Generative Semantic Domain Adaptation for Perception in Autonomous Driving

Amitangshu Mukherjee[1*], Ameya Joshi[2], Anuj Sharma[1], Chinmay Hegde[2] and Soumik Sarkar[1]

[1*]Iowa State University, Ames, IA, USA.
[2*]New York University, New York City, NY, USA.

*Corresponding author(s). E-mail(s): mukher44@purdue.edu;
Contributing authors: ameya.joshi@nyu.edu; anujs@iastate.edu;
chinmay.h@nyu.edu; soumiks@iastate.edu;

Abstract

Autonomous driving systems depend on their ability to perceive and understand their environments for navigation. Neural networks are the building blocks of such perception systems, and training these networks requires vast amounts of diverse training data that includes different kinds of driving scenarios in terms of terrains, object categories, and adverse illumination/weather conditions. However, most publicly available traffic datasets suffer from having been sampled under clean weather and illumination conditions. Data augmentation is often used as a strategy to improve the diversity of training data for training machine learning-based perception systems. However, standard augmentation techniques (such as translation and flipping) help neural networks to generalize over simple spatial transformations and more nuanced techniques are required to accurately combat semantic variations in novel test scenarios. We propose a new data augmentation method called "semantic domain adaptation" that relies on the use of attribute-conditioned generative models. We show that such data augmentation improves the generalization capability of deep networks by analyzing their performance in perception based tasks such as classification and detection on different datasets of traffic objects that are captured (i) at different times of the day and (ii) across different weather conditions, and comparing with models trained using traditional augmentation methods. We further show that GAN based augmented classification models are more robust against parametric adversarial attacks than the non-GAN based augmentation models.

Keywords: Autonomous Driving, Generative Adversarial Networks, Object Classification, Object Detection, Domain Adaptation, Data Augmentation, Robust Machine Learning

1 Introduction

Perception systems for autonomous vehicles are increasingly dependent on deep neural networks that analyze images to detect and classify objects of importance [1]. Such neural networks generally require massive amounts of diverse training data that is representative of various environmental conditions comprising of adverse illumination and weather variations. However, in most publicly available datasets, images captured in clear visibility and daylight conditions dominate as compared to images representing adverse conditions such as night-time and rough weather effects. Such imbalance in training data leads to poor generalization for classification or detection models. On the other hand, manual collection and annotation of traffic data in adverse conditions can be resource-intensive and expensive. While data augmentation based on simple affine image transformations is often used to improve diversity and correct for data imbalance, such techniques seldom capture semantically meaningful variations.

To address this problem, we propose a novel data augmentation method that leverages special *attribute-conditioned generative models* to transform images under modifiable attributes such as illumination due to daylight or weather conditions. These attribute generative models such as the Attribute GAN (AttGAN) [2] are capable of reconstructing an input image into a modified version of itself with a desired attribute. These generative models allow for fine-grained control over the attribute space and generate semantically valid synthetic representation of true data.

In order to measure the efficacy of our "semantic" data augmentation, we analyze the performance of traffic object classifiers based on the ResNet [3] and MobileNet [4] architectures, and show significant improvements in class-wise F_1 scores for BDD++ [5] with day/night and clear/snow images. We also show improvement in performance of the RetinaNet [6] architecture and show improvements in its mAP scores.

Recent work [7] describes a new GAN based data augmenter for day to night image transfer and shows improvements for performances of box-detectors by training on the images generated by their AugGAN model. Although this paper and our work address a similar issue for domain adaptation, our work is significantly different in terms of the data augmentation scheme and how we want to make models robust for less-occurring object classes in real world driving datasets while preserving model performance on the high-occurring object classes. Note, we compare our generative augmentation scheme with traditional computer vision techniques and not with other generative models. The rationale is that we do not propose a new generative model architecture here. Rather we propose a framework that can leverage different generative model architectures to generate images used for robustification of deep discriminative networks. Hence, we perform empirical analysis to show that GAN based methods are better than non-GAN based methods (as opposed to highlighting a particular GAN architecture used in the process).

In summary, we make the following contributions:

- We present a novel approach in training Deep Generative models to generate synthetic night and snow images for Semantic Domain Adaptation. The process is divided into Semantic Domain Translation and Semantic Data Augmentation steps.

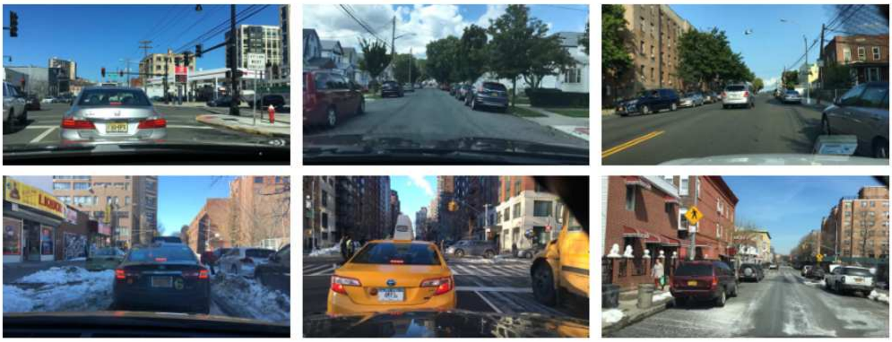

Figure 1 Original images from the BDD dataset. The first row contains images with clear weather attribute. The second row contains images with snowy weather attribute.

- Through Semantic Domain Translation we show how attribute controlled generative models can transfer benign images into their adverse counterparts which capture the semantics of the original adverse conditions.
- With these synthetic images, we introduce Semantic Data Augmentation, a data augmentation strategy that improves model performance under adverse semantic domain shifts.
- Through rigorous empirical analysis, we show that Semantic Data Augmentation works better than other data augmentation schemes for perception tasks in autonomous driving.
- We show that classification models trained with GAN based augmentation technique are more robust against parametric adversarial attacks than non GAN based augmentation approaches.
- We show the detection models trained on the Berkeley DeepDrive Dataset can transfer to other datasets such as Virtual KITTI with comparable performance levels.

1.1 Related Work

In this section we are going to discuss some of the works that are relevant to the problem addressed in the paper. We start by mentioning the prevalent driving datasets used in the literature. We then describe the various approaches taken over the years in an attempt to solve the task of domain translation in Autonomous Driving. We conclude the section by mentioning some of the relevant GAN based data augmentation techniques in the literature and how our approach although addressing a similar task is quite different from the previous approaches.

Datasets for autonomous driving. Dataset preparation has been a key focus recently in autonomous driving research. Due to ever increasing availability of data due to various source, it is also important to make sure that the collected data is not biased [8]. Driving datasets generally are of two types: synthetically generated traffic scenes, and real-world imagery. Synthetic data generation relies on the use of graphics engines [9, 10] and games [11]. CARLA [9] uses the UNITY game engine to simulate traffic behaviour and generate high fidelity data. The Synthia dataset [10] is another dataset built along the same lines with rendered city scenes and corresponding segmentation masks. Datasets such as KITTI [12], CamVID [13], Oxford Robotcar [14],

Table 1 Number of images for each weather subcategory in BDD

Weather category	Clear	Snowy	Rainy	Overcast	Cloudy	Fog
No. of Images	37344	5549	5070	8770	4881	130

Waymo [15], Berkeley Deep Drive (BDD) [16], nuScenes [17], Lyft [18], VKITTI 2.0 [19] and the Boxy [20] by Bosch represent large scale real world data for semantic segmentation, scene recognition, and motion propagation both for urban and highway driving scenarios. Our approach can be used in augmentation of any of these datasets using a generative model trained to transform input images under various semantic attributes. In this work we mainly focus on the Berkeley Deep Drive dataset [16] and we show some additional results on VKITTI 2.0 [19]. Table 1 shows the imbalance in images in the training dataset across the various weather categories. From Fig. 1, we can see that although the snowy images in the second row show presence of snow in the images, they look similar to the ones with clear weather attribute.

Domain Adaptation. Dai *et al.*[21] introduces a novel method to add synthetic fog of variable densities to real clear weather scenes using semi-supervised learning. Sakaridis *et al.*[22] augment the original Cityscapes dataset[23] with synthetic fog. Sakaridis *et al.*[24] focuses on the problem of semantic segmentation on nighttime images providing a novel pipeline to gradually transfer daytime images to nighttime images. Lore *et al.* [25] adaptively brightens images by learning semantic features in low light conditions using a deep autoencoder. Sakaridis *et al.*[24] provides a novel pipeline to gradually transfer daytime images to nighttime images based on segmentation masks. Works such as [26] try to enhance the quality of GAN generated images under various domain shifts in Autonomous Driving. Huang *et al.*[27] use GAN generated images to robustify detectors. DeepTest [28] introduces an automated testing framework for DNNs used for autonomous driving by generating affine transformations of images under illumination and weather conditions. DeepRoad [29] improves upon the results of DeepTest [28] using GAN-generated images under snowy and rainy conditions based on the framework of [30]. CyCADA [31], BicycleGAN [32], Augmented CycleGAN [33], Pix2PixHD [34] and UNIT [35] ensure semantic constraints on the real and generated images through cyclic consistency loss. There has been recent work in domain adaptation such as [36] where the semantic properties of a GAN based de-rained image is evaluated by an object detector. [37] further provides a suitable test-bed for steering models used in Autonomous Driving via Adversarial Perturbations. Recent work, Liu *et al.* [38] approaches a more realistic version of the domain adaptation problem where the networks adapt to compound heterogeneous domains with mixed factors via instance wise curriculum and domain memory. Perceptual GAN [39] improves the detection of smaller object classes by reducing the representation gap from that of the larger objects. Zheng *et al.* [40] proposes a coarse-to-fine feature adaptation method where the foreground regions are transferred via attention mechanism in an adversarial learning setup. Recent work [41] demonstrates the vulnerability of such image to image translation tasks after examining the same under the lens of adversarial perturbations in autonomous driving and robotic applications for both paired and unpaired domain translations. **The key difference between our work and the related work is the novel use of *attribute controlled generative adversarial network* to carry out the task of domain translation from benign to adverse weather and illumination conditions. Specifically, we change the time-of-day attribute for traffic scenes to flip day images into night ones. We also**

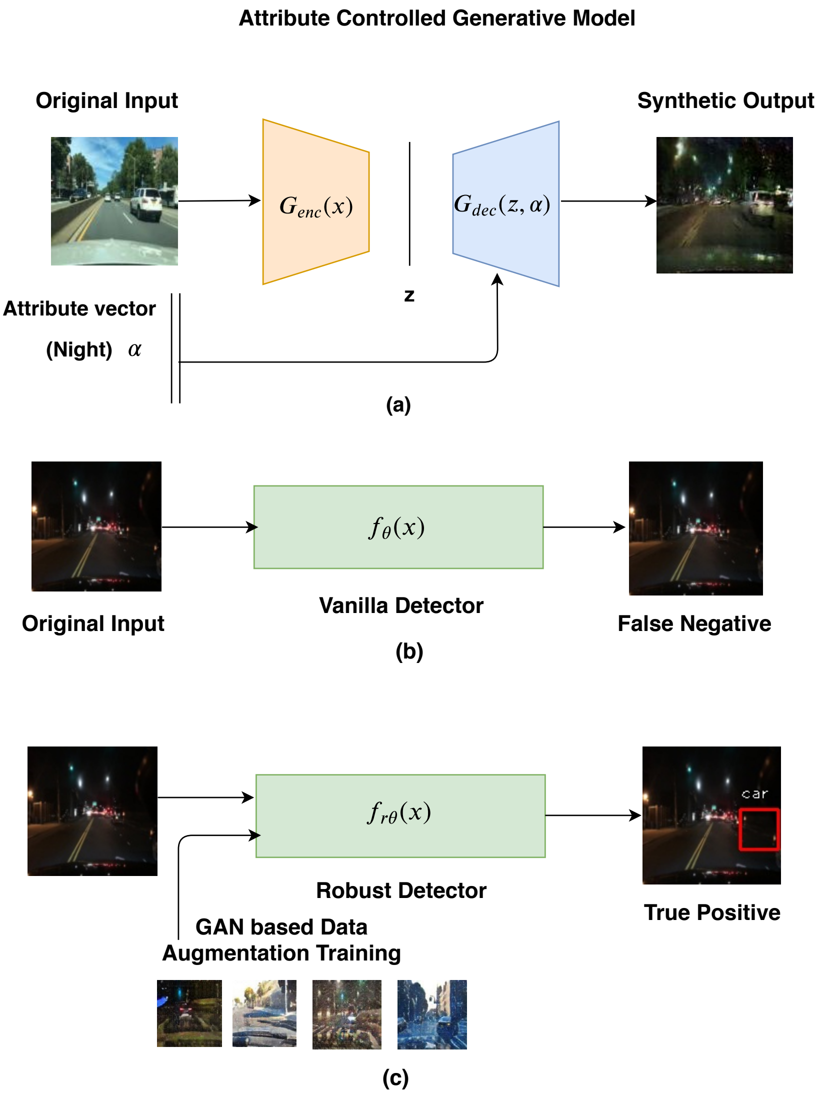

Figure 2 Block Diagram showing the overall methodology of the Semantic Domain Adaptation Framework which consists of (a) Semantic Domain Translation and (c) Semantic Data Augmentation. Part a shows the use of AttGAN to transform benign day images to their adverse night counterparts. Part b shows an instance where a model trained on an imbalanced dataset biased towards clean data points sometimes outputs a false negative when presented with a adverse data point from the test set. Part c shows where a model trained on our GAN based augmented balanced dataset gives a more robust prediction on the adverse data point as opposed to the one trained on an imbalanced dataset.

show such transformation for snowy scenes by using AttGAN to generate snow-like precipitation on clear labelled input driving images. In order to do so we do not require any semantic map information and other object based information but only the illumination and weather labels which we treat as attributes for the driving dataset. This coupled with the target to eradicate the implicit data bias towards dominant classes such as cars over less occurring classes such as trucks and buses, we introduce a data augmentation technique using the GAN generated images to improve performances over baseline classification and detection networks. The details of this technique can be found in Section II.

Attribute Controlled Generative Adversarial Networks. Generative Adversarial Networks (GAN) [42] are popularly used as a method to generate samples from real world image distributions. Fader Networks [43], Attribute GAN [2] and StarGAN [44] extend this to generate facial images with specific attributes which are provided as conditional inputs to autoencoders. There have been similar other papers which apply attribute controlled GANs to edit facial attributes of an image. **We on the other hand have extended the novel application of such attribute controlled models to change the global and local semantic features of driving scenes.**

GAN based Data Augmentation The concept of using generative models to create synthetic data for autonomous driving tasks is not new [45]. Uricár *et al.* [46] present a comprehensive survey of advanced data augmentation techniques using GANs. The paper talks about various applications of GANs in the field of autonomous driving including methods such as domain adaptation and various 2D and 3D synthesis of data for autonomous driving including implications of adversarial attacks in this domain. Lee *et al.* [47] use context aware GANs to construct synthetic scenes for autonomous driving. This work proposes an end to end learning technique where two separate generators decide what kind of object and where the particular object is going to be placed in images. This is a novel method to generate semantic images for augmentation. Wang *et al.* [48] present conditional GANs that allow for semantic manipulation of high resolution images. However, the effectiveness of such synthetic data for training has not been rigorously measured. MUNIT [49] describes an auto labelling pipeline for data augmentation via unsupervised image to image translation. Sallab *et al.* [50] produces realistic LIDAR images from both simulated and low resolution real LIDAR images using CycleGANs in order to augment dataset. SurfelGAN [51] leverages a GAN to synthesize realistic camera images for novel positions and orientations of the autonomous vehicle and moving objects in a scene using texture mapped surfels. PhysGAN [52] generates realistic physical world resilient adversarial examples to attack perception systems of autonomous vehicles in a continuous manner. Choi *et al.* [53] introduces a framework which first shows data augmentation via GANs and then a self-ensembling method to enhance performance of a segmentation network for target scenarios in autonomous driving. Uři *et al.* [54] shows a novel GAN based method to generate artificial soiling data on fisheye lenses along with annotation masks. The various kinds of data augmentation techniques have been studied in details in Khoshgoftaar *et al.* [55].

There have been methods to generate synthetic data without using Generative Adversarial Networks. Gatys *et al.* [56] introduced an algorithm known as the Neural Style Transfer which takes three different images; an input image which is to be

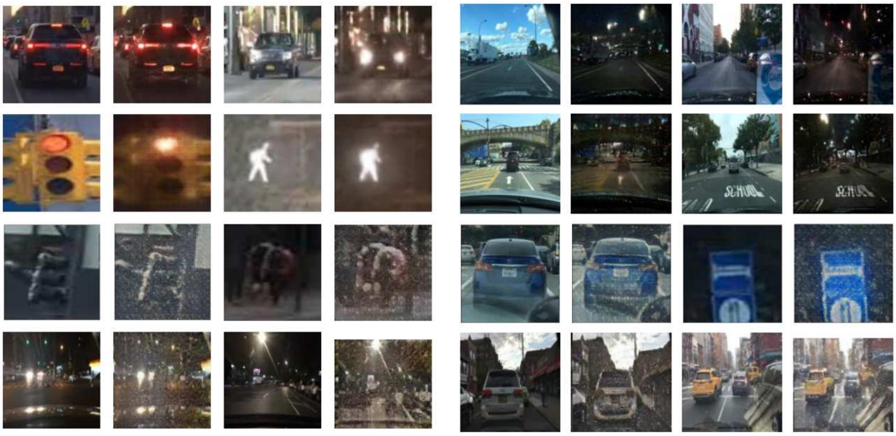

Figure 3 This figure shows the local and global features generated by the two trained deep generative models described before. The first, third, fifth and seventh columns represent the original image of objects of interest. The second, fourth, sixth and eighth columns represent the flipped versions of the images from day to night. The first row shows images where the head and tail lights are lit up at night. Similar glowing effects are seen in the second row in case of traffic lights. The last two rows show similar results for the clear to snowy transformations. The generated snowy precipitations can be seen quite clearly.

transformed, a content image whose contents needs to be preserved in the final transformed image and a style image whose style needs to be blended into the transformed image.

Our approach uses AttGAN, a specific attribute-controlled GAN model to modify semantic benign attributes of input images to their adverse counterparts conditioned on weather and illumination attributes. We then use these GAN generated adverse images to augment our pre-existing dataset in order to eradicate the class imbalance between frequently occurring classes such as cars and less frequently occurring classes such as trucks and buses. For our augmented dataset we do not use original adverse images such as night and snowy images from the BDD dataset. Instead we only use the GAN generated semantic adverse images created to augment the original dataset. So the dataset we train on has benign (day/clear) images from the BDD dataset and the adverse (night/snowy) images generated by the attribute controlled model. Details of the augmentation procedure can be found in Sections III-C and III-E.

2 Semantic Domain Translation

In this section we introduce the Semantic Domain Translation framework to transform benign (day/clear) images to their adverse counterparts (night/snowy). As discussed earlier, most of the training data points consist of clear and clean data and training only on these benign data points are not sufficient to make deep models robust against the visually degraded counterparts of the data.

2.1 Methodology

In order to train the models in a robust manner, we generate synthetic images with the help of a GAN and then use these images to augment our pre-existing dataset. We see a performance improvement for both classification as well detection tasks. Our framework is shown in detail in the block diagram Figure 2.

We train an Attribute Controlled Generative Network known as the Attribute GAN or AttGAN [2] which has been used extensively in the facial attribute editing literature. The network consists of an encoder and a decoder along with a discriminator-classifier pair to generate and analyse the reconstructed images as shown in Figure 2. They optimize over a combination of a reconstruction loss, an adversarial loss and an attribute constraint loss to ensure the editing of the exact desired attribute while preserving the attribute excluding details at the same time. The encoded latent vector is conditioned on the attribute vector during the decoding process. This results in the decoupling of semantic attributes from the underlying identity data. AttGAN takes as input an image and an attribute vector where each element represents an attribute.

$$x^{b'} = G_{dec}(G_{enc}(x^a), b) \tag{1}$$

G_{dec} and G_{enc} are the decoder and encoder respectively. a is the original attribute of the input image x^a and b is the desired attribute to appear in the final image, $x^{b'}$.

Attribute Classification Constraint To generate the reconstructed image with the desired attributes, AttGAN relies on an attribute classifier C as an adversary for the generator. The training objective for C is then:

$$\min_{G_{enc},G_{dec}} L_c = \mathbb{E}_{x^a \sim p_{data}, b \sim p_{\text{attr}}} [l_g(x^a, b)] \tag{2}$$

where p_{attr} denotes the distribution the set of attributes and, $l_g(x^a, b)$ is the summation of binary cross entropy losses of all the attributes present in the image.

Reconstruction Loss. The reconstruction loss preserves the attribute exclusive information in the image and trains the decoder to reconstruct the input image x^a by encoding the latent representation conditioned on the original attribute a.

$$\min_{G_{enc},G_{dec}} L_{rec} = \mathbb{E}_{x^a \sim p_{data}} [x^a - x^{a'}{}_1] \tag{3}$$

The ℓ_1 loss is used to reduce blurriness of the reconstructed image.

Adversarial Loss The adversarial loss is imposed to make the reconstructed images realistic in nature. AttGAN uses the WGAN formulation for training the generator and the discriminator using the objective:

$$\min_{G_{enc},G_{dec}} L_{adv} = -\mathbb{E}_{x^a \sim p_{data}, b \sim p_{attr}} [D(x^{b'})] \tag{4}$$

where $D(\cdot)$ is the discriminator in a classifier-discriminator pair (refer [57] for further details on the architecture.).

In order to train an AttGAN, we optimize over the combination of the three losses given by the summation of the losses (with appropriate coefficients) described above.

In the following subsections, we describe in detail how we train such a generative model and generate synthetic images for Semantic Domain Adaptation in order to train the deep detection and classification neural networks.

Table 2 Data Distribution for the trained model variants.

Model	Day/Night	Clear/Snow
M_1	Original Day Images from BDD Dataset	Clear Images from BDD Dataset
M_2	Original Day and Night Images from BDD Dataset	Original Clear and Snowy Images from BDD Dataset
M_3	Original Day and GAN generated Night Images	Original Clear and GAN generated Snowy Images
M_4	Original Day and synthetic Night Images	Original Clear and synthetic Snowy Images

2.2 Preprocessing

For training the AttGAN, we use a sub-sampled version of the BDD++ dataset [5] where we crop original images of the following four classes: cars, traffic signs, traffic lights, and persons. We select these image crops based on the time of the day label as well as on different weather labels to ensure variety and support further experiments and analysis. Further, we balance the training dataset by oversampling traffic signs and lights that are fewer in number as compared to other classes.

2.3 Training the Generative model

We train two different AttGAN[2] models to generate synthetic datasets to train the classification networks. We train one such attribute model on day and night attributes on the cropped image training dataset and infer on the test and validation datasets. We can see that the attribute-controlled generative model is successful in flipping the attributes of the validation and test image crops. Given an image crop with the "day" attribute, the model can flip the image to the desired "night setting".

In Figure. 3, we see that the generative model transforms the original image with the required semantic shift. We then use these shifted images to augment our training dataset for the classification model described below.

For detection, we are required to generate uncropped full images of the driving scenes. For this, we initialize our *full-image* AttGAN with the pre-trained weights of the model trained for crops as a form of transfer learning [58]. We apply transfer learning due to two major reasons; we would like to (1) leverage the local features learned from the crops for generating full-size driving scene and, (2) compensate for the lack of a large variety of data.

Fig. 3 shows that the model learns both global and local features successfully as seen by the successful semantic shifts from day to night. We also include the night to day translation in the figure to show that our model has learned to domain shift the other way as well. However, we do not include these images in our data augmentation strategy.

To train an AttGAN for simulating snowy occlusion effects on the image crops, we use additional synthetic images generated by *DesnowNet* [59] along with original clear images from the BDD Dataset. This is to compensate for the insufficient amount of original snowy images in BDD [16]. We condition the generative model to learn this snow occlusion mask and transform images with any weather attribute to exhibit snowy precipitation effects. We train the generative model to learn the synthetic snow occlusion mask generated from the DesnowNet dataset, the presence and absence of

Table 3 Dataset composition used for training the classifiers

Classes	Original Day	Original Night	Synthetic Night	Original Clear	Original Snow	Synthetic Snow
Cars	25421	17658	23669	27622	11134	18967
Person	9178	4378	7245	3624	539	2207
Traffic Sign	15786	10348	13468	10660	2224	5324
Traffic Light	8234	3383	7357	2700	613	2178

which in the training set indicates snowy and non-snowy scenarios. This implies train images with snow masks has a label of 1 to indicate the presence of a snow mask and a label of 0 in the absence of the mask.

3 Semantic Data Augmentation

In this section, we introduce the Semantic Data Augmentation framework which is a GAN based augmentation technique to make deep classification and detection models robust under domain shifts from benign to adverse weather/illumination conditions for Autonomous Driving perception applications. We test these models on unseen adverse test sets and compare the performances of these models in terms of F_1 scores for classifiers and MAP scores for box detectors.

3.1 General Setup

We train separate classification and detection models with four different settings:

1. Original images with only the benign attributes (day/clear) referred to as M_1
2. Original images with both the benign (day/clear) and adverse (night/snowy) attributes referred to as M_2
3. Original images with benign attributes and Synthetic images generated by the attribute controlled Generative Network explained in Section 2 referred to as M_3
4. Original images with benign attributes and Synthetic images generated by traditional computer vision frameworks referred to as M_4 (explained in details in Section 3.2)

We also compare models trained with our approach, M_3 (GAN based) and M_4 (graphics based) . We show that our models trained on semantic GAN generated data are more robust to adverse settings than those using synthetic data generated by graphics based methods.

Remark 1 *We do not compare with works such as AugGAN [7] since our goal here is to show that for Autonomous Driving applications, a GAN based data augmentation approach is better than a non GAN based approach. We do not focus on the particular generative models used in the process.*

3.2 Comparisons with Other Augmentation Methods

We also compare our approach with data augmentation performed using other methods. One such approach relies on the use of image processing/graphics based method to simulate weather artifacts. For fair comparison, we train models with the same architecture using datasets augmented with the following methods.

Table 4 Performance of Deep Classification Models trained under four different settings on Day and Night Image crops. The F1-score values represent the performances of these deep classifiers under the two conditions; higher the F1-score, the better is the model. The model having higher F1-score scores between M_3 and M_4 are highlighted

		Resnet-34		Mobilenet-v2	
Setting	Labels	Original Day images (F1-Score)	Original Night images (F1-Score)	Original Day images (F1-Score)	Original Night images (F1-Score)
Original Day Images (M_1)	car	0.95	0.89	0.76	0.71
	person	0.86	0.63	0.52	0.35
	traffic light	0.70	0.50	0.32	0.27
	traffic sign	0.89	0.79	0.60	0.57
Original Day and Night Images (M_2)	car	0.93	0.93	0.73	0.72
	person	0.77	0.66	0.45	0.33
	traffic light	0.69	0.68	0.29	0.32
	traffic sign	0.85	0.86	0.57	0.59
Semantic Data Augmentation **(ours)** (M_3)	car	**0.97**	**0.96**	**0.93**	**0.92**
	person	0.88	**0.80**	0.79	**0.67**
	traffic light	0.76	**0.73**	**0.68**	**0.60**
	traffic sign	0.89	**0.88**	**0.82**	**0.82**
Synthetic Data Augmentation (M_4)	car	0.96	0.93	0.90	0.86
	person	**0.90**	0.78	**0.81**	0.64
	traffic light	**0.79**	0.61	0.64	0.46
	traffic sign	**0.91**	0.83	0.80	0.73

We use a standard image darkening approach using gamma (γ) correction to create synthetic night-time images. For constructing the synthetic dataset, we randomly sample a γ value for each input image from a normal distribution, $\mathcal{N}(1,3)$ and apply a gamma transform as follows,

$$x' = x^{\gamma} \tag{5}$$

To ensure fair comparison, we restrict our augmentation to the same number input images used for training our semantic augmentated models.

For the case of snow, we use the methods presented in [60], where they use alpha blending and randomly generated snow mask images to generate synthetic adversarial examples. We use a similar approach to generate random snow masks on our training images. As before, we ensure fair comparison by generating approximately the same number of synthetic examples as in previously described experiments.

In the following sections, we explain Semantic Data Augmentation for classifiers and detectors in more detail.

3.3 Semantic Data Augmentation for Deep Classifiers

Setup. We train ResNet [3] and MobileNet [4] classifiers under the four different settings mentioned above on training datasets augmented with semantically transformed images. For generating additional training examples, we use the pretrained AttGANs to flip the benign attribute to the corresponding adverse attribute. We consider two examples of adverse attributes: night-time and snow. We train each Resnet [3] and MobileNet [4] model for 25 epochs each and have a test set accuracy above 95% on the respective test data.

We test the models trained on the two categories on an unseen test set sampled from the original images. To augment the dataset with synthetic (adverse) images for classification, we target the classes with less number of images in the original dataset to begin with such as traffic lights and persons. We balance these classes by generating semantic adverse images with out trained AttGANs. After balancing these classes, we then generate adverse examples for more frequent objects such as cars and traffic signs and augment the training dataset. Through this, we make sure

Table 5 Performance of Deep Classification Models trained under four different settings on Snowy and Clear Images. The F1-score values represent the performances of these deep classifiers under the two conditions; higher the F1-score, the better is the model. The model having higher F1-score between M_3 and M_4 are highlighted.

		Resnet-34		Mobilenet-v2	
Setting	Labels	Original snowy images (F1)		Synthetic snowy images (F1)	
Original Clear Images (M_1)	car	0.84	0.85	0.76	0.76
	person	0.71	0.55	0.73	0.63
	traffic light	0.63	0.56	0.56	0.42
	traffic sign	0.71	0.65	0.61	0.54
Original Clear and Snowy Images (M_2)	car	0.90	0.91	0.80	0.81
	person	0.76	0.68	0.74	0.66
	traffic light	0.67	0.59	0.54	0.44
	traffic sign	0.79	0.74	0.62	0.59
Semantic Data Augmentation **(ours)** (M_3)	car	**0.94**	**0.95**	**0.91**	**0.92**
	person	**0.85**	**0.78**	**0.87**	**0.77**
	traffic light	**0.84**	0.78	**0.78**	**0.73**
	traffic sign	**0.90**	**0.87**	**0.85**	**0.80**
Synthetic Data Augmentation (M_4)	car	0.92	0.94	0.90	0.91
	person	0.79	0.77	0.81	0.77
	traffic light	0.81	**0.79**	0.69	0.66
	traffic sign	0.84	0.82	0.82	0.78

that all objects of classes have an *comparable* number of benign and adverse images. The details of each object class per attribute is given in Table 3. Since we use an AttGAN for data augmentation, we can further leverage images with unused weather labels like fog, overcast and partly cloudy labels of the BDD dataset and generate additional synthetic snow images. Table. 2 provides a brief description of our training and evaluation protocols for the two adverse settings.

In order to test the efficacy of our augmentation approach, we analyse classifier performance individually on adverse and non-adverse subsets of our test set. For our augmentation strategy to be successful, it should improve classifier performance on the adverse subset while preserving (improving) the same on benign images. We analyse the class-wise F_1 scores to ensure that the inherent class imbalance does not skew the results.

Results. From Tables 4 and 5, we observe that our GAN based data augmentation strategy is successful at improving classifier performances against adverse images for all four settings. Table 4 demonstrates that our approach (M_3) is able to preserve performance on benign day examples as well as have a major boost in performance on the adverse night images than (M_2) and (M_4). For day images, the performances of models trained on M_3 and M_4 are comparable and in some cases M_4 perform slightly better than M_3 but for adverse night images the models trained on M_3 outperforms that of M_4 for all object classes.

In Table 5, we see that the model performances trained on M_3 and M_4 are comparable for all classes except for the class traffic light where the M_4 outperforms M_3 by some margin for ResNet-34. Also note, testing on synthetic images in Table 5 shows that a model trained on the original dataset shows comparable performance on synthetic images. We therefore infer that the our transformation produces realistic

Table 6 Robustness analysis of the Resnet-34 models trained on four different settings tested against worst-of-10 Random Sampling [62] and Semantic Adversarial Attacks [61].

Model Attacked	Attribute	Semantic Adv (%)	Random Sampling (%)
M_1	Day/Night	81.0	13.5
	Clear/Snowy	23.5	21.0
M_2	Day/Night	82.7	15.0
	Clear/Snowy	58.7	22.8
M_3	Day/Night	**90.5**	**29.5**
	Clear/Snowy	**77.5**	**33.0**
M_4	Day/Night	87.2	21.0
	Clear/Snowy	24.7	24.2

images as compared to that of the original data. Our approach therefore allows for *semantically* augmenting under-represented classes to improve performance.

3.4 Robustness Measures of Deep Classification Models

We also test the adversarial robustness of models trained using various augmentation methods against parametric adversarial attacks [61, 62]. Joshi *et al.* [61] construct untargeted attacks to fool classifiers by optimizing over the semantic attributes of an input image using attribute GANs. We leverage this to construct attacks against models trained with our approach as well those trained with the other approaches. We use these attacks to change the benign (day/clear) attribute to adverse (night/snowy) attributes in an adversarial optimization framework to fool the classifiers. On the other hand, Engstrom *et al.* [62] randomly sample the parameter space and select those corresponding to high cross-entropy loss. We use the same mechanism to sample the parameter space of an AttGAN and select the 10 worst performing parameter vectors to construct our attacks. We test these attacks on the ResNet-34 models trained on the various settings described above.

From Table 6, we can see that the Resnet-34 model trained under setting M_3 is the most robust model for either of the attack attributes against both the parametric attacks. We can see a trend in terms of robustness of the model under each setting. The model trained under setting M_3 is the most robust followed by M_4, M_2 and then ultimately M_1. This is actually intuitive which shows augmentation helps but a GAN based semantic augmentation method is more robust than a synthetic non GAN method. Moreover in one experiment,M_2 performs much better than M_4 which shows the disadvantage of a non GAN based augmentation technique.

3.5 Semantic Data Augmentation for Deep Detectors

While our approach shows promise for classification tasks, autonomous driving systems generally use detection or segmentation models to perceive complex traffic scenery. We therefore analyse the performance of our semantic data augmentation on a standard detector- RetinaNet [6]. All experiments were performed on a single workstation equipped with an NVidia Titan X_p GPU in PyTorch [63] v1.0.0.

RetinaNet uses focal loss to train a single stage detector while considering the data imbalance between foreground and background pixels. We analyse the effectiveness of our approach by training four instances of RetinaNet for each adverse semantic attribute as mentioned in Section 3.1. For example, considering the semantic attribute of snowy precipitation, we train the first model (M_1) with images containing

Figure 4 Detection results shown under six different settings. The first column shows detection results of the model trained only with day images. The second column shows results of the model trained under original day and night images. The third column shows results of the model trained under original day and synthetic night images. Similarly the last three columns show results for the same experiment conducted for the 'snow' attribute.

no precipitation, the second one with the complete dataset (M_2), the third model (M_3) with our augmented dataset with GAN generated precipitation on a subset of the images and the fourth model (M_4) with the same augmented images mentioned in third setting but generated through standard image processing procedures. Note that for the third and fourth settings, we only use adverse examples generated by the GAN and image processing methods instead of the original adverse examples. This is to ascertain the specific improvement due to the addition of the synthetic images. The four models are each tested on two separate test sets corresponding to the adverse and non-adverse semantic attributes. Our analysis remains similar to the experiments done for the classification task except that we consider mAP scores for analysing the detector performance instead of accuracy.

Similarly for training the models under illumination conditions we follow a similar setting as mentioned above. The baseline model is trained only on day images (M_1), the second model is trained on original day and night images (M_2), the third model is trained on original day and synthetic night images which are generated by flipping the original day images via the attribute controlled generative network (M_3). In the third setting for synthetic night images we flip the same number of day images as is the number of night images used to train the model in the second setting keeping the number of original day images under both settings fixed. We repeat the same steps for M_4, but instead of using a GAN, we use traditional computer vision frameworks to generate the synthetic images. We do this to provide a fair comparison of the model trained on the third setting with the models trained on the second and fourth settings.

Table 7 Performance of Semantic Augmentation for detection under Adverse Snow conditions RetinaNet with Resnet 50 as backbone architecture. The model having higher mAP scores between M_3 and M_4 are highlighted.

Image Size (256, 256)		**IoU Threshold = 0.5**			
Test Dataset	Classes	M_1 (Benign Data)	M_2(Orig. Dataset)	M_3 (Semantic Augmentation)	M_4 (Synthetic Data Augmentation)
Benign (Clear)	Bus Car Truck	14.0 18.0 13.0	13.0 20.0 17.0	**14.0** **18.0** **17.0**	9.2 17.2 12.0
	mAP (Overall)	15.0	15.3	**16.3**	12.8
Original (Clear and Snowy)	Bus Car Truck	13.0 17.0 13.0	13.0 18.0 15.0	**13.0** **18.0** **16.0**	9.0 17.0 12.0
	mAP (Overall)	14.3	15.3	**15.7**	12.7
Image Size (512, 512)		**IoU Threshold = 0.5**			
Test Dataset	Classes	M_1 (Benign Data)	M_2(Orig. Dataset)	M_3 (Semantic Augmentation)	M_4 (Synthetic Data Augmentation)
Benign (Clear)	Bus Car Truck	7.0 24.0 8.0	8.0 25.0 12.0	**9.4** **25.0** **13.0**	4.9 23.0 10.0
	mAP (Overall)	13.0	15.0	**15.8**	12.63
Original (Clear and Snowy)	Bus Car Truck	6.0 24.0 6.0	7.0 24.0 11.0	**8.0** **25.0** **13.0**	4.2 24.0 9.9
	mAP (Overall)	12.0	14.0	**15.3**	12.7

Training. We train three RetinaNet models with Resnet-50 as the backbone architectures. Each model has been trained for 35 epochs each and the hyperparameters are chosen as in [6].

Mean Average Precision (mAP) scores Mean Average Precision (mAP) is defined as the mean of the areas under the classwise precision-recall curves. For measuring precision and recall, we define an object as correctly detected if the intersection over union (IoU) of the detected box and true box is greater than 0.5. However, due to the low resolution of our images (128×128), an error of *one* pixel^2 may lead to a large error in the IoU value. Therefore, the absolute values of the mAP scores may not provide us much evidence of detector performance. Instead, we observe the relative performance in mAP scores for our experiments to analyse the efficacy of semantic data augmentation.

Results. Table 7 shows the performance of the four models trained on the three different data distributions. As expected, the model trained only on the benign images shows worse performance as compared to the model trained on both benign and adverse sets. However, the model trained using our approach (original benign images and sysnthetic adverse images) shows comparable performance to the model trained on the full dataset. We also emphasize that our attribute conditioned GAN is trained

Table 8 Performance of Semantic Augmentation for detection under illumination conditions for RetinaNet with Resnet 50 as backbone architecture. The model having higher mAP scores between M_3 and M_4 are highlighted.

Image Size (256, 256)		**IoU Threshold = 0.5**			
Test Dataset	Classes	M_1 (Benign Data)	M_2(Orig. Dataset)	M_3 (Semantic Augmentation)	M_4 (Synthetic Data Augmentation)
Benign (Day)	Bus	14.0	15.0	**19.0**	12.0
	Car	20.0	20.0	**20.0**	19.0
	Truck	18.0	17.0	**18.0**	14.0
	mAP (Overall)	17.33	17.33	**19.0**	15.0
Original (Night)	Bus	7.0	15.0	**12.0**	4.0
	Car	17.0	20.0	**18.0**	15.0
	Truck	8.0	16.0	**12.0**	10.0
	mAP (Overall)	10.7	17.0	**14.0**	9.7
Image Size (512, 512)		**IoU Threshold = 0.5**			
Test Dataset	Classes	M_1 (Benign Data)	M_2(Orig. Dataset)	M_3 (Semantic Augmentation)	M_4 (Synthetic Data Augmentation)
Benign (Day)	Bus	5.0	2.1	**5.0**	4.1
	Car	29.0	30.0	**30.0**	29.0
	Truck	12.0	4.0	**16.0**	13.0
	mAP (Overall)	15.3	12.03	**17**	15.37
Original (Night)	Bus	1.0	2.7	**2.1**	1.3
	Car	19.0	25.0	**21.0**	17.0
	Truck	7.7	2.1	**10.0**	8.4
	mAP (Overall)	9.23	9.93	**11.03**	8.9

with adverse images not found in the original BDD dataset. In spite of this, AttGAN generates valid adverse images with snowy precipitation. We, therefore show that attribute conditioned GANs prove to be a realistic tool for semantic augmentation even in complex tasks such as detection. More importantly, we show that the model trained under setting M_3 performs better than the model trained under setting M_4 for all the object classes.

Table 8 shows the performance of the four models trained on three different data distributions for illumination conditions. From the mAP scores for the three different object classes it can be seen that under night conditions, the model trained under (M_3) performs better than model trained under (M_1). As mentioned earlier, we use the exact same number of synthetic night images in (M_3) as there are original number of night images in (M_2). We do this to show the efficacy of our method where we are able to synthetically generate night images from just the benign day set and still get comparable performance with the model trained under (M_2) without having to do any sort of data augmentation. We again show that the model trained under setting M_3 performs better than the model trained under setting M_4 for all the object classes.

We observe that for both the attributes the models trained under the setting M_4 perform worse than both M_2 and M_3. This shows for perception based tasks in Autonomous Driving, images produced by traditional CV methods fail to capture both

Table 9 Transferability of Semantic Augmentation on VKITTI 2.0. RetinaNets (ResNet50 backbone) are trained on BDD as in Tab. 8 and tested on VKITTI 2.0 [19].

		IoU Threshold = 0.5			
Test Dataset	Classes	M_1	M_2	M_3	M_4
Benign (Day)	Car	24.24	**27.03**	26.54	23.69
	Truck	0.38	0.23	**0.75**	0.28
	mAP (Overall)	12.31	13.63	**13.64**	11.985
Original (Night)	Car	24.17	**26.53**	25.95	23.03
	Truck	0.47	1.15	**1.62**	0.55
	mAP (Overall)	12.32	13.84	**13.78**	11.79

local and global semantic features in a driving scene. On the other hand, Generative Adversarial Networks are capable of learning these semantic features for various driving tasks. Thus these networks generate images that are more suitable for augmentation to make deep classification and detection models robust against domain shifts under adverse illumination and weather conditions. From both tables 8 and 7, we see that our model (M_3) performs better (M_2) on images of size 512×512. This is due to the increase in AP score of the object classes Bus and Truck which is a result of our Data Augmentation strategy. In this way we are able to robustify models on lesser occurring objects in the dataset while keeping the performance on the more recurring object classes in the dataset consistent. **We also show that our data augmentation transfers well across datasets. For this, we test our trained detectors on VKitti 2.0 [19]. As Table. 9 shows, M_3 performs better than M_1 and M_4 in all cases, and is comparable with M_2.**

Fig. 4 shows the vulnerability of the model trained under (M_1) and the robustness of the model trained under (M_3). We can see that in the first and second rows, the model trained under (M_3) detects objects which the models (M_1) and (M_2) detect as false negatives. In the third row, we can see the comparable detection results of models (M_2) and (M_3) while the model trained on (M_1) misses the car right in front.

4 Conclusion and future work

We have shown that semantic data augmentation is a viable approach to tackle the lack of data diversity. Especially for autonomous vehicles, our approach can compensate for the dearth of data captured under adverse conditions. We have empirically analysed the effect of our approach on real-world classification and detection tasks and show promising results. Additionally, while we show experiments for AttGANs that are limited to size constraints, our approach can be extended to better and more sophisticated generative models such as Progressive GANs [64]. Another avenue for future study is to deploy such trained models on real world systems and analyse the effect of data augmented models versus those trained with true data.

5 Conflict of Interest

The authors declare that they have no conflict of interest.

6 Acknowledgments

This work was supported in part by NSF grants CCF-2005804, CNS-1845969, DARPA AIRA grant PA-18-02-02, AFOSR YIP grant FA9550-17-1-0220, an ERP grant from ISU, a GPU gift grant from NVIDIA Corporation, and faculty fellowships from the Black and Veatch Foundation.

References

[1] Haghighat, A.K., Ravichandra-Mouli, V., Chakraborty, P., Esfandiari, Y., Arabi, S., Sharma, A.: Applications of deep learning in intelligent transportation systems. Journal of Big Data Analytics in Transportation (2019)

[2] He, Z., Zuo, W., Kan, M., Shan, S., Chen, X.: Attgan: Facial attribute editing by only changing what you want. IEEE T Image Process (2019)

[3] He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. CVPR (2015)

[4] Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., Adam, H.: Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint **abs/1704.04861** (2017)

[5] Mukherjee, A., Joshi, A., Sarkar, S., Hegde, C.: Attribute-controlled traffic data augmentation using conditional generative models. In: CVPRW (2019)

[6] Lin, T.-Y., Goyal, P., Girshick, R., He, K., Dollar, P.: Focal loss for dense object detection. ICCV (2017)

[7] Lin, C., Huang, S., Wu, Y., Lai, S.: Gan-based day-to-night image style transfer for nighttime vehicle detection. IEEE trans Intell Transp Syst (2020)

[8] Griffin, G.P., Mulhall, M., Simek, C., Riggs, W.W.: Mitigating bias in big data for transportation. Journal of Big Data Analytics in Transportation (2020)

[9] Dosovitskiy, A., Ros, G., Codevilla, F., Lopez, A., Koltun, V.: CARLA: An open urban driving simulator. In: CoRL (2017)

[10] Ros, G., Sellart, L., Materzynska, J., Vazquez, D., Lopez, A.: The SYNTHIA Dataset: A large collection of synthetic images for semantic segmentation of urban scenes. In: CVPR (2016)

[11] Richter, S.R., Vineet, V., Roth, S., Koltun, V.: Playing for data: Ground truth from computer games. In: ECCV (2016)

[12] Geiger, A., Lenz, P., Stiller, C., Urtasun, R.: Vision meets robotics: The KITTI dataset. IJRR (2013)

[13] Fauqueur, J., Brostow, G., Cipolla, R.: Assisted video object labeling by joint

tracking of regions and keypoints. In: ICCVW (2007)

[14] Maddern, W., Pascoe, G., Linegar, C., Newman, P.: 1 Year, 1000km: The Oxford RobotCar Dataset. IJRR (2017)

[15] Sun, P., Kretzschmar, H., Dotiwalla, X., Chouard, A., Patnaik, V., Tsui, P., Guo, J., Zhou, Y., Chai, Y., Caine, B., Vasudevan, V., Han, W., Ngiam, J., Zhao, H., Timofeev, A., Ettinger, S., Krivokon, M., Gao, A., Joshi, A., Zhang, Y., Shlens, J., Chen, Z., Anguelov, D.: Scalability in perception for autonomous driving: Waymo open dataset. In: CVPR (2020)

[16] Yu, F., Chen, H., Wang, X., Xian, W., Chen, Y., Liu, F., Madhavan, V., Darrell, T.: Bdd100k: A diverse driving dataset for heterogeneous multitask learning. In: CVPR (2020)

[17] Caesar, H., Bankiti, V., Lang, A.H., Vora, S., Liong, V.E., Xu, Q., Krishnan, A., Pan, Y., Baldan, G., Beijbom, O.: nuscenes: A multimodal dataset for autonomous driving. In: CVPR (2020)

[18] Houston, J., Zuidhof, G., Bergamini, L., Ye, Y., Jain, A., Omari, S., Iglovikov, V., Ondruska, P.: One Thousand and One Hours: Self-driving Motion Prediction Dataset (2020)

[19] Cabon, Y., Murray, N., Humenberger, M.: Virtual kitti 2. arXiV preprint **abs/2001.10773** (2020)

[20] Behrendt, K.: Boxy vehicle detection in large images. In: ICCVW (2019)

[21] Dai, D., Sakaridis, C., Hecker, S., Gool, L.V.: Curriculum model adaptation with synthetic and real data for semantic foggy scene understanding. IJCV (2019)

[22] Sakaridis, C., Dai, D., Gool, L.V.: Semantic foggy scene understanding with synthetic data. IJCV (2018)

[23] Cordts, M., Omran, M., Ramos, S., Scharwächter, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S., Schiele, B.: The cityscapes dataset. In: CVPR (2015)

[24] Sakaridis, C., Dai, D., Gool, L.V.: Semantic nighttime image segmentation with synthetic stylized data, gradual adaptation and uncertainty-aware evaluation. arXiv preprint **abs/1901.05946** (2019)

[25] Lore, K.G., Akintayo, A., Sarkar, S.: Llnet: A deep autoencoder approach to natural low-light image enhancement. Pattern Recognition (2015)

[26] Yoo, J., Eom, H., Choi, Y.S.: Image-to-image translation using a cross-domain auto-encoder and decoder. In: Applied Sciences (2019)

[27] Huang, S.-W., Lin, C.-T., Chen, S.-P., Wu, Y.-Y., Hsu, P.-H., Lai, S.-H.: Auggan:

Cross domain adaptation with gan-based data augmentation. In: ECCV (2018)

[28] Tian, Y., Pei, K., Jana, S., Ray, B.: Deeptest: Automated testing of deep-neural-network-driven autonomous cars. ICSE (2018)

[29] Zhang, M., Zhang, Y., Zhang, L., Liu, C., Khurshid, S.: Deeproad: Gan-based metamorphic autonomous driving system testing. arXiv preprint **abs/1802.02295** (2018)

[30] Liu, M.-Y., Breuel, T., Kautz, J.: Unsupervised image-to-image translation networks. In: NeurIPS (2017)

[31] Hoffman, J., Tzeng, E., Park, T., Zhu, J.-Y., Isola, P., Saenko, K., Efros, A.A., Darrell, T.: Cycada: Cycle-consistent adversarial domain adaptation. In: ICML (2018)

[32] Zhu, J.-Y., Zhang, R., Pathak, D., Darrell, T., Efros, A.A., Wang, O., Shechtman, E.: Toward multimodal image-to-image translation. In: NeurIPS, (2017)

[33] Almahairi, A., Rajeswar, S., Sordoni, A., Bachman, P., Courville, A.C.: Augmented cyclegan: Learning many-to-many mappings from unpaired data. In: ICML (2018)

[34] Wang, T.-C., Liu, M.-Y., Zhu, J.-Y., Tao, A., Kautz, J., Catanzaro, B.: High-resolution image synthesis and semantic manipulation with conditional gans. In: CVPR (2018)

[35] Zhu, J.-Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. ICCV (2017)

[36] Sudo, F., Hashimoto, Y., Lisi, G.: Object detection to evaluate image-to-image translation on different road conditions. In: Intelligent Human Systems Integration 2020 (2020)

[37] Machiraju, H., Balasubramanian, V.N.: A little fog for a large turn. arXiv preprint **abs/2001.05873** (2020)

[38] Liu, Z., Miao, Z., Pan, X., Zhan, X., Lin, D., Yu, S.X., Gong, B.: Open compound domain adaptation. In: CVPR (2020)

[39] Li, J., Liang, X., Wei, Y., Xu, T., Feng, J., Yan, S.: Perceptual generative adversarial networks for small object detection. CVPR (2017)

[40] Zheng, Y., Huang, D., Liu, S., Wang, Y.: Cross-domain object detection through coarse-to-fine feature adaptation. In: CVPR (2020)

[41] Wang, L., Cho, W., Yoon, K.-J.: Deceiving image-to-image translation networks for autonomous driving with adversarial perturbations. IEEE Robot. Autom. Lett. (2020)

[42] Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A.C., Bengio, Y.: Generative adversarial nets. In: NeurIPS (2014)

[43] Lample, G., Zeghidour, N., Usunier, N., Bordes, A., Denoyer, L., *et al.*: Fader networks: Manipulating images by sliding attributes. In: NeurIPS (2017)

[44] Choi, Y., Choi, M., Kim, M., Ha, J.-W., Kim, S., Choo, J.: Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In: CVPR (2018)

[45] Bowles, C., Chen, L., Guerrero, R., Bentley, P., Gunn, R.N., Hammers, A., Dickie, D.A., del C. Valdés Hernández, M., Wardlaw, J.M., Rueckert, D.: Gan augmentation: Augmenting training data using generative adversarial networks. arXiV preprint **abs/1810.10863** (2018)

[46] Uricar, M., Krizek, P., Hurych, D., Sobh, I.M., Yogamani, S., Denny, P.: Yes, we gan: Applying adversarial techniques for autonomous driving. Electronic Imaging **2019**(15), 48–1 (2019)

[47] Lee, D., Liu, S., Gu, J., Liu, M.-Y., Yang, M.-H., Kautz, J.: Context-aware synthesis and placement of object instances. In: NeurIPS, (2018)

[48] Wang, T.-C., Liu, M.-Y., Zhu, J.-Y., Tao, A., Kautz, J., Catanzaro, B.: High-resolution image synthesis and semantic manipulation with conditional gans. CVPR (2017)

[49] Yu, W., Sun, Y., Zhou, R., Liu, X.: Gan based method for labeled image augmentation in autonomous driving. In: ICCVE (2019)

[50] Sallab, A.E., Sobh, I., Zahran, M., Essam, N.: Lidar sensor modeling and data augmentation with gans for autonomous driving. arXiv preprint **abs/1905.07290** (2019)

[51] Yang, Z., Chai, Y., Anguelov, D., Zhou, Y., Sun, P., Erhan, D., Rafferty, S., Kretzschmar, H.: Surfelgan: Synthesizing realistic sensor data for autonomous driving. In: CVPR (2020)

[52] Kong, Z., Guo, J., Li, A., Liu, C.: Physgan: Generating physical-world-resilient adversarial examples for autonomous driving. In: CVPR (2020)

[53] Choi, J., Kim, T.-K., Kim, C.: Self-ensembling with gan-based data augmentation for domain adaptation in semantic segmentation. ICCV (2019)

[54] Uricar, M., Sistu, G., Rashed, H., Vobecký, A., Krízek, P., Burger, F., Yogamani, S.K.: Let's get dirty: Gan based data augmentation for soiling and adverse weather classification in autonomous driving. arXiv preprint **abs/1912.02249** (2019)

[55] Khoshgoftaar, T.M., Shorten, C.: A survey on image data augmentation for deep

learning. In: Journal of Big Data (2019)

[56] Gatys, L.A., Ecker, A.S., Bethge, M.: Image style transfer using convolutional neural networks. In: CVPR (2016)

[57] He, Z., Zuo, W., Kan, M., Shan, S., Chen, X.: Attgan: Facial attribute editing by only changing what you want. arxiv preprint (2017)

[58] Goodfellow, I., Bengio, Y., Courville, A., Bengio, Y.: Deep Learning vol. 1. MIT Press, ??? (2016)

[59] Liu, Y.-F., Jaw, D.-W., Huang, S.-C., Hwang, J.-N.: Desnownet: Context-aware deep network for snow removal. IEEE T Image Process (2018)

[60] Kang, D., Sun, Y., Hendrycks, D., Brown, T., Steinhardt, J.: Testing Robustness Against Unforeseen Adversaries (2019)

[61] Joshi, A., Mukherjee, A., Sarkar, S., Hegde, C.: Semantic adversarial attacks: Parametric transformations that fool deep classifiers. ICCV (2019)

[62] Engstrom, L., Tsipras, D., Schmidt, L., Madry, A.: A rotation and a translation suffice: Fooling cnns with simple transformations. arxiv preprint **abs/1712.02779** (2017)

[63] Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., Lerer, A.: Automatic differentiation in pytorch. In: NeurIPS-W (2017)

[64] Karras, T., Aila, T., Laine, S., Lehtinen, J.: Progressive growing of gans for improved quality, stability, and variation. arXiv preprint **abs/1710.10196** (2018)