



# Dual-Stream Network for RGB-T Camera

FCN-Pix2Pix & ERFNet

# The Competition



Artifact



Time

4 Round  
1 hour/R



Research

MOBILITY  
AUTONOMY  
PRECEPTION  
COMMUNICATION



Date

Aug.2019  
Feb.2020  
Aug.2020

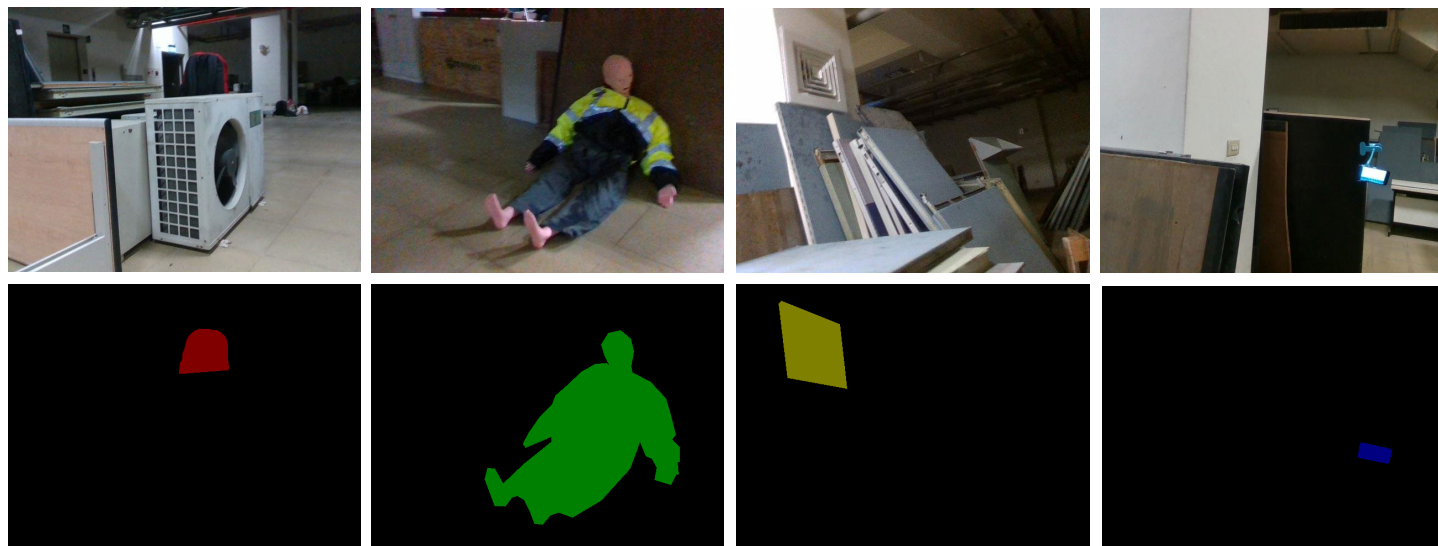


Location

美國 匹茲堡  
美國 西雅圖

# Dataset

SubT-Urban	Backpack	Survivor	Vent	Phone
N of Frames	910	1151	1092	918



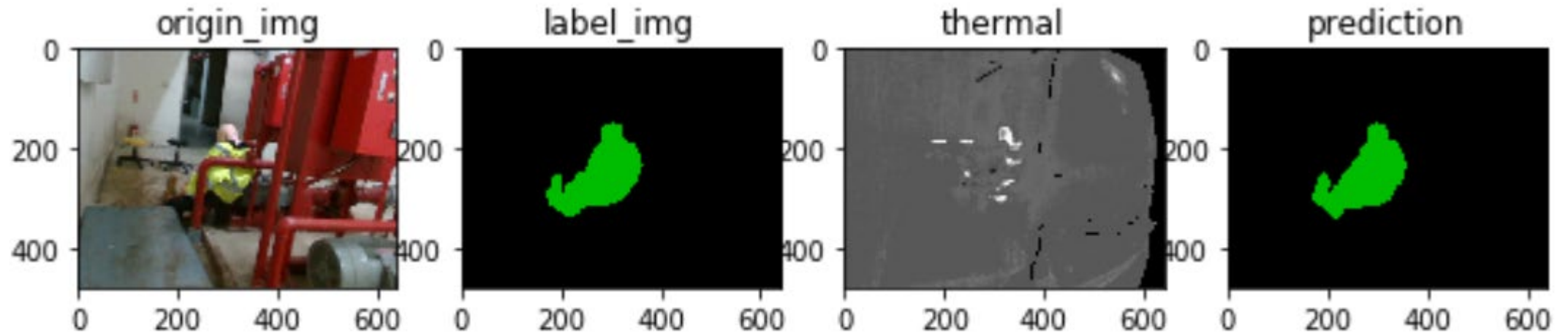
We present our dataset of 4071 RGBD and Thermal images with per-pixel human annotations. Contains four artifacts (survivor, phone, vent, backpack) using handheld Intel RealSense D435 and FLIR Boson 320 thermal camera to collect data.

We then collect data from multiple different environments with varying degrees of lighting and randomly place artifacts, taking into account the artifact placement angle, view angle, line of sight, and the distance, these environments include the cluttered basement in NCTU.

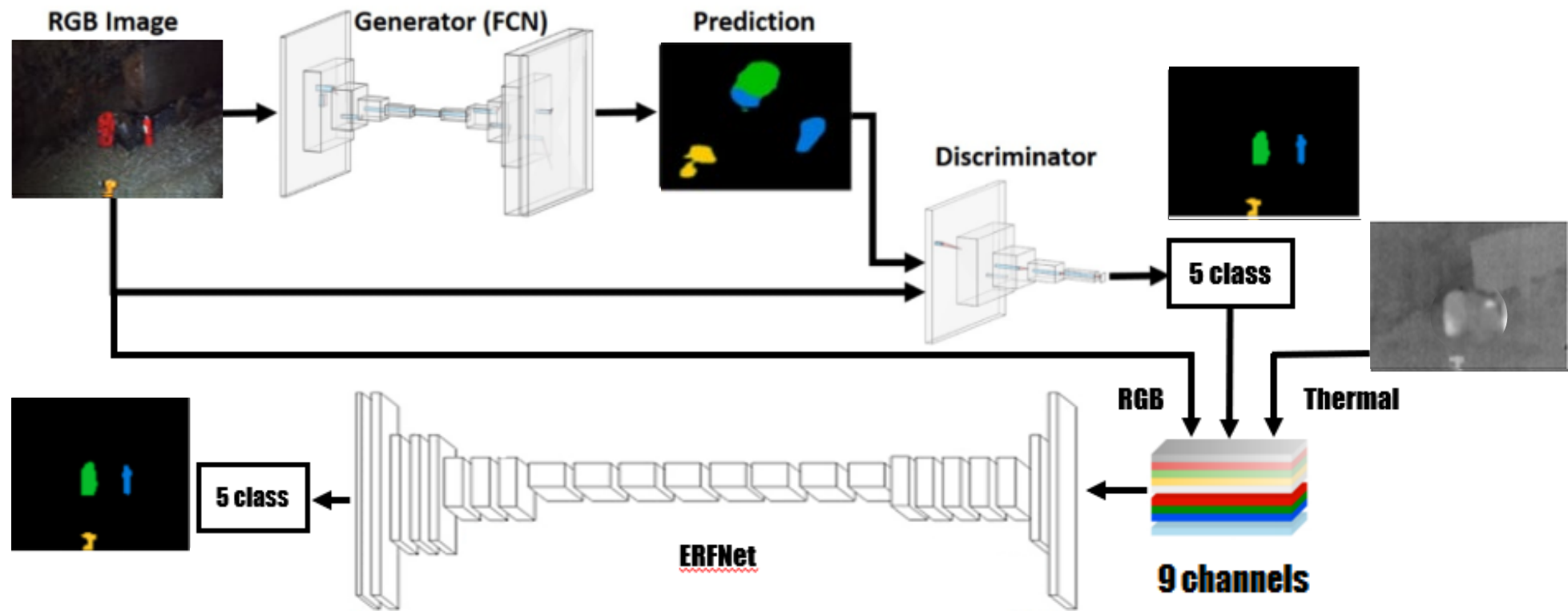
# Results

Dataset : SubT-Urban

Network	Mode	Backpack	Survivor	Vent	Phone	mIoU
FCN	RGB	0.6920	0.7012	0.5360	0.5003	0.6836
FCN-Pix2Pix	RGB	0.9518	0.9139	0.9422	0.8357	0.9109
FCN-Pix2Pix-ERFNet	RGB-T	0.9519	0.9148	0.9503	0.8479	0.9162



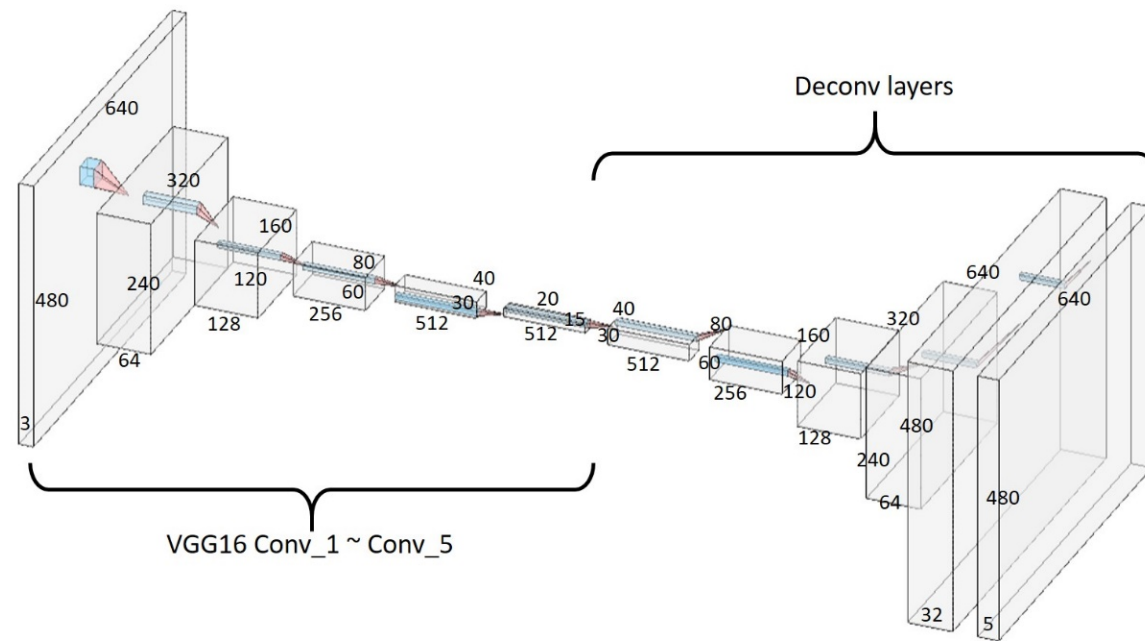
# Dual-Stream Network FCN-Pix2Pix & ERFNet



# FCN-Pix2Pix Network Architectures

## Generator

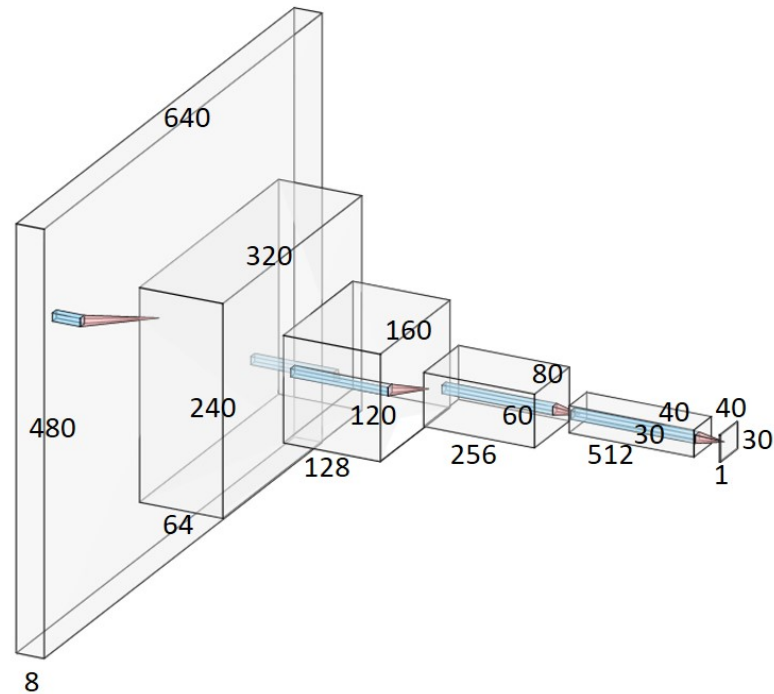
- Fully Convolutional Network (FCN)
- Input :  $480 \times 640 \times 3$
- Output :  $480 \times 640 \times 5$  (vent, backpack, phone, survivor, and background)



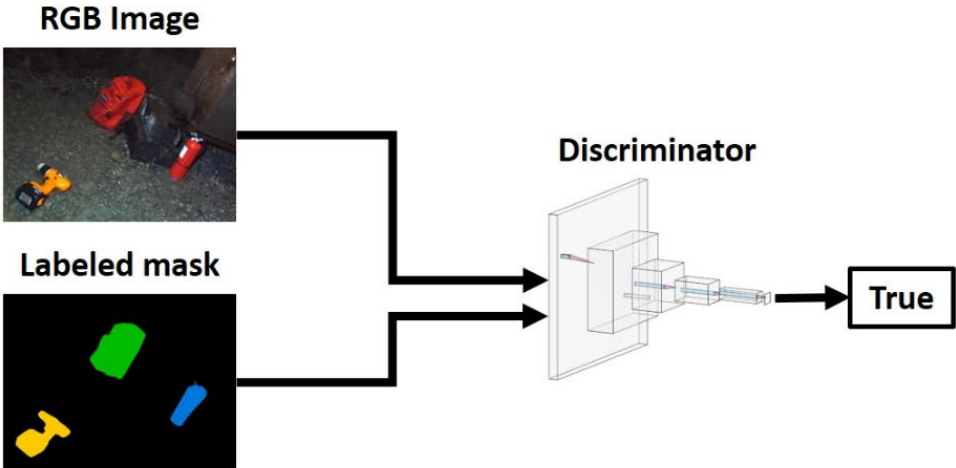
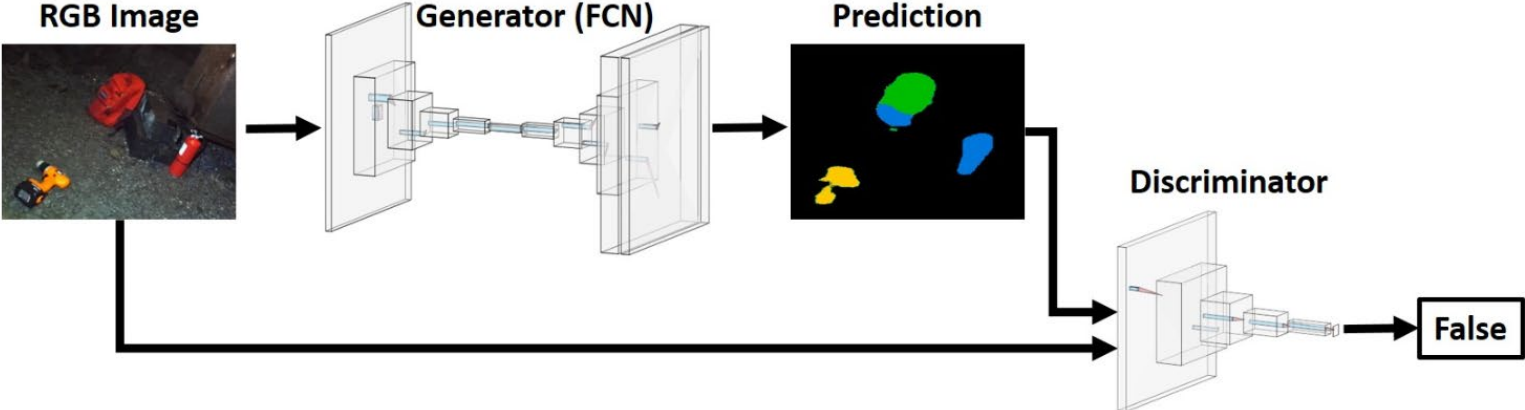
# FCN-Pix2Pix Network Architectures

## Discriminator

- Patch Discriminator
- Input :  $480 \times 640 \times 8$  (3 + 5)
- Output :  $30 \times 40 \times 1$



# FCN-Pix2Pix Network Architectures





# Results

