Semantic Domain Adaptation for Deep Classifiers via GAN-based Data Augmentation

Amitangshu Mukherjee*
Iowa State University
amimukh@iastate.edu

Ameya Joshi
Iowa State University
ameya@iastate.edu

Chinmay Hegde
New York University
chinmay.h@nyu.edu

Soumik Sarkar
Iowa State University
soumiks@iastate.edu

Abstract

Data augmentation is often used as a strategy to improve the diversity of training data for machine learning systems. While standard augmentation techniques (such as translation and flipping) help neural networks to generalize over spatial transformations, more nuanced techniques would be required to capture semantically different variations in data. We propose a new data augmentation method that relies on the use of attribute-conditioned generative models to modify the semantic properties of existing training data. We show that such data augmentation improves the generalization capability of deep classifiers by analyzing their performance on datasets of traffic objects that are captured (i) at different times of the day and (ii) across different weather conditions.

1 Introduction

Perception systems for autonomous vehicles are built around deep neural networks that analyze images to detect and classify objects of importance. Such neural networks generally require massive amounts of diverse training data, required to be representative of various environmental conditions comprising of adverse illumination and weather conditions. While data augmentation based on affine transformations is often used to improve variety and correct for data imbalance, such techniques seldom capture semantically meaningful variations. Specifically, for traffic scenes, images captured in clear daylight conditions dominate in the training data as compared to those 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, collection and annotation of such diverse traffic data can be resource-intensive and expensive.

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 an Attribute GAN (AttGAN) [9] 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 [8] and MobileNet [11] architectures and show

*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 ISU, a GPU gift grant from NVIDIA Corporation, and faculty fellowships from the Black and Veatch Foundation.

Machine Learning for Autonomous Driving Workshop at the 33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada.

Figure 1: The original and the corresponding synthetic images generated from the AttGAN [9]. The first and third columns depict the original images. The second column depicts reconstructed images with flipped night attribute. The fourth column depicts the reconstructed images with flipped snow mask attribute.

significant improvements in class-wise F_1 scores for BDD++ [19] with day/night and clear/snow images.

2 Related Work

Data augmentation for autonomous driving research has been studied with great detail and there has been several recent work. 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, 21] and games [20]. CARLA [4] uses the UNITY game engine to simulate traffic behaviour and generate high fidelity data. The Synthia dataset [21] is another dataset built along the same lines with rendered city scenes and corresponding segmentation masks. Datasets such as KITTI [6], CamVID [5], Oxford Robotcar [18], Waymo [1], Lyft [13] and Berkeley Deep Drive (BDD) [26] 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 [24] introduces an automated testing framework for DNNs used for autonomous driving by generating affine transformations of images under illumination and weather conditions. DeepRoad [27] improves upon the results using GAN-generated images under snowy and rainy conditions based on the framework of [15]. CyCADA [10] and UNIT [28] ensure semantic constraints on the real and generated images through cyclic consistency loss.

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.[22] augment the original Cityscapes dataset[2] with synthetic fog. Sakaridis et al.[23] 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. [17] adaptively brightens images by learning semantic features in low light conditions using a deep autoencoder. Sakaridis et al.[23] provides a novel pipeline to gradually transfer daytime images to nighttime images based on segmentation masks. Huang et al. [12] uses a GAN generated images to robustify detectors.

Generative Adversarial Networks (GAN) [7] are popularly used as a method to generate samples from real world image distributions. Fader Networks [14] and Attribute GANs et al. [9] extend

		Tested on night-time images			Tested on daytime images		
Setting	Labels	Precision	Recall	F1-Score	Precision	Recall	F1-Score
Trained on day images	car	0.95	0.85	0.89	0.97	0.93	0.95
	person	0.49	0.86	0.63	0.81	0.91	0.86
	traffic light	0.40	0.68	0.50	0.58	0.90	0.70
	traffic sign	0.82	0.76	0.78	0.92	0.85	0.89
Trained on day and night images	car	0.96	0.90	**0.93**	0.96	0.89	0.93
	person	0.52	0.88	0.66	0.65	0.93	0.77
	traffic light	0.57	0.85	**0.68**	0.57	0.88	0.69
	traffic sign	0.91	0.81	0.82	0.90	0.81	0.85
Semantic domain adaptation: Day and synthetic night images	car	0.97	0.87	0.92	0.98	0.96	**0.97**
	person	0.68	0.88	**0.77**	0.89	0.93	**0.91**
	traffic light	0.52	0.62	0.57	0.72	0.86	**0.78**
	traffic sign	0.79	0.90	**0.86**	0.94	0.91	**0.92**

Table 1: *Performance of ResNet 34 classification model trained under three different settings on Night and Day 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. Observe that the model trained on original day and synthetic night images outperforms the model trained only on day images for all four classes under both test sets. This model shows comparable performance to that of the model trained with original day and night images from the dataset for cars, outperforms on persons and traffic signs but under-performs on traffic lights.*

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. [25] 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 Semantic Domain Adaptation

3.1 Preprocessing

For training the AttGAN, we use a sub-sampled version of the BDD++ dataset [19] where we consider original images of the following four classes: cars, traffic signs, traffic lights, and persons. We make these image crops conditioned on the time of the day labels as well as on different weather labels. We select these four object classes to create a balanced training dataset since number of image crops of other categories such as buses and trucks across the dataset are comparatively much less than that of other classes. This creates an imbalance in data which results in improper training of deep generative as well as classification and detection models.

3.2 Training the Generative model

We train two different AttGAN[9] 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".

To train an AttGAN to simulate snowy occlusion effects on the image crops, we use additional synthetic images generated by *DesnowNet* [16] along with original images from the BDD Dataset. This is to compensate for the insufficient amount of original "snow" images in BDD [26]. We condition the generative model to learn this snow occlusion mask and transform images with any weather attribute to exhibit snowy precipitation effects.

3.3 Data Augmentation Strategy

We train ResNet classifiers on the four classes mentioned above on training datasets augmented with semantically transformed images. For generating additional training examples, we use the

		Tested on original snowy images			Tested on synthetic snowy images		
Setting	Labels	Precision	Recall	F1-Score	Precision	Recall	F1-Score
Trained on clear images	car	0.86	0.85	0.86	0.87	0.85	0.85
	person	0.64	0.79	0.71	0.52	0.59	0.56
	traffic light	0.61	0.65	0.63	0.57	0.55	0.56
	traffic sign	0.77	0.67	0.71	0.65	0.64	0.65
Trained on clear and snow images	car	0.89	0.91	0.90	0.89	0.92	0.91
	person	0.76	0.82	0.78	0.68	0.66	0.66
	traffic light	0.64	0.70	0.67	0.57	0.61	0.59
	traffic sign	0.85	0.73	0.69	0.79	0.70	0.75
Trained on clear and synthetic snow images	car	0.95	0.92	**0.94**	0.96	0.95	**0.96**
	person	0.76	0.92	**0.83**	0.76	0.91	**0.83**
	traffic light	0.85	0.87	**0.86**	0.83	0.82	**0.82**
	traffic sign	0.92	0.86	**0.89**	0.90	0.86	**0.88**

Table 2: *Performance of ResNet-34 classification model trained under three different settings on snow 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. Observe that the model trained on original clear and synthetic snow images outperforms the models trained on original images across all four object classes.*

pretrained AttGANs to flip the benign attribute (day) to the corresponding adverse attribute (night). We consider two examples of adverse attributes: night-time and snow. We train three separate ResNet-34 [8] models with three different settings: (1) original day/clear images (2) original day/clear and night/snowy images and (3) original day/clear and synthetic night/snowy images. We test the performances of these classifiers on the four selected classes over shift from day to night and from non-snowy to snowy images.

We transform images with other weather labels (foggy and overcast) from the original dataset instead of clear images to add variety to the training data. For the third setting, we choose only one of the original or synthetic image for training. However, we include the flipped versions in the validation dataset. We also avoid using the original adverse condition images to test the efficacy of our approach.

The models trained on the two categories are then tested on an unseen test set sampled from the original images[2]. 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 precision, recall and F_1 scores, to ensure that the inherent class imbalance does not skew the results.

Results. From Tables 1 and 2, we observe that our GAN based data augmentation strategy is successful at improving classifier performance in adverse settings. Table 1 demonstrates that our approach is able to preserve performance on benign examples. Also note, setting (2) in Table 2 shows that a model trained on the original dataset shows comparable performance on synthetic images. We therefore infer that the our transformation produces realistic images as compared to that of the original data. Additionally, comparing the performance across the three settings shows a significant performance boost for all classes. Our approach therefore allows for *semantically* augmenting under-represented classes to improve performance.

4 Discussion and Conclusion

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. Our analysis of semantic domain adaptation demonstrates promising results for deep classifiers. Additional experiments, however, are required to test the effects of such data augmentation on detectors, segmentation networks and multi-modal networks. We defer this analysis to future work. Additionally, while we show experiments for AttGANs, that are limited to size

[2]The test images are created to validate the robustness of the models under shift from day/clear to night/snowy. During training, we create separate validation and test sets as per the training distribution of that particular model.

constraints, our approach can be extended to better and more sophisticated generative models such as Progressive GANs.

References

[1] Waymo open dataset: An autonomous driving dataset, 2019.

[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.

[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.

[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.

[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.

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

[7] 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.

[8] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2015.

[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.

[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.

[11] Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*, 2017.

[12] Sheng-Wei Huang, Che-Tsung Lin, Shu-Ping Chen, Yen-Yi Wu, Po-Hao Hsu, and Shang-Hong Lai. Auggan: Cross domain adaptation with gan-based data augmentation. In *The European Conference on Computer Vision (ECCV)*, September 2018.

[13] R. Kesten, M. Usman, J. Houston, T. Pandya, K. Nadhamuni, A. Ferreira, M. Yuan, B. Low, A. Jain, P. Ondruska, S. Omari, S. Shah, A. Kulkarni, A. Kazakova, C. Tao, L. Platinsky, W. Jiang, and V. Shet. Lyft level 5 av dataset 2019. `https://level5.lyft.com/dataset/`, 2019.

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

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

[16] Yun-Fu Liu, Da-Wei Jaw, Shih-Chia Huang, and Jenq-Neng Hwang. Desnownet: Context-aware deep network for snow removal. *IEEE Transactions on Image Processing*, 27:3064–3073, 2018.

[17] Kin Gwn Lore, Adedotun Akintayo, and Soumik Sarkar. Llnet: A deep autoencoder approach to natural low-light image enhancement. *Pattern Recognition*, 61:650–662, 2015.

[18] 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.

[19] Amitangshu Mukherjee, Ameya Joshi, Soumik Sarkar, and Chinmay Hegde. Attribute-controlled traffic data augmentation using conditional generative models. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 83–87, 2019.

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

[21] 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.

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

[23] 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.

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

[25] 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.

[26] 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.

[27] 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.

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

A Additional Results

		Tested on night-time images			Tested on daytime images		
Setting	Labels	Precision	Recall	F1-Score	Precision	Recall	F1-Score
Trained on day images	car	0.83	0.71	0.76	0.81	0.64	0.72
	person	0.43	0.65	0.52	0.25	0.58	0.35
	traffic light	0.23	0.52	0.32	0.20	0.38	0.27
	traffic sign	0.63	0.58	0.60	0.57	0.57	0.56
Trained on day and night images	car	0.83	0.64	0.73	0.84	0.64	0.72
	person	0.38	0.54	0.45	0.24	0.51	0.32
	traffic light	0.20	0.57	0.29	0.23	0.52	0.32
	traffic sign	0.57	0.56	0.57	0.59	0.58	0.59
Semantic domain adaptation: Day and synthetic night images	car	0.86	0.81	0.83	0.84	0.74	0.78
	person	0.53	0.64	0.65	0.35	0.76	0.48
	traffic light	0.38	0.59	0.46	0.30	0.46	0.37
	traffic sign	0.77	0.61	0.68	0.67	0.61	0.65

Table 3: *Performance of MobileNetV2 classification model trained under three different settings on Night and Day 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. Observe that the model trained on original day and synthetic night images outperforms the model trained only on original day and night images for all four classes under both test sets.*

		Tested on original snowy images			Tested on synthetic snowy images		
Setting	Labels	Precision	Recall	F1-Score	Precision	Recall	F1-Score
Trained on clear images	car	0.76	0.77	0.76	0.78	0.75	0.76
	person	0.65	0.83	0.73	0.64	0.62	0.63
	traffic light	0.54	0.57	0.56	0.37	0.49	0.42
	traffic sign	0.68	0.56	0.61	0.55	0.52	0.54
Trained on clear and snow images	car	0.81	0.78	0.80	0.84	0.78	0.81
	person	0.69	0.80	0.74	0.63	0.69	0.66
	traffic light	0.46	0.67	0.54	0.36	0.57	0.44
	traffic sign	0.72	0.55	0.62	0.63	0.55	0.59
Semantic domain adaptation: Day and synthetic night images	car	0.86	0.87	0.94	0.91	0.90	0.91
	person	0.72	0.83	0.77	0.67	0.78	0.72
	traffic light	0.74	0.72	0.73	0.68	0.71	0.69
	traffic sign	0.81	0.75	0.78	0.81	0.74	0.77

Table 4: *Performance of MobileNetV2 classification model trained under three different settings on snow 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. Observe that the model trained on original clear and synthetic snow images outperforms the models trained on original images across all four object classes.*

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

Table 5: *Dataset composition used for training the classifiers. Note the data imbalance in individual classes with much fewer nighttime/snowy images. Since we use a conditional GAN for data augmentation, we can leverage images with labels of fog, overcast and partly cloudy labels of the BDD dataset and generate synthetic snow images.*