

On the Use of Deep Learning for Open World Person Re-Identification in Thermal Imagery

T.A. Robson - Supervised by T.P. Breckon
Department of Computer Science, Durham University



Introduction

A fundamental task for a distributed multi-camera surveillance system is person re-identification, or to be able to correctly identify people as they transition between different camera views. This has been well researched in the colour spectrum, but there has been very little research done on solving this problem in the thermal spectrum, making our work state of the art. Recently attention of the researchers of re-identification has shifted to deep learning based solutions, with much success. The purpose of this research to replicate this success in thermal imagery.

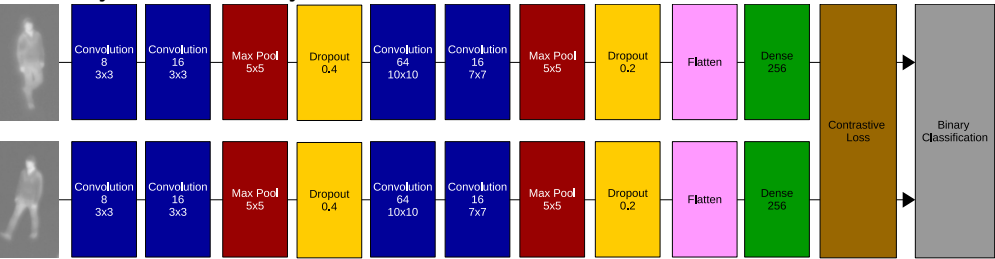
Person Detection and Tracking

The processes of person detection and tracking employed in this project can be broken down into multiple stages.

- ▶ Background Subtraction. The Mixture of Gaussians (MOG) [1] method allows the system to learn a background model and compares each new frame against the background.
- ▶ Person Identification. Performing contour detection on a foreground object and using the Histogram of Oriented Gradients (HOG) [2] to determine whether this object is a person.
- ▶ Track-Learn-Detect (TLD) [3]. This breaks down the person-tracking task into Tracking (following the object between frames), Learning (estimating the positional errors and updating the detector) and Detection (correcting the tracker based on previous observations).

Network Architecture

Our network architecture is a deep siamese Convolutional Neural Network (CNN). The network is trained to determine whether a pair of images show the same person or not, outputting a Euclidean distance which informs a binary classification. Our CNN consists of convolutional layers, pooling layers and fully connected layers.



Network Results

Here we present the performance of our network. We show the confusion matrices and classification reports for the training, valiation and testing splits.

The graphs show the distribution of Euclidean distances between pairs of images of the same person and of different people. The clear split shows that our system is performing well.

		Predicted Label	
		Positive	Negative
Ground Truth	Positive	0.9982	0.0018
	Negative	0.0003	0.9997

(a) Confusion matrix for training set

		Predicted Label	
		Positive	Negative
Ground Truth	Positive	0.9882	0.0118
	Negative	0.0024	0.9976

(b) Confusion matrix for validation set

		Predicted Label	
		Positive	Negative
Ground Truth	Positive	0.9915	0.0085
	Negative	0.0043	0.9957

(c) Confusion matrix for test set

Metric	Accuracy	Precision	Recall	F1 Score
Value	99.89%	99.81%	99.97%	99.89%

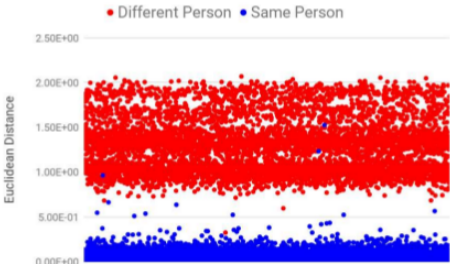
(a) Classification report for training set

Metric	Accuracy	Precision	Recall	F1 Score
Value	99.29%	99.76%	98.82%	99.29%

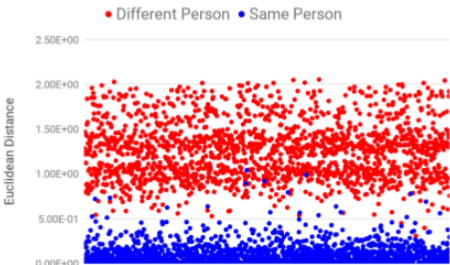
(b) Classification report for validation set

Metric	Accuracy	Precision	Recall	F1 Score
Value	99.36%	99.15%	99.57%	99.36%

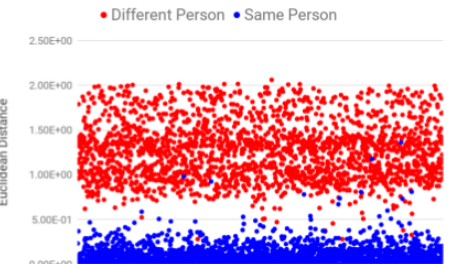
(c) Classification report for test set



(a) Train set results



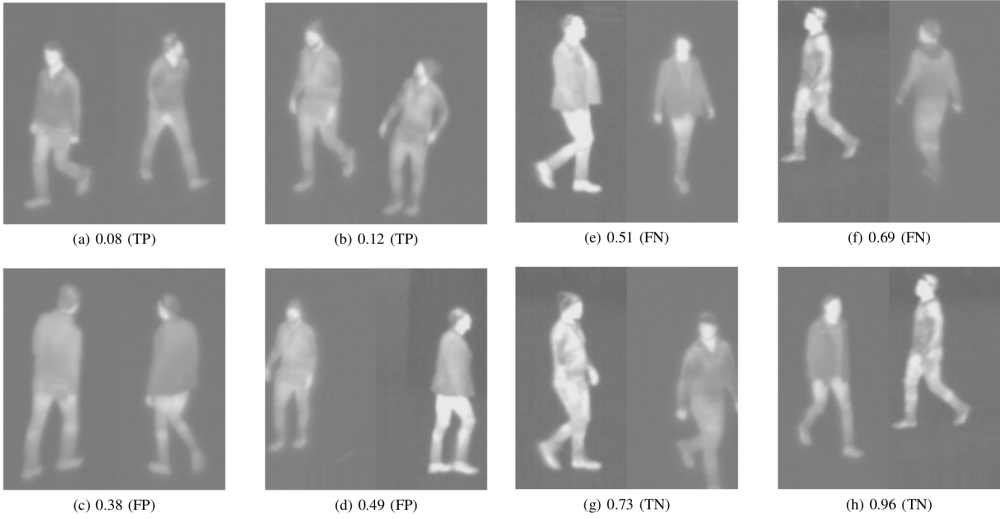
(b) Validation set results



(c) Test set results

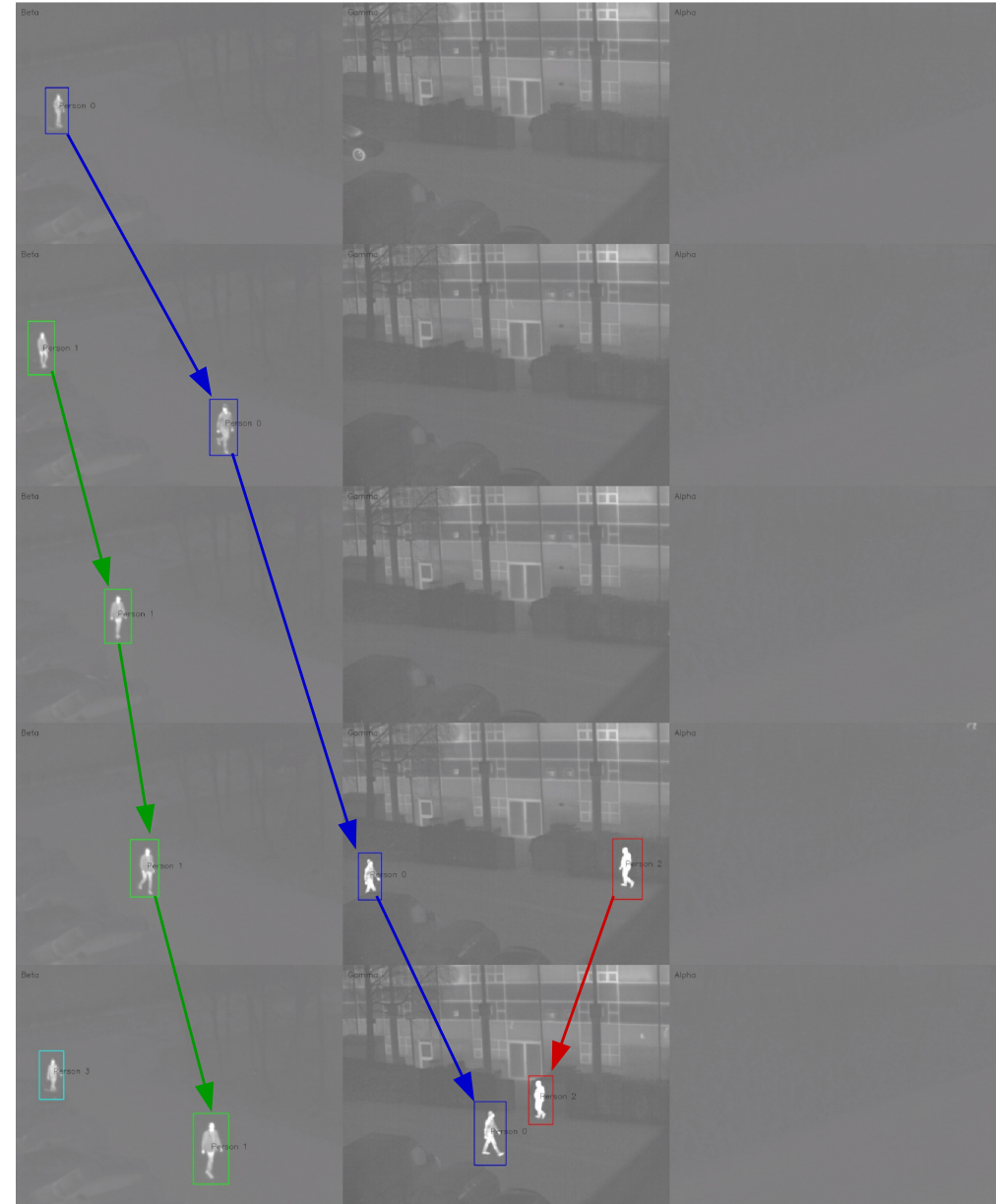
Network Classification Results

The captions of each image shown here give the Euclidean distance between the image pair, as well as whether this classification was a True Positive, False Positive, False Negative or True Negative.



The Re-Identification System

Here we show the performance of our re-identification system across our dataset of 3 cameras. It works well, achieving 96.32% accuracy. A video can be found at <https://tinyurl.com/ybd69py2>.



Conclusion

We have developed a fully functional thermal re-identification system using a track-learn-detect (TLD) tracker and a deep siamese CNN that performs well on a varied dataset. As there has been very little previous work on solving the re-identification problem in thermal imagery, we are defining the state of the art with this work, and have shown that it is possible when using deep learning to accurately discriminate between people as they move through an area covered by a thermal camera network.

References

- [1] Zoran Zivkovic. Improved Adaptive Gaussian Mixture Model for Background Subtraction. Proceedings of the 17th International Conference on Pattern Recognition, 2004.
- [2] Navneet Dalal and Bill Triggs. Histograms of Oriented Gradients for Human Detection. In IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005.
- [3] Zdenek Kalal, Krystian Mikolajczyk, and Jiri Matas. Tracking-Learning-Detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 6, 2010.