TOWARDS SCENE UNDERSTANDING UNDER CHALLENGING ILLUMINATION CONDITIONS FOR ADAS

SRINIVAS K S S, PRATYUSH SAHAY, RAJESH BISWAL

NVIDIA GPU TECHNOLOGY CONFERENCE (GTC) 2017

HARMAN

INTRODUCTION

HARMAN

- 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'.

Source: WHO

- Driver Assist systems using thermal vision

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

Can we provide an affordable Night Vision System for most car segments?

Sources: Road Accident statistics from WHO, NHTSA

INTRODUCTION

HARMAN

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

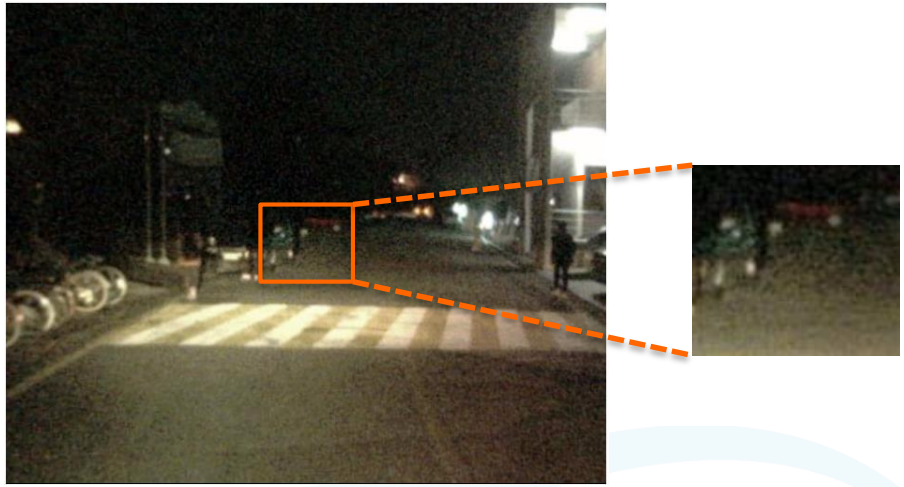
HARMAN

Challenges in low-light scene understanding using RGB systems

Image Source: KAIST Pedestrian Dataset

Challenges in low-light scene understanding using RGB systems

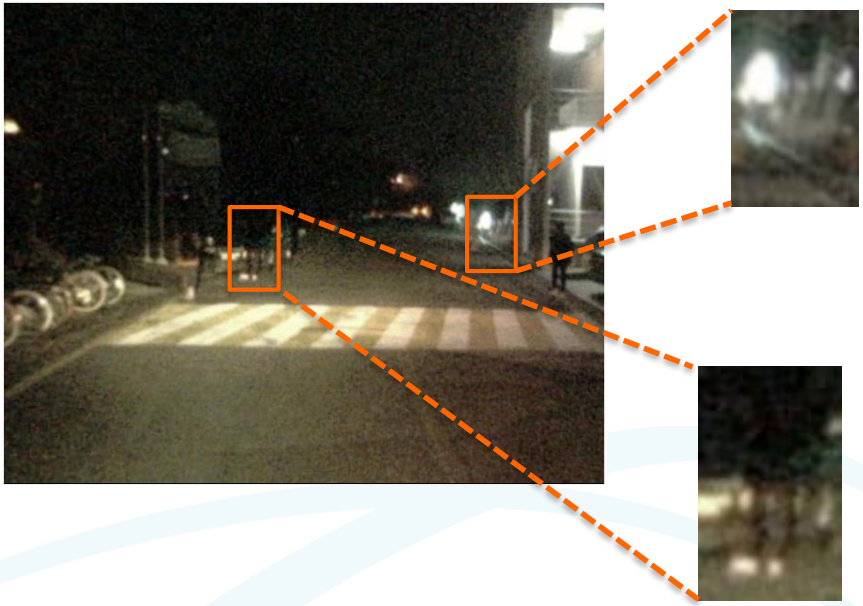
- Sensor Noise



INTRODUCTION

Challenges in low-light scene understanding using RGB systems

- Sensor Noise
- Illumination Variation



INTRODUCTION

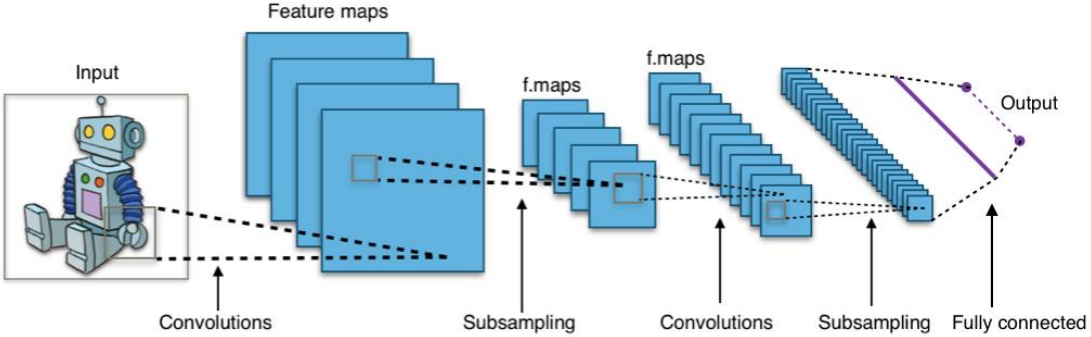
Challenges in low-light scene understanding using RGB systems

- Sensor Noise
- Illumination Variation
- Poor contrast of objects

INTRODUCTION

Challenges in low-light scene understanding using RGB systems

- Sensor Noise
- Illumination Variation
- Poor contrast of objects



Deep Learning based Solution

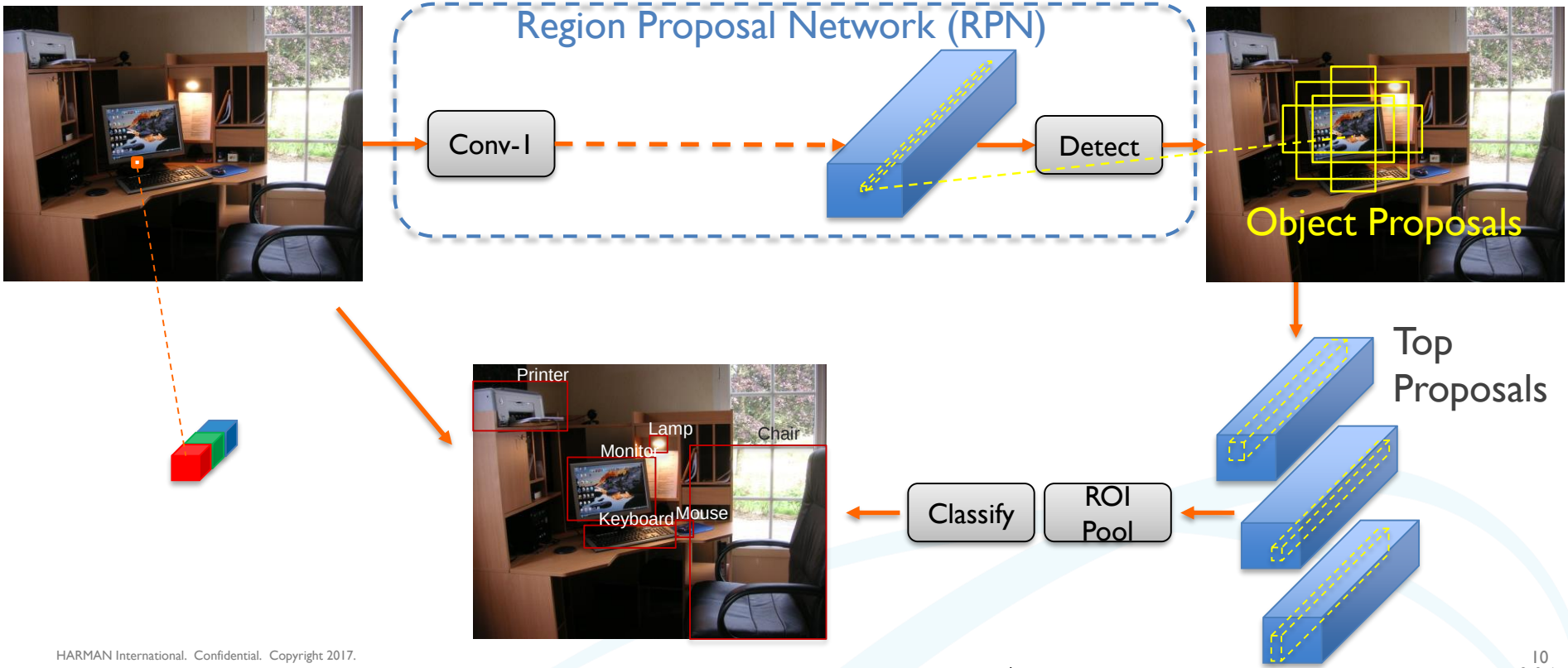
- End-To-End learning systems without the need of hand-crafted features

INTRODUCTION

- **Generic Deep Object Detector: *Faster RCNN***
- **Our Deep Pedestrian Detector**
- **Proposed *Multi-modal Knowledge Distillation***
- **Experimental Evaluation and Results**

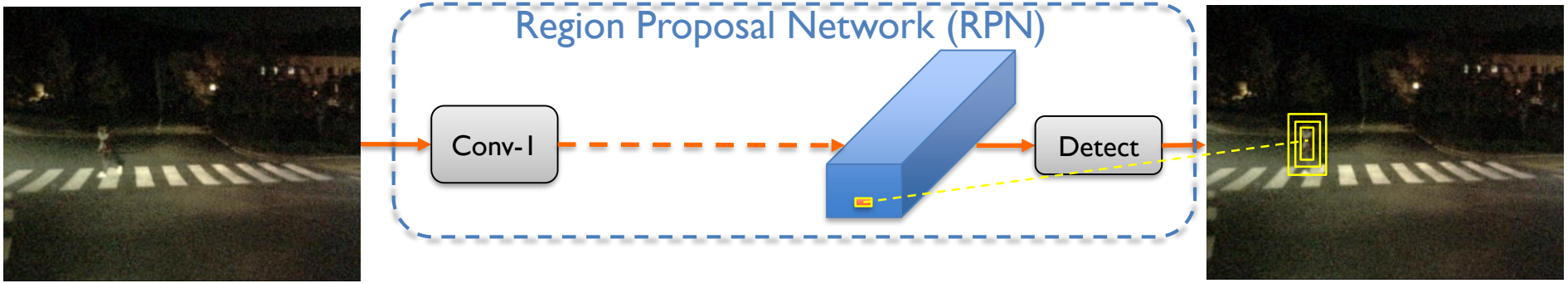
GENERIC DEEP OBJECT DETECTOR

FASTER RCNN



Ref: Ren,Shaoqing, et al., "Faster RCNN …", NIPS 2015

OUR PEDESTRIAN DETECTOR

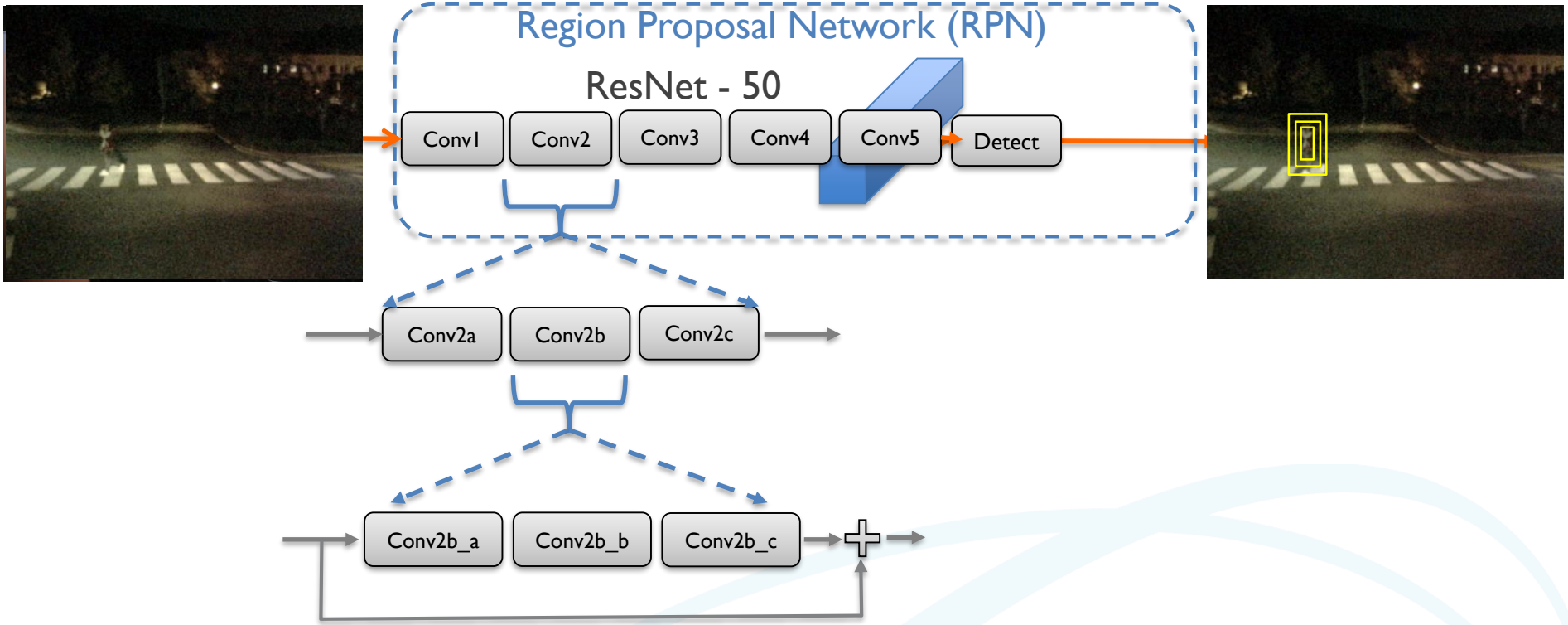


BASED ON RPN ALONE

Ref: Zhang, Liliang, et al. "Is Faster R-CNN …?." ECCV 2016

OUR PEDESTRIAN DETECTOR

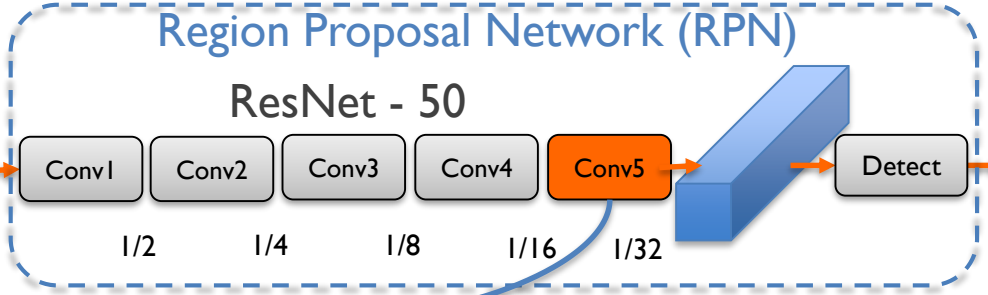
TUNING : THE BASE NETWORK

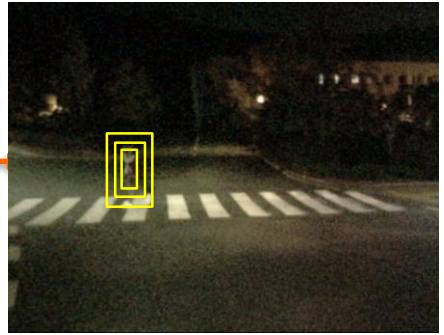


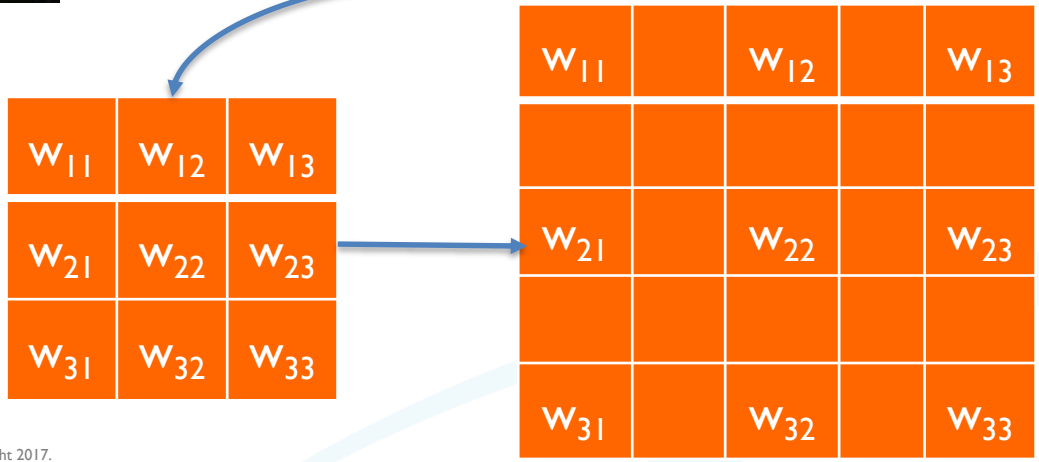
Ref: He, Kaiming, et al. "Deep residual learning … " CVPR,2016

OUR PEDESTRIAN DETECTOR

TUNING : DILATING THE LAST CONV BLOCK



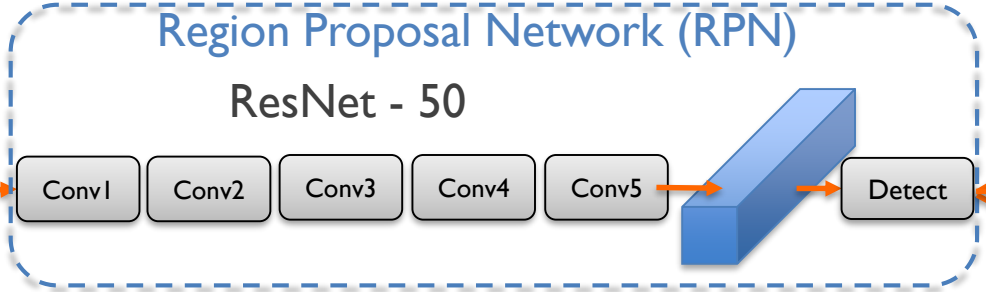




Ref: Chen, Liang-Chieh, et al. "Semantic …" arXiv:1412.7062 (2014)

OUR PEDESTRIAN DETECTOR

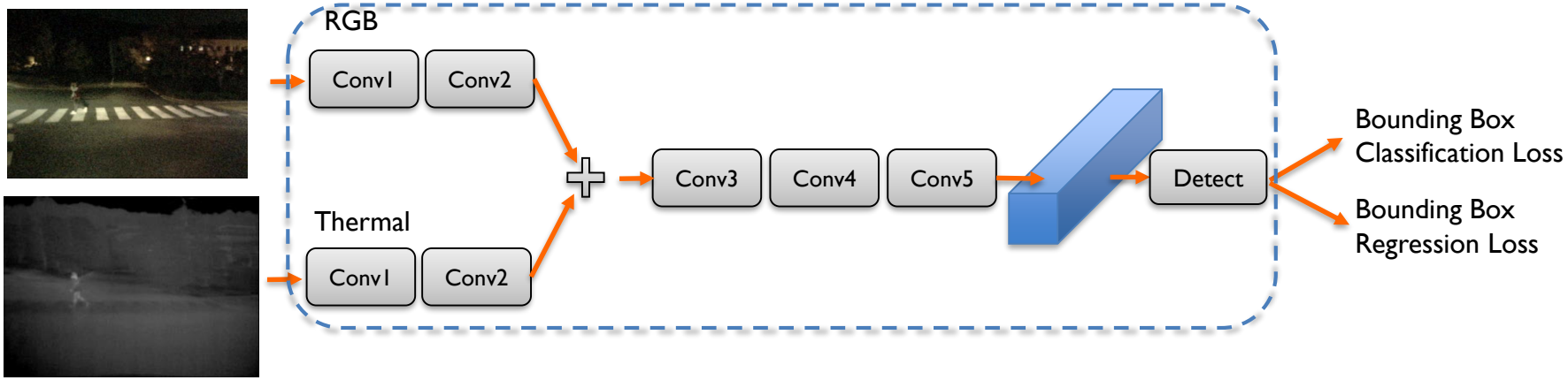
THE NEW BASE NETWORK



- 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?

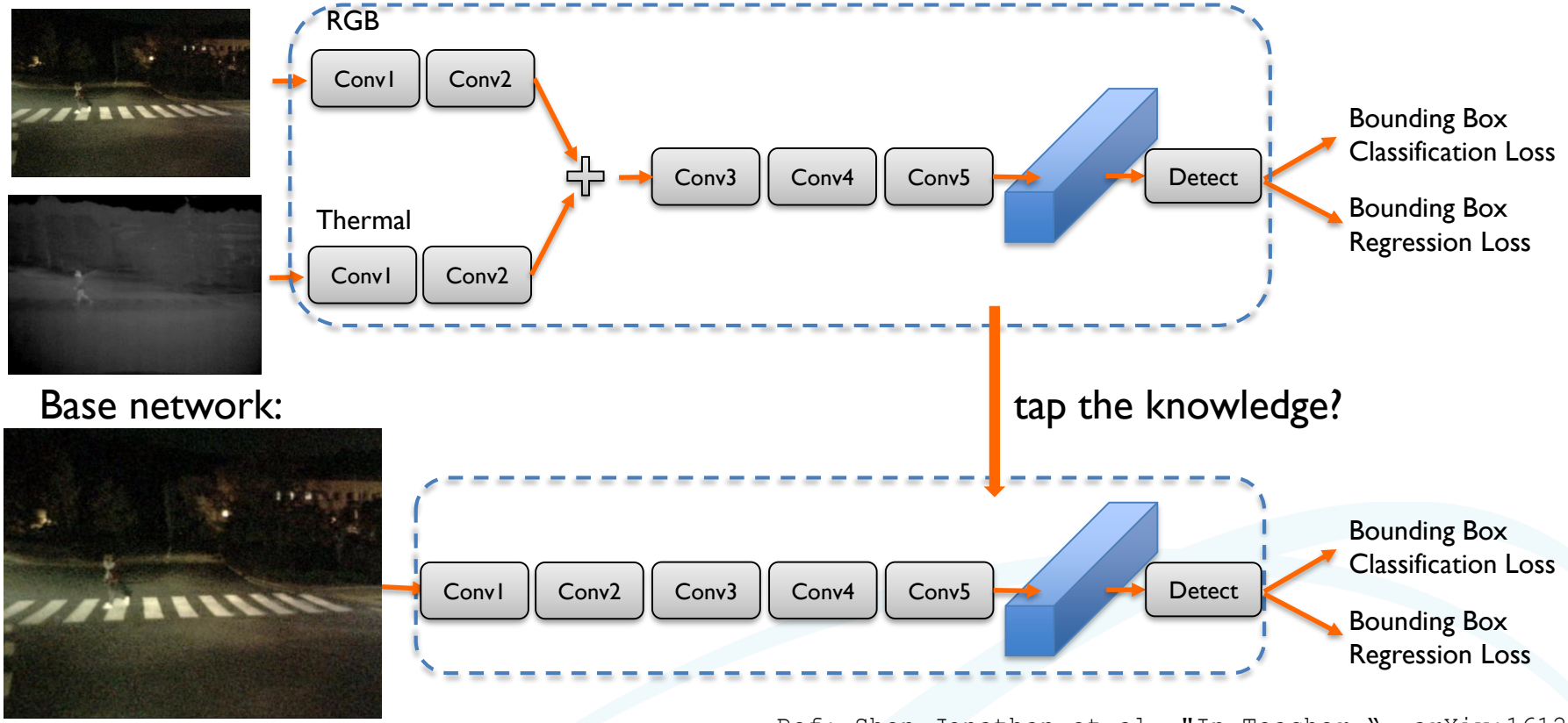
MULTIMODAL LEARNING

TRAINING PHASE #1



MULTIMODAL LEARNING

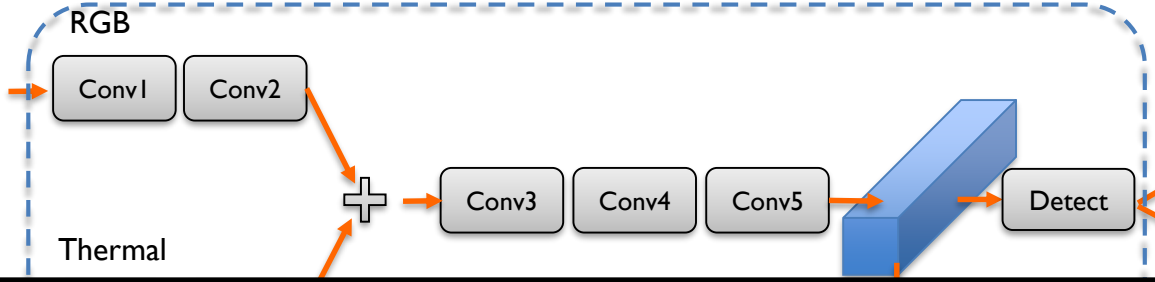
TRAINING PHASE #2



Ref: Shen Jonathan,et al. "In Teacher…", arXiv:1612.00478

MULTIMODAL KNOWLEDGE DISTILLATION

TEACHER-STUDENT LEARNING



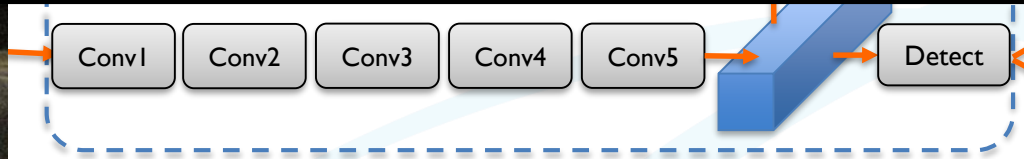
Teacher network

Bounding Box Classification Loss

Bounding Box Regression Loss

Hypothesis: Multi-modal Knowledge Distillation will improve detection

Result: Detection performance degraded ☹

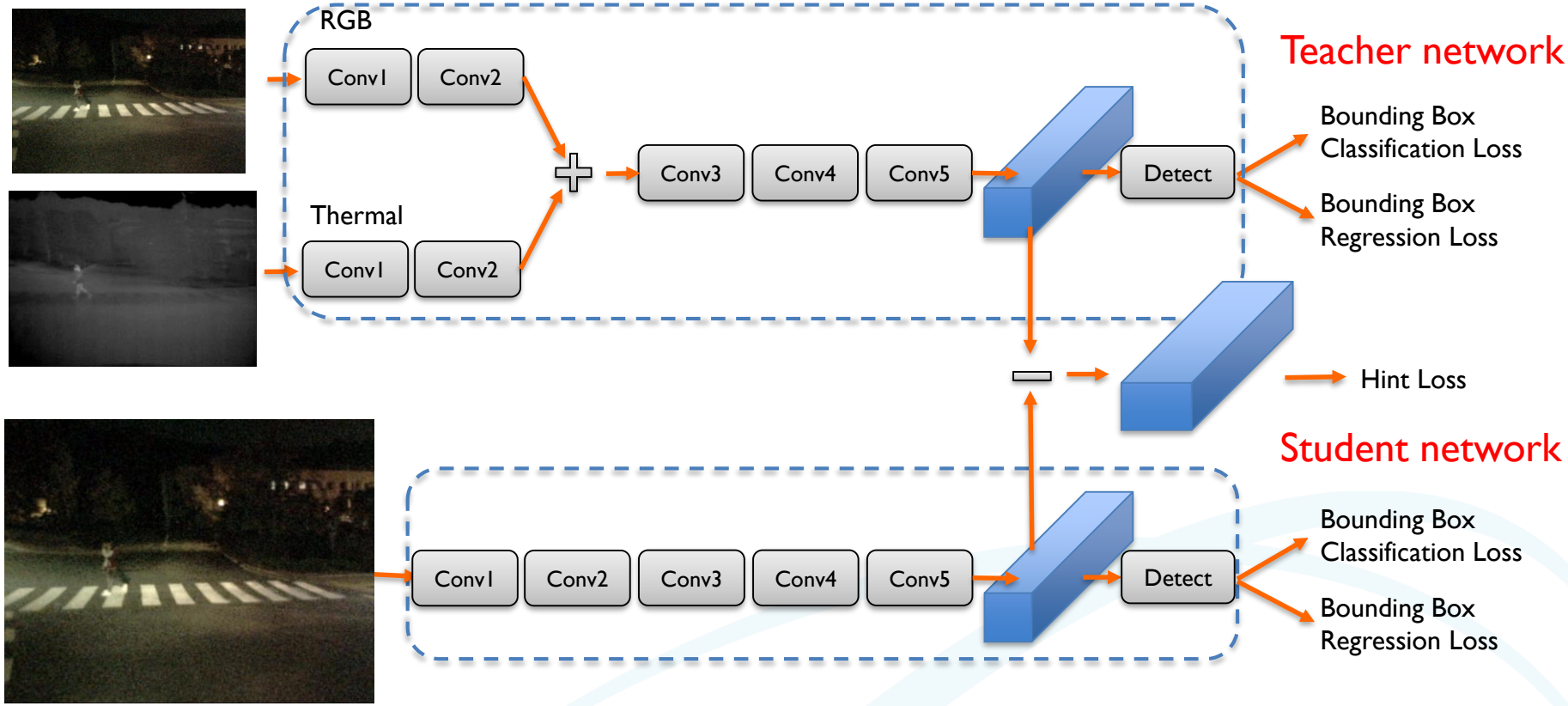


Classification Loss

Bounding Box Regression Loss

MULTIMODAL KNOWLEDGE DISTILLATION

TEACHER-STUDENT LEARNING

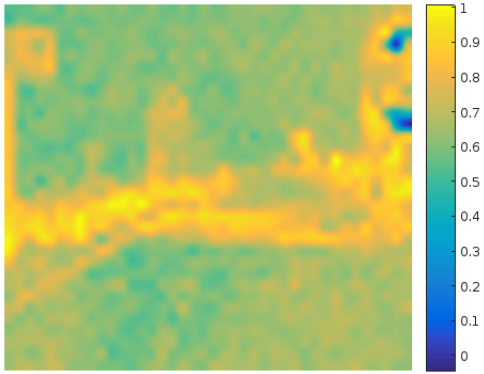


MULTIMODAL KNOWLEDGE DISTILLATION

PEDESTRIAN VISIBILITY

Image

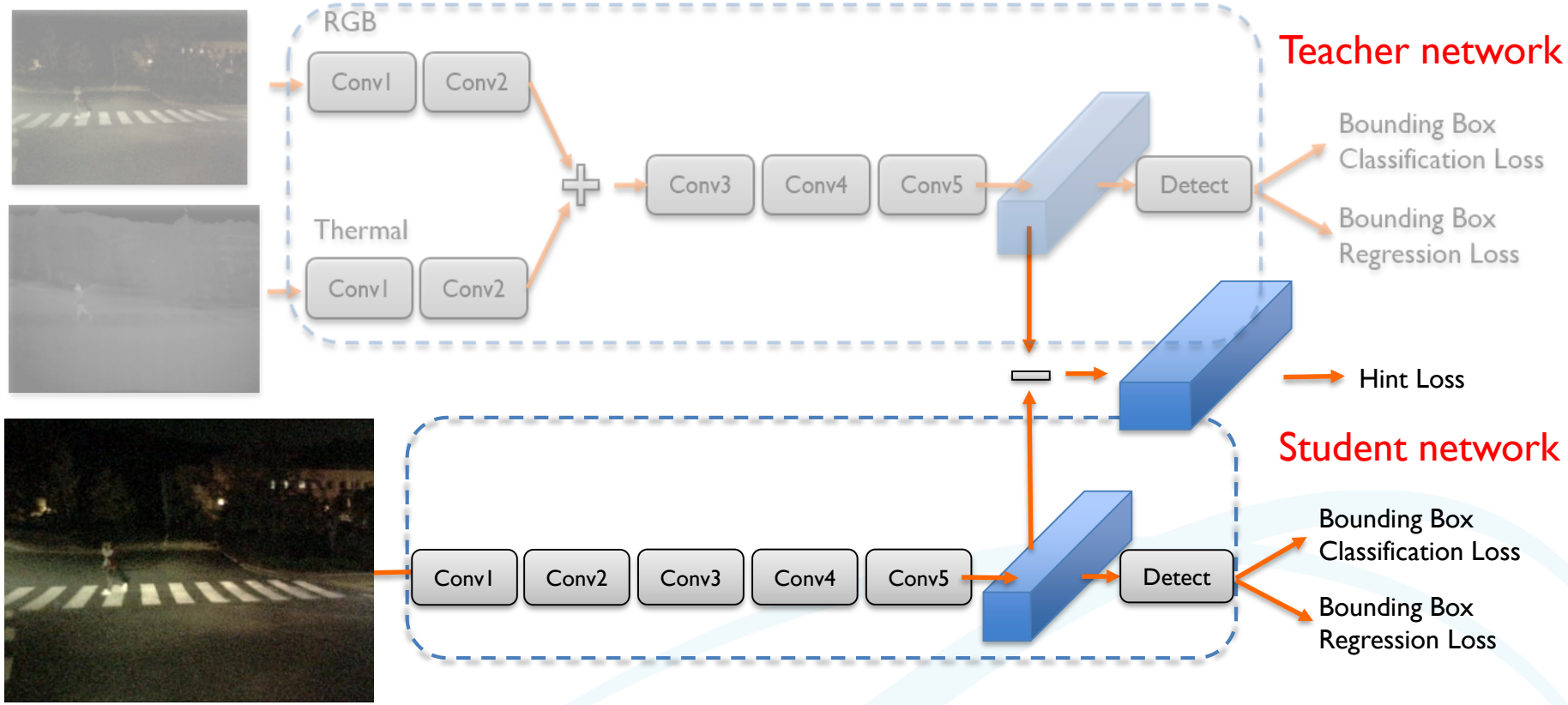
Visibility Map



Patch Visibility: Normalized Entropy of a patch's gray-scale histogram

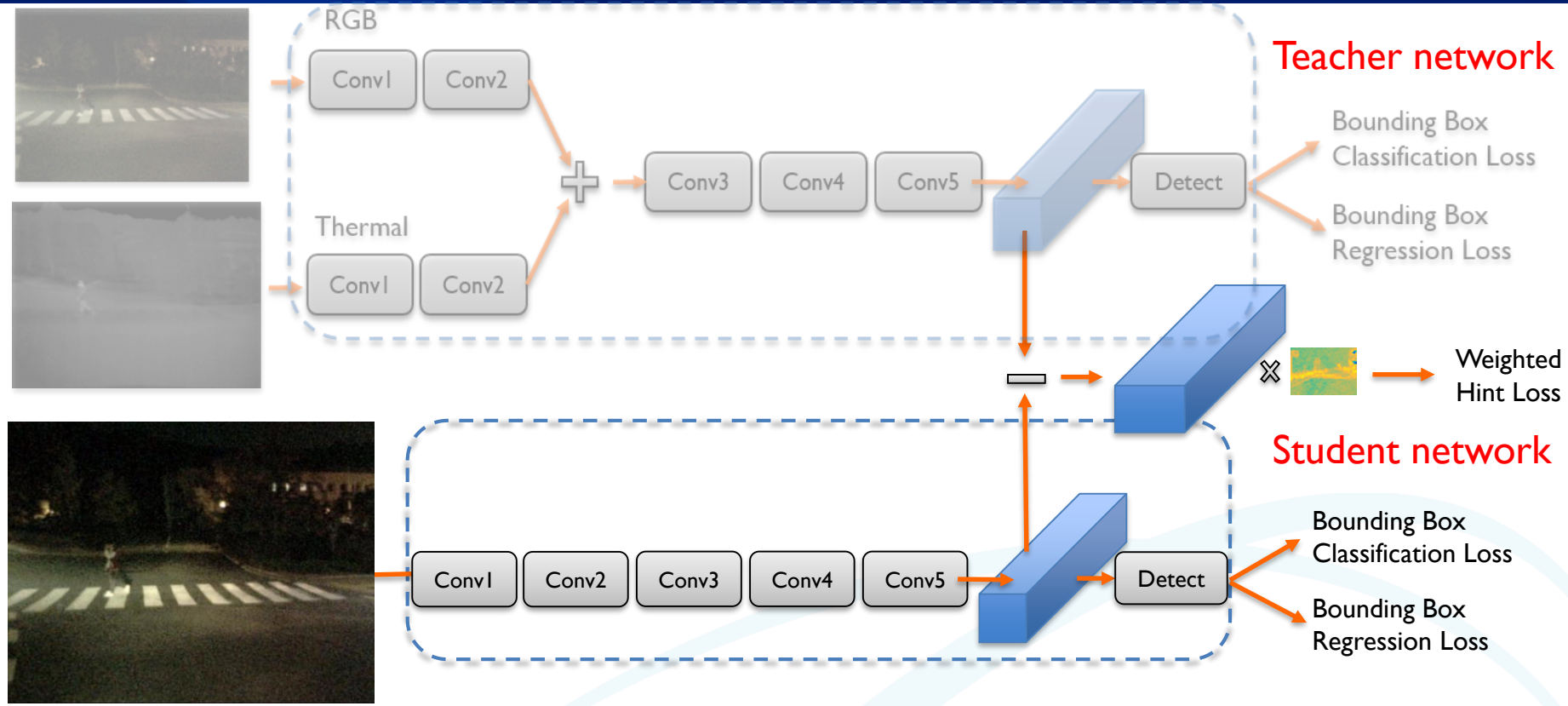
MULTIMODAL KNOWLEDGE DISTILLATION

WEIGHTED TEACHER-STUDENT LEARNING



MULTIMODAL KNOWLEDGE DISTILLATION

WEIGHTED TEACHER-STUDENT LEARNING

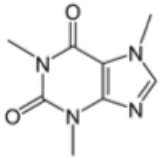


EXPERIMENTAL EVALUATION

HARMAN

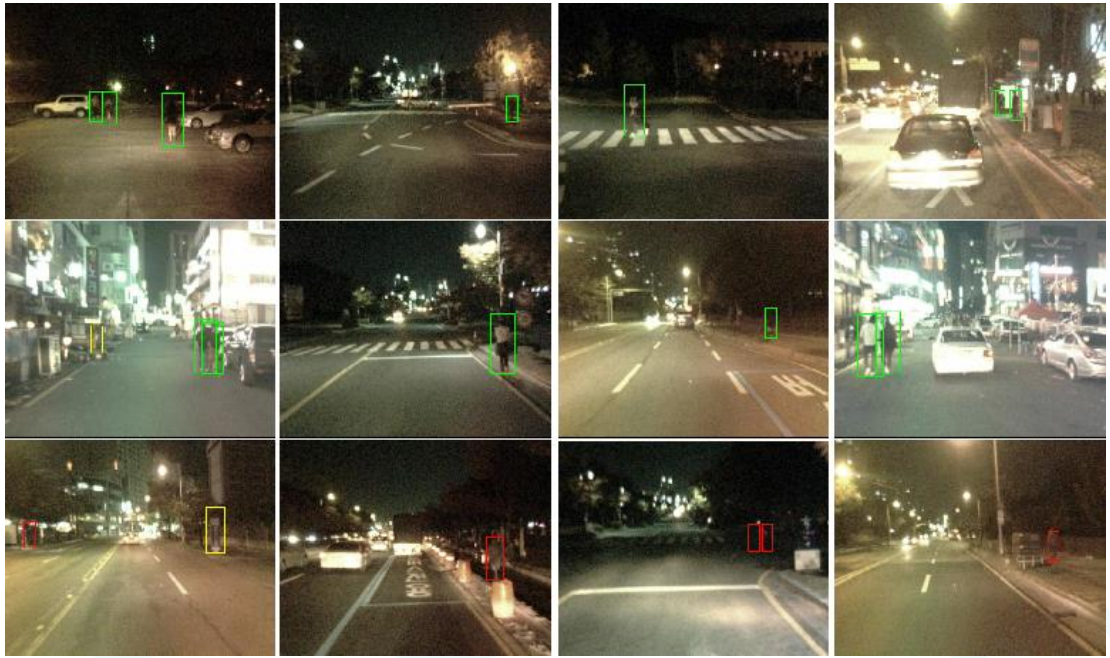
- **Dataset:** KAIST Multi-spectral Pedestrian Dataset
  - Night-time driving in campus, urban and downtown localities
  - Training Data: ~7,200 thermal and RGB pairs

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



RESULTS

Model (based on RGB)	Miss Rate (Lower the better)
Hwang et. al. (CVPR '15)	90.2 %
Liu et. al. (BMVC '16)	64.4 %
Wagner et. al. (ESANN '16)	63.4 %
Ours (Direct Training)	**55.3 %**
Ours (Guided Training)	**52.8 %**



~11 %

CVC DATASET

HARMAN

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

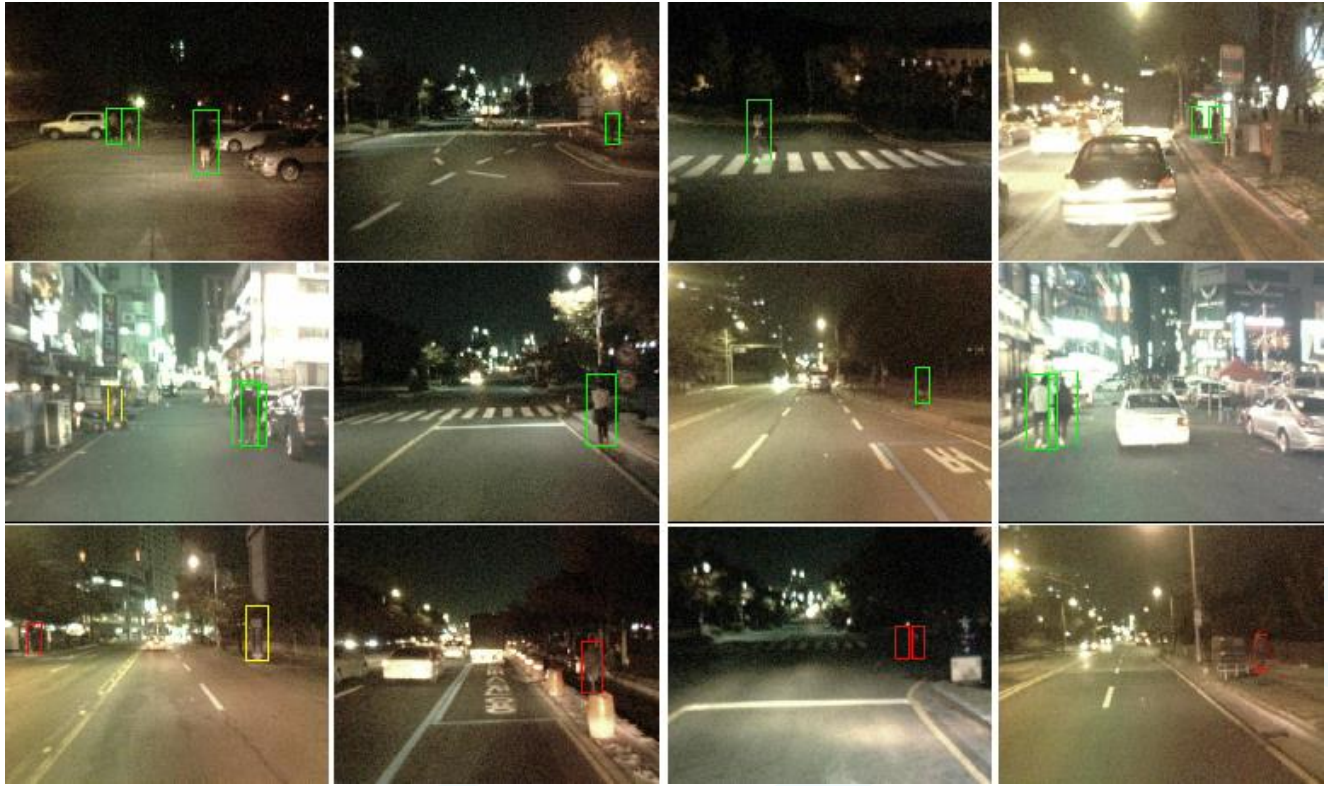
- **Results:** We achieve a **35 %** miss rate on CVC night dataset
  - For visibility > 0.4, a miss rate of ~ 25 %

Model (based on RGB)	Miss Rate (Lower the better)
Hwang et. Al (CVPR '15)	90.2 %
Liu et. Al (BMVC '16)	64.4 %
Wagner et. Al (ESANN '16)	63.4 %
Ours – KAIST Dataset (Direct Training)	55.3 %
Ours – KAIST Dataset (Guided Training)	52.8 %
Ours – CVC Dataset (Guided Training)	35%

18%

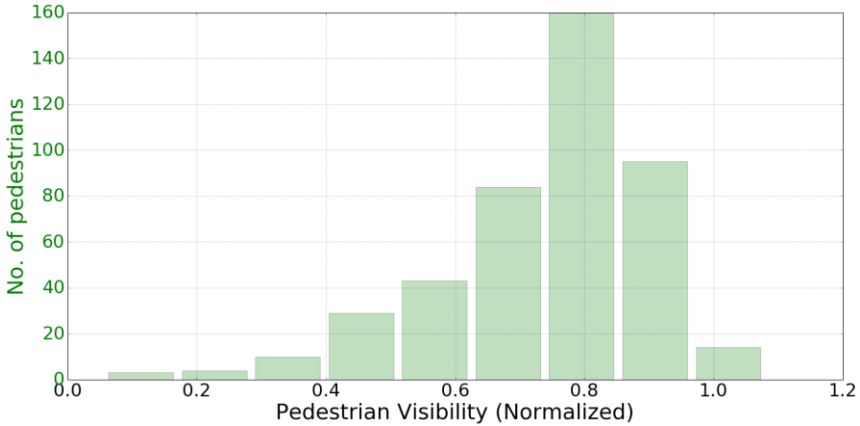
Goal: For visibility > 0.4, a miss rate below10 %

QUALITATIVE RESULTS



ANALYSIS

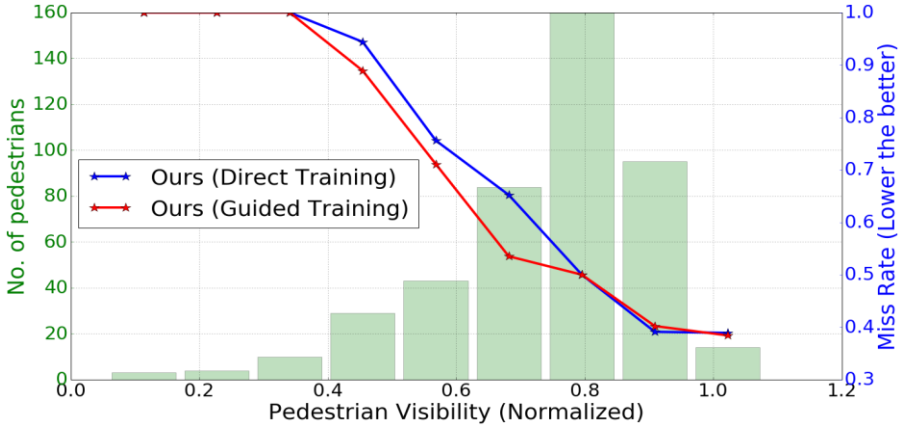
Visibility based Detection Performance



ANALYSIS

Visibility based Detection Performance

Visibility based Detection Performance

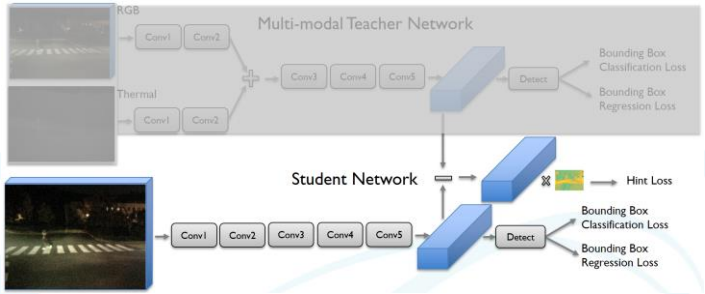


DEMO

SUMMARY

- Majority of car segments could benefit from RGB based scene understanding in low-light conditions

	Training	Deployment
Base Detector	RGB	RGB
Guided Detector	RGB, Thermal	RGB



THANK YOU