FLIR Thermal Starter Dataset Introduction
Version 1.3 August 16, 2019

The FLIR Thermal Starter Dataset provides an annotated thermal image and non-annotated RGB image set for training and validation of object detection neural networks. The dataset was acquired via a RGB and thermal camera mounted on a vehicle. The dataset contains a total of 14,452 annotated thermal images with 10,228 images sampled from short videos and 4,224 images from a continuous 144-second video. All videos were taken on the streets and highways in   Santa Barbara, California, USA from November to May. Videos were taken under generally clear-sky conditions at both day and night.

Thermal images were acquired with a FLIR Tau2 (13 mm f/1.0, 45-degree HFOV and 37-degree VFOV). RGB images were acquired with a FLIR BlackFly at 1280 X 512m (4-8 mm f/1.4-16 megapixel lens with the FOV set to match Tua2). Both   cameras were operated in default mode. The cameras were in a single enclosure 1.9 +/-0.1 inches apart from each other. Images were captured via USB3 video using FLIR-proprietary software. The majority of 10,228 thermal images   were sampled at a rate of two images per second (native videos were 30 frames per second of video). A minority of images that were acquired in less object-rich environments were sampled at a rate of one image per second.

Human annotators labeled and put bounding boxes around ten categories of objects. The MSCOCO labelvector was used for class numbering.
•   Category 1: People
•   Category 2: Bicycles - bicycles and motorcycles (not consistent with coco)
•   Category 3: Cars - personal vehicles and some small commercial vehicles.
•   Category 17: Dogs

Annotators were instructed to make bounding boxes as tight as possible. Tight bounding boxes that omitted small parts of the object, such as extremities, were favored over broader bounding boxes. Personal accessories were not included in the bounding boxes on people. When occlusion occurred, only non-occluded parts of the object were annotated. Heads and shoulders were favored for inclusion in the bounding box over other parts of the body for people   and dogs. When occlusion allowed only parts of limbs or other minor parts of an object to be visible, they were not annotated. Wheels were the important part of the Bicycles category. Bicycle parts typically occluded by riders,   such as handlebars, were not included in the bounding box. People riding the bicycle were annotated separately from the bicycle. When an object was split by an occlusion, two separate annotations were given to the two visible   parts of the object.

Annotations were created only for thermal images. The thermal and RGB camera did not have identical placement on the vehicle and therefore had different viewing geometries, so the thermal annotations do not represent the placement of objects in the RGB image.

The folder structure consists of three folders, each with five subfolders. For the sampled images, a suggested training and validation set have been created via two subfolders (labeled "train" and "val"). Entire videos were assigned to either be in a suggested training or validation set.
•   video: Contains a 144-second video with images given a unique consistent identifier (1 to 4224)
•   train: Contains 8,862 sampled images given a unique consistent identifier number (1 to 8,862)
•   val: Contain 1,366 sampled images given a unique consistent identifier number (8,863 to 10,228)

Baseline accuracy for Training and Validation data was established using the RefineDetect512 neural network designed for 512 X 512 images and pre-trained on MSCOCO data (https://arxiv.org/pdf/1711.06897.pdf and https://github.com/ sfzhang15/RefineDet). The base neural network was trained on 8-bit thermal images and annotations in the training folder. Test data was not used. A mAP IoU(0.5) of 0.587 was achieved for all categories combined for the   Validation data. http://cocodataset.org/#detection-eval was used for accuracy assessment criteria. mAP scores were obtained for People (0.794), Bicycles (0.580), and Cars (0.856) categories.

The following subfolders are used in the file structure:
•   Annotated_thermal_8_bit: The folder contains the 8-bit thermal data processed to have the annotation boundary boxes from the annotations folder overlaid on them.

• thermal annotations.json: These are annotations generally formatted in the MSCOCO annotion style. For both images and individual annotations additional data not present in coco has been added in a field called extra info. • thermal_16_bit   14 bit, 640 X 512 thermal images acquired by a FLIR Tau2 camera, without automatic gain control (AGC) applied. Images are in a 16-bit .tiff format. One tool set capable of viewing 16 bit images is   available at: https://imagej.net

• thermal_8_bit   8 bit, AGC applied, .jpeg formatted images which are otherwise identical to the images in the thermal_16_bit folder.

• RGB: 8 bit RGB (three channel) images.   Note that 499 images in Training, 109 images in Validation, and 29 images in Video do not have RGB counterpart images. The image resolution is commonly 1600 X 1800, but some images are different resolutions, including 480 X 720, 1536 X 2048, and 1024 X 1280.

For all images some minimal bluring has been applied to liscence plates in order to make them illegible. In RGB some minimal bluring was also applied to faces.

Please contact the FLIR ADAS team at ADAS-Support@flir.com for assistance.