ETH zürich

ETH Library

Semantic Understanding of Foggy Scenes with Purely Synthetic Data

Conference Paper

Author(s):
Hahner, Martin; Dai, Dengxin; Sakaridis, Christos; Zaech, Jan-Nico; Van Gool, Luc

Publication date:
2019

Permanent link:
https://doi.org/10.3929/ethz-b-000387150

Rights / license:
In Copyright - Non-Commercial Use Permitted

Originally published in:
https://doi.org/10.1109/itsc.2019.8917518

This page was generated automatically upon download from the ETH Zurich Research Collection.
For more information, please consult the Terms of use.

Semantic Understanding of Foggy Scenes with Purely Synthetic Data

Martin Hahner[1], Dengxin Dai[1], Christos Sakaridis[1], Jan-Nico Zaech[1], and Luc Van Gool[1,2]

***Abstract*— This work addresses the problem of semantic scene understanding under foggy road conditions. Although marked progress has been made in semantic scene understanding over the recent years, it is mainly concentrated on clear weather outdoor scenes. Extending semantic segmentation methods to adverse weather conditions like fog is crucially important for outdoor applications such as self-driving cars. In this paper, we propose a novel method, which uses purely synthetic data to improve the performance on unseen real-world foggy scenes captured in the streets of Zurich and its surroundings. Our results highlight the potential and power of photo-realistic synthetic images for training and especially fine-tuning deep neural nets. Our contributions are threefold, 1) we created a purely synthetic, high-quality foggy dataset of 25,000 unique outdoor scenes, that we call *Foggy Synscapes* and plan to release publicly 2) we show that with this data we outperform previous approaches on real-world foggy test data 3) we show that a combination of our data and previously used data can even further improve the performance on real-world foggy data.**

I. Introduction

The last years have seen tremendous progress in tasks relevant to autonomous driving [1]. It has also been hyped that autonomous vehicles of multiple companies have driven for several millions of miles by now. This evaluation or measurement, however, is mainly performed under favorable weather conditions such as the typically great weather in California. In the meanwhile, the development of computer vision algorithms are also focused and benchmarked with clear weather images. As argued in [2], [3], outdoor applications such as automated cars, however, still need to function well in adverse weather conditions. One typical example of an adverse weather condition is fog, which degrades data quality and thus the performance of popular perception algorithms significantly. The challenge exists for both Cameras [4], [3] and LiDAR sensors [5], [6]. This work investigates semantic understanding of foggy scenes with Camera data.

Currently, in the era of deep learning, the most popular and best performing algorithms addressing semantic scene understanding are neural networks trained with many annotations of real images [7], [8]. While this strategy seems to be promising as many algorithms still benefit from having more data, applying the same protocol to all adverse conditions (e.g. fog, rain, snow and nighttime) and their combinations (e.g. foggy night) is problematic. The manual annotation part is hardly scalable to so many domains. This cost of manual annotation is more pronounced for adverse weather conditions, where it is—due to the poor visibility—much harder to provide precise human annotations. This paper aims at improving semantic understanding of real foggy scenes without using additional human annotations of real foggy images.

To overcome this problem, Sakaridis et al. [3] has recently proposed an approach to imposing synthetic fog into real clear weather images and learning with those partially synthetic data. While it generates state of the art results for the task, the method has a few drawbacks. First, the size of the generated partially synthetic dataset is limited by the size of existing datasets created for clear weather condition as the ground truth labels are inherited from the latter. Furthermore, their method requires depth completion and de-noising of real-world scenes which itself is a very challenging and unsolved problem. The imperfect depth maps lead to artifacts in simulated fog. In order to address these two issues, this work takes a step further and develops a method for semantic foggy scene understanding with purely synthetic data. By purely synthetic data, we mean that both the underlying images and the imposed fog are synthetic. Due to the synthetic nature, the images, its corresponding semantic labels and its corresponding depth maps can be obtained easily via running rendering algorithms. More importantly, the depth maps are accurate, which leads to realistic fog simulation. While our method addresses the two problems of [3], the drawback lies in its synthetic nature of underlying scenes, which may lack the richness of real-world scenes.

The main aim of this work is to answer the following two questions: (1) whether purely synthetic data can outperform partially synthetic data for semantic understanding of real foggy scenes; and (2) whether these two complementary data sources (one with better underlying images and the other with better fog effect) can be combined and boost the performance further. The short answer to both questions is yes and the detailed answers are given in the following sections.

To summarize, the main contributions of the paper are: 1) proposing a new purely synthetic dataset for semantic foggy scene understanding which features 25000 high-resolution foggy images; 2) demonstrating that purely synthetic data of foggy scenes with accurate depth information can outperform partially real data with imperfect depth information when tested on real foggy scenes; and 3) demonstrating that the combination of purely synthetic fog data and partially synthetic fog data gives the best results than either of the method alone.

[1]Martin Hahner, Dengxin Dai, Christos Sakaridis, Jan-Nico Zaech and Luc Van Gool are all with the Toyota TRACE-Zurich team at the Computer Vision Lab, ETH Zurich, 8092 Zurich, Switzerland firstname.lastname@vision.ee.ethz.ch

[2]Luc Van Gool is also with the Toyota TRACE-Leuven team at the Dept. of Electrical Engineering ESAT, KU Leuven, 3001 Leuven, Belgium luc.vangool@kuleuven.be

Our work takes advantage of the recent progress in computer graphics for generating realistic synthetic driving data [9], [10], [11]. It also further reinforces the current belief that there is a great potential of learning with high-quality synthetic data.

Foggy Synscapes will be publicly available at `trace.ethz.ch/foggy_synscapes`.

II. Related Work

A large body of recent literature is dedicated to semantic understanding of outdoor scenes under *normal* weather conditions, developing both large-scale annotated datasets [12], [8], [13] and end-to-end trainable models [14], [15], [16], [17] which leverage these annotations to create discriminative representations of the content of such scenes. However, most outdoor vision applications, including autonomous vehicles, also need to remain robust and effective under *adverse* weather or illumination conditions, a typical example of which is fog [2]. The presence of fog degrades the visibility of a scene significantly [4], [18] and the problem of semantic understanding becomes more severe as the fog density increases. Even though the need for specialized methods and datasets for semantic scene understanding under adverse conditions has been pointed out early on in the literature [8], only very recently has the research community responded to this need both in the dataset [19], [20] and in the methodological direction [3], [21], [22], [23], [6], [24]. Our work also answers this need and specifically targets the condition of fog, by constructing a large-scale photo-realistic synthetic foggy dataset that originates from a synthetic clear weather counterpart to enable adaptation to fog.

The SYNTHIA [9] and GTA [10] datasets are the first examples of purely synthetic data—rendered with video game engines—that were used for training in combination with real data to improve semantic segmentation performance on real outdoor scenes, while similar work concurrently considered indoor scenes [25]. The main advantage of these approaches is the drastically reduced cost of ground-truth generation compared to real-world datasets, which require demanding manual annotation. The utility of purely synthetic training data has been further emphasized in [26] for the task of vehicle detection, where a model trained *only* on a massive synthetic dataset is shown to outperform the corresponding model trained on the real large-scale Cityscapes [8] dataset. While all aforementioned works on synthetic data pertain to *normal* conditions, Sakaridis et al. [3] generated Foggy Cityscapes, a partially synthetic foggy dataset created by simulating fog on the original clear weather scenes of Cityscapes [8] and inheriting its ground-truth semantic annotations at no extra cost. The fog simulation pipeline of [3] is improved in [21] by leveraging semantic annotations to increase the accuracy of the required depth map, resulting in the Foggy Cityscapes-DBF dataset. Both of these synthetic foggy datasets have been utilized in [3], [21] to improve semantic segmentation performance of state-of-the-art CNN models [14], [15] on *real* foggy benchmarks. We are inspired by both lines of research and combine the fully controlled setting of purely synthetic data with the synthetic fog generation pipeline of [3], [21]. In particular, we exploit the very recent large-scale photo-realistic Synscapes [11] dataset, which comes with ground-truth depth maps, to generate Foggy Synscapes. The ground-truth depth in Synscapes drastically simplifies the fog simulation pipeline and completely eliminates artifacts in the resulting synthetic images of Foggy Synscapes. By contrast, potentially incorrect estimation of depth values in the complex pipeline of [3], [21] introduces artifacts to Foggy Cityscapes and Foggy Cityscapes-DBF.

Besides the above recent works on pixel-level parsing of foggy scenes, there have also been earlier works on fog detection [27], [28], [29], [30], classification of scenes into foggy and fog-free [31], and visibility estimation both for daytime [32], [33], [34] and nighttime [35], in the context of assisted and autonomous driving. The closest of these works to ours is [32], which generates synthetic fog and segments foggy images to free-space area and vertical objects. However, our semantic segmentation task lies on a higher level of complexity and we employ state-of-the-art CNN architectures, exploiting the most recent advances in this area.

Our work also bears resemblance to domain adaptation methods. [36] focuses on adaptation across weather conditions to parses simple road scenes. More recently, adversarial domain adaptation approaches have been proposed for semantic segmentation, both at pixel level and feature level, adapting models from simulated to real environments [37], [38], [39], [23]. Our work is complementary, as we generate high-quality synthetic foggy data from photo-realistic synthetic clear weather data to enable adaptation to the foggy domain.

III. Method

In this section, we outline our synthetic fog generation pipeline displayed in Figure 1. For more details, we refer the reader to [3], where our pipeline is adapted from.

The standard optical model for fog that forms the basis of this pipeline was introduced in [40] and is expressed as

$$\mathbf{F}(\mathbf{x}) = t(\mathbf{x})\mathbf{R}(\mathbf{x}) + (1 - t(\mathbf{x}))\mathbf{L}, \tag{1}$$

where $\mathbf{F}(\mathbf{x})$ is the observed foggy image at pixel $\mathbf{x}$. $\mathbf{R}(\mathbf{x})$ is the clear scene radiance and $\mathbf{L}$ the atmospheric light, which is assumed to be globally constant. For the atmospheric light estimation we use the same approach as in [3]. For homogeneous fog, the transmittance depends on the distance $\ell(\mathbf{x})$ between the camera and the scene through

$$t(\mathbf{x}) = \exp\left(-\beta \ell(\mathbf{x})\right). \tag{2}$$

Note that the distance $\ell(\mathbf{x})$ in (2) is *not* equivalent to the depth information provided in datasets that focus on automated driving like Cityscapes [8] or Synscapes [11]. The provided depth in those datasets commonly measure the distance between the image plane and the scene. See Figure 2 for an illustration of the difference.

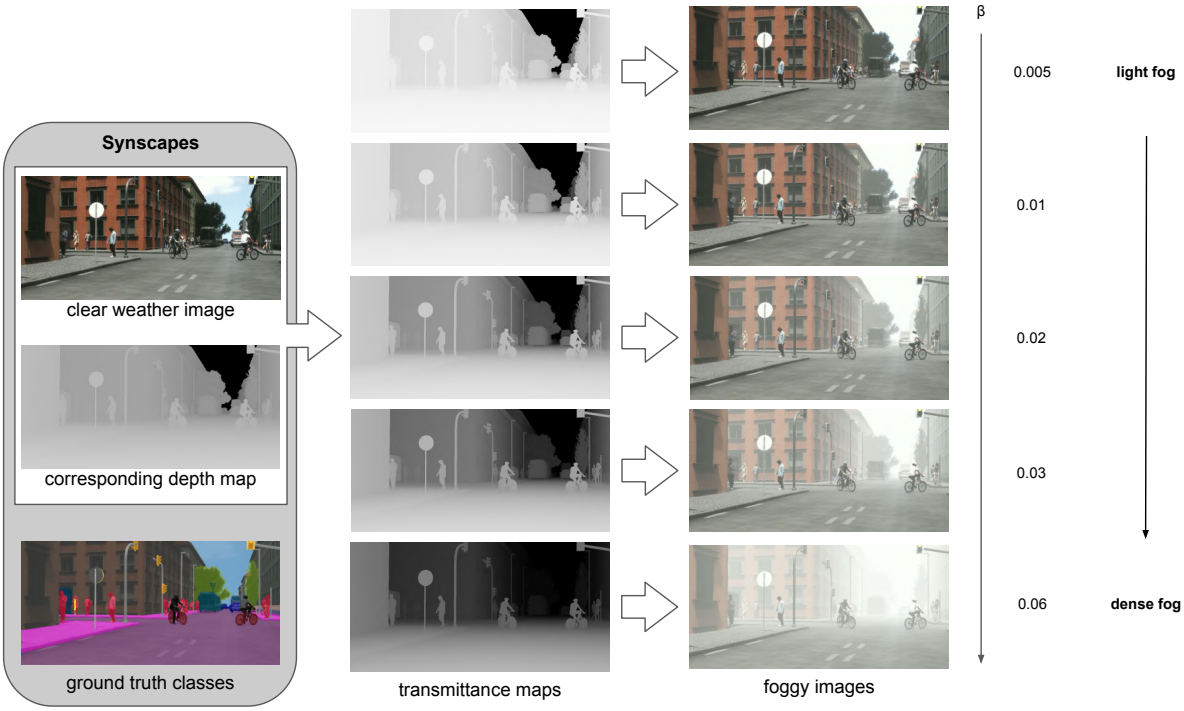

Fig. 1. Our synthetic fog generation pipeline using the gapless depth information provided by the Synscapes [11] dataset.

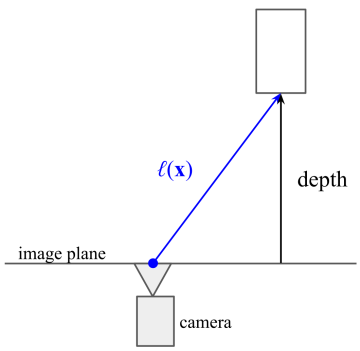

Fig. 2. Difference between $\ell(\mathbf{x})$ and the commonly provided depth information in datasets like Cityscapes [8] and Synscapes [11].

Further, the attenuation coefficient β controls the density of the fog where larger values of β correspond to denser fog. By definition of the National Oceanic and Atmospheric Administration within the U.S. Department of Commerce [41] fog is called fog if it decreases the visibility, among meteorologist more formally called the *meteorological optical range* (MOR), to less than 1 km. For homogeneous fog $\mathrm{MOR} = \frac{2.996}{\beta}$ always holds.

So by the aforementioned definition [41], the lower bound $\beta \geq 0.002996$ corresponds to the lightest fog configuration possible and as a matter of course is always obeyed in our synthetic fog generation pipeline, where $\beta \in [0.005, 0.01, 0.02, 0.03, 0.06]$ is used. These β-values correspond to a visibility of approximately 600m, 300m, 150m, 100m, and 50m respectively.

In contrast to prior work [3], [21] only using real input images (with incomplete and imperfect depth information) to the synthetic fog generation pipeline, our work focuses on synthetic input images (with complete and perfect depth information). Hence, our resulting images are purely synthetic.

For our experiments, we chose the recently released Synscapes [11] dataset. Synscapes is created by an end-to-end approach focusing on photo-realism using the same physically based rendering techniques that power high-end visual effects in the film industry. Those rendering techniques accurately capture the effects of everything from illumination by sun and sky, to the scene's geometry and material composition, to the optics, sensor and processing of the camera system. The dataset consists of 25000 procedural and unique clear weather images that do not follow any path through a given virtual world. Instead, an entirely unique scene is generated for each and every individual image. As a result, the dataset contains a wide range of variations and unique combinations of features.

Results of the presented pipeline for synthetic fog generation on example images from Synscapes [11] are provided in Figure 3 for $\beta = 0$ (which is equivalent to the clear weather input image), 0.02 and 0.06. β-values of 0.02 and 0.06 corresponds to a visibility of approximately 150m and 50m respectively. The required inputs in (1) are the clear weather image $\mathbf{R}$, the atmospheric light $\mathbf{L}$ and the corresponding transmittance map t. We call this new dataset Foggy Synscapes, where the ground truth annotations are inherited as is from Synscapes.

In the rightmost column in Figure 3, where there are

Fig. 3. Comparison of clear weather images from Synscapes [11] against images from our adapted Foggy Synscapes for $\beta = 0$, 0.02, and 0.06.

no clouds present in the sky of the clear weather image, we can also see a rare failure case of our synthetic fog generation pipeline. Due to the missing clouds, in this image, a pixel of the blue sky will be selected as atmospheric light constant, which leads to the bluish tint in the synthetic fog. This is where our assumption of the air being totally homogeneous breaks. In images like the one in the second-rightmost column in Figure 3, when there is blue sky and clouds, our pipeline does not break since it will pick the atmospheric light constant from a pixel in the clouds.

In Figure 4, we qualitatively compare our Foggy Synscapes to Foggy Cityscapes [3]. One can see that the synthetically added fog in our Foggy Synscapes looks far more realistic than the synthetically added fog in Foggy Cityscapes. To a great extent, this is due to the perfect depth information provided by the original Synscapes [11] dataset. If the provided depth was not accurate, the quality of the synthesized foggy images would degrade and we would have artifacts similar to the ones present in Foggy Cityscapes.

IV. Experiments

In this section, we present our findings on two real-world datasets that contain foggy scenes of various densities. The first one is called Foggy Driving [3]. A dataset which is exclusively meant for testing. It contains 101 annotated images for all 19 evaluation classes of Cityscapes [8]. While 33 images are finely annotated for every pixel in the image, the majority of images (68 images) are annotated at a coarser level. 51 of the images were captured with a cell phone camera in foggy conditions at various areas of Zurich and the remaining 50 images were collected from the web. Foggy Driving [3] contains more than 500 annotated vehicles and almost 300 annotated humans.

The second dataset is called Foggy Zurich [21], containing 3808 real-world foggy road scenes captured while driving in the city of Zurich and its suburbs using a GoPro Hero 5 attached to the inside of a car's windshield. Initially, its test split Foggy Zurich-test consisted of 16 images with pixel-level semantic annotations for 18 out of 19 evaluation classes of Cityscapes [8] (the train class is missing). In a more recent work [42], Foggy Zurich-test has been extended and now includes pixel-level semantic annotations for in total 40 images. These 40 images form the test set that we used for our evaluation. Compared to Foggy Driving, Foggy Zurich-test only includes foggy images of uniform and high resolution that are all annotated at a *fine* level.

On the network architecture side, we chose RefineNet [15] to compare with previous work [3] and confirmed our findings with a state of the art real-time semantic segmentation architecture BiSeNet [43]. Regarding RefineNet, we used the same training and fine-tuning policy as in [3]. As second architecture we chose BiSeNet [43] because we believe that lightweight real-time network architectures like BiSeNet [43], which have an order of magnitude less parameters than RefineNet [15], could be the way to tackle various adverse weather conditions such as haze and fog in the future.

The baseline model of the BiSeNet [43] architecture was trained for 80 epochs on $2,975$ clear weather images from the Cityscapes [8] training set using stochastic gradient descent, an initial learning rate of 0.01 and polynomial learning rate decay with power 0.9 as described in [16]. Further, we used momentum 0.09 and weight decay 0.0005. The training always was carried out with a batch size of 4 on a single NVIDIA Titan Xp. Fine-tuning with foggy images produced by our pipeline for all experiments described below was carried out with a $10\times$ lower initial learning rate (0.001) and the remaining hyper-parameters unchanged.

A. Partially vs. Purely Synthetic Data

For this experiment we fine-tuned the baseline models of RefineNet [15] and BiSeNet [43] once with the partially synthetic Foggy Cityscapes [3] and once with our purely synthetic Foggy Synscapes. For fine-tuning on Foggy Cityscapes, we chose the *refined* subset of 498 images, which are of better quality than the complete set of $2,975$

Fig. 4. Qualitative comparison between our Foggy Synscapes (top) and Foggy Cityscapes [3] (bottom) for $\beta = 0.02$.

training images. Fine-tuning on this *refined* set of Foggy Cityscapes was carried out for 50 epochs and validation was executed every epoch (498 images). Only the model with the lowest validation loss was saved for testing on real-world foggy data. Fine-tuning on our Foggy Synscapes was conducted for one epoch of 24500 training images and 500 images were excluded from training and kept as validation set. Validation using our dataset was executed every 125 iterations (500 images) and only the model snapshot with the lowest validation loss was saved for testing here, too.

In both, Table I for Foggy Driving and in Table II for Foggy Zurich-test, we see that fine-tuning on purely synthetic data outperforms fine-tuning on partially synthetic data for both network architectures.

B. Quantity vs. Quality

To even get better numbers on Foggy Synscapes, one could try to fine-tune not only for one, but maybe more epochs. For this paper, however, we wanted to go another way and wished to answer the question whether the benefit of our Foggy Synscapes lies only in its much larger quantity of images or whether it actually lies within its superior fog quality. Therefore we fine-tuned the baseline models of both network architectures only on the first 498 images of Foggy Synscapes. Results on both datasets, presented in Table I for Foggy Driving and Table II for Foggy Zurich-test, illustrate that the benefit truly lies within the quality of the synthetic fog and not just in the much larger scale of the dataset.

Why some experiments with less images are even outperforming the experiment on the full size of Synscapes is a bit surprising. It could be that the first 498 images we selected contain less error cases as the one visualized in the rightmost column of Figure 3. This investigation of why exactly this is happening is left for future work. One could also imagine to explicitly filter out those failure cases by using the provided meta-data parameter *sky_contrast* of the original Synscapes [11] dataset. This parameter defines the contrast of the sky, where values between 2-3 indicate fully overcast sky and higher values between 5-6 indicate direct sunlight (which notably increase the chance of such failure cases).

C. Combination of Partially & Purely Synthetic Data

Finally, we also investigate what happens if we fine-tune on the combination of both datasets, Foggy Cityscapes [3] and our Foggy Synscapes. Therefore we mixed the two datasets with a $2:1$ ratio favouring our Foggy Synscapes, meaning for every two images of Foggy Synscapes, there is one Foggy Cityscapes image in the training and validation set. Ratios of $1:1$ and $1:5$ were also tested, but were not as beneficial as the ratio $2:1$. Note that in this setting all images from Foggy Cityscapes had to be used multiple times since its *refined* set of 498 images is significantly smaller than our Foggy Synscapes with 24500 training images. Results of this experiment can also be seen in Table I for Foggy Driving and Table II for Foggy Zurich-test. Improved performance on both datasets generally indicate that the mixture of both datasets is in parts significantly more powerful than one on its own.

D. Discussion

All results presented are consistently achieved using $\beta = 0.01$ for Foggy Cityscapes [3] and $\beta = 0.005$ for our Foggy Synscapes. For the experiment where we combined both datasets, best results are achieved with $\beta = 0.005$. Higher β-values were also explored, but its results were not as beneficial as the β-values 0.005 and 0.01. Experiments using $\beta \in [0.03, 0.06]$ sometimes even showed a drop in performance compared to the clear weather baseline model.

Figure 5 shows four exemplary images of Foggy Zurich-test that show representative predictions using our Foggy Synscapes dataset. The top two rows illustrate results using the RefineNet [15] and the bottom two rows illustrate results using the BiSeNet [43] architecture. While improvements can be observed, the performance on foggy scenes is still a lot worse compared to what other papers may illustrate on clear weather scenes. This supports our claim that foggy scenes are indeed (way) more challenging than clear weather scenes.

Model	Training	Fine-Tuning (# of images)	road	sidewalk	building	wall	fence	pole	tr.light	tr.sign	vegetation	terrain	sky	person	rider	car	truck	bus	train	motorcycle	bicycle	**mean IoU**
RefineNet [15]	C	-	90.1	29.3	68.3	27.3	**16.7**	41.3	54.2	59.6	68.0	6.8	88.7	**60.9**	45.4	66.4	5.5	9.6	45.4	9.8	48.4	44.3
	C	FC (498)	91.7	29.7	73.0	**29.0**	14.8	43.4	54.0	**61.6**	71.2	6.9	85.7	59.3	46.7	67.3	8.4	17.2	53.7	13.1	48.9	46.1
	C	FS (24,500)	**92.4**	32.9	**76.1**	16.8	14.6	43.3	55.0	60.8	74.0	9.3	90.8	49.8	36.0	**72.2**	17.5	**51.3**	65.0	11.1	50.3	48.4
	C	FS (498)	91.1	**34.9**	74.4	18.4	15.7	43.7	**57.0**	60.0	75.8	10.0	**92.7**	55.5	**47.3**	71.8	**24.8**	30.0	48.0	**17.3**	**54.4**	48.6
	C	FC + FS (498)	**92.4**	34.0	**76.1**	23.9	16.2	**45.6**	55.9	**61.6**	**76.4**	**11.1**	92.2	57.5	45.6	69.9	13.7	42.3	**82.2**	14.1	52.6	**50.7**
BiSeNet [43]	C	-	84.2	13.6	**75.9**	40.9	**12.4**	7.2	14.8	25.3	56.8	4.0	50.7	56.5	13.6	83.5	0.2	9.8	0.7	0.0	26.7	27.2
	C	FC (498)	**90.4**	18.1	67.3	36.3	6.4	10.2	26.9	23.2	**72.1**	1.4	66.9	47.4	11.4	**84.4**	0.1	13.2	4.5	0.0	11.8	30.3
	C	FS (24,500)	75.0	15.7	64.5	30.7	8.7	24.6	27.3	34.7	50.0	12.6	88.8	56.4	33.3	74.2	8.0	40.8	**35.6**	0.0	30.3	30.9
	C	FS (498)	76.5	14.9	62.4	30.1	10.2	**28.8**	34.4	40.6	57.0	4.4	**92.6**	**61.7**	**48.6**	77.0	**9.3**	58.6	13.2	0.0	29.0	31.8
	C	FC + FS (498)	79.0	**19.0**	67.0	**55.9**	12.3	27.1	**36.6**	**41.1**	59.9	**18.4**	83.9	60.4	33.8	83.6	6.1	**73.3**	16.1	0.0	**40.0**	**35.2**

TABLE I

TEST RESULTS ON FOGGY DRIVING.

C = CITYSCAPES, FC = FOGGY CITYSCAPES, FS = FOGGY SYNSCAPES

Model	Training	Fine-Tuning (# of images)	road	sidewalk	building	wall	fence	pole	tr.light	tr.sign	vegetation	terrain	sky	person	rider	car	truck	bus	train	motorcycle	bicycle	**mean IoU**
RefineNet [15]	C	-	74.3	56.5	35.5	20.2	23.8	39.6	54.4	58.3	58.3	28.9	66.8	1.6	**27.4**	81.7	0.0	1.9	-	21.1	**6.2**	34.6
	C	FC (498)	81.2	56.7	36.5	27.5	24.6	44.2	59.6	57.8	48.2	33.6	50.2	5.3	25.3	81.9	0.0	**29.2**	-	36.0	3.1	36.9
	C	FS (24,500)	83.6	60.0	46.6	31.9	33.6	45.1	62.2	61.5	68.3	35.2	79.0	4.3	21.5	**82.0**	0.0	0.2	-	44.7	5.1	40.3
	C	FS (498)	**87.9**	59.5	**54.5**	40.9	**44.8**	47.8	**63.6**	**62.5**	**74.1**	**39.3**	**84.9**	4.5	24.6	75.3	0.0	0.1	-	43.5	4.7	**42.7**
	C	FC + FS (498)	87.5	**60.6**	46.0	**41.1**	38.5	**48.2**	62.4	61.9	67.3	38.1	74.4	**6.2**	22.5	80.8	0.0	1.7	-	**45.9**	3.8	41.4
BiSeNet [43]	C	-	63.9	22.7	61.1	9.5	31.6	3.9	4.1	10.6	45.9	12.4	27.6	0.0	0.0	75.6	0.0	0.0	-	0.0	0.0	16.1
	C	FC (498)	72.3	32.5	**73.3**	7.0	22.6	9.5	**31.2**	20.7	66.9	23.6	50.5	0.0	19.3	80.8	0.0	0.1	-	**14.3**	0.0	25.0
	C	FS (24,500)	60.7	37.9	71.9	23.4	21.7	17.0	17.4	36.9	58.7	31.8	81.6	**14.7**	**27.5**	74.2	0.0	0.0	-	0.0	0.0	27.8
	C	FS (498)	56.4	31.8	62.9	22.4	11.6	20.3	20.8	40.4	63.4	41.7	**88.1**	1.7	**27.5**	77.6	0.0	1.1	-	0.0	0.0	27.6
	C	FC + FS (498)	**73.7**	**38.4**	72.1	**26.8**	**34.5**	**24.5**	27.1	**42.2**	**69.5**	**47.8**	81.2	3.5	9.3	**84.9**	0.0	**3.2**	-	0.0	0.0	**30.9**

TABLE II

TEST RESULTS ON FOGGY ZURICH-TEST.

C = CITYSCAPES, FC = FOGGY CITYSCAPES, FS = FOGGY SYNSCAPES

Since our Foggy Synscapes is so much larger in scale than Foggy Cityscapes [3], we also tried to train models directly on Foggy Synscapes (starting from ImageNet [7] pretrained weights). But unfortunately the resulting models could not outperform the clear weather *Cityscapes* baseline models on real-world foggy data. So we concluded that the impact of real-world texture as present in Cityscapes [8] seems to be still more important for the model "core" than the appearance of photo-realistic fog.

V. CONCLUSION

In this paper, we were able to demonstrate that purely synthetic data of foggy scenes with accurate depth information can outperform partially real data with imperfect depth information when tested on real foggy scenes. We also showed that a combination of both types of data, purely synthetic and partially real, can further improve the scene understanding of real world foggy scenes. This is because in the setting where we merge the two datasets, the best of both worlds is combined. On the one side we have Foggy Cityscapes [3], which contains real textures from Cityscapes [8] but imperfect fog and on the other side we have our Foggy Synscapes, which has imperfect texture but much better looking fog than Foggy Cityscapes. We also learned that the power of Foggy Synscapes does not lie in its larger quantity, but mainly in its better fog quality. This further supports the assumption by the research community that there is great potential of evermore realistic looking synthetic data.

REFERENCES

[1] S. Hecker, D. Dai, and L. Van Gool, "End-to-end learning of driving models with surround-view cameras and route planners," in *European Conference on Computer Vision (ECCV)*, 2018.

[2] S. G. Narasimhan and S. K. Nayar, "Vision and the atmosphere," *Int. J. Comput. Vision*, vol. 48, no. 3, pp. 233–254, Jul. 2002.

[3] C. Sakaridis, D. Dai, and L. Van Gool, "Semantic foggy scene understanding with synthetic data," *International Journal of Computer Vision*, 2018.

[4] S. G. Narasimhan and S. K. Nayar, "Contrast restoration of weather degraded images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 6, pp. 713–724, Jun. 2003.

[5] M. Kutila, P. Pyyknen, W. Ritter, O. Sawade, and B. Schufele, "Automotive lidar sensor development scenarios for harsh weather conditions," in *International Conference on Intelligent Transportation Systems (ITSC)*, 2016.

[6] M. Bijelic, T. Gruber, and W. Ritter, "Benchmarking image sensors under adverse weather conditions for autonomous driving," in *Intelligent Vehicles Symposium (IV)*, 2018, pp. 1773–1779.

[7] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein *et al.*, "Imagenet large scale visual recognition challenge," *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, 2015.

[8] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The Cityscapes dataset for semantic urban scene understanding," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[9] G. Ros, L. Sellart, J. Materzynska, D. Vazquez, and A. M. Lopez, "The SYNTHIA dataset: A large collection of synthetic images for semantic segmentation of urban scenes," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

[10] S. R. Richter, V. Vineet, S. Roth, and V. Koltun, "Playing for data: Ground truth from computer games," in *European Conference on Computer Vision*. Springer, 2016.

[11] M. Wrenninge and J. Unger, "Synscapes: A photorealistic synthetic dataset for street scene parsing," *CoRR*, vol. abs/1810.08705, 2018. [Online]. Available: http://arxiv.org/abs/1810.08705

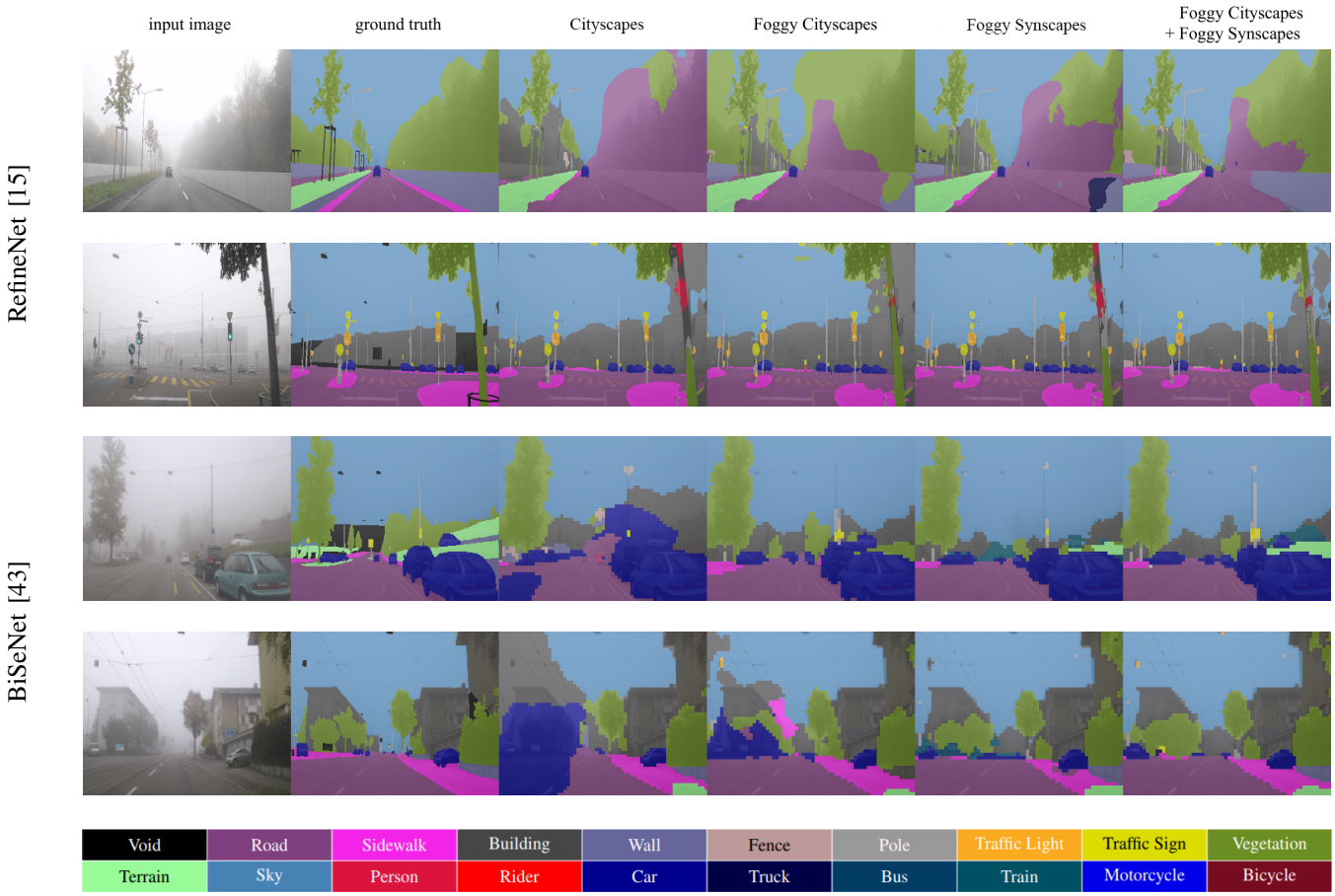

Fig. 5. Representative RefineNet [15] predictions (top two rows) and BiSeNet [43] predictions (bottom two rows) on input images from Foggy Zurich predicted by the Cityscapes baseline model, alongside the predictions of the same model fine-tuned on Foggy Cityscapes, Foggy Synscapes, and the combination of both. This figure is best viewed in color and on screen.

[12] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? The KITTI vision benchmark suite," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012.

[13] G. Neuhold, T. Ollmann, S. Rota Bul, and P. Kontschieder, "The Mapillary Vistas dataset for semantic understanding of street scenes," in *The IEEE International Conference on Computer Vision (ICCV)*, 2017.

[14] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," in *International Conference on Learning Representations*, 2016.

[15] G. Lin, A. Milan, C. Shen, and I. Reid, "Refinenet: Multi-path refinement networks with identity mappings for high-resolution semantic segmentation," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[16] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[17] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *The European Conference on Computer Vision (ECCV)*, 2018.

[18] R. T. Tan, "Visibility in bad weather from a single image," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2008.

[19] O. Zendel, K. Honauer, M. Murschitz, D. Steininger, and G. Fernandez Dominguez, "WildDash - creating hazard-aware benchmarks," in *The European Conference on Computer Vision (ECCV)*, September 2018.

[20] C. Sakaridis, D. Dai, , and L. Van Gool, "Guided curriculum model adaptation and uncertainty-aware evaluation for semantic nighttime image segmentation," in *International Conference on Computer Vision (ICCV)*, 2019.

[21] C. Sakaridis, D. Dai, S. Hecker, and L. Van Gool, "Model adaptation with synthetic and real data for semantic dense foggy scene understanding," in *European Conference on Computer Vision (ECCV)*, 2018.

[22] D. Dai and L. Van Gool, "Dark model adaptation: Semantic image segmentation from daytime to nighttime," in *IEEE International Conference on Intelligent Transportation Systems*, 2018.

[23] M. Wulfmeier, A. Bewley, and I. Posner, "Incremental adversarial domain adaptation for continually changing environments," in *International Conference on Robotics and Automation (ICRA)*, 2018.

[24] M. Bijelic, F. Mannan, T. Gruber, W. Ritter, K. Dietmayer, and F. Heide, "Seeing through fog without seeing fog: Deep sensor fusion in the absence of labeled training data," *CoRR*, vol. abs/1902.08913, 2019. [Online]. Available: http://arxiv.org/abs/1902.08913

[25] A. Handa, V. Patraucean, V. Badrinarayanan, S. Stent, and R. Cipolla, "Understanding real world indoor scenes with synthetic data," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[26] M. Johnson-Roberson, C. Barto, R. Mehta, S. N. Sridhar, K. Rosaen, and R. Vasudevan, "Driving in the matrix: Can virtual worlds replace human-generated annotations for real world tasks?" in *IEEE International Conference on Robotics and Automation*, 2017.

[27] S. Bronte, L. M. Bergasa, and P. F. Alcantarilla, "Fog detection system based on computer vision techniques," in *International IEEE Conference on Intelligent Transportation Systems*, 2009.

[28] M. Pavlić, H. Belzner, G. Rigoll, and S. Ilić, "Image based fog detection in vehicles," in *IEEE Intelligent Vehicles Symposium*, 2012.

[29] R. Gallen, A. Cord, N. Hautire, and D. Aubert, "Towards night fog detection through use of in-vehicle multipurpose cameras," in *IEEE Intelligent Vehicles Symposium (IV)*, 2011.

[30] R. Spinneker, C. Koch, S. B. Park, and J. J. Yoon, "Fast fog detection for camera based advanced driver assistance systems," in *International IEEE Conference on Intelligent Transportation Systems (ITSC)*, 2014.

[31] M. Pavlić, G. Rigoll, and S. Ilić, "Classification of images in fog and fog-free scenes for use in vehicles," in *IEEE Intelligent Vehicles Symposium (IV)*, 2013.

[32] J. P. Tarel, N. Hautire, A. Cord, D. Gruyer, and H. Halmaoui, "Improved visibility of road scene images under heterogeneous fog," in *IEEE Intelligent Vehicles Symposium*, 2010, pp. 478–485.

[33] R. C. Miclea and I. Silea, "Visibility detection in foggy environment," in *International Conference on Control Systems and Computer Science*, 2015.

[34] N. Hautire, J.-P. Tarel, J. Lavenant, and D. Aubert, "Automatic fog detection and estimation of visibility distance through use of an onboard camera," *Machine Vision and Applications*, vol. 17, no. 1, pp. 8–20, 2006.

[35] R. Gallen, A. Cord, N. Hautire, É. Dumont, and D. Aubert, "Nighttime visibility analysis and estimation method in the presence of dense fog," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 1, pp. 310–320, 2015.

[36] E. Levinkov and M. Fritz, "Sequential bayesian model update under structured scene prior for semantic road scenes labeling," in *IEEE International Conference on Computer Vision*, 2013.

[37] A. Shrivastava, T. Pfister, O. Tuzel, J. Susskind, W. Wang, and R. Webb, "Learning from simulated and unsupervised images through adversarial training," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[38] S. Sankaranarayanan, Y. Balaji, A. Jain, S. Nam Lim, and R. Chellappa, "Learning from synthetic data: Addressing domain shift for semantic segmentation," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[39] J. Hoffman, E. Tzeng, T. Park, J.-Y. Zhu, P. Isola, K. Saenko, A. Efros, and T. Darrell, "CyCADA: Cycle-consistent adversarial domain adaptation," in *International Conference on Machine Learning*, 2018.

[40] H. Koschmieder, "Theorie der horizontalen Sichtweite," *Beitrage zur Physik der freien Atmosphre*, 1924.

[41] *Federal Meteorological Handbook No. 1: Surface Weather Observations and Reports*. U.S. Department of Commerce / National Oceanic and Atmospheric Administration, 2005.

[42] D. Dai, C. Sakaridis, S. Hecker, and L. Van Gool, "Curriculum model adaptation with synthetic and real data for semantic foggy scene understanding," *International Journal of Computer Vision (IJCV)*, 2019.

[43] C. Yu, J. Wang, C. Peng, C. Gao, G. Yu, and N. Sang, "Bisenet: Bilateral segmentation network for real-time semantic segmentation," *CoRR*, vol. abs/1808.00897, 2018. [Online]. Available: http://arxiv.org/abs/1808.00897