THERMAL IMAGE SEGMENTATION

Author: Vipul V Suresh

• Abstract

This project aims at achieving high accuracy segmentation on Infrared based Thermal images. As a first step, Single class segmentation is done for obstructions (pedestrians and cars classified into one class) identified on a dataset acquired by KAIST[8] with segmentation masks created by Ghose. Et al[5]. Using this a benchmark, another dataset acquired by the Northeastern University's robotics department using the NUANCE[9] car to collect dataset around the Boston area is used to test the accuracy of segmentation by the network and show the efficiency of segmentation masks created using different mathematical models. U-Net which is Convolutional Neural Network architecture developed by Olaf et al, for biomedical image segmentation is considered the framework to solve this problem which with further improvement proves to be a viable model. Upon testing different types of masks, to Zero binary threshold method[3] was used to create masks segmenting the obstructions with a thermal reading. The predicted segmentation on the NUANCE[9] validation dataset has a binary cross entropy loss of 0.118

• Introduction

Deep neural networks have introduced state of the art architectures to aid the rapid development of automation technologies among which autonomous driving vehicles has seen an increased popularity due to the works of Tesla, Waymo and other organizations' successful implementation of Deep neural networks in real-time. Backed by the concept of Computer vision which deals with understanding the information present in a digital video or images and the ever-increasing computational resource available right around the corner, Deep learning has opened possibilities ranging across a multitude of domains for its application. One such specific task that stands out among the various methods of prediction and classification is the Segmentation of Images and real time videos which in its basic form classifies every pixel available on the screen to a certain class, allowing for the ability to distinguish between various aspects of the scene and draw distinctions from one another. The robust neural engines available now runs on a typical camera operating under the visible wavelength of the spectrum. Outside the visible spectrum is the Infrared wavelengths which has a distinct advantage of reading the temperature of an object in view. This allows for a whole new approach to solve certain problems faced by vision cameras.

Identifying the semantics of a scene is essential for a wide range of applications ranging from autonomous cars, medical imaging, industrial automation, and law enforcement. Current state of the art vision camera does a particularly good job when there is enough light and no occlusion between the camera and the object in question. This is where thermal imaging comes into play which can aid a vision camera by reading the heat signatures in the form of Infrared radiation emitted by most of the objects. This intrinsic quality of an IR based camera can help detect objects during night and through materials of a certain thickness. Thermal cameras are proven to be more effective when there is no light for the sole reason that the temperature differences between two different objects are high. This quality of thermal imaging can be exploited in various applications. In this project, we start with detecting pedestrians and vehicles around normal driving conditions.

Considering the application of thermal images in the Advanced Driver Assistance Systems and segmenting pixel labels classified as an obstruction acts as a first step in this endeavor. Based on research conducted on several approaches, the performance of the model depends upon the way it is trained on the segmentation masks. Most part of this project dealt with finding the right kind of segmentation masks for the *NUANCE[9]* dataset.

KAIST[8] pedestrian datasets was used as benchmark for the model. The reason for this being the efforts by Ghose et al[5] has yielded neatly annotated segmentation masks for the KAIST[8] dataset which produces high accuracy prediction of segmentation masks. Using the preprocessed images from *NUANCE[9]* dataset, segmentation masks for pedestrians and other objects with thermal readings is used to compare the accuracy in prediction based on the UNet[1] architecture.

- **Background**

Considering Semantic segmentation as a way for scene understanding, there has been numerous types of architectures with accuracy in segmentation up to production ready standards. Some of which being; EfficientNet-L2+NAS-FPN, DeepLabv3+ , CFNet, PSP Net, SSDD etc., which performs exceptionally on high end graphic processor units. For the sake of this project however, we study the lesser known yet efficient architectures which has shown promising results in the Thermal Image domain as well.

UNet [1]: Proposed by Olaf Ronneberger et al, [1] this architecture based on the Convolutional neural network with modifications based on up-sampling and down-sampling the convolutions, and adding skip connections (inspired from ResNet) disproves the notion that for successful evaluation of a Deep Neural Network architecture, there is always a need for perfectly annotated training dataset. “The architecture consists of a contracting path to capture context and a symmetric expanding path that enables precise localization”[1].

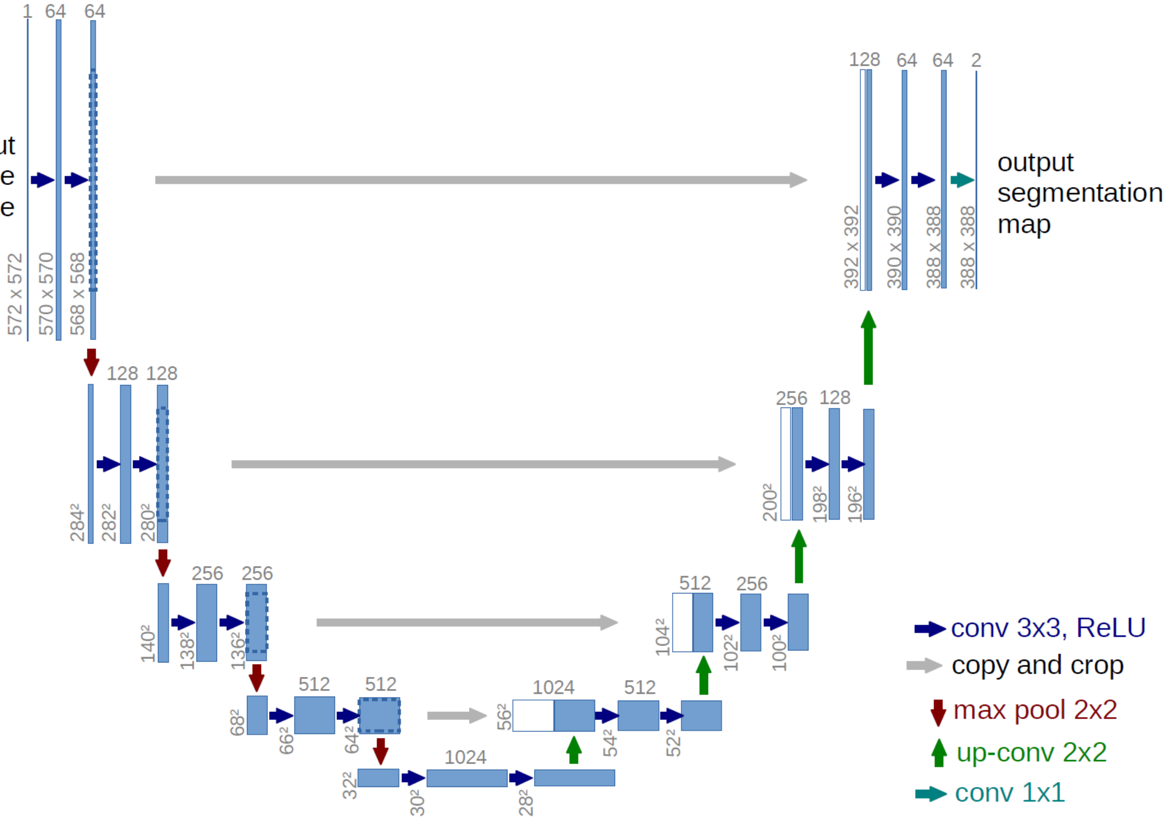
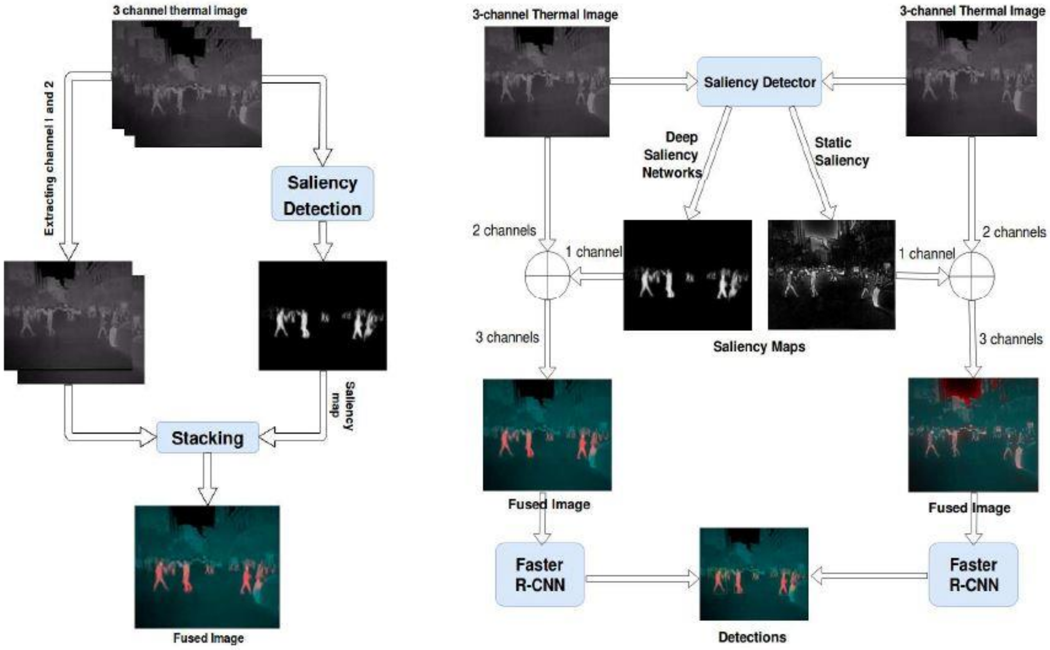


Figure 1[1]

Multispectral Deep Networks for pedestrian detection[2]: Work done by Liu et al analyze a modified version of Faster R-CNN, ConvNet for pedestrian detection on a combination of thermal and vision scene fusion techniques.

Segnet[4]: Alex Kendall et al proposed SegNet with a novel architecture involving encoder-decoder based Deep convolutional architecture for segmentation which performs extremely well with aid from boosted computational resources.

Saliency maps: Ghose et al[5] proposes the augmentation of thermal images with its saliency maps generated by static saliency and Deep saliency networks such as PiCA-Net and R^3 Net.



(a) (b)

Figure 1. (a) Procedure for augmenting thermal images with saliency maps, (b) Faster R-CNN training procedure on augmented images

• **Approach**

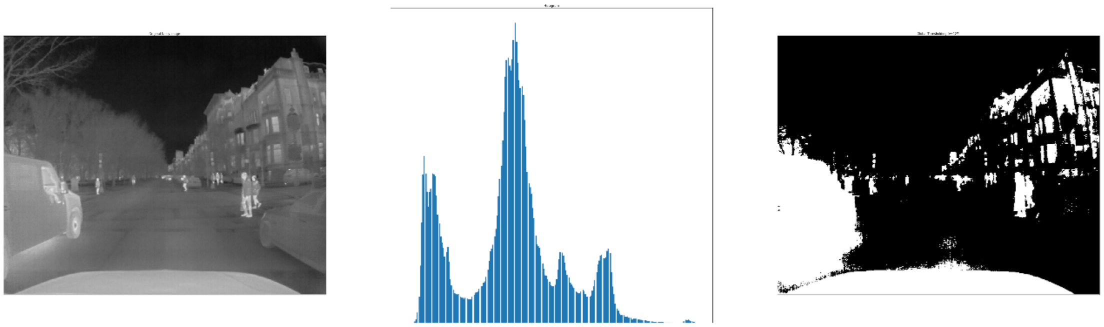
Preprocessing for Generating masks:

To find the best type of segmentation masks, various types of mathematical models translated into computer vision models compiled under the OpenCV library was used to test multiple ways of creating segmentation masks for our thermal image dataset. Out of many other models tested, these stood out to yield promising results.

Threshold based mask:

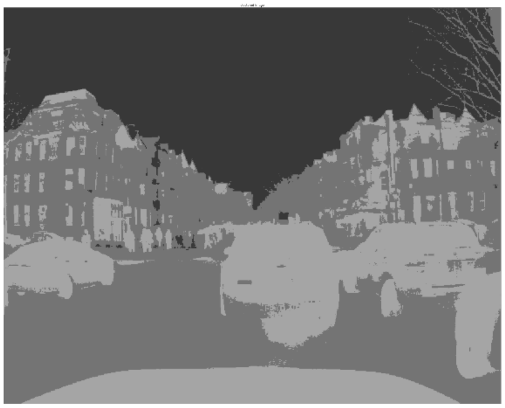
Method 1: Otsu's method[6], is used to perform automatic image thresholding. In the simplest form, the algorithm returns a single intensity threshold that separate pixels into two classes, foreground, and background. This threshold is determined by minimizing intra-class intensity variance, or equivalently, by maximizing inter-class variance. Otsu's method is a one-dimensional discrete analog of Fisher's Discriminant Analysis, is related to Jenks optimization method, and is equivalent to a globally optimal k-means performed on the intensity histogram.

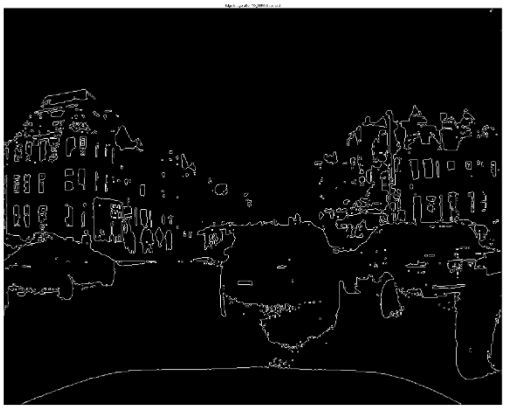
Method 2: Global threshold[9] with k-means: Global thresholding is based on the assumption that the image has a bimodal histogram and, therefore, the object can be extracted from the background by a simple operation that compares image values with a threshold value *T*[7]. For additional filtering of pixels around the highlighted vales, K-means clustering was done to remove noise in the image.



Method 3: to Zero: This operation is done by setting the pixel value to zero if the source pixel values is less than the threshold

Edge based mask: Canny edge detection algorithm is used to detect edges along the pixels above a threshold. This is a multi-step algorithm that can detect edges with noise suppressed at the same time. A Gaussian filter is used to reduce the noise and unwanted textures in an image which is then used to find the gradient between the pixel values using known gradient operators. One such gradient operators called Sobel gradient gives another type of filter named the watershed algorithm.





- **Results**:

Out of various thresholding methods used to create segmentation masks, to Zero threshold method proved to qualitatively yield better results among its counterparts. Applying a binary filter to it gave a mask which captured most of the pixel values with temperature readings that would be considered as obstacles in real world scenarios.

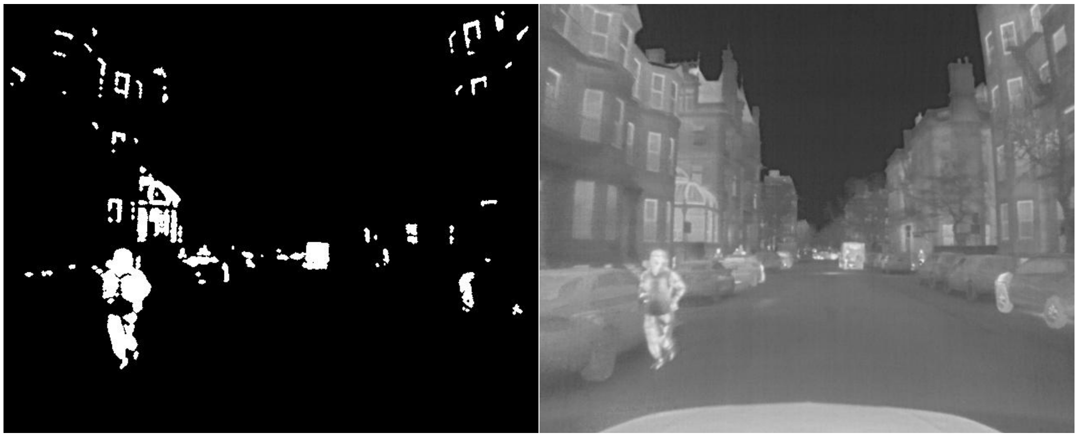
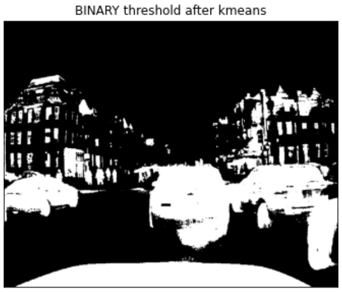
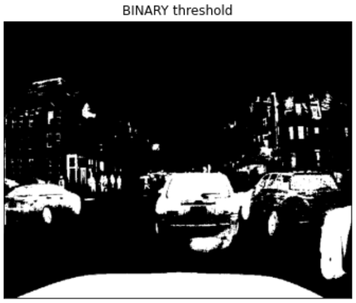
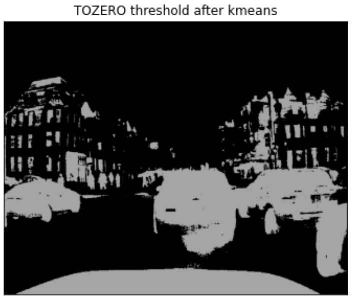


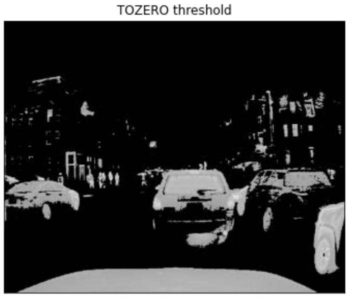
Fig (3). Segmented mask vs the original image

Below are the results of some of the other thresholding techniques tested iteratively over several images. Although most of these methods yield similar results, to Zero threshold shows cleaner clustering away from the noise in the image.



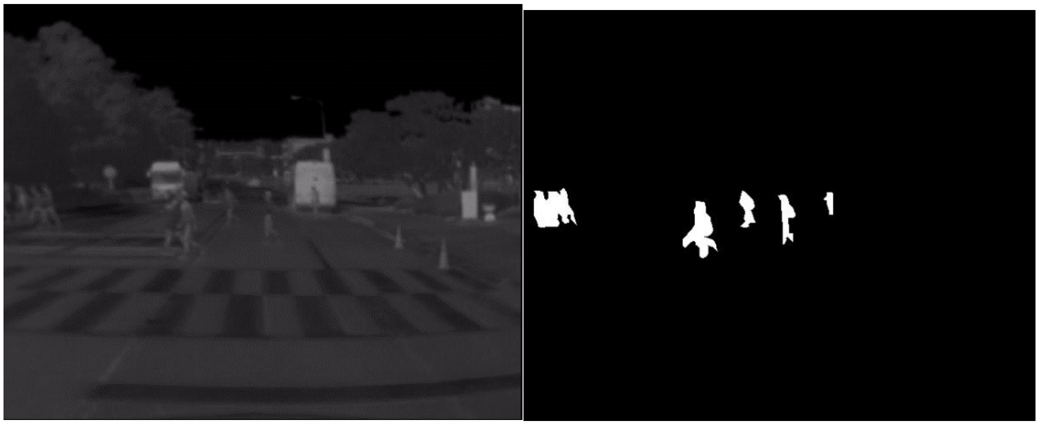




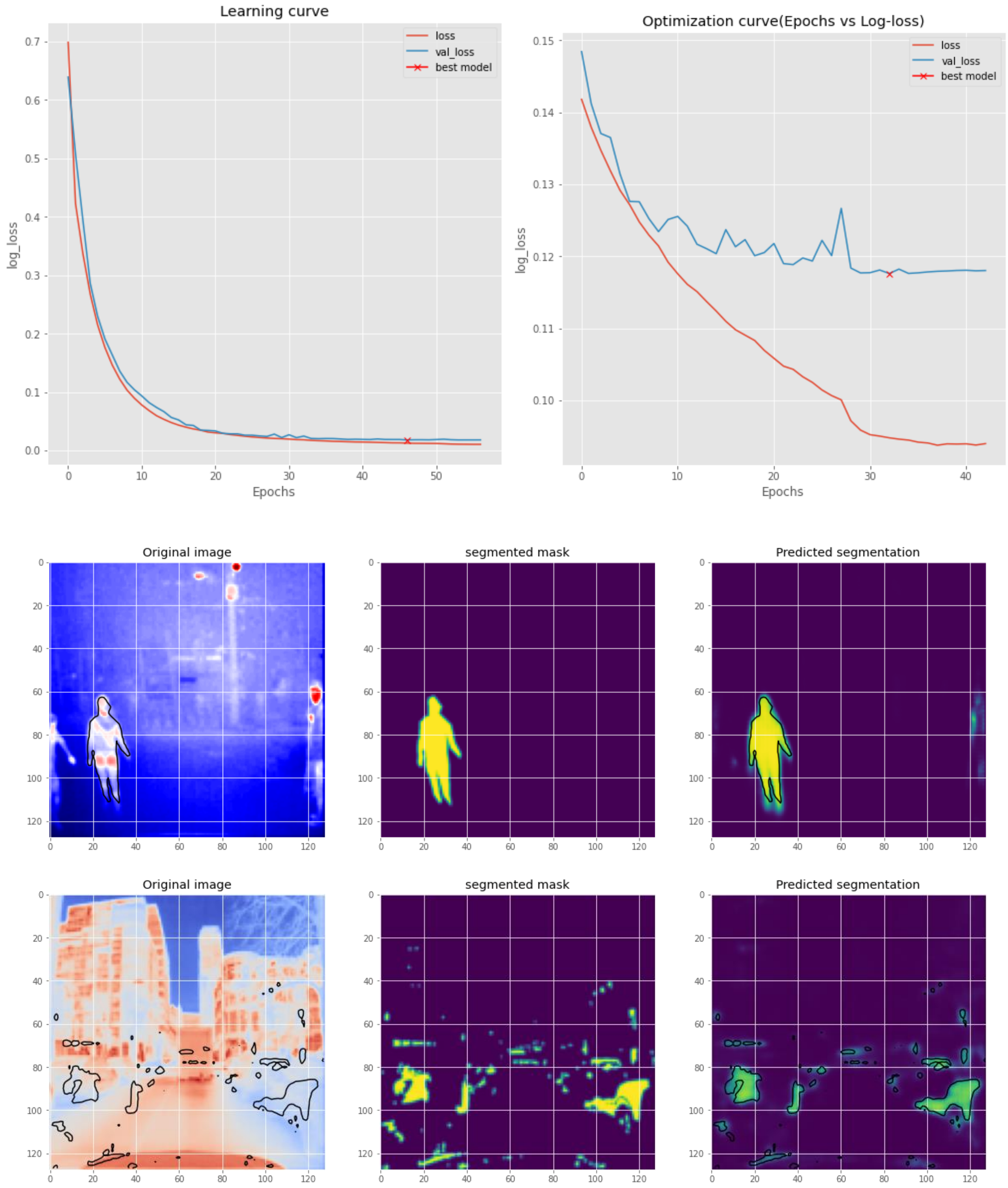


Benchmark results:

To test the robustness of UNet[1] architecture and its ability to perform better even with limited and medium quality Segmentation masks as annotations to the images, a benchmark score with loss curve is calculated using the KAIST[8] pedestrian dataset with segmentation masks provided by Ghose et al[5]. These masks surround all the pixels around pedestrians on a thermal image. Figure below shows one such mask and its corresponding original image from the dataset.



Upon using this dataset in the UNet[1] algorithm, a high accuracy of Segmentation was predicted with the best model achieving a binary cross entropy value of 0.0172 as shown below along with the predicted segmentation masks. The same experiment was done on the dataset from *NUANCE[9]* with the to Zero Binary masks generated which gave a binary cross entropy of 0.118 on the validation dataset.



L-R: loss on KAIST[8] dataset, loss on *NUANCE[9]* data. Line2: segmentation results on KAIST[8], Line3: Segmentation results on NUANCE

• ***Dataset***

- *NUANCE[9]* Autonomous car data collected 36120 sensor messages recorded by the FLIR Boson camera mounted along with a camera and LiDAR array. Specific topics were extracted from the ROSBAG file using MATLAB. For this project every 20th frame from the FLIR BOSON camera was extracted with a collection of 1,806 thermal images.

  These images were then processed to create a segmentation mask which acts as pixel labels for the raw image.

- KAIST[8] dataset: Out of the available images in this dataset, 1,702 images was used along with its segmentation masks were used with a train to test split of 70:30, which is almost the same amount of data being used for the *NUANCE[9]* data.

Experiment steps:

Step1: Using the Annotations for KAIST[8] Salient Pedestrian Dataset[2] from Ghose et al[5], which does not need preprocessing, with a train to validation split of 70: 30, the model was trained using the UNet[1] architecture. As mentioned in the architecture, A convolutional block with 2 layers were added with a kernel size of 3x3 with both the layers activated using a ReLU function and passed through Batch normalization. This is then passed through a series of two different path with the first path going through down sampling and up sampling to end with a convoluted segmentation map.

Step 2: Results from this step is used as a benchmark to test the second dataset.

Step 3: The same architecture is used to test the second dataset from *NUANCE[9]*

- **Conclusion**

The above approach taken towards segmentation of thermal images gives promising results which with more refinement over different datasets paired with other approaches to generate ground truth labels could supplement other state of the art computer vision infrastructures being used to tackle real world problems. With the little available well annotated *NUANCE[9]* dataset, the UNet[1] architecture proves to be an efficient model for the purpose of image segmentation.

To summarize, the project aimed at providing a qualitative as well as quantitative evaluation to the fact that thermal imagery is essential in certain applications to aid the vision cameras and other sensors to understand the scene better and make accurate predictions, eventually solving their shortcomings. Any supervised Deep network is as good as its ground truth annotations and pixel labels, focusing more on automating different approaches to have a robust dataset is essential and over a couple of years down the line, this could benefit other tasks which would need unsupervised Neural networks to be implemented.

- **References**

[1] https://arxiv.org/abs/1505.04597

[2] J. Liu, S. Zhang, S. Wang, and D. N. Metaxas. Multispectral deep neural networks for pedestrian detection. arXiv preprint arXiv:1611.02644, 2016

[3] https://docs.opencv.org/3.4/db/d8e/tutorial_threshold.html

[8][4] Soonmin Hwang, Jaesik Park, Namil Kim, Yukyung Choi, and In So Kweon. Multispectral pedestrian detection: Benchmark dataset and baseline. In CVPR, pages 1037– 1045, 2015.

[5]https://openaccess.thecvf.com/content_CVPRW_2019/papers/PBVS/Ghose_Pedestrian_Detection_in_Thermal_Images_Using_Saliency_Maps_CVPRW_2019_paper.pdf

[6] https://drive.google.com/drive/folders/12a5OxlFF3ZNcAWARumASsiRHnp0l7bRu

[7] https://en.wikipedia.org/wiki/Otsu%27s_method

[9]https://www.sciencedirect.com/topics/engineering/global-thresholding#:~:text=A%20global%20thresholding%20technique%20is,obtained%20from%20the%20whole%20image. https://docs.opencv.org/master/da/d22/tutorial_py_canny.html

[10]https://drive.google.com/drive/folders/1xowIku0Wp7PKvMO_sHw53Z-0vY0mfTxt?usp=sharing

[11]https://www.google.com/search?q=thermal+images&safe=off&rlz=1C1CHBF_enUS916US916&sxsrf=ALeKk00p1Er8VsP8FC7bIiD3JUisl6KHCw:1607610963516&source=lnms&tbm=isch&sa=X&ved=2ahUKEwij9arD0cPtAhURac0KHQQmC9cQ_AUoAXoECBIQAw&biw=1536&bih=731&dpr=1.25#imgrc=vuzb1beLCgLZsM&imgdii=F3ri454zfRYXvM

[12]https://github.com/Information-Fusion-Lab-Umass/Salient-Pedestrian-Detection/blob/master/README.md#pedestrian-detection-from-thermal-images-using-saliency-maps

[13] http://www.ams.sunysb.edu/~lindquis/3dma/man_3dma/manual/node19.html

[14] https://en.wikipedia.org/wiki/Otsu%27s_method

[15]http://scholar.google.com/scholar_url?url=http://danida.vnu.edu.vn/cpis/files/Refs/LAD/Algorithm%2520AS%2520136-%2520A%2520K-Means%2520Clustering%2520Algorithm.pdf&hl=en&sa=X&ei=GV7SX6SREo6CywTRt52ACQ&scisig=AAGBfm19Hvy-Iv9iKUAnxtyGWznvyLBsrw&nossl=1&oi=scholarr