicle; SurfelGAN.

**Reference** to this paper should be made as follows: Babu, D.V., Umasankar, A., Somasundaram, K., Velu, C.M., Nisha, A.S.A. and Karthikeyan, C. (2024) 'Image quality estimation based on visual perception using adversarial networks in autonomous vehicles', *Int. J. Engineering Systems Modelling and Simulation*, Vol. 15, No. 1, pp.37–46.

**Biographical notes:** D. Vijendra Babu obtained his PhD from the Jawaharlal Nehru Technological University, Hyderabad, India and he has 22 years of experience in the field of education, research and administration at various levels. He is currently designated as the Vice Principal and a Professor in the Department of Electronics and Communication Engineering at Aarupadai Veedu Institute of Technology, Vinayaka Mission's Research Foundation. He has obtained grant for three Australian patents, one Indian design patent and published five Indian patents.

A. Umasankar earned his BE from the Coimbatore Institute of Technology, Bharathiar University, Coimbatore, India, ME from the Karunya Institute of Technology, Bharathiar University, Coimbatore, India and PhD from the St. Peter's University, Chennai, India. He is currently working as a Lecturer (Grade 17/11) and technical committee coordinator in the Department of Electrical and Electronics Engineering Technology, at Yanbu Industrial College, Saudi Arabia.

K. Somasundaram is having about 25 years of experience in industry and teaching. He served in various positions in industry and teaching. He is currently serving as a Professor in the Institute of Information Technology, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai.

C.M. Velu is a Professor in the Department of Networking and obtained his MS in Systems and Information from BITS, PILANI in 1994 with CGPA 8.0/10.0. He obtained his ME (CSE) from the Sathyabama University, Chennai in 2007. He obtained his PhD in Artificial Intelligence using data visualisation techniques from the Anna University, Chennai in 2015.

A. Sahaya Anselin Nisha is working as an Associate Professor in the Department of Electronics and Communication Engineering, Sathyabama Institute of Science and Technology, Chennai since 2003. She has 18 years of teaching and research experience in the field of Microstrip Antenna, Microstrip Filter and Signal Processing. She is also an active MIE (corporate member) in the Institution of Engineers, India.

C. Karthikeyan has been awarded a PhD in Computer Science and Engineering from the Jawaharlal Nehru Technological University (JNTU) Hyderabad. He has more than 20 years of teaching experience in computer science and engineering background in India and abroad. He has published more than 36 papers in SCI/SCOPUS/WoS journals and book chapters.

## 1 Introduction

Autonomous vehicles (AVs) are expected to reduce traffic congestion by increasing throughput, enhance road safety by reducing human error, and relieve drivers from the stress of driving, allowing for increased efficiency and/or time for rest, among other things. Over the last three decades, technological activities in designing self-driving automobile technology have gradually increased, fuelled in part by developments in sensor and computing technologies that have culminated in smaller and cheaper hardware. In today's vehicles, automated vehicles are becoming more popular. Truly automated driving is already a difficult challenge, and there is a lot of ongoing testing underway to address technical issues. Perception is the first step, which is accomplished using a variety of sensors. Cameras, radar, ultrasonics, and lidar are examples of basic sensors that have already been deployed. Perception is the first level of processing, in which semantic objects such as lanes/vehicles and structural objects such as free space and standard obstacles are observed. They're then combined into a generalised abstract representation, which is usually a 2D or 3D map of objects oriented toward the self-vehicle.

This map is used by a driving policy model to determine the vehicle's trajectory. The standard approach is to construct the modules individually, but moment learning is now being attempted. Deep learning advancements have accelerated the maturation of autonomous driving applications. Deep learning models have established themselves as a baseline in perception and are increasingly

gaining steam in other modules such as fusion and driving strategy. Convolutional neural networks (CNNs) and recurrent neural networks (RNN) are two types of discriminative models used to solve grouping and regression problems. Discriminative models collect features that are necessary to solve the labelling problem but they rarely capture all of the data's content. On the other hand, generative prototypes attempt to capture the information distribution and thus form a more powerful depiction. This group includes generative adversarial networks (GAN), which have proven to be an important generative paradigm in a variety of domains and activities.

Unsupervised learning approaches with strong semantic object representation properties, such as generative models, are gaining popularity. The variational auto-encoder (VAE) (Kingma and Welling, 2014) is one of them, but it cannot provide images that are plain enough. The glow (Kingma and Dhariwal, 2018) is a flow-based generation method that has not gotten much attention yet. GANs has produced significant improvement in image analysis and are gaining growing educational and industrial interest. Image processing (Joo et al., 2018), inpainting of image (Iizuka et al., 2017), generation of the text (Nie et al., 2019), processing of medical image (Majurski et al., 2019), segmenting semantically (Zhu et al., 2019), colourisation of image (Nazeri et al., 2018; Liu et al., 2018), image-to-image conversion (Azadi et al., 2018), and generation of art (Azadi et al., 2018; Elgammal et al., 2017) are all examples of GAN techniques. GANs are still commonly used in facial reconstruction and editing, including face maturity (Yang et al., 2018) and gender transition (Wang et al., 2019).

At this time, image applications are being used in a wider variety of applications. A successful image quality assessment (IQA) criterion can track image efficiency refine algorithms, and direct parameter setting, also, to automatically evaluating their qualities. However, although many IQA parameters have been created, they still fall short of meeting the functional requirements. This is because quantitative IQA ratings cannot be perfectly compatible with subjective picture interpretation. As a result, it is critical to research an accurate and realistic IQA measure that is as close to subjective experience as possible (Wang et al., 2004). The illumination and colour detail is what humans interpret in photographs. The vision luminance and chroma will then be used as the initial details of the visual image. The local distinction and tolerance of HVS, on the other hand, affect the visual perception of the luma or chroma frequency of any point in the scene. The tolerance and the local typical contrast are all used as weights in the amount of the initial perception detail. The summed knowledge is referred to as the vision picture contents, and it is represented by a model.

Our work also has a strong link to the 3D restoration of outdoor settings. The standard methodology is to restore a dense 3D point cloud through image collections using a form from movement (Ullman, 1979; Wang et al., 2018) and multi-view stereo (Wu, 2013), then use Poisson restoration (Furukawa and Ponce, 2010) to produce a mesh depiction. This paradigm is well suited to situations in which we have several photographs covering a similar area from various angles, which is not usually the case in our situation. We will now provide precise depth knowledge to supplement the camera digital images thanks to the rapid development of LiDAR technology. This is used in our research to add fine-grained picture textures to the standard surfel (Kazhdan et al., 2006) portrayal, which not only eases the reconstruction method but also improves the portrayal of the colour natural world. Substitutes to the surfel include truncated signed distance functions (Pfister et al., 2000) and their more recent derivatives (Curless and Levoy, 1996). We have mentioned a recent paper (Park et al., 2019) that learns a point-wise dense description from a position cloud for the sake of rendering. While they have shown positive results, their method assumes a static atmosphere and is thus unsuitable for outdoor scenarios where a scene typically comprises tens of millions of points.

GANs (Aliev et al., 2019) have piqued the attention of academics and business alike. If Goodfellow et al. (2014) attempts to create realistic images, Isola et al. (2017) extend the concept to a more functional dependent image synthesis scenario. Following research (Zhu et al., 2017); great strides have been made in upgrading the quality of images produced by GAN methods. For more information on this topic, see Creswell et al. (2018) and Wang et al. (2018) used Cityscape to train a video synthesiser that can turn a video of a graphical representation into a video of images that are realistic (Sabarinath and Markkandan, 2015). Our method, on the other hand, employs labelled 3D training samples for many complex object types, making it more cost-effective. When ground-truth 3D boxes are swapped out for higher quality offline 3D perception pipelines, we think that this need may be reduced. Finally, we discuss the common issue of GAN assessment by introducing two new criteria that are useful for this purpose.

## 2 Related works

Adversarial networks are the most significant network model that efficiently generates the realistic samples in the numerous contemporaneous applications in which the images play a predominant role (Upasana et al., 2015). GANs is associated with the RNN for the enhancement of the sensor modelling for the detection of sequence generation. Though GAN can generate the sequential, it is utilised to obtain the desired output in language processing (Leonid and Jayaparvathy, 2021). Image generation produces the large continues true time series values rather than GANs model but still, the GAN achieves the continuous-time series by utilising the RNN network model based on GANs (Arnelid, 2018). Rather than the GAN model, a modified image generation method is utilised to promote the functioning of the audio waveforms (Donahue et al., 2018; ListoZec et al., 2018). GAN with the extended support of AIO-HMM can produce the time series value via the improvised RC-GAN model in the image domain.

The generator and classifier compete to improve their methods until the duplicate samples cannot be differentiated from genuine ones, based on a two-player maximum likelihood game (Goodfellow et al., 2014; Creswell et al., 2017). "The discriminator's ability to detect them improves in parallel with the quality of the samples generated throughout the practise session", says 36. To train a comparison classifier, the discriminator must extract characteristics, and the generator may have been any programme that can learn the array of training data. These algorithms include RNN, LSTM, and CNN. Commonly, de CNNs are employed as generators, where the generator must produce precise probabilities, and as an inverse operation of CNN to give detailed distributions. Combining GAN with other versions has seen steady progress in recent years, as we'll see in a moment. Besides enabling the training and analysis of embodied agents, the simulation model environment may provide a significant amount of data for teaching deep neural networks. For urban scenery segmentation, LeCun et al. (1989), Thangamani et al. (2020) and Narmatha et al. (2020) generates an enormous amount of labelled training data using a simulated environment. Improve object recognition by creating new pictures with dynamic objects.

Qualitative evaluations of images (IQA). Photo recovery, compression, reconstruction, and post-processing may impair the visual accuracy of photos. As a result of their application situations, IQA methods are categorised as full-reference or no-reference (NR-IQA). Image/video coding, reconstruction, and contact accuracy are frequently measured using FR-IQA methods because they evaluate similarities between two pictures from the perspective of information or perceived feature similarity. FR-IQA methods may be traced back to SSIM (Thangamani et al., 2021; Ganesh Babu et al., 2021; Sridevi et al., 2021; Zhang et al., 2014), which was the first to include structural information. Several FR-IQA techniques have been proposed to bridge this gap between IQA methods and human judgements (Zhang et al., 2018). Researchers are looking at IQA applications because of advanced data-driven methods that include other machine-vision problems. Other than those mentioned above, NR-IQA techniques have been proposed for evaluating image quality without the requirement for a reference picture. A popular NR-IQA technique is NIQE (Mittal et al., 2012; Blau and Michaeli, 2018) or PI. In many recent research, NR-IQA and FR-IQA methods have been combined to compute IR algorithms. IQA techniques have advanced, however IR approaches are still only tested using a few of IQA methods (e.g., PSNR/SSIM/PI).

Creating a new image quality evaluation dataset that incorporates the findings of GAN-based algorithms is the first obstacle to be overcome in this project. For example, in an IQA dataset, there are many skewed photos with human-annotated quality ranges. It may be used to compare the estimate accuracy of the IQA technique against human judgement. A fundamental computer vision problem, the IR team is tasked with restoring a high-quality picture from degraded data. For example: smelling, blurring and downward sampling are some of the usual degradation processes. Different types of picture deterioration need different types of IR assignments. When it comes to image SR and denoising, for example, image SR tries to recover high quality images from low resolution observations. For decades, a variety of IR algorithms have been suggested to maintain the status quo.

## 3 Proposed model

The proposed GAN model is utilised to improve performance of the AVs. The object in the road and other vehicles are analysed and predicted through the image sequence that is being fed as an input to the proposed system model. This image based on the visual perception can control both the speed and steering angle together in the AV and thus, the consequence arising during the travelling can be avoided. GAN is the best approach in extracting the visual features and has the potential to directly control the steering angle from raw pixels and controls the motion of the vehicle as per the surrounding environmental condition. The camera fixed at the front portion provides the image at high quality that enables us a clear vision of the road. The most essential parameter that needs to be determined is the speed of the AV to smoothly travel in the road and therefore, it is necessary to determine via visual perception instead of predicting directly. The AV performance on the road is controlled only through visual perception of the image quality obtained. The vehicle is controlled to slow down during the obstacle or when nearer to other vehicles via the analysis of visual perception.

The proposed GAN model enables autonomous driving by estimating the captured images based on visual perception. The enhancement of the AV initially includes data augmentation. GAN model is generally familiar for the image-to-image transition. GAN can convert the black and white images into a colour image and is has the stuff to produce realistic images even at various weather conditions. The realistic images are extensively utilised for the AV from the dataset; the dataset includes numerous pedestrian images. The image translation promotes the consistent movement of the vehicle using GAN. The discriminator's task described for the effective performance of the AV remains unaltered. The generator includes both the true and a false sample image which is being compared with the ground-truth image includes some amount of traditional loss $L_1$ in association with the objective loss in. The traditional loss in images leads to the image blurring.

$$lL_1 = E_{x,y\sim Pdata(x,y)z\sim} P_z(z)\left[\left\|y - G(x,y)\right\|_1\right] \tag{1}$$

$$G^* = \arg\min_G \max_D L_{CGAN}(G,D) + \lambda l_{L_1} \tag{2}$$

Here, the loss function $lL_1$ is included in determining the final objective. The result being obtained is much effective in enabling the performance of the AV.

GAN is also applicable for an unpaired image translation purpose and promotes the map learning $G: X \rightarrow Y$, $G(X)$ images face the difficulty in separating the images from the $Y$ distribution part. This type of mapping is not much effective and therefore the inverse mapping is introduced $F: Y \rightarrow X$ produces the consistent loss of $F(G(X)) \approx X$ and vice versa and this approach are known as CycleGAN. The CycleGAN performs the inverse mapping which is quite varying from $L_{GAN}$ is given below:

$$L_{GAN}(G, D_Y, X, Y) = E_{y \sim Pdata(y)}[\log D_Y(y)] + E_{x \sim Pdata(x)}\left[\log\left(1 - D_Y(G(x))\right)\right] \quad (3)$$

The consistency loss during the cyclic process is estimated as follow:

$$L_{cyc}(G, F) = E_{x \sim Pdata(x)}\left[\left\|F(G(x)) - x\right\|_1\right] + E_{y \sim Pdata(y)}\left[\left\|F(G(y)) - y\right\|_1\right] \quad (4)$$

The expression for the cycleGAN is given as follows,

$$L(G, F, D_X, D_Y) = L_{GAN}(G, D_Y, X, Y) + L_{GAN}(F, D_X, Y, X) + \lambda L_{cyc}(G, F) \quad (5)$$

The components involved in the estimation process are controlled by $\lambda$.

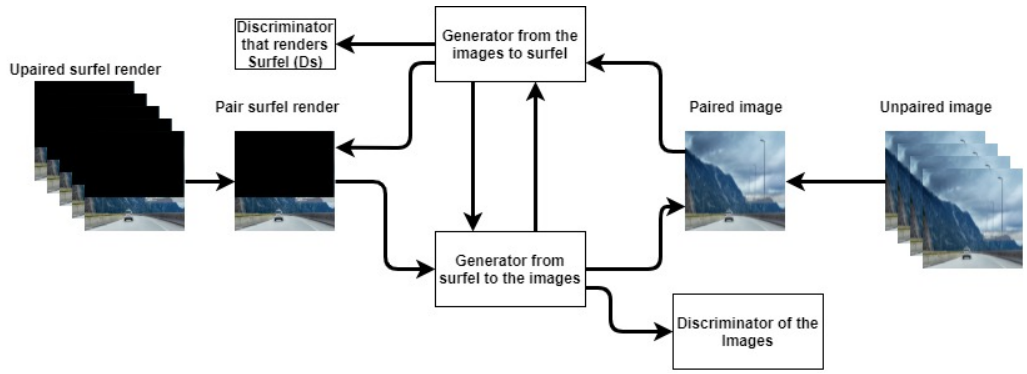
Surfel map: for example, in the GAN system, the generator and classifier fight to improve their methods until duplicate samples cannot be distinguished from the genuine ones (Goodfellow et al., 2014; Creswell et al., 2017). The discriminator's ability to detect them improves in line with the sample quality throughout the practise session. If the generator can learn the array of training data, the discriminator must extract characteristics and train a comparative classifier utilising those parameters. In addition to RNNs (LeCun et al., 1989; Thangamani et al., 2020; Narmatha et al., 2020), there are also LSTMs (Thangamani et al., 2020; Narmatha et al., 2020) and CNNs (Narmatha et al., 2020). It is common to employ de CNNs both as generators and as an inverse operation of CNN to produce detailed distributions. When coupled with other versions, GAN has progressed rapidly in recent years, as will be described briefly below. Additionally, the simulation model environment facilitates end-to-end learning and analysis for embodied entities by providing a huge amount of data for teaching deep neural networks. By simulating an urban environment, Thangamani et al. (2021) generates a large amount of fully labelled training data for urban scenery segmentation. to enhance item recognition accuracy, generate pictures with new locations of dynamic objects.

Rating the quality of your images (IQA). Photos that have been retrieved, compressed, reconstructed, or post-processed may be assessed for visual correctness using IQA techniques. Full-reference and no-reference methods to IQA are categorised depending on their application situations (NR-IQA). Many image/video coding, reconstruction, and contact accuracy measurements are conducted using FRA-IQA methods, which evaluate similarities in information or perceived feature similarity between two images from a distance. As an alternative to the widely used PSNR, FR-IQA methods may be traced back to SSIM (Sridevi et al., 2021), which was the first to incorporate structural information. Numerous FR-IQA techniques have been proposed to bridge the gap between the impacts of IQA methods and human choices (Zhang et al., 2014). Researchers are looking at IQA applications because of advanced data-driven methods, including other machine vision problems (Zhang et al., 2018). Beyond the above-discussed FR-IQA techniques to evaluate image quality, NR-IQA methods have been proposed to assess image quality without a reference picture. In addition to NIQE (Mittal et al., 2012) and PI (Blau and Michaeli, 2018), there are other NR-IQA techniques. To compute IR algorithms, many recent research combined NR-IQA and FR-IQA methods. To date, most IR techniques are tested using PSNR, SSSIM, and PI despite the advances in IQA methodologies.

Initial objective is to build a new image quality evaluation database that incorporates findings from techniques based on GAN. There are a significant number of distorted photos in the IQA dataset that have been annotated by humans with a variety of graphic quality levels on them. Use it to compare the IQA method's estimate accuracy against human judgement. Basic computer vision challenge: Restoring a picture of good quality after it has degraded. These are involved in processes such as sniffing, blurring, downward sampling, and other types of degradation. When an image is degraded, there are a variety of IR assignments that may be made. Picture SR, for example, tries to recover a high-resolution image from a low-resolution observation, whereas image denoising aims to eliminate undesirable noise from a high-resolution image. Many IR algorithms have been suggested in the last few decades to preserve the status quo. The radius of the surfel disk is $\sqrt{3}v$, in which $v$ denotes voxel size. The surfel hue is measured using the analogous colours from the camera image of the LiDAR points binned in a voxel. Traditional surfel maps suffer from a trade-off between geometric correctness and fine-grained features. A new method is thus used, with emphasis on geometric correctness and texture information. Based on each surfel disk's centroid, akk grids are created. Each grid centre's colour is used to encode reasonably large texture data. Figure 1 illustrates the training procedure for the SurfelGAN model's paradigm.

The surfel representation is improved by constructing a code generator of $k \times k$ grid for different 'n' distances, as each surfel can have a distinctive look across various frames depending on the change both in lighting conditions and relative pose. The relative pose is obtained at the various distance and view angles. The colour of each bin is evaluated initially is critical in obtaining a streamlined rendering image. Depending on the camera pose, a $k \times k$ patch is used during the rendering stage. The standard surfel map incorporates numerous objects at object boundaries and produces non-smooth colour in non-boundary regions. The textural surfel map is improved by reducing artefacts and thus, produces more vibrant images.

**Figure 1** Training of the SurfelGAN prototype (see online version for colours)




Synthesis: GAN-based data synthesis is used for a diverse range of autonomous driving applications. GAN performs visual perception by producing a supervised classification of input camera frames using data augmentation to allow better prediction over adverse climatic and illumination variations. Sensor correction, like resolving noisy inputs and sensor modelling, is also available. Synthesis can take place in two-and three-dimensional spaces, and spatiotemporal spaces like images and videos.

LiDAR is one of the most important sensors being used in autonomous driving based on its physical properties. LIDAR generates 3D point clouds to perceive precise depth irrespective of the lighting situation. The majority of GAN's methods Point Clouds are irrelevant to 3D point clouds, but still, SurfelGAN suggested a two-fold modification to the GAN learning algorithm to generating point clouds. Furthermore, the transformation of images into point clouds has been studied. Probabilistic mapping of low-dimensional domain to the space of 3D images is achieved by implementing 3D-GAN. The task of representing 3D shapes also produces 2D images of numerous artefacts taken from unidentified viewpoints. This method generates 3D images of comparable quality to train 3D data and permit the input image to obtain unsupervised generation views.

Although the surfel image reconstruction captures diverse environment visualisation resulting in surfel renderings with a significant realism gap while compared to real-world renderings. The imperfect reconstruction and poor symmetry and texturing images are overcome by implementing SurfelGAN. The proposed SurfelGAN includes a generator that converts surfel renderings created by the surfel scene reconstruction into realistic-looking images. The surfel rendering includes semantic and instance classification maps. The encoder-decoder model of the GAN generator is $G_{\theta_S}^{S\to I}$ with the trainable parameter $\theta_S$. The supervised loss is being used to train the generator with provided combinations of surfel renderings Sp and images Ip. SurfelGAN has only been trained through supervised learning. Furthermore, an adversarial loss determined via a real image discriminator $D_{\phi_I}^{I}$ along with the trainable variable $\phi_I$.

The inadequacy number of paired training data undergoes training in which the surfel rendering is compared to a true image. However, data that is not paired is problematic. The proposed systematic approach obtains unpaired data easily based on two main reasons: strengthening discriminator generalisation by training with more unlabelled samples and normalising the generator by maintaining loop continuity. The encoder-decoder model namely reverse generator $G_{\theta_I}^{I\to S}$ is like the framework $G_{\theta_S}^{S\to I}$ but with more output channels for both semantic and instance maps. Both paired $S_p$ and unpaired $S_u$ are being converted to a true image where the surfel rendering suffers from loop accuracy loss. This holds for every paired $I_p$ or unpaired $I_u$ true image. At last, we provide the surfel rendering discriminator $D_{\phi_S}^{S}$, which assesses the quality of created surfel images. SurfelGANs with increased period consistency is referred to as SurfelGAN-SACs. The following goal is optimised:

$$\begin{aligned}\max_{\phi_S,\phi_I}\min_{\phi_S,\phi_I} & L_r\left(G_{\theta_S}^{S\to I},S_p,I_p\right)+\lambda_1 L_r\left(G_{\theta_I}^{I\to S},I_p,S_p\right)\\ &+\lambda_2 L_a\left(G_{\theta_I}^{S\to I},D_{\phi_I}^{I},S_{p,u}\right)+\lambda_3 L_a\left(G_{\theta_I}^{I\to S},D_{\phi_S}^{I},I_{p,u}\right)\\ &+\lambda_4 L_c\left(G_{\theta_S}^{S\to I},G_{\theta_I}^{I\to S},S_{p,u}\right)+\lambda_5 L_c\left(G_{\theta_I}^{I\to S},G_{\theta_S}^{S\to I},I_{p,u}\right)\end{aligned} \tag{6}$$

Here, supervised reconstruction is denoted as $L_r$, the adversarial network is denoted as $L_a$, and cycle consistency loss is denoted as $L_c$.

The hinged Wasserstein loss is used for adversarial testing since it efficiently maintains the training. To display and reconstruct images, a cycle-consistency loss is determined using $l^1 - loss$ and to represent semantic and instance maps, a cross-entropy loss is evaluated.

Typically, AV technologies necessitate gathering and composing a large amount of training data. Depending on the virtual environments the data is collected and trained in virtual environments frequently struggle to categories the images in corresponding to the real-world situations. A deep learning model trained on samples from a source domain to a target domain is attained using domain adaptation. The proposed system model includes a GAN-based pixel-level domain adaptation approach that produces convincing samples and utilises well to object classes. Reinforcement learning for self-driving vehicles is developed in a simulated environment has been shown to work well in a real-world setting. The proposed system method enables the autonomous driving of a vehicle by performing the visual

perception from simulated to real-world captured images. This model for end-to-end driving is built by learning simulated and real photos and then together studied to control measures using labels from a simulated expert driver. The proposed framework is demonstrated by establishing a driving strategy that can then be applied to real-world scenarios.

Algorithm for the proposed system model:

Step 1 Start.

Step 2 For every epoch, sample the size $M$ {$x1$, $x2$, ….., $xm$} minibatch for every image in the set $X$ without having a replacement.

Step 3 For every epoch, sample the size $M$ {$y1$, $y2$, ….., $ym$} minibatch for every image in the set $Y$ without having a replacement.

Step 4 Using the $G$ incumbent synthesised $M$ of $Y$ images namely {$G(x1)$, $G(x2)$, ….., $G(xM)$}.

Step 5 Using the $F$ incumbent synthesised $M$ of $X$ images namely {$F(y1)$, $F(y2)$, ….., $F(yM)$}.

Step-6: Update the DY model for $DY(yi)$ = true and DY($G(xi)$) = False, for every $i$, $i$ = 1, 2, …, $M$.

Step 7 Update the DX model for $DX(xi)$ = true and $DX(F(yi))$ = False, for every $i$, $i$ = 1, 2, …, $M$.

Step 8 Update $F$ and $G$ model for $xi \approx F(G(xi))$ and $yi \approx G(F(yi))$ for every $i$, $i$ = 1, 2, …, $M$.

Step 9 Repeat the above steps until the image set becomes empty.

Step 10 Stop.

- Object detection: AV s are seen as static complex structures, and the proposed system model utilised Waymo open dataset's high-quality 3D bounding box metadata to combine the LiDAR points from different samples in recognising each object. The iterative closest point (ICP) algorithm is used to optimise point cloud classification, resulting in a dense point cloud that enables a more precise and improved surfel reconstruction for each vehicle. Our method does not necessitate 3D box groundtruth but still uses cutting-edge vehicle detection and tracking algorithm to obtain preliminary ICP estimates. This experiment, though, will be saved for future research. The rebuilt vehicle model is applicable at any position of choice when simulating our climate. The deformable objects are recreated using a surfel model for each LiDAR scan individually and the reconstructed pedestrian is placed anywhere within the scene of the LIDAR scan. The process of accurately reconstructing a deformable model from several scans will be left to future research.
- Super-resolution: the low-resolution sensors are used in the autonomous driving domain and obtain the corresponding high-resolution images will help to improve systems that have been trained on high-resolution data. Still, it is quite challenging to obtain a high-resolution representation from low-resolution images. SurfelGAN is an image super-resolution architecture that uses a GAN to infer images naturally. An adversarial loss for natural output and a content loss for perceptual similarity is proposed as part of a perceptual loss function.
- Distance weighted loss: due of the limited breadth of the surfel map, our depiction contains large portions of unknown areas. When compared to other regions, the degree of inequality there has risen dramatically. Apart from that, there is a lot of uncertainty since the surfel is so far away from the camera. A distance-weighted loss is used to stabilise our GAN training. Distance information is used to magnify reconstruction loss by creating a distance map during data pre-processing and using it as a weighting coefficient.
- Adversarial training and testing: the adversarial teaching was created, through which sample points were applied to training to strengthen the model. Adversarial training is also known as loss function learning, which means that we will use an adversarial loss to improve the final classifier's reliability. GAN structure is being used for loss embedding, which alleviates the fault of ill-posed mission formulation. Since there are some strict safety standards in AD, adversarial sample generation could be used to measure corner cases and reliability. There is an automatic research mechanism for autonomous driving that uses deep learning, although it does not use GAN. The Adam optimiser was used to train the proposed framework model. Both the generator and the discriminator are provided with an initial learning rate to $2e^{-4}$ and followed by Relu activation, batch normalisation is utilised. In this analysis, the value of $\lambda_1$, $\lambda_2$, $\lambda_3$, $\lambda_4$, $\lambda_5$ are assigned and trained depending on one Nvidia Titan V100 GPU.

## 4 Result and discussion

The majority of the tests are carried out on three different versions of our proposed model. The surfel image-to-image model is trained by organised methodology (S) to minimise the $l^1$ – *los*s in the real-world picture. This method of training necessitates the use of paired results. As a result, only WOD-TRAIN can be used to train. WOD-TRAIN is utilised to train data in Supervised + Adversarial (SA). The adversarial loss is represented as $l^1$ – *los*s. The existing

camera image dataset in Supervised + Adversarial + Cycle (SAC) is trained using WOD-TRAIN. They are not paired. In comparison to the adversarial loss, we consider using a cycle-consistency loss as shown in Figure 2. The direct surfel renderings that are fed as an input to the proposed model are the baseline for the enhancement of the AV.

The downstream module is involved in analysing the primary simulation data acquired from camera data is being evaluated. The AV detector does on the generated images is independent of fine-tuning and the detector perceives the generated images similarly to actual images. The captured images are resized to 512 × 512 resolution pixels undergo the training and testing process in the proposed network which further includes a ResNet architecture and an SSD detection as a vehicle detector. WOD-TRAIN is utilised to train the datasets. The accuracy of the detector is quantitatively compared to the original surfel renderings in which the proposed system is fed with input images from SurfelGAN variants.

**Figure 2** Comparison of the various SurfelGAN with respect to Waymo open dataset training novel view (see online version for colours)

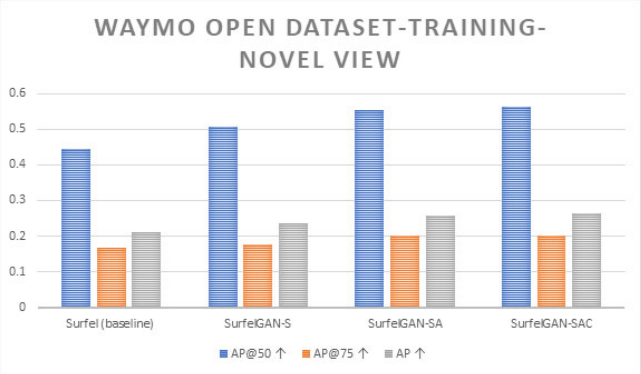
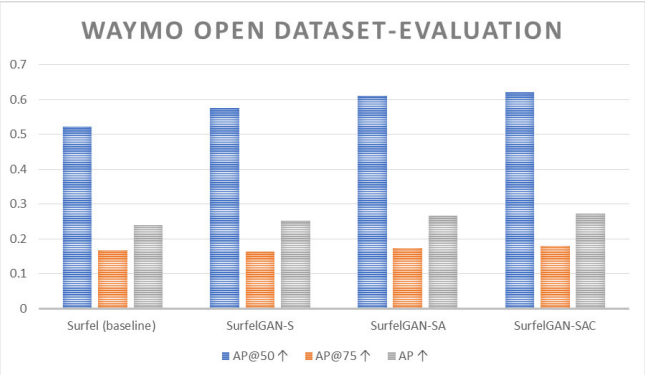



**Figure 3** Comparison of the various SurfelGAN with respect to Waymo open dataset evaluation (see online version for colours)



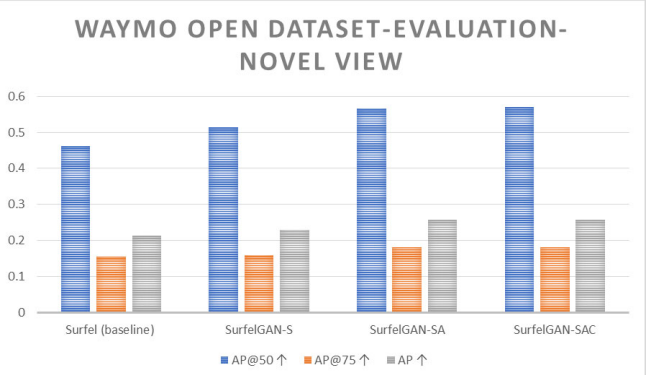

The texture surfel image reconstruction is developed and trained in the WOD-EVAL set achieves the accuracy precision of 51.5% at AP@50 corresponding involved in decent detection accuracy for the surfel images in Figure 3. SurfelGAN's suggested SurfelGAN work has a 62.8% discrepancy between these surfel renderings and actual pictures. SurfelGAN-S, SurfelGAN-SA, and SurfelGAN-SAC variations continue to improve upon surfel representations. To compare, SurfelGAN-SAC increases the AP@50 measure from 51.5% to 63.0% on WOD-EVAL (Figure 4). SurfelGAN-SAC representations are comparable to actual pictures in the detector's field of view, which was the primary motivation for this study. There are two distinct generalisation techniques included in the SurfelGAN model.

$$d\left((t,R),(t',R')\right) = \|t-t'\| + \lambda_R \frac{\left\|\log\left(R^T R'\right)\right\|}{\sqrt{2}} \tag{7}$$

Here, translational pose and rotational pose are denoted as $t$ and $R$ in WOD-EVAL-NV. In WOD-EVAL, $t'$ and $R'$ is the translational and rotational pose. $\lambda_R$ value is set as 1.

The surfel renderings in terms of rendering direction acquire a better quality bias is deviating too far from the actual poses. The surfel scene is reconstructed from different various surfel scene runs with higher precision, this issue will be solved. The proposed AV is moving on the road is controlled with the exploration of direction. According to our planned SurfelGAN work, the gap between these surfel renderings and real pictures is 62.8%. Variants of SurfelGAN-S, SurfelGANSA, and SurfelGANSAC continue to improve upon the basic representations of SurfelGAN. As demonstrated in Figure 4, SurfelGAN-SAC increases the WOD-EVAL AP@50 measure from 51.5% to 62.8%. SurfelGAN-SAC representations are comparable to actual pictures in the detector's field of view, which was the primary motivation for this study. In the suggested SurfelGAN model, the two distinct generalisation techniques are combined.

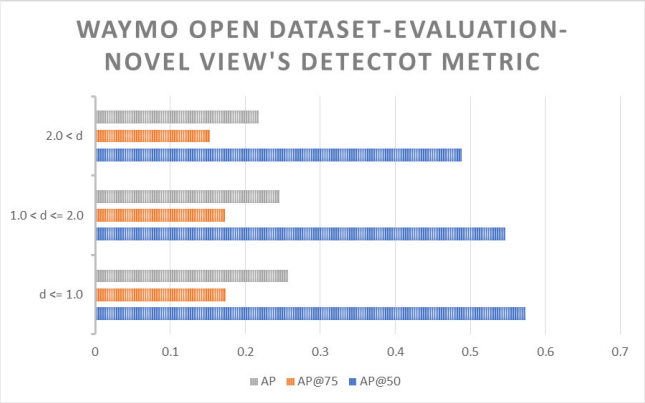
**Figure 4** Comparison of the various SurfelGAN with respect to Waymo open dataset evaluation novel view (see online version for colours)




Data augmentation generates the data that undergoes the training process to detect the object in the pathway of the AV. The baseline used WOD-TRAIN to train a vehicle detector and WOD-EVAL to assess its consistency. WOD-TRAIN and surfel images produced by WOD-TRAIN-NV are used to train and examine another vehicle detector and even WOD-EVAL can also be utilised for determining the vehicle in the pathway. WOD-TRAIN-NV mainly acquires WOD-3D bounding boxes rather than fitting 2D bounding boxes during the training process. The surfels in the 3D bounding boxes are

shown into a 2D novel image using the axis-aligned bounding box. The data augmentation improved the detector's overall accuracy metric greatly, raising the AP@50 score from 20.5% to 24.3%, the AP@75 score from 9.9% to 11.8%, and the average AP from 10.5% to 12.4%. The key cause of the disparity is that images are resized significantly by using an off-the-shelf detector and direct training is provided to the surfel renderings. A major difference is encountered in augmenting data with SurfelGAN synthesised images, demonstrating the SurfelGAN model's realism.

**Figure 5** Waymo open dataset evaluation novel view's detector metric break at various levels of perturbation (see online version for colours)




## 5 Conclusions

In this paper, image quality estimation based on visual perception using adversarial networks in AVs is enhanced. SurleyGAN is an efficient data-driven method is utilised in autonomous driving simulations by extracting captured images. The captured images in the 3D model and LIDAR data are acquired for the development of surfel map representation. The vehicle passing by the AV, objects, and other moving objects are detected by the captured images and sensor data enable us to visually analyse the position of the AV in the pathway to the target termination point. The proposed SurfelGAN image synthesis helps in reconstructing or rendering artefacts based on the various points of view in the environment. Experimental results showed that our synthesised sensor data not only has a high degree of realism but that is used to supplement training datasets for deep neural networks. The image extraction can be improved further in the future via the development of a dynamic approach to detecting an object. Thus, the AV is controlled via the visual perception of captured images by implementing a novel SurfelGAN network to identify the vehicles and moving object locations and orientations in the environmental scenario.

## References

Aliev, K-A., Ulyanov, D. and Lempitsky, V. (2019) *Neural Point-Based Graphics*, arXiv preprint arXiv: 1906.08240.

Arnelid, H. (2018) *Sensor Modelling with Recurrent Conditional Gans*, Master's thesis, Chalmers University of Technology.

Azadi, S., Fisher, M., Kim, V., Wang, Z., Shechtman, E. and Darrell, T. (2018) 'Multi-content GAN for few-shot font style transfer', in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, June, pp.7564–7573.

Blau, Y. and Michaeli, T. (2018) 'The perception-distortion tradeoff'', in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.6228–6237.

Creswell, A., White, T., Dumoulin, V., Arulkumaran, K., Sengupta, B. and Bharath, A.A. (2018) 'Generative adversarial networks: an overview', *IEEE Signal Processing Magazine*.

Creswell, T., White, V., Dumoulin, K., Arulkumaran, B., Sengupta, A. and Bharath, A. (2017) 'Generative adversarial networks: an overview', *IEEE Signal Process*., January, Vol. 35, No. 1, pp.53–65.

Curless, B. and Levoy, M. (1996) 'A volumetric method for building complex models from range images', in *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH '96*.

Donahue, C. McAuley, J. and Puckette, M. (2018) 'Synthesizing audio with generative adversarial networks', *CoRR*, Vol. abs/1802.04208.

Elgammal, A., Liu, B., Elhoseiny, M. and Mazzone, M. (2017) *CAN: Creative Adversarial Networks, Generating 'Art' by Learning About Styles and Deviating from Style Norms*, arXiv: 1706.07068 [online] https://arxiv.org/abs/1706.07068 (accessed 21 June 2017; 18:05:13 UTC).

Furukawa, Y. and Ponce, J. (2010) 'Accurate, dense, and robust multi-view stereopsis', *TPAMI*.

Ganesh Babu, O., Danya, T.R., Sakthieswaran, N., Suresh, P. and Muthuraman, U. (2021) 'Sustainable characteristics of fly ash based geopolymer concrete incorporating alccofine, zeolite and rubber fibers', *Revista Română de Materiale/Romanian Journal of Materials*, Vol. 51, No. 1, pp.17–24.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. and Bengio, Y. (2014) 'Generative adversarial nets', in *NeurIPS*.

Goodfellow, J. et al. (2014) 'Generative adversarial nets', in *Proc. Int. Conf. Neural Inf. Process. Syst.,* pp.2672–2680.

Iizuka, S., Simo-Serra, E. and Ishikawa, H. (2017) 'Globally and locally consistent image completion', *ACM Trans. Graph.*, July, Vol. 36, No. 4, pp.1–14.

Isola, P., Zhu, J-Y., Zhou, T. and Efros, A.A. (2017) 'Image-to-image translation with conditional adversarial networks', in *CVPR*.

Joo, D., Kim, D. and Kim, J. (2018) 'Generating a fusion image: one's identity and another's shape', in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, June, pp.1635–1643.

Kazhdan, M., Bolitho, M. and Hoppe, H. (2006) 'Poisson surface reconstruction', in *Proceedings of the Fourth Eurographics Symposium on Geometry Processing*, Vol. 7.

Kingma, D.P. and Dhariwal, P. (2018) 'Glow: generative flow with invertible 1×1 convolutions', in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, pp.236–245.

Kingma, D.P. and Welling, M. (2014) 'Auto-encoding variational Bayes', in *Proc. Int. Conf. Learn. Represent. (ICLR)*, pp.1–14 [online] https://arxiv.org/abs/1312.6114 (accessed 1 May 2014).

LeCun, Y. et al. (1989) 'Backpropagation applied to handwritten zip code recognition', *Neural Comput.*, Vol. 1, No. 4, pp.541–551.

Leonid, T.T. and Jayaparvathy, R. (2021) 'Statistical-model based voice activity identification for human-elephant conflict mitigation', *J. Ambient Intell. Human Comput.*, Vol. 12, pp.5269–5275, https://doi.org/10.1007/s12652-020-02005-y.

ListoZec, E., Mohammadiha, N. and Schliep, A. (2018) 'Statistical sensor modelling for autonomous driving using autoregressive input-output HMMS', in the *21st IEEE International Conference on Intelligent Transportation Systems*.

Liu, Y., Qin, Z., Wan, T. and Luo, Z. (2018) 'Auto-painter: cartoon image generation from sketch by using conditional Wasserstein generative adversarial networks', *Neurocomputing*, October, Vol. 311, No. 9, pp.78–87.

Majurski, M., Manescu, P., Padi, S., Schaub, N.J., Hotaling, N.A., Simon, G.C. and Bajcsy, P. (2019) 'Cell image segmentation using generative adversarial networks, transfer learning, and augmentations', in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp.1–22.

Mittal, A., Soundararajan, R. and Bovik, A.C. (2012b) 'Making a 'completely blind' image quality analyzer', *IEEE Signal Processing Letters*, Vol. 20, No. 3, pp.209–212.

Narmatha, C., Thangamani, M. and Jafar Ali Ibrahim, S. (2020) 'Research scenario of medical data mining using fuzzy and graph theory', *International Journal of Advanced Trends in Computer Science and Engineering*, February, Vol. 9, No. 1, pp.349–355.

Nazeri, K., Ng, E. and Ebrahimi, M. (2018) *Image Colorization with Generative Adversarial Networks*, arXiv: 1803.05400 [online] http://arxiv.org/abs/1803.05400 (accessed 16 May 2018).

Nie, W., Narodytska, N. and Patel, A. (2019) 'Relgan: Relational generative adversarial networks for text generation', in *Proc. Int. Conf. Learn. Represent. (ICLR)*, August, pp.1–20 [online] https://openreview.net/forum?id=rJedV3R5tm.

Park, J.J., Florence, P., Straub, J., Newcombe, R. and Lovegrove, S. (2019) 'Deepsdf: learning continuous signed distance functions for shape representation', in *CVPR*.

Pfister, H., Zwicker, M., Van Baar, J. and Gross, M. (2000) 'Surfels: surface elements as rendering primitives', in *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*.

Sabarinath, R. and Markkandan, S. (2015) 'Complexity analysis and performance comparison of channel decomposition methods used in MIMO precoder design', *International Journal of Applied Engineering Research (IJAER)*, Vol. 10, No. 72, pp.332–335.

Sridevi, R., Kumar, C., Suresh, P., Ganesan, S., Subramanian, S. and Suresh, A. (2021) 'Development of a pedagogical framework to analyze the performance of induction machines', *International Journal of Electrical Engineering and Education*, Vol. 58, No. 2, pp.501–516.

Thangamani, M., Ganthimathi, M., Sridhar, S.R., Akila, M., Keerthana, R. and Ramesh, P.S. (2020) 'Prediction of dengue fever using intelligent classifier', *International Journal of Emerging Trends in Engineering Research*, April, Vol. 8, No. 4, ISSN: 2347–3983.

Thangamani, M., Kavitha Bharathi, S. and Suresh Kumar, N. (2021) 'Exploring nano biotechnology for detecting specific disease in medical diagnosis and therapeutic', *OP Conference Series: Materials Science and Engineering*, Vol. 1091, No. 1, pp.1–5.

Ullman, S. (1979) 'The interpretation of structure from motion', *Proceedings of the Royal Society of London. Series B. Biological Sciences*.

Upasana, S., Markkandan, S. and Venkateswaran, N. (2015) 'Centralized and decentralized precoding frame work in multiuser – MIMO wireless communication', *Advances in Natural and Applied Sciences, AENSI Journal*, special edition, Vol. 9, No. 6, pp.478–485.

Wang, T-C., Liu, M-Y., Zhu, J-Y., Liu, G., Tao, A., Kautz, J. and Catanzaro, B. (2018) *Video-to-Video Synthesis*, arXiv preprint arXiv: 1808.06601.

Wang, Y., Gonzalez-Garcia, A., van de Weijer, J. and Herranz, L. (2019) 'SDIT: scalable and diverse cross-domain image translation', in *Proc. 27th ACM Int. Conf. Multimedia*, October, pp.1267–1276.

Wang, Z., Bovik, A.C., Sheikh, H.R. and Simoncelli, E.P. (2004) 'Image quality assessment: From error visibility to structural similarity', *IEEE Trans. on Image Processing*, Vol.13, No. 4, pp.600–612.

Wu, C. (2013) 'Towards linear-time incremental structure from motion', in *3DV 2013, 2013 International Conference on 3D Vision*, Seattle, Wash., pp.127–134, DOI: 10.1109/3DV. 2013.25.

Yang, H., Huang, D., Wang, Y. and Jain, A.K. (2018) 'Learning face age progression: a pyramid architecture of GANs', in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, June, pp.31–39.

Zhang, L., Shen, Y. and Li, H. (2014) 'VSI: a visual saliency induced index for perceptual image quality assessment', *IEEE Transactions on Image Processing*, Vol. 23, No. 10, pp.4270–4281.

Zhang, R., Isola, P., Efros, A.A., Shechtman, E. and Wang, O. (2018) 'The unreasonable effectiveness of deep features as a perceptual metric', in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.586–595.

Zhu, J-Y., Park, T., Isola, P. and Efros, A.A. (2017) 'Unpaired image-to-image translation using cycle-consistent adversarial networks', in *ICCV*.

Zhu, X., Zhang, X., Zhang, X-Y., Xue, Z. and Wang, L. (2019) 'A novel framework for semantic segmentation with generative adversarial network', *J. Vis. Commun. Image Represent.*, January, Vol. 58, No. 4, pp.532–543.