TOWARDS SCENE UNDERSTANDING UNDER CHALLENGING ILLUMINATION CONDITIONS FOR ADAS

SRINIVAS K S S, PRATYUSH SAHAY, RAJESH BISWAL

NVIDIA GPU TECHNOLOGY CONFERENCE (GTC) 2017
INTRODUCTION

• Fatality rate per mile of travel is three times higher in night-time compared to day
  Almost half of all road traffic deaths are among ‘pedestrians, cyclists and motorcyclists’.

• Driver Assist systems using thermal vision
  Source: WHO
  • Range (200-300m)
  • $$$
  • Available in luxury segment - ~14%
  • Less effective in warmer temperatures

Can we provide an affordable NightVision System for most car segments?

Sources: Road Accident statistics from WHO, NHTSA
INTRODUCTION

Night Vision System using RGB camera only….

- Range (200-300m)
- $$$
- Available in luxury segment - ~14%
- Less effective in warmer temperatures
- High Fidelity
- Thermal Image

- Range (30-60m)
- $
- Available in multiple segment - ~75%
- Effective across temperatures
- Accuracy needs to be improved
- More Natural Image

Sources: Road Accident statistics from WHO, NHTSA
Challenges in low-light scene understanding using RGB systems

Image Source: KAIST Pedestrian Dataset
Challenges in low-light scene understanding using RGB systems

- Sensor Noise
Challenges in low-light scene understanding using RGB systems

- Sensor Noise
- Illumination Variation
Challenges in low-light scene understanding using RGB systems

- Sensor Noise
- Illumination Variation
- Poor contrast of objects
Challenges in low-light scene understanding using RGB systems

- Sensor Noise
- Illumination Variation
- Poor contrast of objects

Deep Learning based Solution

- End-To-End learning systems without the need of hand-crafted features
INTRODUCTION

• Generic Deep Object Detector: *Faster RCNN*

• Our Deep Pedestrian Detector

• Proposed *Multi-modal Knowledge Distillation*

• Experimental Evaluation and Results
GENERIC DEEP OBJECT DETECTOR

FASTER RCNN

Region Proposal Network (RPN)

Conv-1 → Detect

Object Proposals

Top Proposals

Classify → ROI Pool

Ref: Ren, Shaoqing, et al., "Faster RCNN ..., NIPS 2015
OUR PEDESTRIAN DETECTOR

BASED ON RPN ALONE

OUR PEDESTRIAN DETECTOR

TUNING : THE BASE NETWORK

Detecting Region Proposal Network (RPN)

ResNet - 50

Ref: He, Kaiming, et al. "Deep residual learning ... " CVPR, 2016
OUR PEDESTRIAN DETECTOR

TUNING: DILATING THE LAST CONV BLOCK

Region Proposal Network (RPN)

ResNet - 50

Conv1  Conv2  Conv3  Conv4  Conv5  Detect

\[ \frac{1}{2} \quad \frac{1}{4} \quad \frac{1}{8} \quad \frac{1}{16} \quad \frac{1}{32} \]

• Is there a way to extract better performance from this base network?
• KAIST dataset has RGB and thermal channels both – can the extra data be utilized?
MULTIMODAL LEARNING

TRAINING PHASE #1

RGB

Conv1 → Conv2

Thermal

Conv1 → Conv2

Conv3 → Conv4 → Conv5 → Detect

Bounding Box Classification Loss

Bounding Box Regression Loss
MULTIMODAL LEARNING

TRAINING PHASE #2

Base network:

Hypothesis: Multi-modal Knowledge Distillation will improve detection

Result: Detection performance degraded 😞
MULTIMODAL KNOWLEDGE DISTILLATION

TEACHER-STUDENT LEARNING

RGB

Teacher network

Bounding Box Classification Loss
Bounding Box Regression Loss

Hint Loss

Student network

Bounding Box Classification Loss
Bounding Box Regression Loss

Conv1  Conv2
Conv1  Conv2
Conv1  Conv2
Detect

Detect

Bounding Box Classification Loss
Bounding Box Regression Loss

Conv1  Conv2  Conv3  Conv4  Conv5
Conv1  Conv2  Conv3  Conv4  Conv5
MULTIMODAL KNOWLEDGE DISTILLATION

PEDESTRIAN VISIBILITY

Visibility 10%
Visibility 30%
Visibility 40%
Visibility 60%
Visibility 80%
Visibility 90%

Patch Visibility: Normalized Entropy of a patch’s gray-scale histogram
MULTIMODAL KNOWLEDGE DISTILLATION

WEIGHTED TEACHER-STUDENT LEARNING

Teacher network
- Bounding Box Classification Loss
- Bounding Box Regression Loss
- Hint Loss

Student network
- Bounding Box Classification Loss
- Bounding Box Regression Loss
MULTIMODAL KNOWLEDGE DISTILLATION
WEIGHTED TEACHER-STUDENT LEARNING

Teacher network
- Bounding Box Classification Loss
- Bounding Box Regression Loss
- Weighted Hint Loss

Student network
- Bounding Box Classification Loss
- Bounding Box Regression Loss

Bounding Box Classification Loss
Bounding Box Regression Loss

RGB
Conv1 Conv2

Thermal
Conv1 Conv2

Conv1 Conv2 Conv3 Conv4 Conv5 Detect

Conv1 Conv2 Conv3 Conv4 Conv5 Detect

Weighted Hint Loss
**EXPERIMENTAL EVALUATION**

- **Dataset:** KAIST Multi-spectral Pedestrian Dataset
  - Night-time driving in campus, urban and downtown localities
  - Training Data: ~7,200 thermal and RGB pairs

- **Network Training:** Nvidia Titan X GPU with Caffe
  - Pre-training on Caltech Ped Dataset (~14 hours)
  - Training on KAIST Dataset (~3 hours)
## RESULTS

<table>
<thead>
<tr>
<th>Model</th>
<th>Miss Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>(based on RGB)</td>
<td>(Lower the better)</td>
</tr>
<tr>
<td>Hwang et. al.</td>
<td>90.2 %</td>
</tr>
<tr>
<td>(CVPR ‘15)</td>
<td></td>
</tr>
<tr>
<td>Liu et. al.</td>
<td>64.4 %</td>
</tr>
<tr>
<td>(BMVC ‘16)</td>
<td></td>
</tr>
<tr>
<td>Wagner et. al.</td>
<td>63.4 %</td>
</tr>
<tr>
<td>(ESANN ‘16)</td>
<td></td>
</tr>
<tr>
<td><strong>Ours</strong> (Direct Training)</td>
<td><strong>55.3 %</strong></td>
</tr>
<tr>
<td><strong>Ours</strong> (Guided Training)</td>
<td><strong>52.8 %</strong></td>
</tr>
</tbody>
</table>
### CVC Dataset

- **Dataset:** CVC Multi-spectral Pedestrian Dataset
  - Gray-scale and Thermal Image Pairs
  - Better Quality Images with less noise

- **Results:** We achieve a **35%** miss rate on CVC night dataset
  - For visibility > 0.4, a miss rate of ~ 25%

### Model Performance

<table>
<thead>
<tr>
<th>Model</th>
<th>Miss Rate (Lower the better)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hwang et. Al (CVPR ’15)</td>
<td>90.2 %</td>
</tr>
<tr>
<td>Liu et. Al (BMVC ’16)</td>
<td>64.4 %</td>
</tr>
<tr>
<td>Wagner et. Al (ESANN ’16)</td>
<td>63.4 %</td>
</tr>
<tr>
<td>Ours – KAIST Dataset</td>
<td></td>
</tr>
<tr>
<td>(Direct Training)</td>
<td>55.3 %</td>
</tr>
<tr>
<td>Ours – KAIST Dataset</td>
<td></td>
</tr>
<tr>
<td>(Guided Training)</td>
<td>52.8 %</td>
</tr>
<tr>
<td>Ours – CVC Dataset</td>
<td></td>
</tr>
<tr>
<td>(Guided Training)</td>
<td><strong>35%</strong></td>
</tr>
</tbody>
</table>

**Goal:** For visibility > 0.4, a miss rate below **10%**
QUALITATIVE RESULTS
ANALYSIS

RESULTS

Visibility based Detection Performance

![Graph showing visibility based detection performance](image_url)

- Visibility 10%
- Visibility 30%
- Visibility 40%
- Visibility 60%
- Visibility 80%
- Visibility 90%
RESULTS

ANALYSIS

Visibility based Detection Performance

![Graph showing visibility based detection performance with different visibility percentages (10%, 30%, 40%, 60%, 80%, 90%) and corresponding images with varying visibility levels (10%, 30%, 40%, 60%, 80%, 90%) showing better detection at lower visibility.]
RESULTS

ANALYSIS

Visibility based Detection Performance

![Graph showing the performance of pedestrian detection based on visibility.

- Ours (Direct Training)
- Ours (Guided Training)

Visibility levels: 10%, 30%, 40%, 60%, 80%, 90%.

Images representing different visibility levels:
- Visibility 10%
- Visibility 30%
- Visibility 40%
- Visibility 60%
- Visibility 80%
- Visibility 90%]
• Majority of car segments could benefit from RGB based scene understanding in low-light conditions

<table>
<thead>
<tr>
<th></th>
<th>Training</th>
<th>Deployment</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Base Detector</td>
<td>RGB</td>
<td>RGB</td>
<td></td>
</tr>
<tr>
<td>Guided Detector</td>
<td>RGB, Thermal</td>
<td>RGB</td>
<td></td>
</tr>
</tbody>
</table>
THANK YOU