A multiple of object tracking algorithm based on YOLO detection

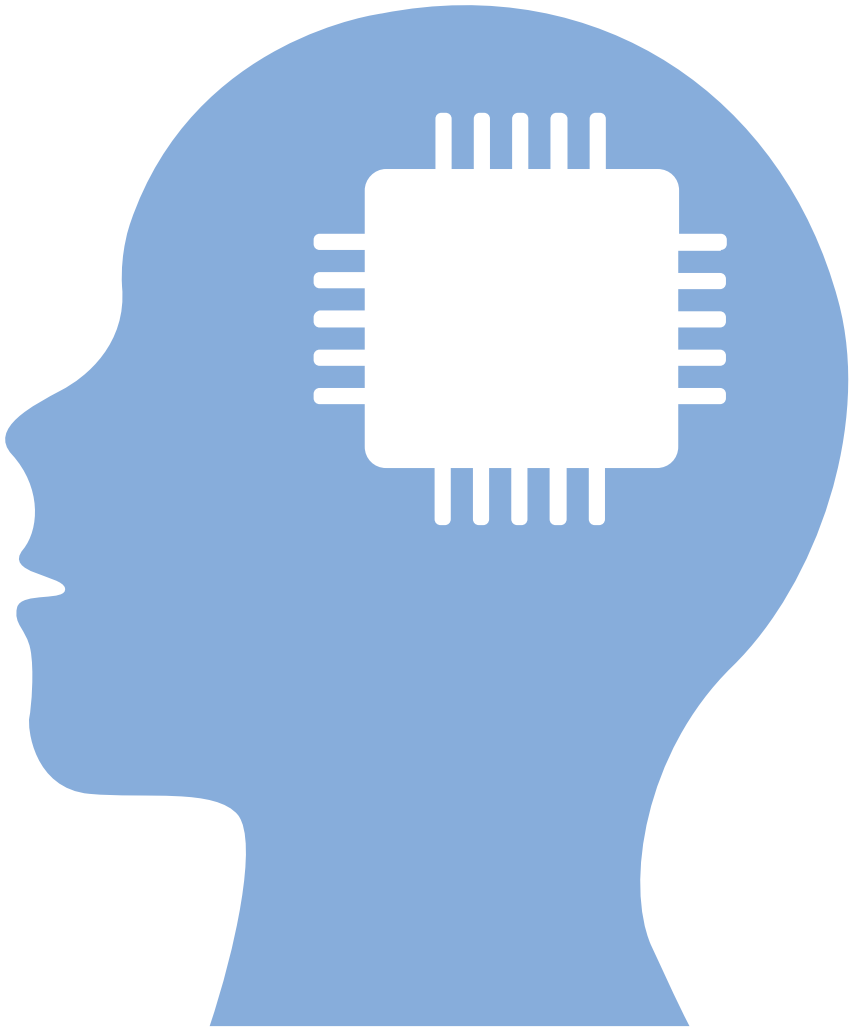
Online Realtime multi target tracking method

Introdcution

Goal: improve the accuracy and efficiency of multi tracking

Implementations:

- Intelligent monitoring
- Human computer interaction
- Virtual reality



Definition

Multi target tracking: A process which **positions target in the image sequence** and finds the most similar candidate target area.

Difficulties faced : complex scenarios such as blocking and eclipse

Comparative value :
Previous methods only track through pixels to generate coordinate motion trajectory therefore they are unable to detect trailing and wandering.

1. Sequential training method based on CNN

Purpose : Effective transfer of the deep features of online application.

- It regards the process as a sequentially training an optimal ensembles of base learners
- And a convolution with mask layer is done to reduce overfitting further.

2.Online Realtime multi target tracking method

- Frame-matching based on frame by frame matching and appearance does not rely on online learning,
- Two stage drift handling method with novel confidence **to** correct drifting tracks due to abrupt motion, change of object under occlusion, prolonged in accurate detection
- In addition a fragmentation handling method based on track to track association is proposed to solve issues when object trajectory is broken due to long term occlusions

3. Novel online multi target tracking framework

- The framework uses top level features to be trained to target class classifier and uses low level to accomplish target matching and association with lower layer containing more details.
- To avoid computational cost by online fine tuning

1. The frame retains historical appearance characteristic for each target
2. The depth model is trained through and offline training strategy

Challenges faced by previous method

1. Extraction of target feature
2. Speed of the multi target tracking process

Proposed Method : YOLO framework for target detection,

- Speed 45fps
- The size and position of a target is first acquired.
- Then indepth extraction is performed to remove noise data of unrelated regions
- This is to reduce complexity and time complexity
- LSTM : Used to get temporal relationship between frames
- Euclidean distance : used to measure similarity as to match and associate targets
- Experiment was done based on MOT-16 and Microsoft MSR data sets

Main Steps

a. Real-time target detection of video streaming by YOLO and select the detection target;

b. The LSTM method is used to obtain continuous target positions in the time series, and the depth features of the detection targets are extracted;

c. Using the Euclidean Metric algorithm to calculate the similarity of different targets;

d. Match and correlate targets to complete tracking of multiple targets.

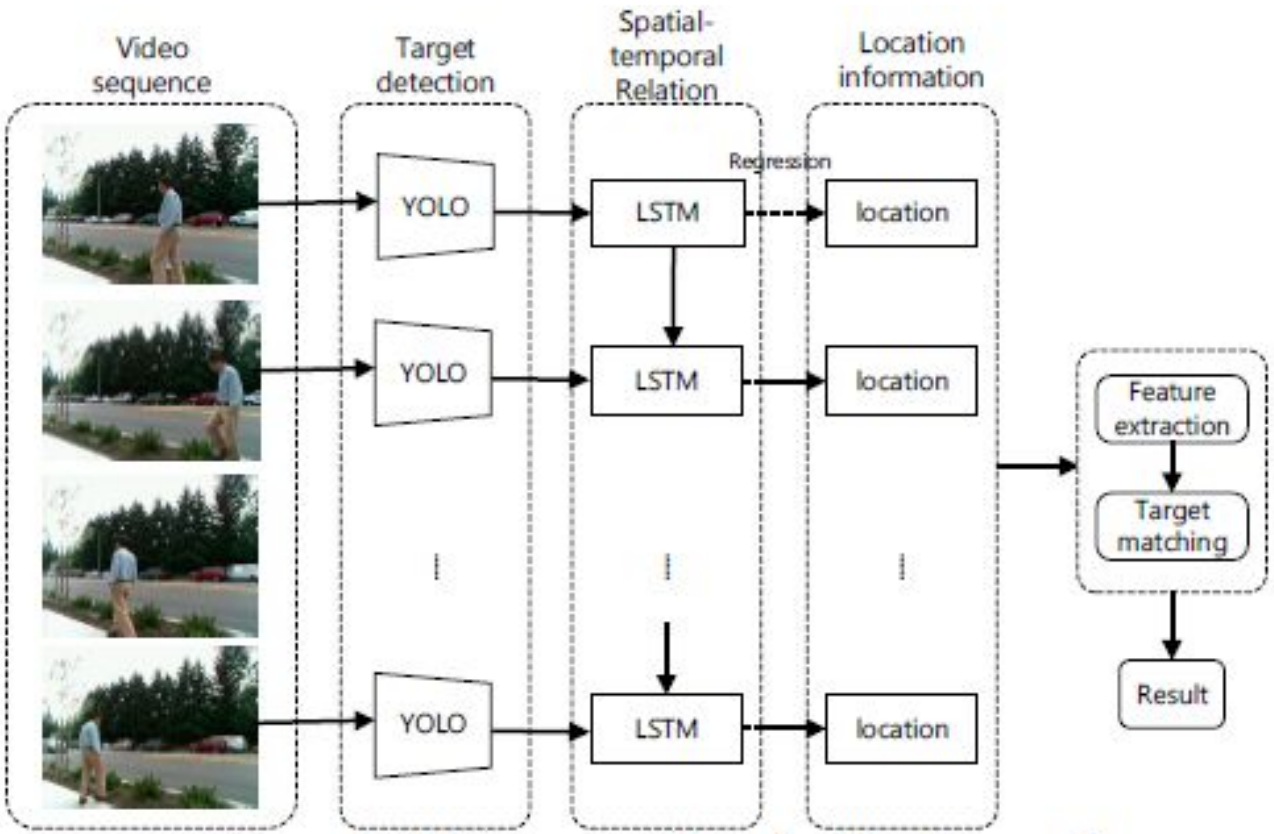


Figure 1. Multi-target tracking algorithm framework

$$p(B_1, B_2, \ldots, B_T \mid X_1, X_2, \ldots, X_T) = \prod_{t=1}^{T} p(B_t \mid B_{<t}, X_{\leq t})$$

The multi tracking probability is expressed as stated above

Bt = positions

Xt = frame

At a certain time t ,

Respectively, B< t is all the historical in positions before t , and X<= t

are the historical input frames until t .

YOLO Method : , a deep learning target detection technology based on regression method,

- Integrates target area prediction and target category prediction into a single neural network model
- 24 convolution layers for image features are extraction
- 2 fully connected layers where image positions and class accuracy are predicted

The YOLO algorithm

a) Firstly, divide the input image into a S*S grid;

b) For each grid, predict there will be B borders which including the boarders representing the target confidence and the probability that each border region is over multiple categories;

c) Based on, the target S*S*B windows can be predicted and then the target windows can be removed if they have the lower probability than the threshold.

d) redundant window can be removed by Non- Maximum Suppression (NMS)

LSTM-based target tracking

- Made to solve vanishing gradient problem
- LSTM introduces three gates hold states
- It can receive the output of the previous moment,
- The current system state current system input can be updated
- Then outputted through the input gate, forgetting gate and output gate
- It is added to the network for tracking module training and it is combined with YOLO

- After regressing the target coordinate positions in successive frames, the target matching algorithm is used to perform target correlation, and then multiple moving targets are continuously tracked
- There are two data streams entering the LSTM in depth networks namely,
  1. target feature map from the convolutional layer
  2. detection information *Bt,i* from the fully connected layer

Feature extraction and target association
1. Feature extraction

- Multi-target tracking algorithm of paper is based on YOLO target detection
- It applies VGG-Net model to extract the depth features
- It is based of convolutional neural **network** that is 19 layers deep
- ROI region is first divided according to the target detection result and then, pool and extract the 1024- dimensional depth features of all targets in detection model.
- Because target feature is highly abstract, the model can accurately characterize each target

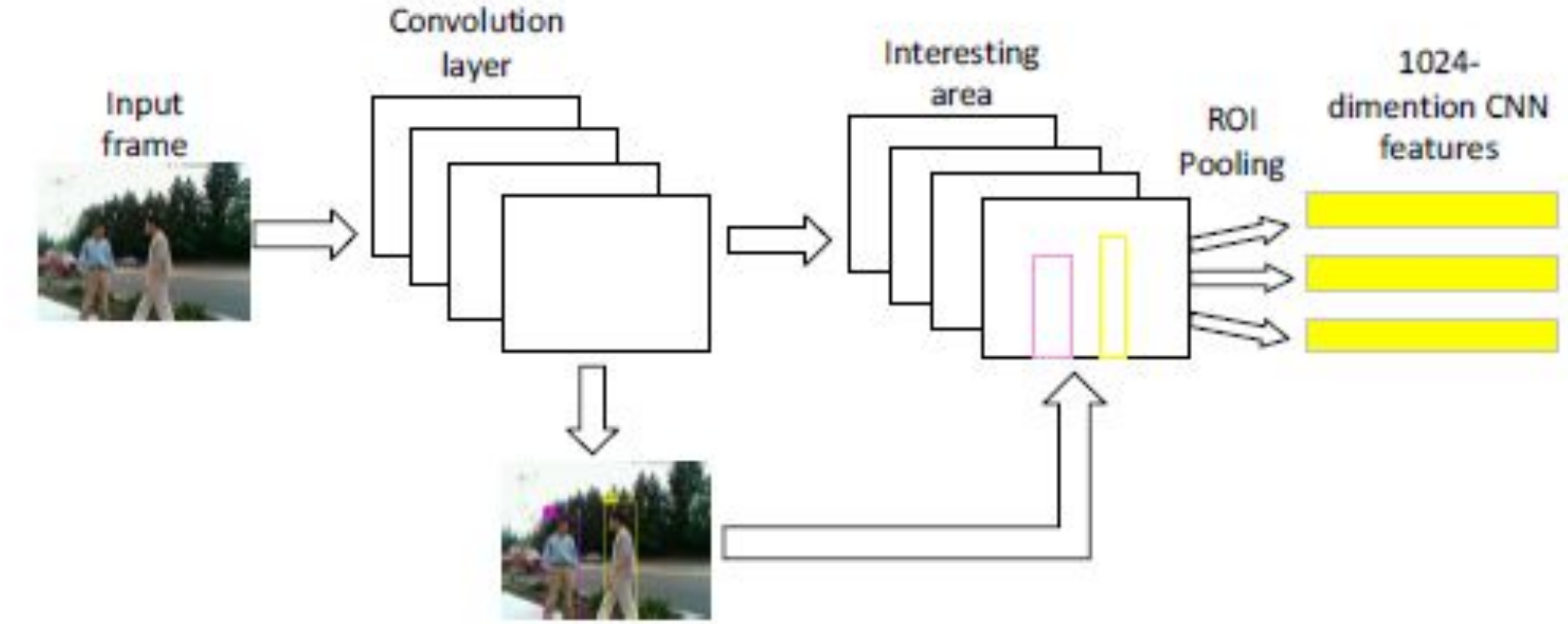


Figure 3. Prepare Your Pape Feature extraction

Target association

Euclidean metric algorithm is the most widely used similarity measurement algorithm.

The distance formula between two points X1 and X2 in *N* dimensional Euclidean space is

$$d = \sum_{i=1}^{N} \sqrt{(x_{1,i} - x_{2,i})^2}$$

N = dimension, *d* = Euclidean distance between two points X1 and X2 .
The smaller the *d* value, the greater the correlation and similarity between the targets.

Euclidean metric to establish the relationship between frames and find the optimal association and matching

$$dist = \sum_{i=1}^{N} \sqrt{(f_{m,i} - f_{n,i})^2}$$

N = dimension, m,n = random two frames,

fm,i = a feature matrix obtained by the No. m frame.

When the Euclidean distance dist of two frames is smaller, the similarity between the two frames is larger, and the correlation between the two frames is stronger.

Experiment Result

MOTA = the number of targets and the accuracy of the target-related attributes

MOTP = the precision in finding the target position.

In general it evaluate the ability of the algorithm to track the target continuously.

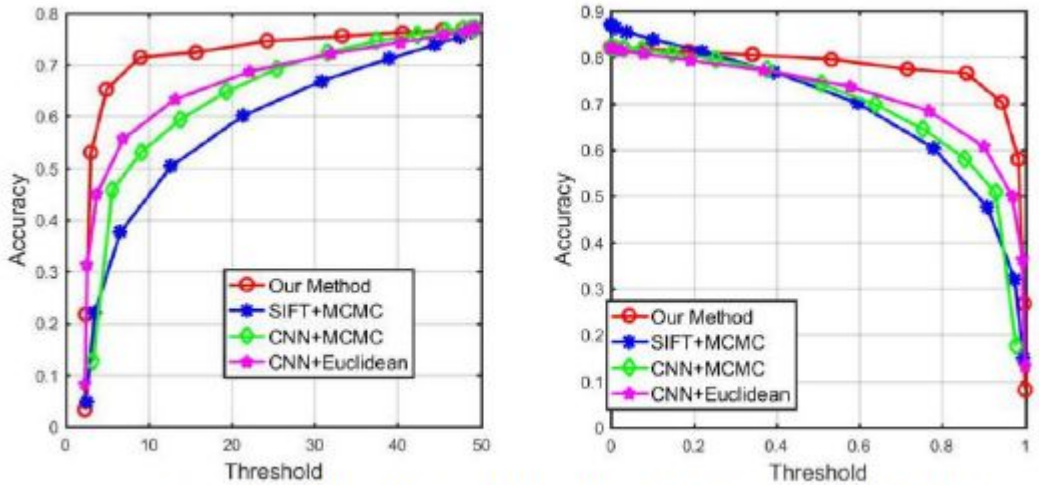


Figure 7. Precision plot (left) and success plot(right)

- The method suggested can express the apparent features well and adapt to the initial appearance change of the objects.
- Compared with the other three algorithms, the proposed algorithm has the highest correct rate under different thresholds,
- So it indicates that the overall performance of the proposed algorithm is better than other tracking algorithms

TABLE I. DIFFERENT ALGORITHMS COMPARE RESULTS IN MSR

Method	Different Evaluation			
	MOTA	***MOTP***	***FN***	***FP***
SIFT+ MCMC	35.0%	75.3%	865	442
CNN+ MCMC	47.5%	74.5%	652	367
CNN+ Euclidean	48.5%	75.5%	534	311
Our Method	50.3%	75.5%	506	320

TABLE II. DIFFERENT ALGORITHMS COMPARE RESULTS IN MOT

Method	Different Evaluation			
	MOTA	***MOTP***	***FN***	***FP***
SIFT+ MCMC	36.2%	73.6%	6871	91173
CNN+ MCMC	45.3%	73.7%	6895	91117
CNN+ Euclidean	46.5%	74.2%	6373	90914
Our Method	47.2%	74.0%	6412	89368

Thank you ☺