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 associate people across different camera views. This has been well researched in colour, but there has been very little research done on solving this problem in thermal, making our work state of the art. Since our work in [1], the attention of the researchers of re-identification has shifted to deep learning based solutions, with much success. We are attempting to replicate this success in thermal imagery.

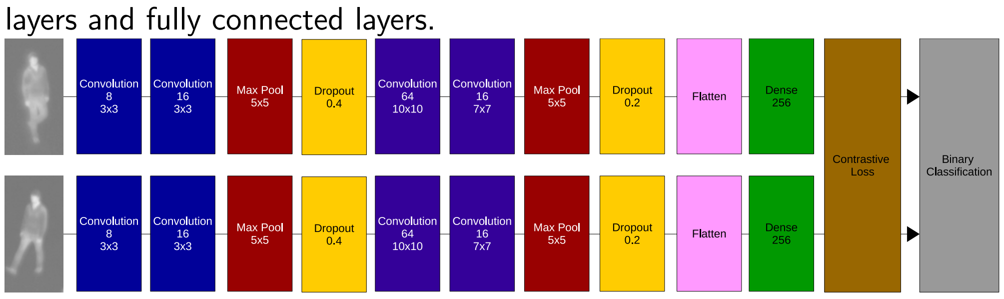
Person Detection and Tracking

The process of Person detection and Tracking employed in this Project can be broken down into multiple stages.

- Background Subtraction. This is done using the Mixture of Gaussians (MOG) [1] method, allowing the system to learn a background model and comparing each new frame to this.
- Person Identification. This is done by performing contour dectection on a foreground target and using the Histogram of Oriented Gradients (HOG) [2] to determine whether this target is a person.
- TLD Tracker [3]. This breaks down the person-tracking task into tracking (following the object between frames), learning (estimating the errors made by the detector and updating it) and detection (correcting the tracker if necessary based on previous observations).

Network Architecture

Our network architecture is a Deep Siamese CNN. This means that the network is trained on pairs of images, and determines whether these 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 the same person and different people. The clear split shows that our system is performing well.

		Predicted Label	
		Positive	Negative
Ground Truth	Positive	TP = 5924	FN = 11
	Negative	FP = 2	TN = 5933

(a) Confusion matrix for training set

		Predicted Label	
		Positive	Negative
Ground Truth	Positive	TP = 1675	FN = 20
	Negative	FP = 4	TN = 1691

(b) Confusion matrix for validation set

		Predicted Label	
		Positive	Negative
Ground Truth	Positive	TP = 2097	FN = 18
	Negative	FP = 9	TN = 2106

(c) Confusion matrix for test set

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

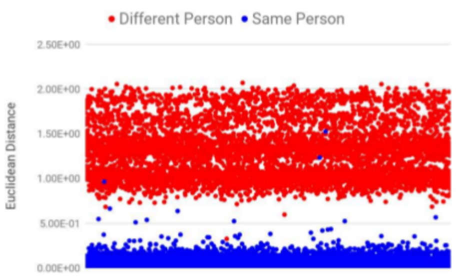
(a) Classification report for training set

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

We present some of the classification results of our network. 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.

(a) 0.082241 (TP)

(b) 0.119061 (TP)

(e) 0.515469 (FN)

(f) 0.697845 (FN)

(c) 0.376771 (FP)

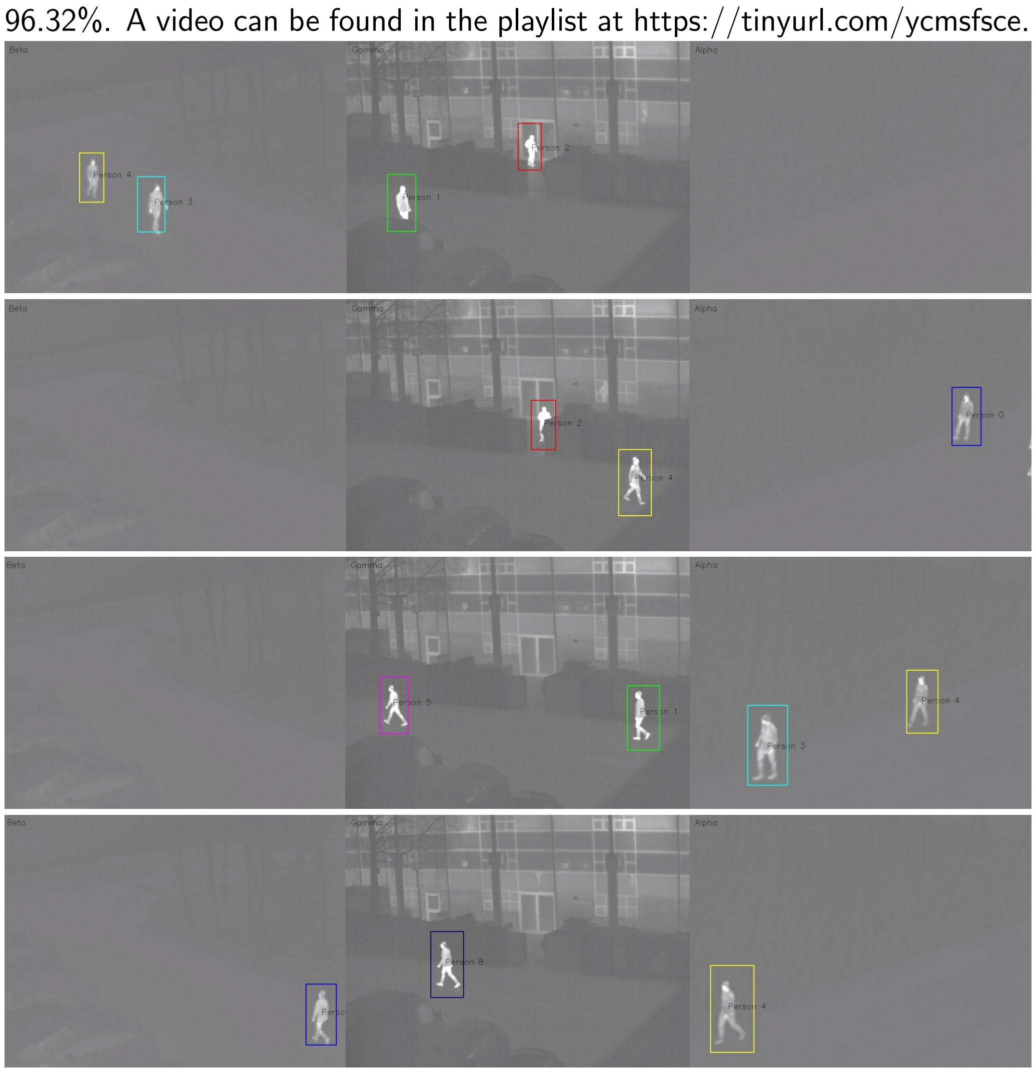
(d) 0.498734 (FP)

(g) 0.734032 (TN)

(h) 0.960652 (TN)

The Re-Identification System

Here we show the performance of our re-identification system across our dataset of 3 cameras. It works very well, achieving and accuracy rate of 96.32%. A video can be found in the playlist at https://tinyurl.com/ycmsfsce.



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 very well on a varied dataset.As the 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 extract features that would not necessarily occur to a human observer.

References

[1] Thomas A. Robson and Toby P. Breckon. Camera-to-Camera Tracking for Person Re-identification within Thermal Imagery. Bachelors thesis, Durham University, 2017.

[2] Zoran Zivkovic. Improved Adaptive Gaussian Mixture Model for Background Subtraction. Proceedings of the 17th International Conference on Pattern Recognition, 2004.

[3] Navneet Dalal and Bill Triggs. Histograms of Oriented Gradients for Human Detection. In IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005.

[4] Zdenek Kalal, Krystian Mikolajczyk, and Jiri Matas. Tracking-Learning-Detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 6, 2010.