International Journal of Engineering, Mathematical and Physical Sciences
ISSN: 2517-9934
Vol:14, No:11, 2020

# Spatial-Temporal Awareness Approach for Extensive Re-Identification

Tyng-Rong Roan, Fuji Foo, Wenwey Hseush

***Abstract*—**Recent development of AI and edge computing plays a critical role to capture meaningful events such as detection of an unattended bag. One of the core problems is re-identification across multiple CCTVs. Immediately following the detection of a meaningful event is to track and trace the objects related to the event. In an extensive environment, the challenge becomes severe when the number of CCTVs increases substantially, imposing difficulties in achieving high accuracy while maintaining real-time performance. The algorithm that re-identifies cross-boundary objects for extensive tracking is referred to Extensive Re-Identification, which emphasizes the issues related to the complexity behind a great number of CCTVs. The Spatial-Temporal Awareness approach challenges the conventional thinking and concept of operations which is labor intensive and time consuming. The ability to perform Extensive Re-Identification through a multi-sensory network provides the next-level insights – creating value beyond traditional risk management.

***Keywords*—**long-short-term memory, re-identification, security critical application, spatial-temporal awareness.

## I. Introduction

IN today's security industry, data generated by CCTV, IoT sensors or social media are the important elements for many security-critical applications. Recent development of AI and edge computing plays a critical role to capture single-sourced sensory events (e.g., detecting an unattended suitcase with a CCTV), which, nevertheless, tell partial and probably imprecise information, rather than the whole truth. Putting together multiple streams of events for cross examination is the way to detect and figure out what actually happened on the field.

One of the key issues immediately following the detection of a meaningful event is to track and trace the moving objects related to the event in a vast area. The problem is referred to as extensive tracking, to differ from the traditional tracking problem that associates an object from one frame to another in one single CCTV. The algorithm that re-identifies cross-boundary objects in extensive tracking problem is referred to extensive re-identification, which emphasizes the issues related to the complexity behind a great number of CCTVs. The goal of extensive tracking is to project the trajectory of every object moving within a surveillance area in an accurate and efficient way.

As an industrial practitioner, we have developed a bionic brain framework to capture sensory events and reconstruct the pictures of what happened and when/where in an immense and persistent memory space. The framework contains five levels, as shown in Fig. 1:

- (L1) perception to capture sensory events such as object detection and tracking,
- (L2) attention to detect complex events and perform re-identification with transient memory,
- (L3) working memory for spatial-temporal awareness, which functions as a mind map to show relationships among pieces of the whole in time and in space,
- (L4) long-term memory system 1 (fast thinking) for prediction, and
- (L5) long-term memory system 2 (slow thinking) for reasoning.

For extensive re-identification, the challenge becomes severe when the number of CCTVs increases substantially, imposing difficulties in achieving high accuracy while maintaining real-time performance. A real-world example is to monitor thousands of passengers passing under hundreds of CCTVs during rush hours in a bus station. In such an extensive environment, the key issue leading to inaccuracy is the large number of candidates to be screened by feature comparison. In addition to feature-based re-identification model, this study employs spatial-temporal relations to narrow down the potential candidates in two ways: (1) by the exclusion principle where two distinct persons cannot show up in the same location at the same time, and (2) by projected movement with probability of persons showing up from the neighbor view areas. Once candidates are screened out, the feature-based model is applied for identification.

Part of this study is to conduct a proof-of-concept (POC) demonstration using extensive re-identification to solve the problem of unattended suitcase detection in a public place. The POC answers three questions: (1) what happened and when/ where, (2) who dropped the suitcase and (3) where is the person now and next.

## II. Related Work

Object re-identification is a challenge task which requires matching objects across disjoint cameras views at different times and locations. This technique has been widely applied on person [1]-[3] and vehicles [4], [5] for the purpose of public safety monitoring and business data analysis.

Person re-identification commonly uses appearance features, determined by clothes or accessories, extracting color, texture or shape to describe an individual's identities [6]-[8]. However, the accuracy of person re-id is influenced by various factors,

T. Roan (PhD) is a senior researcher and W. Hseush (PhD) is a computer scientist at BigObject, Taiwan (e-mail: jennyroan@bigobject.io, wenwey@bigobject.io).

F. Foo is Chief Digital Officer of Certis, 1958 20 JALAN AFIFI CERTIS CISCO CENTRE Singapore 409179 (e-mail: fujuss_foo@certisgroup.com).

such as camera properties [9], complex background and lighting [10] and individual pose changes [11]. Wojke et al. [12] have improved the original SORT algorithm [13] with deep appearance descriptor to track objects throughout longer period and reduce the number of identify switches.

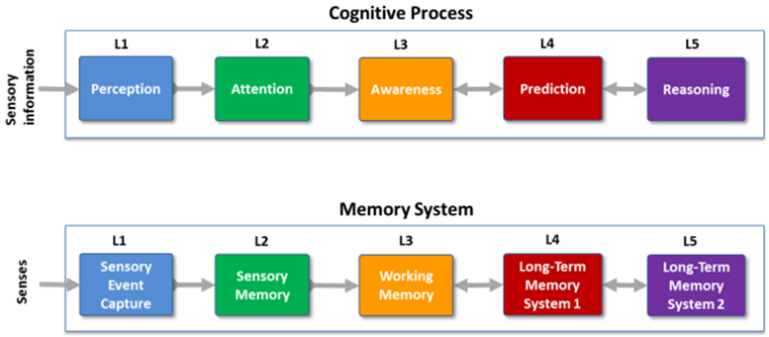


Fig. 1 Bionic Brain Framework

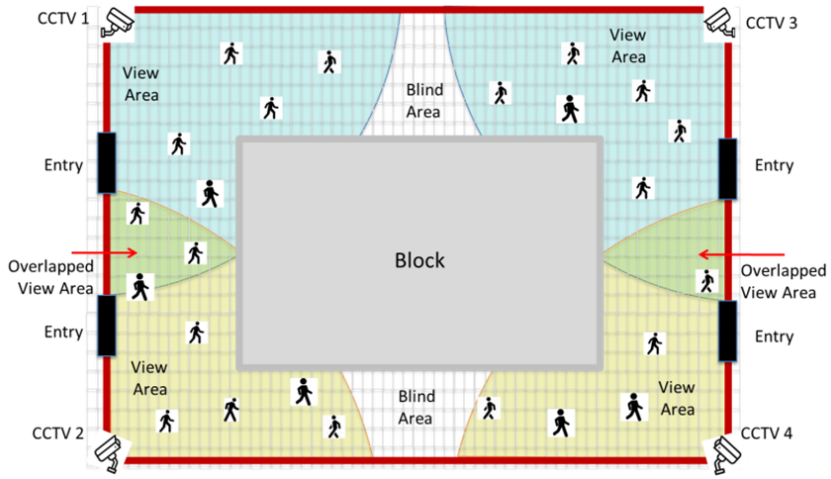


Fig. 2 CCTV View Areas and Blind Areas

## III. Background and Assumption

We assume a large number of CCTVs over a vast public area and numerous moving objects such as persons passing through the area, where all moving objects are anonymous (i.e., without personal ID). A bionic computing system (referred to as the bionic brain) is designed to track every moving object within the area, starting with the first appearance of an object captured by a CCTV and ending with the last appearance of the object. The area that a CCTV can "see" is referred to as a *view area*, where two view areas may overlap in portion. The area not viewed by any CCTV is referred to a *blind area*. With proper CCTV calibration and positioning, the world coordinates (i.e., latitude/longitude) for every pixel in a frame captured by a CCTV can be calculated and derived. A simple example of public area with four CCTVs (therefore four view areas) is shown in Fig. 2, where two blind areas are not covered by the CCTVs. The space where no people can enter or pass through is labeled as "block".

While a moving object disappears into a blind area, the bionic brain strives to project the possible trajectories inside the area in order to keep track of the object movement throughout the vast space. The space is gridded into a finite set of rectangular cells (latitude/longitude rounded up to the *n*th decimal place), identified as geohashes, a convenient way of expressing locations (anywhere in the world) using a short alphanumeric string, with greater precision obtained with longer strings. For a blind area, the set of geohashes that are parts of or adjacent to a view area are referred to as *margin cells of the blind area*.

To achieve real-time performance, the bionic brain adopts a scalable architecture for extensive tracking, as shown in Fig. 3.

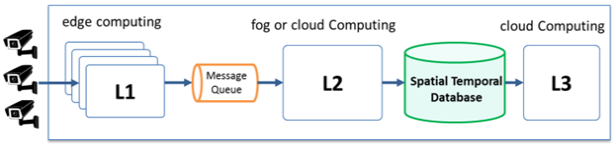


Fig. 3 Scalable Architecture

L1 includes a set of GPU-powered edge devices to perform object detection such as Yolo [15], followed by a local tracking algorithm and feature extraction. L2 runs on a cluster of fog/

cloud computers to perform complex event detection and re-identification. L3 is a cloud computing subsystem to retain and trace the trails of all objects in a huge spatial-temporal database, that is, the mind map of the bionic brain.

Three levels of algorithms developed for extensive tracking are shown in Fig. 4, each of which is for L1, L2 or L3, respectively.


Fig. 4 Extensive Tracking Algorithms

L1 tracking algorithm is designed for tracking objects within one CCTV, while L2 re-identification algorithm is for tracking moving objects across CCTVs. L3 tracing algorithm is designed to reason out and construct a traceable graph by connecting the paths built in L2.

Although it takes all three levels of algorithms to work for a real-world track-and-trace application, the key issue we address in this paper is the one in L2, where, when an object shows up in a view area for the first time, the bionic brain attempts to figure out which CCTV view area it came from and its trajectory passing in that CCTV view area. A combination of feature-based model and spatial-temporal computing model is adopted for L2. For the first one, we adopt *Re-Identification Network* [14], a learning-based model with 2048 dimensions of human appearance features. The spatial-temporal model is designed to address the issues that only arise from the large-scale re-identification problem, aiming to significantly reduce the numbers of candidate objects for feature matching.

## IV. Extensive Tracking

The re-identification algorithm shown in Fig. 4 is further divided into two parts, one for overlapped view areas and the other for non-overlapped view areas. The algorithm designed for overlapped view areas is referred to as geo-location join and the second one is referred to as re-identification (by features). Together with the tracking algorithm such as Deep Sort [12], the bionic brain delivers a reliable and accurate solution to cover different situations in extensive tracking as shown in Fig. 5.

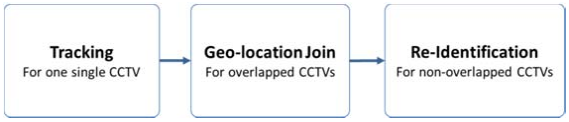


Fig. 5 Tracking and Re-Identification

### *A. Geo-Location Join for Overlapped View Areas*

For the situation of overlapped view areas, geo-location join algorithm associates one object viewed by one CCTV with another according to the exclusion principle, which states that two distinct persons cannot show up at the same location in the same time, and therefore two instances with the same geohashes can be considered as the same person.

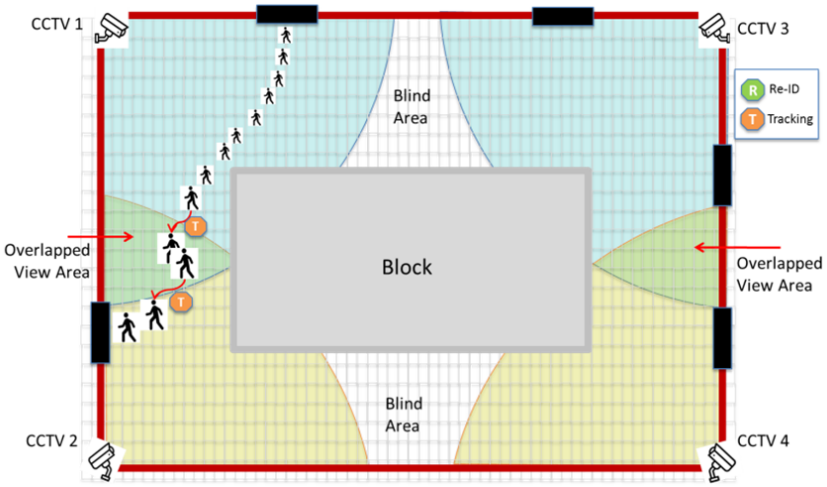


Fig. 6 Geo-Location Join (Overlapped)

Fig. 6 shows one person walking from one view area into the overlapped portion, where the person is captured by two CCTVs, both with identical coordinates (geohashes). The bionic brain treats these two instances as the same person and merges two into one.

### *B. Re-Identification for Non-Overlapped View Areas*

In a situation with a person showing up from a bind area, the algorithm projects the movement of persons coming from the view areas adjacent to the blind area. Once candidate persons are narrowed down, the feature-based model is applied to re-identify the person (i.e., find out from which view area the person comes.)

Fig. 7 shows a public space crowded with pedestrians, where, for the sake of example, two persons are walking through a blind area. Once a person W shows up in the upper-left view

area, the system first looks for potential candidates in a shared memory system, which retains records for every pedestrian seen in the past 30 seconds. The size of the initial set of candidates is potentially large. The system then applies a spatial-temporal awareness model, called *Blind Area Model* (stated in the next section) to substantially narrow down the potential candidates to a very limited set, saying two in this example. By comparing against the two candidates with feature matching, the system learns the one in the upper-right view area showing a higher confidence level than the other one (in the lower-right area) and considers it as the "predecessor" of W. For a few seconds later, a person J shows up in the lower-left view area, the system again screens and narrows down the potential candidates and re-identifies the one who came from the lower-right view area. At this moment, the bionic brain demonstrates a successful tracking task.

### *C. Re-Identification Anomalous Conditions*

For the sake of discussion, while tracing the upper-left appearance, the previous example in Fig. 7 turns out to be the lower-right one with a higher confidence level. It leads to an anomalous situation where both W and J are determined from the lower-right view area. The phenomenon is referred to as a *split* for tracking. When a split condition occurs, the bionic brain treats both (W and J) as the "suspects" for the person to be tracked. The only way to determine the real one is by L3 tracing algorithm, which uses long-term historical data and re-examination with additional information or with the help of human.

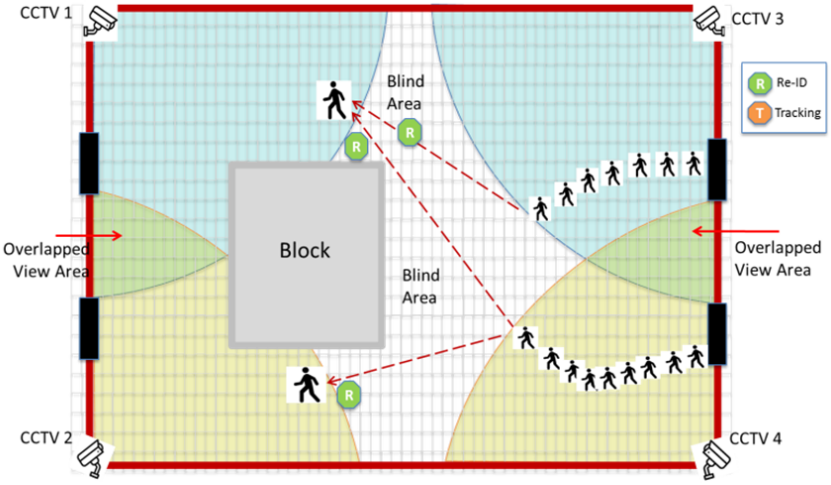


Fig. 7 Re-Identification (Non-Overlapped)

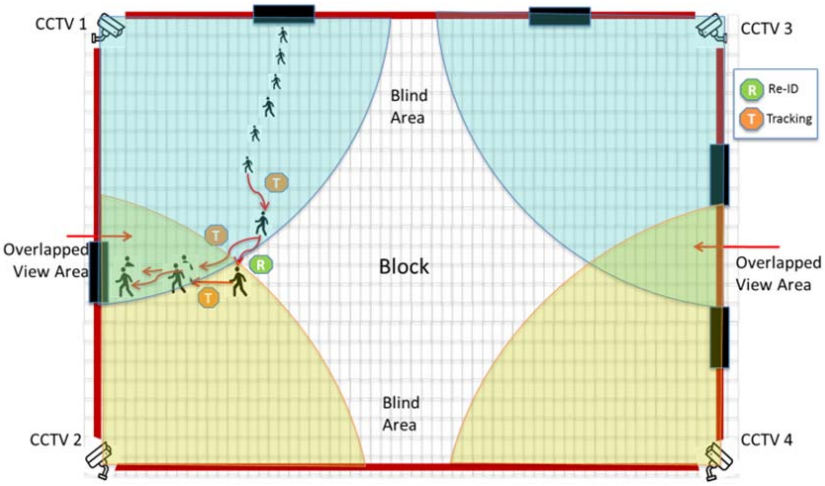


Fig. 8 Anomalous Conditions (Split and Merge)

Another anomalous condition is shown in Fig. 8, where a person W walks from the upper-left view area, footprint by footprint, exits and then enters to the lower-left view area. While the system successfully re-identifies W, he immediately makes a turn and back to the overlapped area as the trajectory shown in Fig. 8. At this moment, unfortunately, both CCTV-1 and CCTV-2 capture W, successfully tracking the person with the L1 tracking algorithm. And then L2 realizes these two

instances with the identical coordinates and merges two into one. This anomaly is referred to as a *merge* condition.

## V. Blind Area Model

In order to precisely project the trajectories of the objects moving throughout the vast space, we propose a model called Blind Area Model (BAM) or sometimes called Alibi Model to effectively manage every blind area. A blind area is essentially a bag, where an object can enter into the area and exit from the area at a later time. Two core questions are answered by BAM as follows:

1. How to know when an object enters into a blind area
2. How to know when the object exits from the blind area.

Remember that all objects are anonymous and can be only seen from the view areas by CCTVs. By the definition of blindness, there is no way to know the real status of objects in a blind area. To answer the first question is to determine whether the appearance of an object is the last one seen in a view area and moving near and toward a blind area next to the view area. It is not easy to determine when an object disappears into a blind area. For an object tracked by the L1 tracking algorithm, losing track does not mean disappearing into a blind area. It may lose track temporarily and re-appear in a few frames. A mathematical model is trained to determine the probability of entering into a blind area (i.e., the probability of whether the appearance is the last one).

The second issue is addressed by detecting the first appearance of an object ***o*** in a view area, which is geo-spatially located very close to a blind area. The detection of the first appearance does not prove ***o*** exiting from a blind area if its lineage cannot be traced, that is, its origin, what happens to it and where it moves over time. This leads to the next question: Where does ***o*** come from? That is, to re-identify ***o***.

In an extensive environment, one of the critical factors for effective re-identification is the number of candidates selected for feature matching. Reducing the potential candidates from dozens to a few will significantly increase the overall accuracy of extensive tracking. With the prior knowledge of the blind space, the goal can be achieved by projecting the probable trails of objects inside the blind area and therefore filtering out those spatially and temporally impossible. Finally, the feature comparison algorithm is applied to re-identify ***o*** out of the potential candidates. That is, linking ***o*** with the most likely one that disappeared into the blind area a few moments earlier. Effective re-identification assures both accuracy and performance.

A hidden Markov Chain model is adopted to deduce the potential candidates out of all possible objects that entered into a blind area. With data collected on the field, the model is trained to determine the probability distribution, P(x, y, t), where t is the duration between the time entering at x (a geohash) and the time exiting at y (a geohash). With the model, when a first-time object that shows up at y from a blind area, all potential candidates that disappeared into the blind area in a reasonable time range from possible x's are selected and matched for re-identification.

Fig. 9 shows a person who first-time shows up in the upper-right view area. With the model, the bionic brain narrows down the suspects into five candidates. Reducing the number of candidates into a controllable size will greatly increase the accuracy of re-identification.

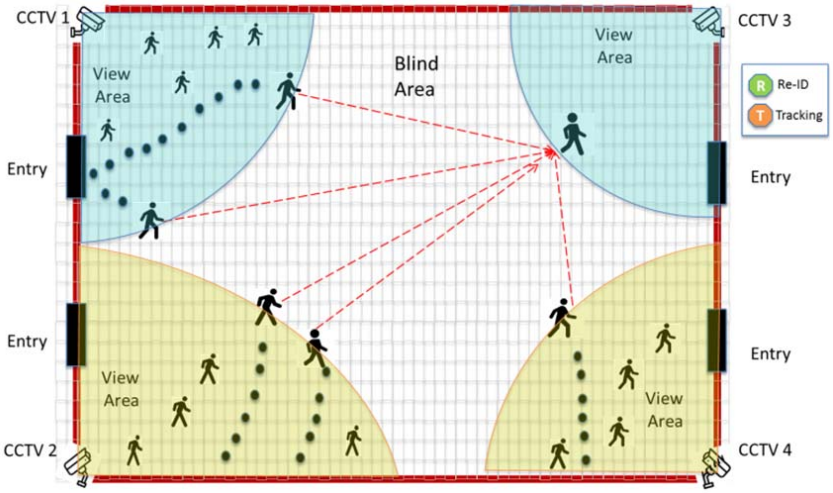


Fig. 9 BAM with Hidden Markov Chain

## VI. Conclusion

The convergence of digital capabilities across the digital and physical environment presents great opportunities for security industry to embrace a new frontier of digital transformation as the society continues to evolve with the rapidly shifting technological landscape. To address the leading physical security operation challenges of reactive threat management and intuition-led decision-making based on subjectivity, the security industry should adopt an operational design first approach, beyond technology & system implementation.

The Spatial-Temporal Awareness approach covered in this

paper challenges the conventional thinking and concept of operations which is labor intensive and time consuming. The ability to perform Extensive Re-Identification through a multi-sensory network provides the next-level insights– creating value beyond traditional risk management. The application extends beyond security operations and can be potentially deployed to transform passengers' experience in aviation hub, commuters' experience in transport hub, consumers' experience in retail malls, patient's experience in hospitals and finally users' experience in precincts.

## REFERENCES

[1] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian, "Scalable Person Re-identification: A Benchmark," in 2015 IEEE International Conference on Computer Vision (ICCV), Dec. 2015, pp. 1116–1124.
[2] E. Fendri, M. Frikha, and M. Hammami, "Multi-level semantic appearance representation for person re-identification system," Pattern Recognit. Lett., vol. 115, pp. 30–38, Nov. 2018.
[3] X. Bai, M. Yang, T. Huang, Z. Dou, R. Yu, and Y. Xu, "Deep-Person: Learning discriminative deep features for person Re-Identification," Pattern Recognit., vol. 98, p. 107036, Feb. 2020.
[4] X. Liu, S. Zhang, Q. Huang, and W. Gao, "RAM: A Region-Aware Deep Model for Vehicle Re-Identification," in 2018 IEEE International Conference on Multimedia and Expo (ICME), Jul. 2018, pp. 1–6.
[5] T.-W. Huang, J. Cai, H. Yang, H.-M. Hsu, and J.-N. Hwang, "Multi-View Vehicle Re-Identification using Temporal Attention Model and Metadata Re-ranking," 2019, pp. 434–442, Accessed: Nov. 02, 2020.
[6] A. Bedagkar-Gala and S. K. Shah, "A survey of approaches and trends in person re-identification," Image Vis. Comput., vol. 32, no. 4, pp. 270–286, Apr. 2014.
[7] A. Li, L. Liu, K. Wang, S. Liu, and S. Yan, "Clothing Attributes Assisted Person Re-identification," IEEE Trans. Circuits Syst. Video Technol., vol. 25, pp. 1–1, Jan. 2014.
[8] M. Munaro, S. Ghidoni, D. T. Dizmen, and E. Menegatti, "A feature-based approach to people re-identification using skeleton keypoints," in 2014 IEEE International Conference on Robotics and Automation (ICRA), May 2014, pp. 5644–5651.
[9] G. Zhang, P. Jiang, K. Matsumoto, M. Yoshida, and K. Kita, "Reidentification of Persons Using Clothing Features in Real-Life Video," Applied Computational Intelligence and Soft Computing, Jan. 11, 2017.
[10] L. Wei, S. Zhang, W. Gao, and Q. Tian, "Person Transfer GAN to Bridge Domain Gap for Person Re-Identification," 2018, pp. 79–88, Accessed: Nov. 02, 2020.
[11] L. Zheng, Y. Huang, H. Lu, and Y. Yang, "Pose-Invariant Embedding for Deep Person Re-Identification," IEEE Trans. Image Process., vol. 28, no. 9, pp. 4500–4509, Sep. 2019.
[12] N. Wojke, A. Bewley, and D. Paulus, "Simple online and realtime tracking with a deep association metric," in 2017 IEEE International Conference on Image Processing (ICIP), Sep. 2017, pp. 3645–3649.
[13] A. Bewley, Z. Ge, L. Ott, F. Ramos, and B. Upcroft, "Simple online and realtime tracking," in 2016 IEEE International Conference on Image Processing (ICIP), Sep. 2016, pp. 3464–3468.
[14] Y.H. Liu"Person Re-Identification Robut to Illumination Change with Clustering-based Loss Fonction", Master Thesis, National Taiwan University.
[15] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection", 2016 IEEE Conference on Computer Vision and Pattern Recognition