Attribute-Controlled Traffic Data Augmentation Using Conditional Generative Models

Amitangshu Mukherjee Ameya Joshi Soumik Sarkar Chinmay Hegde*
Iowa State University
{amimukh, ameya, soumiks, chinmay}@iastate.edu

Abstract

Perception systems of self-driving vehicles require large amounts of diverse data to be robust against adverse lighting and weather conditions. Collection and annotation of such traffic data is resource-intensive and expensive. To circumvent this challenge, we introduce an approach where we train attribute-based generative models conditioned on the time-of-day labels to reconstruct semantically valid transformed versions of the original data. We further show the generalization capabilities of our model where they are able to reconstruct full traffic scenes despite having only being trained on constrained crops of the original images. Finally, we present a new dataset derived from an original traffic scene dataset augmented with data generated by our attribute-based conditional generative models.

1. Introduction

Autonomous vehicular systems rely on the use of real-world data to train perception systems. This data generally consists of color traffic images taken by cameras placed on cars and are further annotated manually. Data collection is an arduous and error-prone process. Additionally, such datasets are generally imbalanced and do not span all possible environmental conditions.

Training perception systems on such imbalanced data would result in undefined behaviour in rare, yet critical, environmental changes. Therefore, an approach to augment an existing dataset with environmental transformations is necessary. Existing approaches to this problem rely on the use of 3D simulations [4, 16] to generate synthetic data. However, these approaches are often not realistic and perception models trained on these augmentation methods are susceptible to synthetic artifacts. On the other hand, a well-trained generative model captures the intricacies of natural images while being able to generate natural, realistic transformations of input images. Additionally, generative adversarial models allow for latent space interpolation. This allows for generating data that may be expensive to obtain naturally while also reducing redundancy in data acquisition.

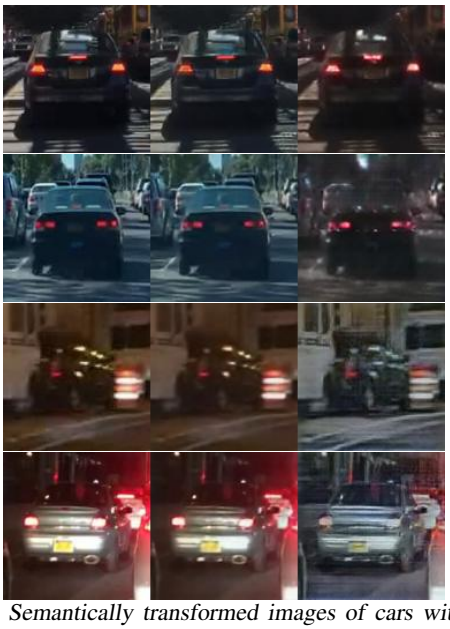

Figure 1. *Semantically transformed images of cars with day or night attributes. The first column contains the original images. The second column and third column show images reconstructed by the AttGAN under the original and flipped attributes. The trained AttGAN successfully learns to decouple the 'day/night' attribute from the underlying invariant data. Note the varying appearance of taillights under the day and night attributes.*

We present a novel application of attribute conditioned generative models to transform street and traffic images under various attributes. The generative model, *Attribute GAN* (AttGAN) [9] reconstructs an input image with various environmental attributes while allowing fine-grained control over the intensity of the effect. We train an AttGAN conditioned on the time-of-day attribute between *Day* and *Night* using crops of objects of interest namely cars and roadside

*This work was supported in part by NSF grants CCF-1750920, CNS-1845969, DARPA AIRA grant PA-18-02-02, AFOSR YIP Grant FA9550-17-1-0220, an ERP grant from Iowa State University, a GPU gift grant from NVIDIA corporation, and faculty fellowships from the Black and Veatch Foundation.

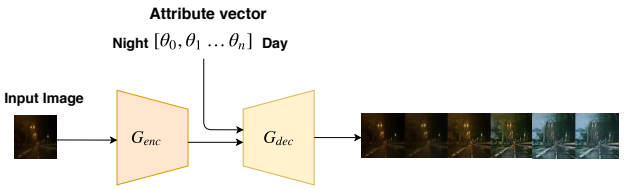

Figure 2. *Block Diagram for interpolation on the time-of-the-day attribute. By uniformly sampling θ_i over the attribute line segment $[0, 1]$, the autoencoder is used to generate semantically valid transformations corresponding to times-of-day between day and night.*

signs, from the Berkeley Deep-Drive Dataset (BDD) [23]. The trained model is then used to generate images with flipped attributes which means that a daytime image is transformed to a night-time image. Using this generative model, we are able to interpolate over the input attribute vectors to generate images for various times-of-day. We also provide examples showing that the generative model learns to capture semantic information about the image that a style transformation or a graphics-based approach cannot capture. We also present *BDD++*, a new dataset of images generated with day and night attributes from the BDD dataset.

Outline. We first provide a brief literature survey of conditional generative models and synthetic data augmentation methods for traffic data images. We then describe our method for training the AttGAN with a modified version of the BDD dataset in Section 3 followed by experiments and results in Section 4. We finally analyze our results and conclude in Section 5.

2. Related Work

The challenge of data augmentation for autonomous driving has been addressed in several recent works. Driving datasets generally are of two types: synthetically generated traffic scenes and real-world data. Synthetic data generation relies on the use of graphics engines [4, 18] and games [16]. CARLA [4] uses the UNITY game engine to simulate traffic behaviour and generate high fidelity data. The Synthia dataset [18] is another dataset built along the same lines with rendered city scenes and corresponding segmentation masks. Datasets such as KITTI [7], CamVID [5], Oxford Robotcar Dataset [14] and Berkeley Deep Drive(BDD) [23] present large scale real world data for semantic segmentation, scene recognition and motion propagation. Our approach enables augmentation of any of these datasets using a generative model trained to transform input images under various attributes.

DeepTest [21] introduces an automated testing framework for DNNs used for autonomous driving by generating affine transformations of images under illumination and weather conditions. DeepRoad [24] improves upon the results using GAN-generated images under snowy and rainy conditions based on the framework of [13]. CyCADA [10] and UNIT [25] ensure semantic constraints on the real and generated images through cyclic consistency loss. Gatys *et al*. [6] introduce a neural algorithm to combine style of one image and the content of another and jointly optimize over the style and content losses to generate a new image while preserving the content of the former and style of the latter image.

Dai *et al*. [3] introduces a novel method to add synthetic fog of variable densities to real clear weather scenes using semi-supervised learning. Sakaridis *et al*. [19] augment original Cityscapes dataset[2] with synthetic fog. Sakaridis *et al*. [20] focuses on the problem of semantic segmentation on nighttime images providing a novel pipeline to gradually transfer daytime images to nighttime images.

Generative Adversarial Networks (GAN) [8] are popularly used as a method to generate samples from real world image distributions. Chen *et al*. [1] present InfoGAN where stylistic factors of the output image are controlled using specific dimensions of the input latent vector. Fader Networks [12] and Attribute GANs *et al*. [9] extend this to generate facial images with specific attributes which are provided as conditional inputs to autoencoders. The concept of using generative models to create synthetic data for autonomous driving tasks is not new. Uricár *et al*. [22] presents a comprehensive survey of advanced data augmentation techniques using GANs.

Our approach uses AttGANs, a specific attribute controlled generative model to modify environmental attributes of input data. Specifically, we change the time-of-day attribute for traffic scenes using an AttGAN trained on a processed version of the BDD dataset.

3. Attribute Interpolation with Conditional Generative Models

3.1. Architecture

The Attribute GAN (AttGAN) [9] is an encoder-decoder architecture which is be used for editing attributes by manipulating the encoded latent representation. An AttGAN disentangles the semantic attributes from the underlying invariances of the data by considering both the original and the flipped labels while training. This is achieved by training a latent discriminator and classifier pair to classify both the original and the transformed image to ensure invariance. The model is trained to jointly optimize over an attribute classification loss, an adversarial loss and a reconstruction loss to ensure proper switching of the desired attributes while still maintaining the realness of the image and preserving the attribute excluding details at the same time. The architecture of the AttGAN ensures skip connections between the decoder and encoder like a U-Net [17] to ensure high quality of reconstructed images in the application

Figure 3. *Examples of images from the BDD++ dataset. The first row contains image crops generated with the daytime label. The second row correspondingly shows images generated with the night attribute. The samples show two objects of interest generated by the trained attribute encoder: cars and traffic signs.*

of image translation.

The AttGAN architecture additionally allows for attribute style manipulation where one controls the style and expression of the desired attribute in the reconstructed image. This is achieved by maximizing the mutual information through optimization of the encoder-decoder learning by binding a set of style controllers and the generated output images, thus making them highly correlated. He *et al*. [9] show an example of such style intensity control. Thus keeping the attribute style manipulation in mind as well as the attribute preserving learning of the model, we select AttGAN as our choice of conditional generative model which we train on a comprehensive driving dataset explained in the following sections.

3.2. Data

For our experiments, we create a modified version of the Berkeley Deep Drive (BDD) Dataset [23] as our dataset for both training and testing. We use the BDD dataset as it contains comprehensive annotations of various driving scenes taken at various time of the day and across different seasons. The annotations include both global features of an image such as drive-able area, the time of the day, the particular weather setting under which the scene has been taken as well as abundant local features which include 2D bounding boxes for object classes of importance, lane markings and segmentation masks.

In the following sections we introduce a non-traditional approach to train a conditional generative model to generate traffic scenery with modifiable environmental attributes.

3.3. Preprocessing

We introduce a non-traditional approach in which we choose a fair representation of the entire dataset and then train an AttGAN [9] conditioned on the features of this subset. Since the goal of our experiments is to reconstruct the same image with flipped day and night attributes, we segment the dataset based on the time of the day label into two classes: *Day* and *Night*. To tackle the data imbalance between the number of day and night images in the original dataset, we decide to crop objects of important classes from the original dataset conditioned on the two labels and create a dataset of our own. We crop the images using the 2D box annotations provided with the dataset and by constraining the aspect ratio of the image crops under an empirically decided upper bound[1]. For the purposes of this work, we only consider the crops of cars and traffic signs, though the concept extends to all given classes in the dataset.

3.4. Training

AttGANs are trained in a supervised manner by training the generative auto-encoder to reconstruct input images under the original and the flipped attribute. For our application, we use our augmented cropped dataset along with the corresponding attribute labels for training. The cropping ensures that the model learns to reconstruct objects of interest over redundant background.

In general, AttGANs are trained by simply flipping the input attribute vector. We observe that in case of our chosen attributes, the transition between night and day is not abrupt. In addition, the primary motivation of our using such attribute conditioned models is that we would like to interpolate over attribute space to reconstruct examples under various times of the day. Thus we train an AttGAN with the attribute space multiplied with a truncated normal distribution. Intuitively, the truncated normal distribution is a more natural prior as compared to an uniform distribution in our case as the data is centered around the day and night attributes. Multiplying a truncated normal distribution to the actual attribute distribution ensures the final flipped attributes are more uniformly distributed.

4. Experiments and Results

We train an AttGAN with 70% of the dataset as training data and 20% for validation. The remaining 10% of the dataset is used for inference. The autoencoder architecture contains five encoding and decoding layers. As mentioned earlier in Section 3.1 the network ensures a U-Net architecture with skip connections to ensure good quality images. The non-linear activation functions for the encoder and the discriminator are Leaky Relu and Rectified Linear Unit for the decoder. We use the ADAM optimizer [11] to optimize over the binary cross-entropy loss. We use the same coefficients for the reconstruction loss, attribute classification loss and the adversarial loss as mentioned in [9]. We resize each crop to a size of 128×128 for training with a

[1]The images produced by cropping the dataset with the 2D box annotations of cars and traffic signs are of varying shapes. We constrain each image crop to be within an aspect ratio of 4:3. Resizing of crops with non-standard aspect ratio induce unnatural sheer in either dimensions which result in improper training of the AttGAN. Additionally we increase the size of each image crop by 30 pixels on either sides than the provided 2D box annotations to create the dataset with perceptible day/night effects.

Figure 4. *Style manipulation of time of day attributes through a trained Attribute GAN with day and night labels. The first column shows the original image. The next ten columns denote the gradual change in style which is the time of day in this case. The first two rows show gradual change from night to day on images from the test set of crop images. The last two rows exhibit gradual change from night to day on images from the original BDD dataset. This shows that the model trained on crop images from the original dataset generalizes on the original uncropped BDD images. Note that the model learns intuitive transformations by dimming taillights or clearing the sky.*

batch size of 32. All experiments were performed on a single workstation equipped with an NVidia Titan X_p GPU in PyTorch [15] v1.0.0.

4.1. Single Attribute flip

In these experiments, we input an attribute value indicating the intensity of attribute that we want to ensure in the reconstructed images. This gives a certain degree of control as to how much we enforce the day and night changes on the input images. Given the label in the test dataset corresponding to each image, the attribute will be flipped so that if the initial label is daytime the image will be flipped to a night image proportional to the flag we pass in. In Figure 1 we can see that our trained model successfully flips the day images to night and vice versa. We augment BDD++ with the images generated with the corresponding flipped attributes of our test set so as to provide day-night pairs.

4.2. Style interpolation

In style interpolation the model reconstructs each image corresponding to each interpolation value and gradually increases this value to iterate over the entire range passed during inference. The model starts interpolating from the least value in the array and gradually increases this value which reflects in the reconstructed output image. In Figure 4 we can see that our trained model successfully changes the style of the image from night to day. Apart from style interpolation this method provides a degree of control over the time-of-day attribute. This method allows for augmentation with additional adverse condition data for autonomous driving research.

Attributes	Day		Night	
	Original	Generated	Original	Generated
Cars	54563	19178	19178	54563
Traffic Signs	7358	5003	5003	7358

Table 1. *Dataset distribution for BDD++. The dataset contains paired images with labels for the time-of-the-day attribute and if they were synthetically generated.*

5. Discussions and Conclusion

In this paper, we present a new approach to train a conditional attribute model to reconstruct traffic scenes with day and night labels. We successfully demonstrate the capability of our trained model to flip and interpolate attributes to change a day traffic scene into night and vice versa. Using the trained attribute model, we create a new dataset, **BDD++** which contains additional reconstructed day and night images. Figure 1 represents this as we see the contrasting appearance of car tail-lights under day and night conditions. Additionally, the choice of using a truncated normal distribution to smooth the attribute samples allows for smoother interpolations as compared to using a simple uniform distribution. We also emphasize that though our model is trained on cropped images, it generalizes to generate full-scale images as seen in the last two rows of Figure 4.

Overall, our work shows that conditional attribute models such as AttGAN can be successfully trained to generate semantically valid traffic scenes to augment existing datasets conditioned on various weather and day/night attributes, thereby facilitating training and testing for safety-critical autonomous driving research.

References

[1] Xi Chen, Yan Duan, Rein Houthooft, John Schulman, Ilya Sutskever, and Pieter Abbeel. Infogan: Interpretable representation learning by information maximizing generative adversarial nets. In *NeurIPS*, 2016. 2

[2] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Scharwächter, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset. In *CVPR*, 2015. 2

[3] Dengxin Dai, Christos Sakaridis, Simon Hecker, and Luc Van Gool. Curriculum model adaptation with synthetic and real data for semantic foggy scene understanding. *IJCV*, 2019. 2

[4] Alexey Dosovitskiy, German Ros, Felipe Codevilla, Antonio Lopez, and Vladlen Koltun. CARLA: An open urban driving simulator. In *Proceedings of the 1st Annual Conference on Robot Learning*, pages 1–16, 2017. 1, 2

[5] Julien Fauqueur, Gabriel Brostow, and Roberto Cipolla. Assisted video object labeling by joint tracking of regions and keypoints. In *ICCV Interactive Computer Vision Workshop.*, 2007. 2

[6] Leon A. Gatys, Alexander S. Ecker, and Matthias Bethge. Image style transfer using convolutional neural networks. In *CVPR*, 2016. 2

[7] Andreas Geiger, Philip Lenz, Christoph Stiller, and Raquel Urtasun. Vision meets robotics: The KITTI dataset. *IJRR*, 2013. 2

[8] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron C. Courville, and Yoshua Bengio. Generative adversarial nets. In *NeurIPS*, 2014. 2

[9] Zhenliang He, Wangmeng Zuo, Meina Kan, Shiguang Shan, and Xilin Chen. Attgan: Facial attribute editing by only changing what you want. *arxiv preprint*, 2017. 1, 2, 3

[10] Judy Hoffman, Eric Tzeng, Taesung Park, Jun-Yan Zhu, Phillip Isola, Kate Saenko, Alexei A. Efros, and Trevor Darrell. Cycada: Cycle-consistent adversarial domain adaptation. In *ICML*, 2018. 2

[11] Diederik Kingma and Jimmy Ba. Adam: a method for stochastic optimization (2014). In *ICLR*, 2015. 3

[12] Guillaume Lample, Neil Zeghidour, Nicolas Usunier, Antoine Bordes, Ludovic Denoyer, et al. Fader networks: Manipulating images by sliding attributes. In *NeurIPS*, 2017. 2

[13] Ming-Yu Liu, Thomas Breuel, and Jan Kautz. Unsupervised image-to-image translation networks. In *NeurIPS*, 2017. 2

[14] Will Maddern, Geoff Pascoe, Chris Linegar, and Paul Newman. 1 Year, 1000km: The Oxford RobotCar Dataset. *The International Journal of Robotics Research (IJRR)*, 36, 2017. 2

[15] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. Automatic differentiation in pytorch. In *NeurIPS-W*, 2017. 4

[16] Stephan R. Richter, Vibhav Vineet, Stefan Roth, and Vladlen Koltun. Playing for data: Ground truth from computer games. In *ECCV*, 2016. 1, 2

[17] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *MICCAI*, 2015. 2

[18] German Ros, Laura Sellart, Joanna Materzynska, David Vazquez, and Antonio Lopez. The SYNTHIA Dataset: A large collection of synthetic images for semantic segmentation of urban scenes. In *CVPR*, 2016. 2

[19] Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Semantic foggy scene understanding with synthetic data. *IJCV*, 2018. 2

[20] Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Semantic nighttime image segmentation with synthetic stylized data, gradual adaptation and uncertainty-aware evaluation. *arXiv preprint arXiv:1901.05946*, 2019. 2

[21] Yuchi Tian, Kexin Pei, Suman Jana, and Baishakhi Ray. Deeptest: Automated testing of deep-neural-network-driven autonomous cars. *ICSE*, 2018. 2

[22] Michal Uricár, Pavel Krízek, David Hurych, Ibrahim M. Sobh, Senthil Yogamani, and Patrick Denny. Yes, we gan: Applying adversarial techniques for autonomous driving. *arXiv preprint arXiv:1902.03442*, 2019. 2

[23] Fisher Yu, Wenqi Xian, Yingying Chen, Fangchen Liu, Mike Liao, Vashisht Madhavan, and Trevor Darrell. Bdd100k: A diverse driving video database with scalable annotation tooling. *arXiv preprint arXiv:1805.04687*, 2018. 2, 3

[24] Mengshi Zhang, Yuqun Zhang, Lingming Zhang, Cong Liu, and Sarfraz Khurshid. Deeproad: Gan-based metamorphic autonomous driving system testing. *arXiv preprint arXiv:1802.02295*, 2018. 2

[25] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. *ICCV*, 2017. 2