Semantic Scene Understanding under Adverse Weather and Lighting Conditions

Dengxin Dai (by Simon Hecker)

Computer Vision Lab, ETH Zurich
- Toyota Research on Automated Cars in Europe (TRACE)
- All TRACE partners: Cambridge, Leuven, Saarbrücken, Prague, Zurich
- TRACE-Zurich team is dedicated to exploratory research
Computer Vision Datasets

Cityscapes

Pascal VOC

KITTI

Mapillary vistas

MSCOCO

NYU Depth Dataset
Autonomous Driving (TRACE)
Motivation

https://www.postauto.ch/en/smartshuttle-video-clips

https://www.aptiv.com/autonomous-mobility

https://vimeo.com/296595830

https://www.postauto.ch/en/smartshuttle-video-clips

https://www.youtube.com/watch?v=x4jd4E7rZFE

https://www.youtube.com/watch?v=tlThdr3O5Qo

https://www.youtube.com/watch?v=MaGb3570K1U

https://www.youtube.com/watch?v=OjDLwnTxyh0

https://www.youtube.com/watch?v=OjFWN18CEKQU

https://www.youtube.com/watch?v=xK77K2QqJ8E

Clear Weather
Vision for all seasons
Robust Perception under Adverse Conditions

Knowledge Learning

Large-scale annotations
➤ Expensive to create
➤ But exist for clear weather

Adverse Driving Conditions

• Synthetic data
• Semi-supervised learning

• Transfer learning
• Regularization
Robust Perception for All Conditions: Talk Structure

Techniques:

• Synthetic Data
• Curriculum Domain Adaptation
• Power of Place
• Image Translation

Applications

• Semantic Foggy Scene Understanding
• Semantic Nighttime Scene Understanding
Semantic Foggy Scene Understanding with Synthetic Data

a) Fog Simulation

left image  right image  depth  fog

labeled  fog

unlabeled  fog

b) Training with Synthetic Fog

c) Foggy Scene Understanding
Fog simulation method – using semantics

- Real scenes + valid physical model
- Simulation is stable, controllable, accurate
- Semantic masks are optional
Simulated foggy scene with varying fog densities

Data and Code are publicly available!
Foggy Cityscapes are available on Cityscapes’ server.

Fog Simulation

- **Image (clear weather)**
- **Human annotation**
- **Semi-synthetic fog**
- **Human annotation**

**Segmentation Model**
- (clear weather)
- (foggy weather)

**Supervised Learning**
- **Object Recognition**

**Unlabeled image**
- (clear weather)

**Recognition result**

**Real foggy image**

- **Recognition result**

**Fog Simulation**

**Object Recognition**

**Table:**
- **Fence**
- **Pole**
- **Traffic Light**
- **Traffic Sign**
- **Vegetation**
- **Truck**
- **Bus**
- **Train**
- **Motorcycle**
- **Bicycle**
- **Void**
- **Road**
- **Sidewalk**
- **Building**
- **Wall**
- **Terrain**
- **Sky**
- **Person**
- **Rider**
- **Car**
Supervised Learning with Semi-Synthetic Fog

- Image (clear weather)
- Human annotation
- Semi-synthetic fog
- Human annotation

Supervised Learning

Segmentation Model (clear weather)

Object Recognition

Unlabeled image (clear weather)

Recognition result

Supervised Learning

Segmentation Model (foggy weather)

Object Recognition

Real foggy image

Recognition result

Fog Simulation

<table>
<thead>
<tr>
<th>Fence</th>
<th>Pole</th>
<th>Traffic Light</th>
<th>Traffic Sign</th>
<th>Vegetation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Truck</td>
<td>Bus</td>
<td>Train</td>
<td>Motorcycle</td>
<td>Bicycle</td>
</tr>
<tr>
<td>Void</td>
<td>Road</td>
<td>Sidewalk</td>
<td>Building</td>
<td>Wall</td>
</tr>
<tr>
<td>Terrain</td>
<td>Sky</td>
<td>Person</td>
<td>Rider</td>
<td>Car</td>
</tr>
</tbody>
</table>
Supervision Transfer from Clear Weather to Foggy Weather

- Image (clear weather)
- Human annotation
- Semi-synthetic fog
- Human annotation

Supervised Learning

Segmentation Model (clear weather)

Unlabeled image (clear weather)

Recognition result

Fog Simulation

Segmentation Model (foggy weather)

Real foggy image

Recognition result

<table>
<thead>
<tr>
<th>Fence</th>
<th>Pole</th>
<th>Traffic Light</th>
<th>Traffic Sign</th>
<th>Vegetation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Truck</td>
<td>Bus</td>
<td>Train</td>
<td>Motorcycle</td>
<td>Bicycle</td>
</tr>
<tr>
<td>Void</td>
<td>Road</td>
<td>Sidewalk</td>
<td>Building</td>
<td>Wall</td>
</tr>
<tr>
<td>Terrain</td>
<td>Sky</td>
<td>Person</td>
<td>Rider</td>
<td>Car</td>
</tr>
</tbody>
</table>
Supervision Transfer from Clear Weather to Foggy Weather

- Image (clear weather)
- Human annotation
- Semi-synthetic fog
- Human annotation

Supervised Learning

Object Recognition

Segmentation Model (clear weather)

Unlabeled image (clear weather)

Fog Simulation

Recognition result

Real foggy image

Segmentation Model (foggy weather)

Recall

Object Recognition

- Supervised Learning
- Object Recognition

Recognition result

Real foggy image

Object Recognition

- Supervised Learning
- Object Recognition

Unlabeled image (clear weather)

Supervision Transfer from Clear Weather to Foggy Weather

- Image (clear weather)
- Human annotation
- Semi-synthetic fog
- Human annotation

Supervised Learning

Object Recognition

Segmentation Model (clear weather)

Unlabeled image (clear weather)

Fog Simulation

Recognition result

Real foggy image

Segmentation Model (foggy weather)

Recall

Object Recognition

- Supervised Learning
- Object Recognition

Recognition result

Real foggy image

Object Recognition

- Supervised Learning
- Object Recognition

Unlabeled image (clear weather)

Supervision Transfer from Clear Weather to Foggy Weather

- Image (clear weather)
- Human annotation
- Semi-synthetic fog
- Human annotation

Supervised Learning

Object Recognition

Segmentation Model (clear weather)

Unlabeled image (clear weather)

Fog Simulation

Recognition result

Real foggy image

Segmentation Model (foggy weather)

Recall

Object Recognition

- Supervised Learning
- Object Recognition

Recognition result

Real foggy image

Object Recognition

- Supervised Learning
- Object Recognition

Unlabeled image (clear weather)

Supervision Transfer from Clear Weather to Foggy Weather

- Image (clear weather)
- Human annotation
- Semi-synthetic fog
- Human annotation

Supervised Learning

Object Recognition

Segmentation Model (clear weather)

Unlabeled image (clear weather)

Fog Simulation

Recognition result

Real foggy image

Segmentation Model (foggy weather)

Recall

Object Recognition

- Supervised Learning
- Object Recognition

Recognition result

Real foggy image

Object Recognition

- Supervised Learning
- Object Recognition

Unlabeled image (clear weather)

Supervision Transfer from Clear Weather to Foggy Weather

- Image (clear weather)
- Human annotation
- Semi-synthetic fog
- Human annotation

Supervised Learning

Object Recognition

Segmentation Model (clear weather)

Unlabeled image (clear weather)

Fog Simulation

Recognition result

Real foggy image

Segmentation Model (foggy weather)

Recall

Object Recognition

- Supervised Learning
- Object Recognition

Recognition result

Real foggy image

Object Recognition

- Supervised Learning
- Object Recognition

Unlabeled image (clear weather)

Supervision Transfer from Clear Weather to Foggy Weather

- Image (clear weather)
- Human annotation
- Semi-synthetic fog
- Human annotation

Supervised Learning

Object Recognition

Segmentation Model (clear weather)

Unlabeled image (clear weather)

Fog Simulation

Recognition result

Real foggy image

Segmentation Model (foggy weather)

Recall

Object Recognition

- Supervised Learning
- Object Recognition

Recognition result

Real foggy image

Object Recognition

- Supervised Learning
- Object Recognition

Unlabeled image (clear weather)
Fog Simulation

Image (clear weather)  Human annotation  Semi-synthetic fog  Human annotation

Supervised Learning

Segmentation Model (clear weather)

Unlabeled image (clear weather)  Recognition result

Supervised Learning

Object Recognition

Segmentation Model (foggy weather)

Real foggy image  Recognition result

Supervised Learning

Object Recognition

Fog Simulation

Table:

<table>
<thead>
<tr>
<th>Fence</th>
<th>Pole</th>
<th>Traffic Light</th>
<th>Traffic Sign</th>
<th>Vegetation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Truck</td>
<td>Bus</td>
<td>Train</td>
<td>Motorcycle</td>
<td>Bicycle</td>
</tr>
<tr>
<td>Void</td>
<td>Road</td>
<td>Sidewalk</td>
<td>Building</td>
<td>Wall</td>
</tr>
<tr>
<td>Terminus</td>
<td>Sky</td>
<td>Person</td>
<td>Rider</td>
<td>Car</td>
</tr>
</tbody>
</table>
Semantic Foggy Scene Understanding with Synthetic Data

- The method is simple, yet effective
- The baseline of image dehazing + semantic segmentation not working well
- Synthetic foggy data still has artifacts
  - The estimation of atmosphere light is not always correct, e.g. images with direct sunlight
  - Depth maps of real images are also noisy
  - The artifacts are more severe for dense fog

Addressing the artifacts:
1. Imposing fog on synthetic images which have perfect depth maps
2. Using unlabeled real foggy data
Foggy Synscapes dataset: impose fog on synthetic images

- 25000 HR images, less artefacts, synthetic texture.
- Purely synthetic data along works better than semi-synthetic data so far
- The combination of purely synthetic data and the semi-synthetic data yield the best results

Data and Code available
Curriculum Domain Adaptation

Using unlabeled real fog

Curriculum Model Adaptation with Synthetic and Real Data for Semantic Dense Foggy Scene Understanding, ECCV 2018, IJCV 2019
## Results

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>RefineNet [40]</strong></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td>34.6</td>
<td>35.8</td>
<td>44.3</td>
</tr>
<tr>
<td>SFSU [59]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td>35.7</td>
<td>35.9</td>
<td>46.3</td>
</tr>
<tr>
<td>AdSegNet [68]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td>25.0</td>
<td>15.8</td>
<td>29.7</td>
</tr>
<tr>
<td>CMAda2 [58]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>42.9</td>
<td>37.3</td>
<td>48.5</td>
</tr>
<tr>
<td>CMAda3+</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>46.8</td>
<td>43.0</td>
<td>49.8</td>
</tr>
</tbody>
</table>

### Ablation Study

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Baseline [40]</strong></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td>34.6</td>
<td>35.8</td>
<td>44.3</td>
</tr>
<tr>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td>36.3</td>
<td>36.1</td>
<td>46.3</td>
</tr>
<tr>
<td>CMAda1</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>37.5</td>
<td>36.4</td>
<td>45.7</td>
</tr>
<tr>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>39.8</td>
<td>35.7</td>
<td>47.5</td>
</tr>
<tr>
<td>CMAda2</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>41.5</td>
<td>37.0</td>
<td>48.5</td>
</tr>
<tr>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>40.6</td>
<td>35.5</td>
<td>47.7</td>
</tr>
<tr>
<td>CMAda3</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>42.9</td>
<td>37.3</td>
<td>48.5</td>
</tr>
<tr>
<td>CMAda3+</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>43.7</td>
<td>40.6</td>
<td>48.9</td>
</tr>
<tr>
<td>CMAda2+</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>43.4</td>
<td>40.1</td>
<td>49.9</td>
</tr>
<tr>
<td>CMAda3+</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>46.8</td>
<td>43.0</td>
<td>49.8</td>
</tr>
</tbody>
</table>
Nighttime Semantic Image Segmentation

Challenges (large domain gap):
- Over-/under-exposure
- Headlight, street lamp
- Reflection
- Motion blur …

Transfer Learning

Curriculum Model Adaptation and Uncertainty-aware Evaluation for Nighttime Semantic Image Segmentation, ICCV 2019
(15:30 Thursday, Poster ID: 196)
Nighttime Semantic Image Segmentation

- Method
  - Power of Time
  - Power of Place
  - Power of Data
- Uncertainty-aware Annotation Evaluation
- Dark Zurich Dataset: a high-quality nighttime semantic seg dataset
Daytime to Nighttime, what is in between?

Areas of the Earth in daylight, twilight and night. (Source: Wikipedia)
Daytime to Nighttime, what is in between?

Twilight

Twilight categories

10/26/19
Curriculum Model Adaptation: Power of Time

Daytime

Nighttime

Civil Twilight

Nautical Twilight

Astronomical Twilight
Curriculum Model Adaptation

- Daytime images
- Human Annotation
- Nighttime images
- Results

Train

Segmentation Model (daytime)

Fine-Tune

Infer

Segmentation Model (twilight)

Infer

Twilight images

Generated Semantic Labels
Images taken under the ‘same’ 6D camera pose but at different time of the day
Annotated day-time images can be converted to nighttime style and be used for supervised training.

- The three strategies: the power of time, the power of place, and the power of data have their own strengths and weaknesses.
- Using them together can compensate the weaknesses.

Curriculum Model Adaptation and Uncertainty-aware Evaluation for Nighttime Semantic Image Segmentation, ICCV 2019
(15:30 Thursday, Poster ID: 196)
Uncertainty-aware Evaluation

Uncertainty-aware predictions: confidence threshold $\theta$ on soft predictions to invalidate pixels

$$\tilde{H}(p) = \arg \max_{c \in C} \{s_c(p)\}$$

$$\theta \in [1/C, 1]$$

$$\tilde{H}(p) = \begin{cases} 
\tilde{H}(p) & \text{if } s_{\tilde{H}(p)}(p) \geq \theta, \\
\text{invalid} & \text{otherwise.}
\end{cases}$$

$$\text{UIoU}(\theta) = \text{UIoU}(1/C) = \text{IoU}$$

*Dark Zurich-test*

151 images

UIoU incorporates indiscernible regions with labels
# Results

Comparison on **Dark Zurich-test** with daytime baselines and competing adaptation methods

<table>
<thead>
<tr>
<th>Method</th>
<th>road</th>
<th>sidewalk</th>
<th>build.</th>
<th>wall</th>
<th>fence</th>
<th>pole</th>
<th>light</th>
<th>sign</th>
<th>veget.</th>
<th>terrain</th>
<th>sky</th>
<th>person</th>
<th>rider</th>
<th>car</th>
<th>truck</th>
<th>bus</th>
<th>train</th>
<th>motorc.</th>
<th>bicycle</th>
<th>mIoU</th>
</tr>
</thead>
<tbody>
<tr>
<td>RefineNet [15]</td>
<td>68.8</td>
<td>23.2</td>
<td>46.8</td>
<td>20.8</td>
<td>12.6</td>
<td>29.8</td>
<td>30.4</td>
<td>26.9</td>
<td>43.1</td>
<td>14.3</td>
<td>0.3</td>
<td>36.9</td>
<td>49.7</td>
<td>63.6</td>
<td>6.8</td>
<td>0.2</td>
<td>24.0</td>
<td>33.6</td>
<td>9.3</td>
<td>28.5</td>
</tr>
<tr>
<td>AdaptSegNet-Cityscapes [24]</td>
<td>79.0</td>
<td>21.8</td>
<td>53.0</td>
<td>13.3</td>
<td>11.2</td>
<td>22.5</td>
<td>20.2</td>
<td>22.1</td>
<td>43.5</td>
<td>10.4</td>
<td>18.0</td>
<td>37.4</td>
<td>33.8</td>
<td>64.1</td>
<td>6.4</td>
<td>0.0</td>
<td>52.3</td>
<td>30.4</td>
<td>7.4</td>
<td>28.8</td>
</tr>
<tr>
<td>AdaptSegNet-Cityscapes→DZ-night [24]</td>
<td>86.1</td>
<td>44.2</td>
<td>55.1</td>
<td>22.2</td>
<td>4.8</td>
<td>21.1</td>
<td>5.6</td>
<td>16.7</td>
<td>37.2</td>
<td>8.4</td>
<td>1.2</td>
<td>35.9</td>
<td>26.7</td>
<td>68.2</td>
<td>45.1</td>
<td>0.0</td>
<td>50.1</td>
<td>33.9</td>
<td>15.6</td>
<td>30.4</td>
</tr>
<tr>
<td>DMAAda [7]</td>
<td>75.5</td>
<td>29.1</td>
<td>48.6</td>
<td>21.3</td>
<td>14.3</td>
<td>34.3</td>
<td>36.8</td>
<td>29.9</td>
<td>49.4</td>
<td>13.8</td>
<td>0.4</td>
<td>43.3</td>
<td>50.2</td>
<td>69.4</td>
<td>18.4</td>
<td>0.0</td>
<td>27.6</td>
<td>34.9</td>
<td>11.9</td>
<td>32.1</td>
</tr>
<tr>
<td>Ours: GCMA</td>
<td>81.7</td>
<td>46.9</td>
<td>58.8</td>
<td>22.0</td>
<td>20.0</td>
<td>41.2</td>
<td>40.5</td>
<td>41.6</td>
<td>64.8</td>
<td>31.0</td>
<td>32.1</td>
<td>53.5</td>
<td>47.5</td>
<td>75.5</td>
<td>39.2</td>
<td>0.0</td>
<td>49.6</td>
<td>30.7</td>
<td>21.0</td>
<td>42.0</td>
</tr>
</tbody>
</table>

Nighttime image

Semantic GT

DMAAda

GCMA – ours
Summary

- Adverse weather conditions and nighttime are very challenging
- Learning with synthetic data of adverse conditions is promising
- A combination of synthetic and (unlabeled) real data gives the best results
- The same learning strategy Curriculum Model Adaptation has been successfully applied to Fog and Nighttime
- Data and code for foggy weather and nighttime are made available

- Vision for All Seasons Workshop in CVPR’19 (applied for CVPR’20)
- Autonomous Driving Workshop in ICCV’19 (28th Oct., Room 401)
- IJCV special issue on Vision for all seasons: deadline: 10th Dec. 2019
Reference:


2. ”Curriculum Model Adaptation with Synthetic and Real Data for Semantic Foggy Scene Understanding”, Sakaridis, Dai, Hecker, Van Gool, IJCV, 2019

3. “Semantic foggy scene understanding with synthetic data”, Sakaridis, Dai, Van Gool, IJCV, 2018

4. “Model Adaptation with Synthetic and Real Data for Semantic Dense Foggy Scene Understanding”, Sakaridis, Dai, Hecker, Van Gool, ECCV, 2018


A complete list of publications in this direction will be provided as well.
Thank you!