Person Re-ID for Follow-Me Task

Sarah Brennan
sbrenn@Stanford.edu

Motivation and Overview

Motivation - This project focuses on the task of re-identifying a target after it is lost from view. The task is intended to be implemented on the JackRabbot (JR), the social navigation robot. The goal is to acquire a target and then follow the target around different spaces at a set distance. This project addresses the specific challenge of re-identifying the target after they are lost from view.

Model - The models are built using video data with bounding boxes to learn feature descriptors which are then clustered to identify the target.

Data

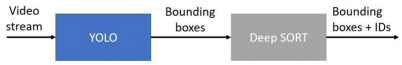
The data used in these tests came from the Motion Analysis and Re-identification Set (MARS) [1], which included video data with over a million bounding boxes marked with the corresponding IDs of over 1,000 people.

The video data came from six different cameras viewing one area. The data is used to evaluate a model's ability to re-identify individuals from different cameras. The challenge of different lighting, time stamps, and viewing angles from different cameras is also relevant to the JR follow-me task because the robot is moving through a changing environment as it tracks the target.

Architecture

Currently JR uses a YOLO system to identify humans and generate bounding boxes. These bounding boxes are passed to a Deep SORT model which generates IDs for each individual. However this model could be more robust against ID switching and occlusions.

For this project, the decision was made to stick with a YOLO to Deep SORT architecture because it is lightweight enough to run real time. Modifications were made to the Deep SORT model to address the issue of occlusion and ID switching to provide a more robust model for follow-me tasks.



Models

Deep SORT – NN Clustering
Nearest neighbor clustering was used to create IDs based on the feature descriptors. The model used a Kalman filter to monitor tracks and assign bounding box detections to specific tracks. The assignment used a metric that took into account short term motion prediction ($d^{(1)}$) and similarity of the features ($d^{(2)}$) [2].

$$c_{i,j} = \lambda d^{(1)}(i,j) + (1\text{-}\lambda) d^{(2)}(i,j)$$

Cosine Metric Learning – Feature Descriptors
A CNN was trained offline to learn unique feature descriptors for the bounding boxes. The CNN used a wide-residual network architecture and output a 128d descriptor vector [3].

Results

Deep SORT – Gating and Gallery
Both the assignment gating and gallery length were modified to try to address the problem of occlusions. The gallery was expanded to store more feature vectors so the model could re-identify after longer periods of absence.

MOT16 Metrics
The total number of ID switches during a segment was important for comparing the different models.

Cosine Metric Learning
The original architecture used two convolutional layers followed by six residual blocks with 2.8 million parameters. The number of convolution layers was modified and a shortcut was added to learn features at different stages similar to OSNet [4].

Layer	Filter Size	Stride	Output
Conv 1	3 x 3	1	32 x 128 x 64
Conv 2	3 x 3	1	32 x 128 x 64
Max Pool 3	3 x 3	2	32 x 64 x 32
Residual 4	3 x 3	1	32 x 64 x 32
Residual 5	3 x 3	1	32 x 64 x 32
Residual 6	3 x 3	2	64 x 32 x 16
Residual 7	3 x 3	1	64 x 32 x 16
Residual 8	3 x 3	2	128 x 16 x 8
Residual 9	3 x 3	1	128 x 16 x 8
Dense 10			128
Batch and l_2 norm			128

Future

Future work includes testing the new models and comparing against the MOT16 evaluation. Also trying transfer learning using data collected from JR, and incorporating target identification. Additionally, attempting some of the unsupervised learning models, and trying an end-to-end approach rather than stacking many custom features.

References

[1] MARS: A Video Benchmark for Large-Scale Person Re-identification. Zheng, Liang and Bie, Zhi and Sun, Yifan and Wang, Jingdong and Su, Chi and Wang, Shengjin and Tian, Qi
[2] Simple Online and Realtime Tracking with a Deep Association Metric. Wojke, Nicolai and Bewley, Alex and Paulus, Dietrich.
[3] Deep Cosine Metric Learning for Person Re-identification. Wojke, Nicolai and Bewley, Alex.
[4] Omni-Scale Feature Learning for Person Re-Identification. Zhou, Kaiyang and Yang, Yongxin and Cavallaro, Andrea and Xiang, Tao.