

Fusing Complementary Trackers for Long-Term Visual Tracking

Matteo Dunnhofer



Kristian Simonato



Christian Micheloni



Motivation



— STARK

— SuperDiMP+MU

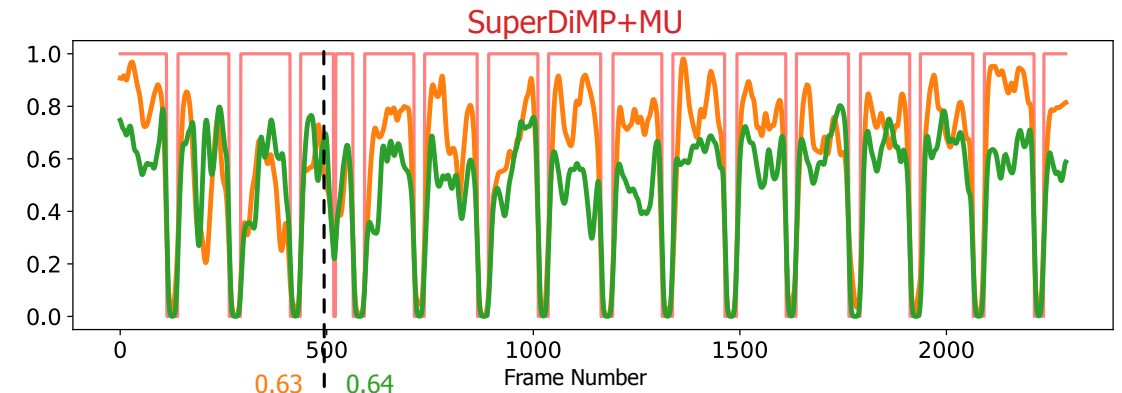
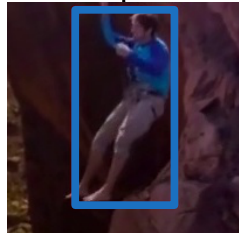
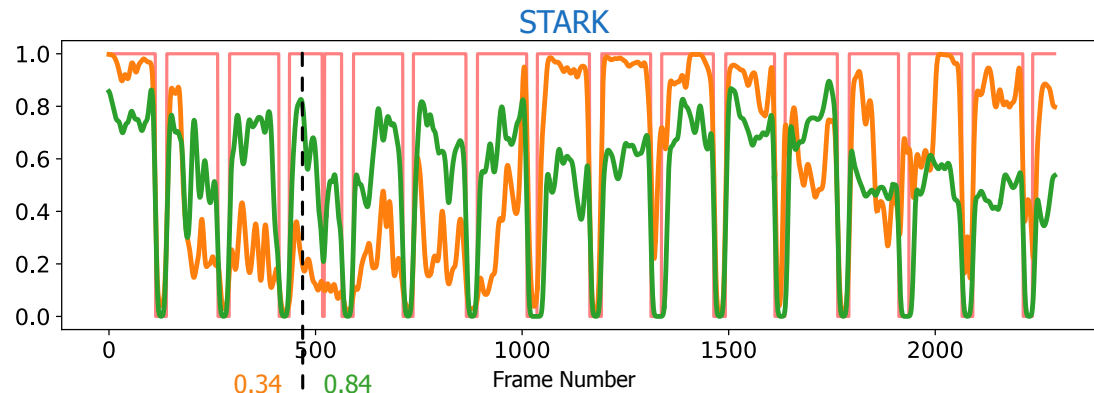
[1] 'Learning Spatio-Temporal Transformer for Visual Tracking', ICCV 2021

[2] 'The Eight Visual Object Tracking Challenge Results', VOT2020



Motivation

Ground-truth Confidence IoU



Fusion of Trackers

ECCV 2012

ECCV 2012

ICML 2014

CVIU 2016

ECCV 2014

NeurIPS 2020

ACCV 2020

MEEM: Robust Tracking via Multiple Entropy Minimization

Jianming Zhang, Shuang

Department of Computer Science
{jzhang, shuang}@hkust.hk

Abstract. We propose a multi-expert model to solve the drift problem in online tracking. In the practical snapshots constitute an expert ensemble to restore the current tracker when needed. In order to correct undesirable model updates, we exploit an online SVM on a budget algorithm for efficient model update and a method achieves substantially better performance on a benchmark dataset of 50 video sequences. In addition, in experiments with a newly collected dataset, we show that the proposed multi-expert model outperforms the baseline tracker, especially in the case of appearance variations.

1 Introduction

In this paper, we focus on the problem of given only the object's initial position and the detection framework. In many online tracking scenarios, which are intended to account for the process of updating the model also being challenge in online visual tracking.

Model drift occurs because factors like training samples can lead to bad model updates, relying on a fixed model prior tends to object appearance changes. Other trackers where an update is prevented when certain conditions of good or bad updates usually rely upon appearance changes, which are often violated when the object's appearance changes. These trackers are trapped in a background region, due to the fact that without a mechanism to correct for past mistakes, instead of trying to prevent bad updates, they can correct the effects of bad updates and build a multi-expert tracking framework, which

D. Fleet et al. (Eds.): ECCV 2014, Part VI, LNCS 8695, pp. 1–12, 2014.
© Springer International Publishing Switzerland 2014.

Visual Tracking via AdaBoost

Ju Hong Yoon, Du Ye

¹ Computer Vision Lab, Owingin
² Applied Computing Lab, Owingin
{jhyoon, kyyoon}@ajou.ac.kr

Abstract. In this paper, a robust visual tracking algorithm is proposed to deal with pose variations, and occlusions. To deal with different feature descriptors are level of robustness to certain changes in independent trackers, we propose two action. The tracker interaction is achieved (TPM) in a probabilistic manner. The from among multiple tracker outputs, the tracker probability. According to variational Bayesian method, the TPM and tracker probability are updated using each tracker's reliability, which function (TLF). When the tracking is performed, the state is obtained and fed into the strategy, which retains the robustness of the trackers. The experimental results demonstrate that the proposed method outperforms the baseline trackers in various benchmark scenarios.

Keywords: Visual tracking, multiple hypothesis tracking, tracker interaction.

1 Introduction

Visual tracking is an important research topic with wide application in surveillance, robotics, medical imaging, and so on. Due to the complexity of visual tracking in dynamic environments, visual tracking is required to deal with dynamic circumstances such as object pose variation, occlusions, and motion blur. In many cases, researchers have discussed how to improve tracking performance in an efficient manner by multiple features to achieve a robust visual tracking. In this paper, we propose a new visual tracking algorithm that can correct the effects of bad updates and build a multi-expert tracking framework, which

D. Fleet et al. (Eds.): ECCV 2014, Part VI, LNCS 8695, pp. 1–12, 2014.
© Springer International Publishing Switzerland 2014.

Ensemble-Based Tracking via Aggregating Crowdsourced Str

Naiyan Wang
Dit-Yan Yeung

Department of Computer Science and Engineering, Hong Kong Baptist University
Clear Water Bay, Hong Kong

Abstract

We study the problem of aggregating the contributions of multiple contributors in a crowdsourcing setting. The data involved is in a form not typically considered in most crowdsourcing tasks, in that the data is structured and has a temporal dimension. In particular, we study the visual tracking problem in which the unknown data to be estimated is in the form of a sequence of bounding boxes representing the trajectory of the target object being tracked. We propose a factorial hidden Markov model (FHMM) for ensemble-based tracking by learning jointly the unknown trajectory of the target and the reliability of each tracker in the ensemble. For efficient online inference of the FHMM, we devise a conditional particle filter algorithm by exploiting the structure of the joint posterior distribution of the hidden variables. Using the largest open benchmark for visual tracking, we empirically compare two ensemble methods constructed from five state-of-the-art trackers with the individual trackers. The promising experimental results provide empirical evidence for our ensemble approach to "get the best of all worlds".

1. Introduction

Visual tracking is a fundamental problem in video semantic analysis. Although it is not a new research problem in computer vision, the challenging requirements of many new applications such as terrorist detection, self-driving cars and wearable computers require that some objects of interest possibly with fast and abrupt motion in uncontrolled environments be tracked as they move around in a video. This has led to a resurgence of interest in visual tracking within the machine learning and computer vision communities. In

Proceedings of the 31st International Conference on Machine Learning, Beijing, China, 2014. JMLR: W&CP volume 32. Copyright 2014 by the author(s).

Online Adaptive Hidden Markov Model for Multi-Target Tracking

Tomas Vojnir*, Jiri Matas

^{*}The Center for Machine Perception
^{*}Karlovna naměstí 13, 121
^{*}Faculty of Civil Engineering
Thakurova 7/2077, 16621

Abstract

In this paper, we propose a novel HMMTxD. The method fuses observations from multiple trackers and a detector by utilizing a binary vector expression to correspond to a binary vector expression. The Markov model is trained in an unsupervised manner to provide a source of tracking. Baum-Welch algorithm that updates the parameters of the model. We show the effectiveness of the proposed method on two standard benchmarks (CVPR2013 publicly available sequences, The KITTI dataset). The HMMTxD often significantly, on all datasets in all

Keywords: visual tracking, on-line learning, hidden Markov model

*corresponding author
Email address: vojnir@cmp.felk.cvut.cz (Jiri Matas),
matas@cmp.felk.cvut.cz (Jiri Matas),
URL: http://cmp.felk.cvut.cz/~matas/
http://cmp.felk.cvut.cz/~matas/

Preprint submitted to Elsevier

A Superior Tracking Algorithm for Building a Strong Tracker

Christian Bailer¹, Alain Pagani¹, and

¹ German Research Center for Artificial Intelligence
(Christian.Bailer,Alain.Pagani,Didier.Dachwald)
² University of Kaiserslautern, Germany

Abstract. General object tracking is a challenging task. Tracking algorithms perform well on different sequences, but each of them has different strengths and weaknesses. The fact that can be utilized to create a fusion approach, the best tracking algorithms in tracking performance programming based trajectory optimization. The proposed algorithm performs well on different sequences, but each of them has different strengths and weaknesses. The fact that can be utilized to create a fusion approach, the best tracking algorithms in tracking performance programming based trajectory optimization.

Keywords: Object Tracking, Data Fusion.

1 Introduction

Visual object tracking is an important problem in a wide range of applications such as surveillance, robotics, and interactive video production. Nowadays, the performance of many specific scenarios like car tracking [18] or pedestrian tracking [19] has improved. However, object tracking in the general case i.e. when the target is not known in advance, remains a challenging task. In many scenarios, the target can still be considered as a point in the scene. The diverse appearances of the target in a frame which represents a challenge for the tracker. The template tracker utilizes the initial template matching and addresses the challenge that can occur in an unknown scene. The detection tracker continuously updates the target by itself. The diverse appearances of the target in a frame which represents a challenge for the tracker. The template tracker utilizes the initial template matching and addresses the challenge that can occur in an unknown scene.

As the evaluation in [29] and our comparison in Table 1, the proposed algorithm performs well on different sequences. A template tracker utilizes the initial template matching and addresses the challenge that can occur in an unknown scene. The detection tracker continuously updates the target by itself. The diverse appearances of the target in a frame which represents a challenge for the tracker. The template tracker utilizes the initial template matching and addresses the challenge that can occur in an unknown scene.

D. Fleet et al. (Eds.): ECCV 2014, Part VII, LNCS 8695, pp. 1–12, 2014.
© Springer International Publishing Switzerland 2014.

Online Decision-Making for Reinforcement Learning

Ke Song, Wei

School of Control Science and Engineering, Tsinghua University
Beijing 100084, China
{songke, wsl}@sem.tsinghua.edu.cn

A deep visual tracker is typically trained on a specific dataset. It is straightforward to consider tracking. However, this is not the case for online tracking. Unlike previous fusion-based methods, we propose a novel hierarchical reinforcement learning framework for online tracking. The framework consists of a policy network and a value network. The policy network is trained to output the tracking parameters. The value network is trained to output the tracking performance. The framework is trained on a large dataset of online tracking sequences. The framework is evaluated on a benchmark dataset of online tracking sequences. The framework achieves state-of-the-art performance on the benchmark dataset.

1 Introduction

As a fundamental task in computer vision, object tracking has attracted much attention in recent years. Inspired by the success of deep learning in image classification, many deep learning-based trackers have been proposed. These trackers achieve state-of-the-art performance on many benchmarks. However, these trackers are trained on a specific dataset and are not suitable for online tracking.

In the past, the community has proposed solutions emphasizing different aspects of the problem. Processing speed was pursued by algorithms like correlation filters [12, 13] or offline methods such as siamese convolutional neural networks (CNNs) [4, 5, 6, 7, 8, 9, 10, 11]. Improved performance was attained by online target adaptation methods [3, 14, 15, 16, 17]. Higher tracking accuracy and robustness were achieved by methods built on top of other trackers [18, 19, 20, 21, 22]. All these characteristics belong to an optimal tracker but they were studied one independently from the other. The community currently lacks a general framework to tackle them jointly. In this view, a single model should be able to (i) track an object in a fast way, (ii) implement simple and effective online adaptation mechanisms, (iii) apply decision-making strategies to select tracker outputs. It is a matter of fact that a large number of tracking algorithms have been produced so far, with different principles exploited. Preliminary solutions were

*Corresponding author
34th Conference on Neural Information Processing Systems



This ACCV 2020 paper, provided here by the Computer Vision Foundation, is the author-created version. The content of this paper is identical to the content of the officially published ACCV 2020 LNCS version of the paper as available on SpringerLink: <https://link.springer.com/conference/accv>

Tracking-by-Trackers with a Distilled and Reinforced Model

Matteo Dunnhofer^{0000-0002-1072-667X}, Niki Martini⁰⁰⁰⁰⁻⁰⁰⁰²⁻⁶⁰⁶²⁻⁸⁶⁴³,
Christian Micheloni⁰⁰⁰⁰⁻⁰⁰⁰³⁻⁴⁵⁰³⁻⁷⁴⁸⁸

{matteo.dunnhofer, niki.martini, christian.micheloni}@uniud.it

Machine Learning and Perception Lab, University of Udine, Italy

Abstract. Visual object tracking was generally tackled by reasoning independently on fast processing algorithms, accurate online adaptation methods, and fusion of trackers. In this paper, we unify such goals by proposing a novel tracking methodology that takes advantage of other visual trackers, offline and online. A compact student model is trained via the marriage of knowledge distillation and reinforcement learning. The first allows to transfer and compress tracking knowledge of other trackers. The second enables the learning of evaluation measures which are then exploited online. After learning, the student can be ultimately used to build (i) a very fast single-shot tracker, (ii) a tracker with a simple and effective online adaptation mechanism, (iii) a tracker that performs fusion of other trackers. Extensive validation shows that the proposed algorithms compete with real-time state-of-the-art trackers.

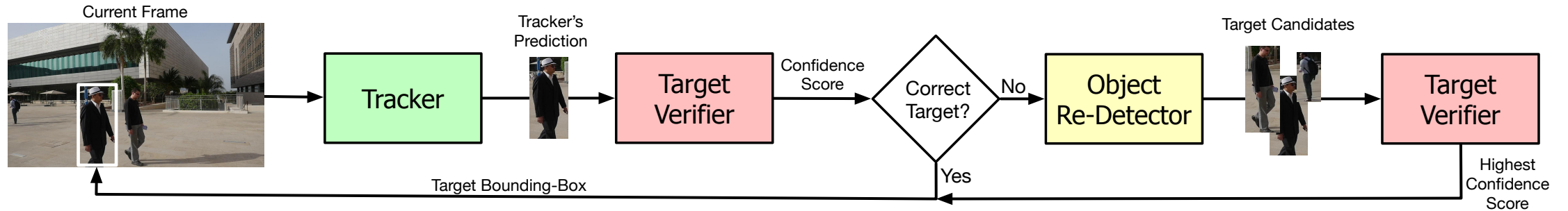
1 Introduction

Visual object tracking corresponds to the persistent recognition and localization by means of bounding boxes of a target object in consecutive video frames. This problem comes with several different challenges including object occlusion and fast motion, light changes, and motion blur. Additionally, real-time constraints are often posed by the many practical applications, such as video surveillance, behavior understanding, autonomous driving, and robotics.

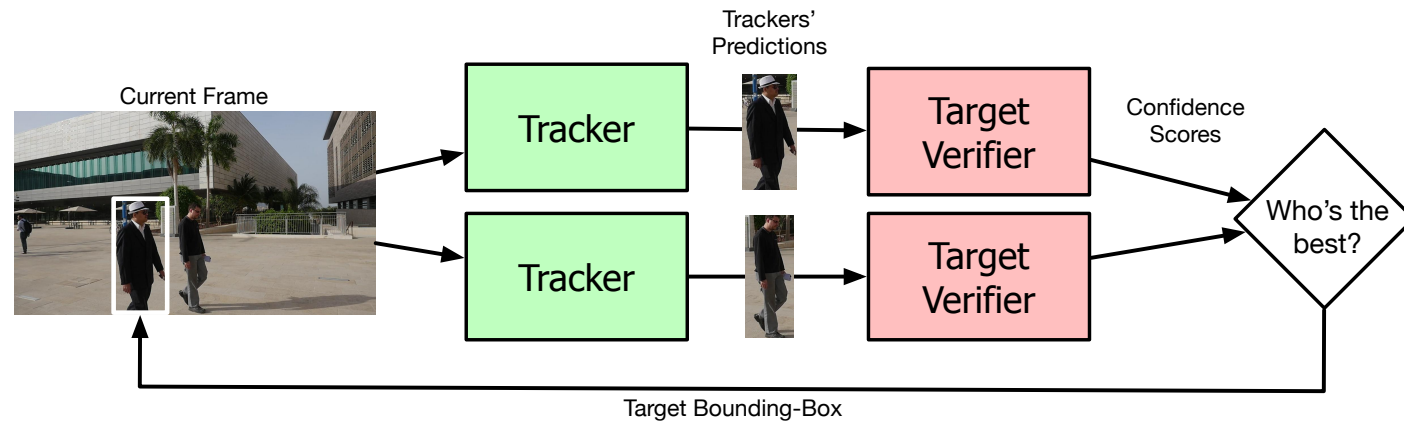
In the past, the community has proposed solutions emphasizing different aspects of the problem. Processing speed was pursued by algorithms like correlation filters [12, 13] or offline methods such as siamese convolutional neural networks (CNNs) [4, 5, 6, 7, 8, 9, 10, 11]. Improved performance was attained by online target adaptation methods [3, 14, 15, 16, 17]. Higher tracking accuracy and robustness were achieved by methods built on top of other trackers [18, 19, 20, 21, 22]. All these characteristics belong to an optimal tracker but they were studied one independently from the other. The community currently lacks a general framework to tackle them jointly. In this view, a single model should be able to (i) track an object in a fast way, (ii) implement simple and effective online adaptation mechanisms, (iii) apply decision-making strategies to select tracker outputs. It is a matter of fact that a large number of tracking algorithms have been produced so far, with different principles exploited. Preliminary solutions were



A Different Long-Term Pipeline

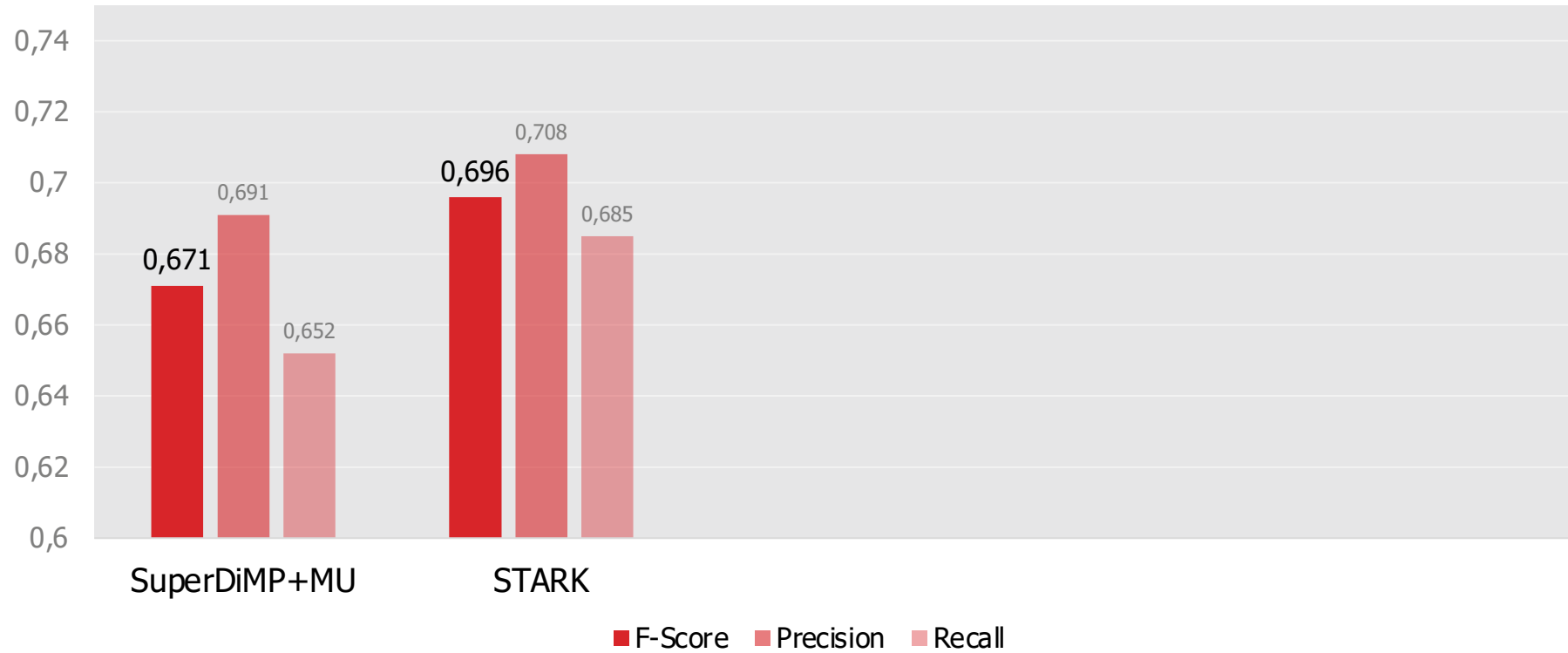


Scheme of the most successful standard long-term solutions

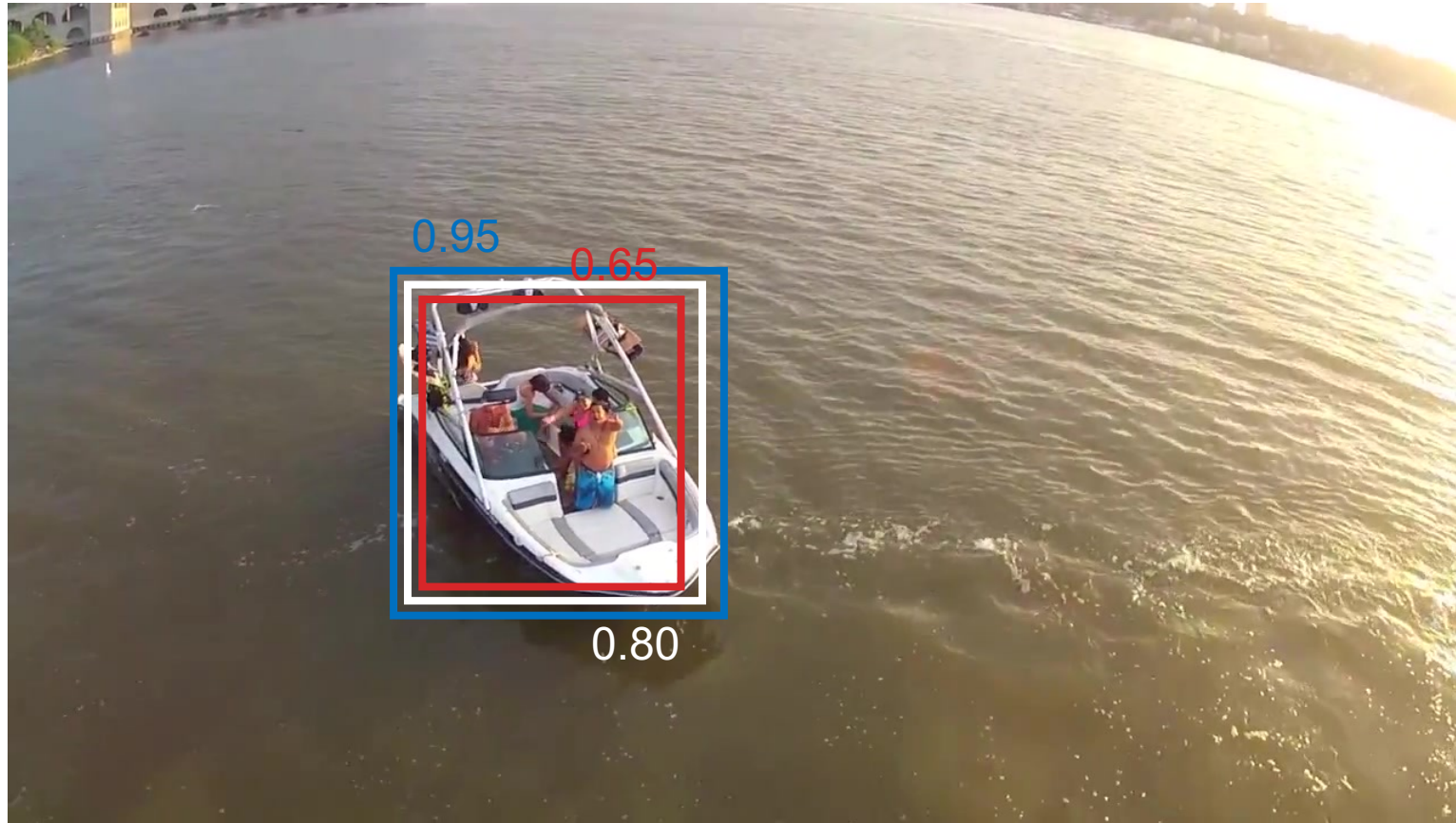


Scheme of our approach

Fusion Baselines



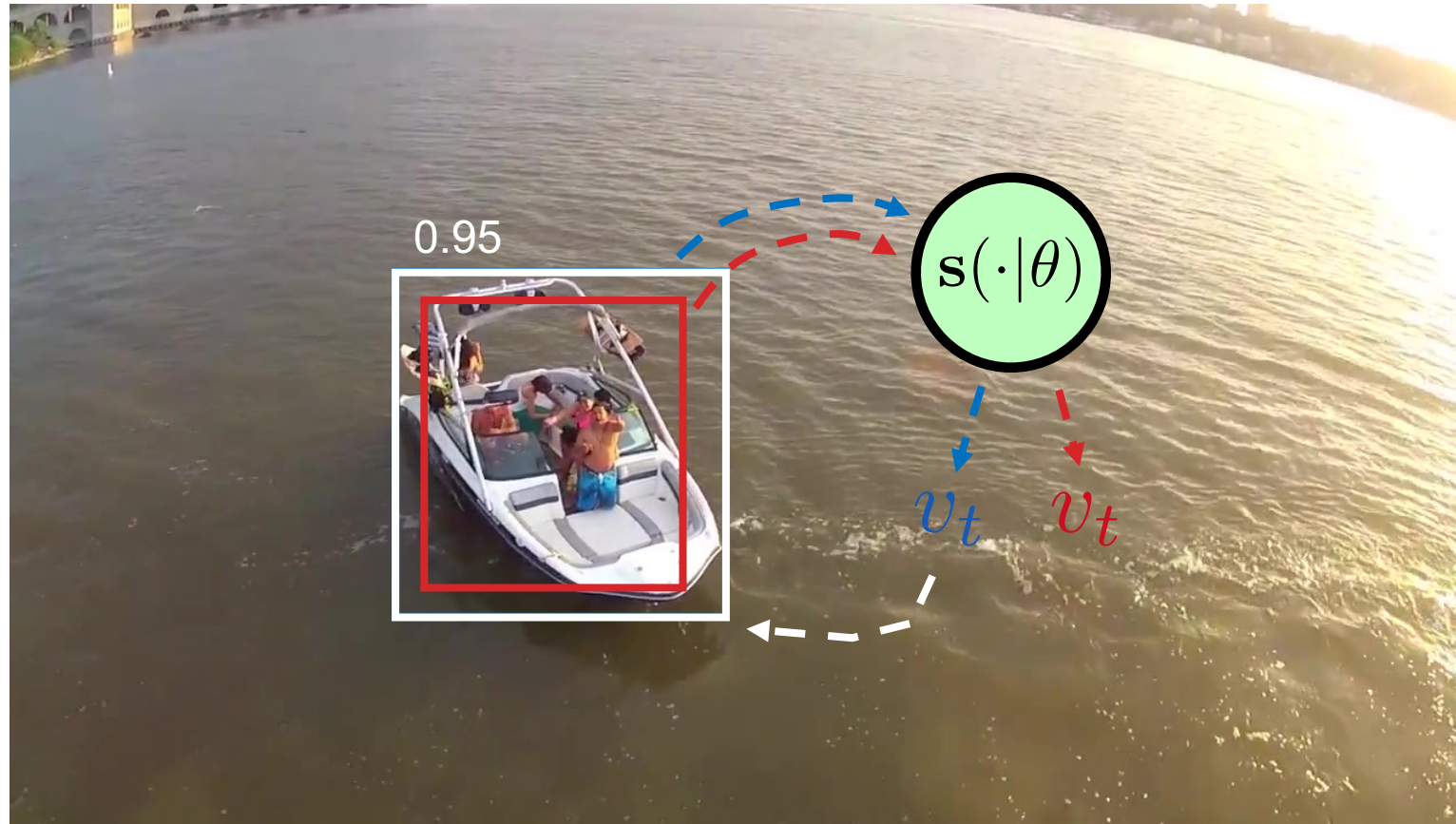
Fusion Baselines - Avg



Fusion Baselines - Max

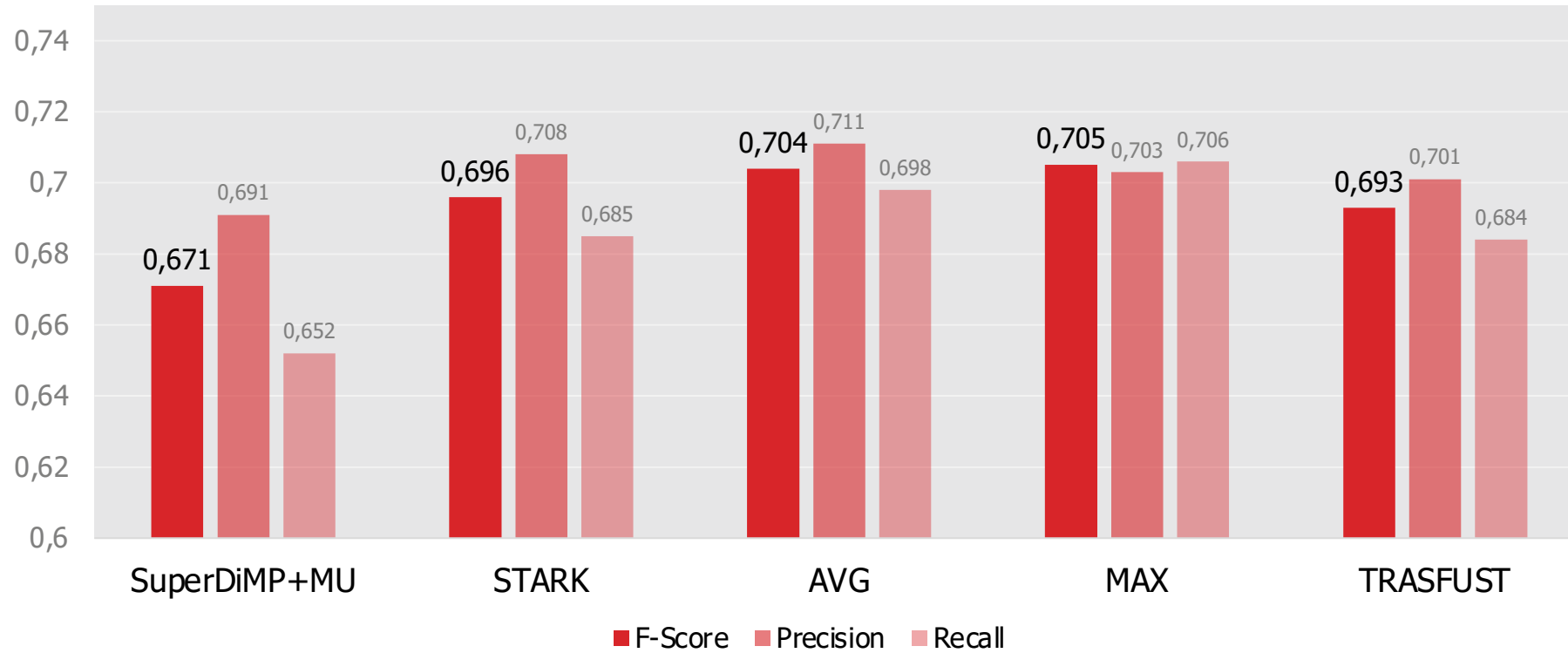


Fusion Baselines - TRASFUST

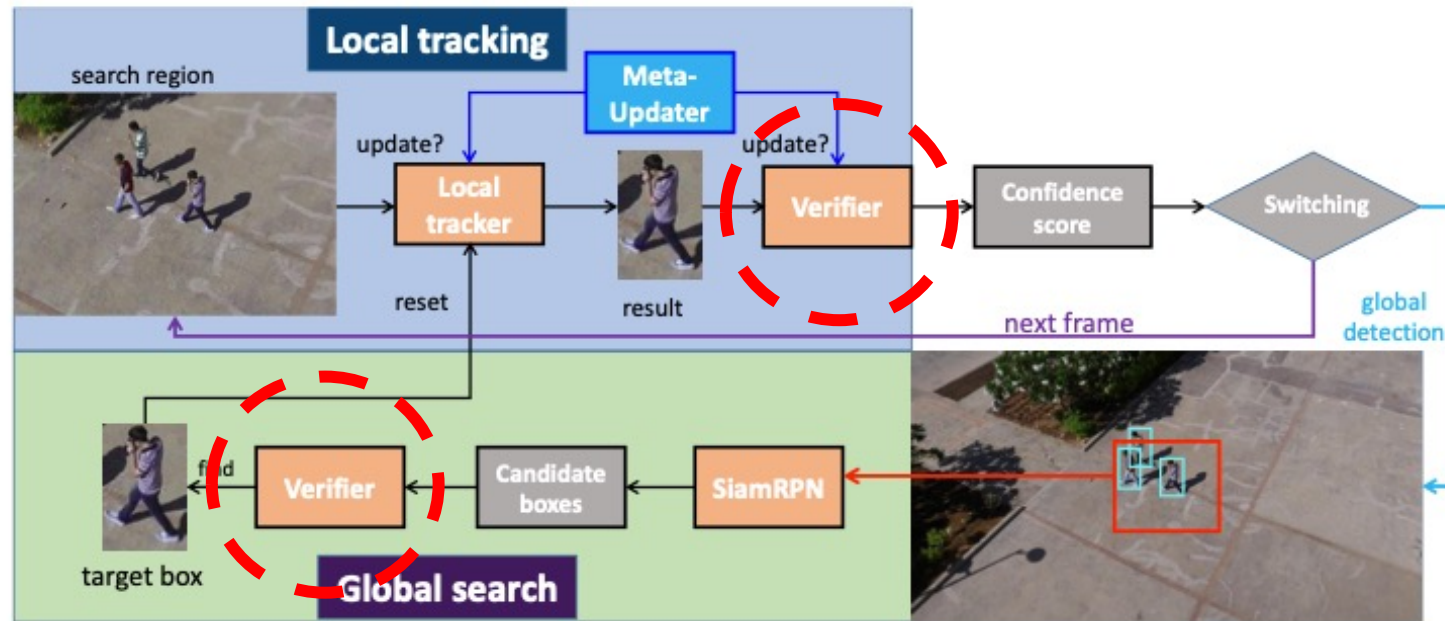


[3] 'Tracking-by-Trackers with a Distilled and Reinforced Model', ACCV 2020

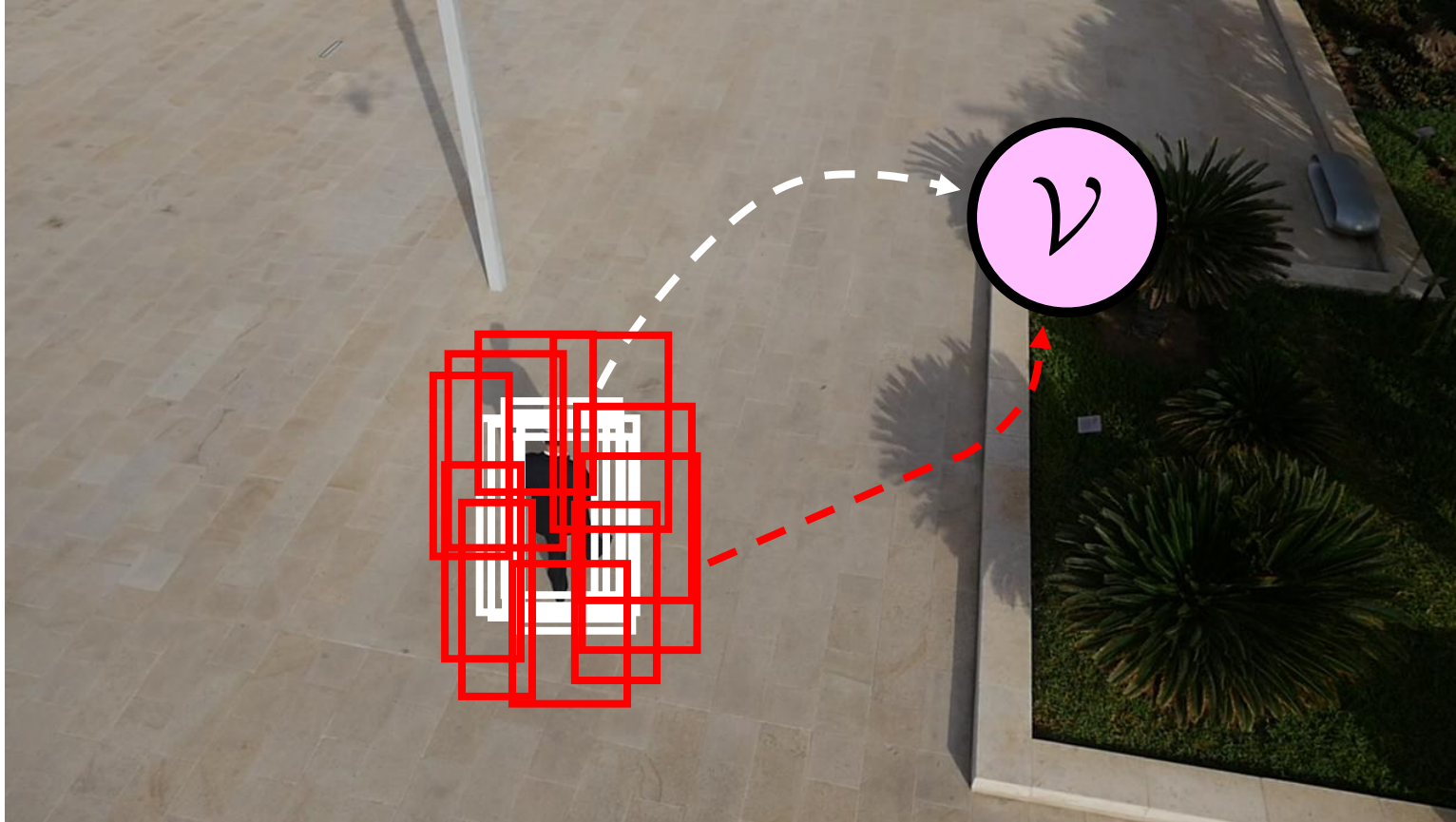
Fusion Baselines



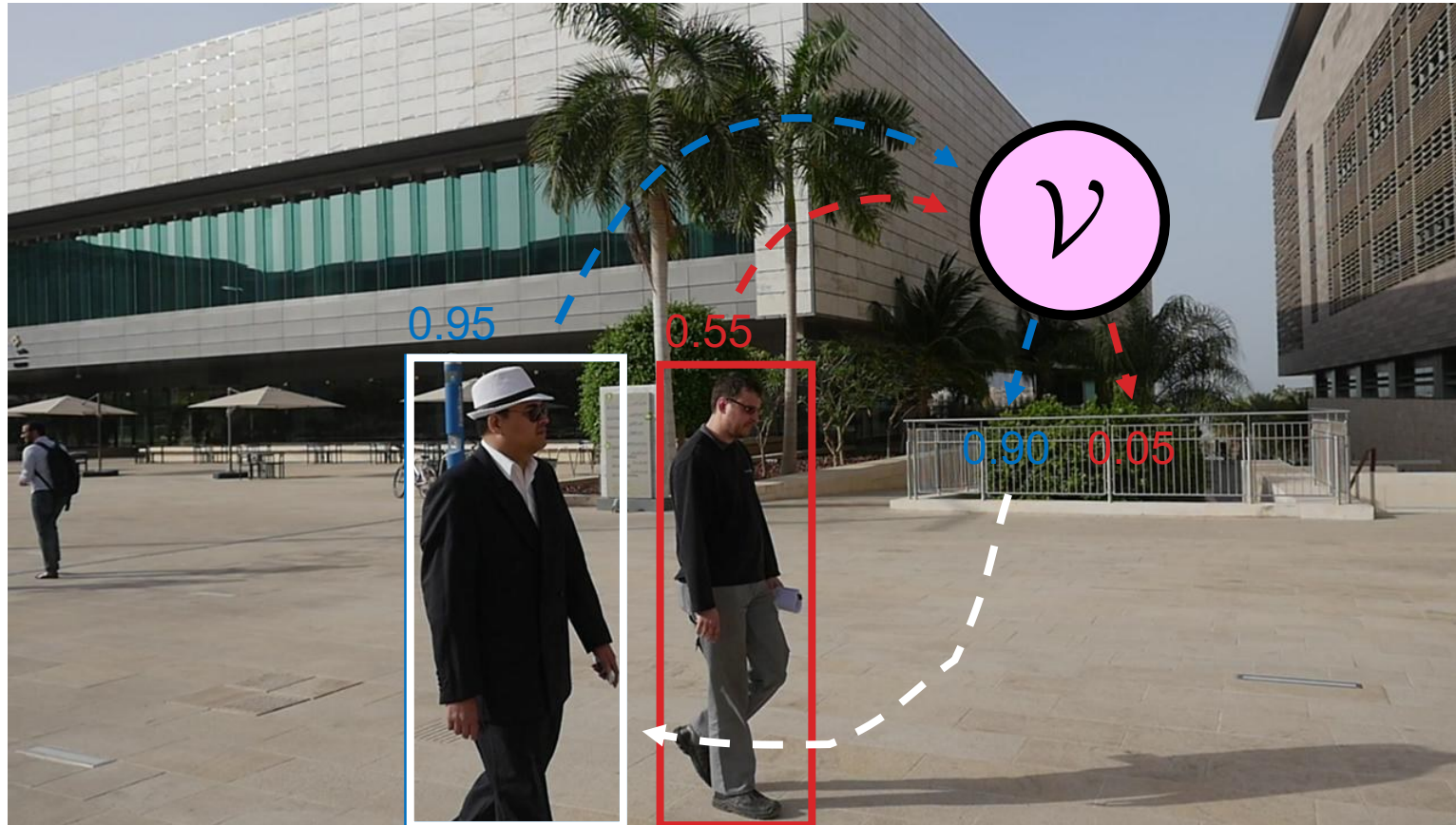
mipLT - Verifier



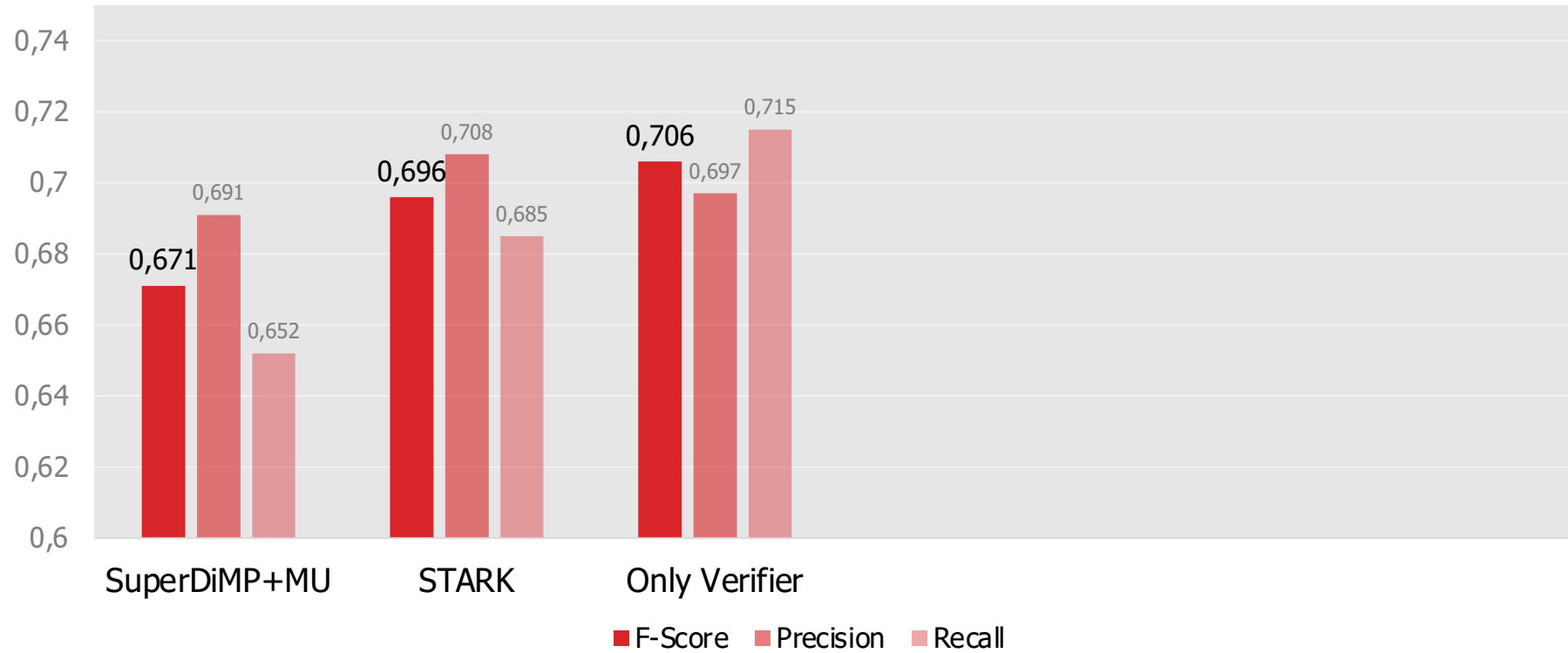
mipLT – Verifier initialization



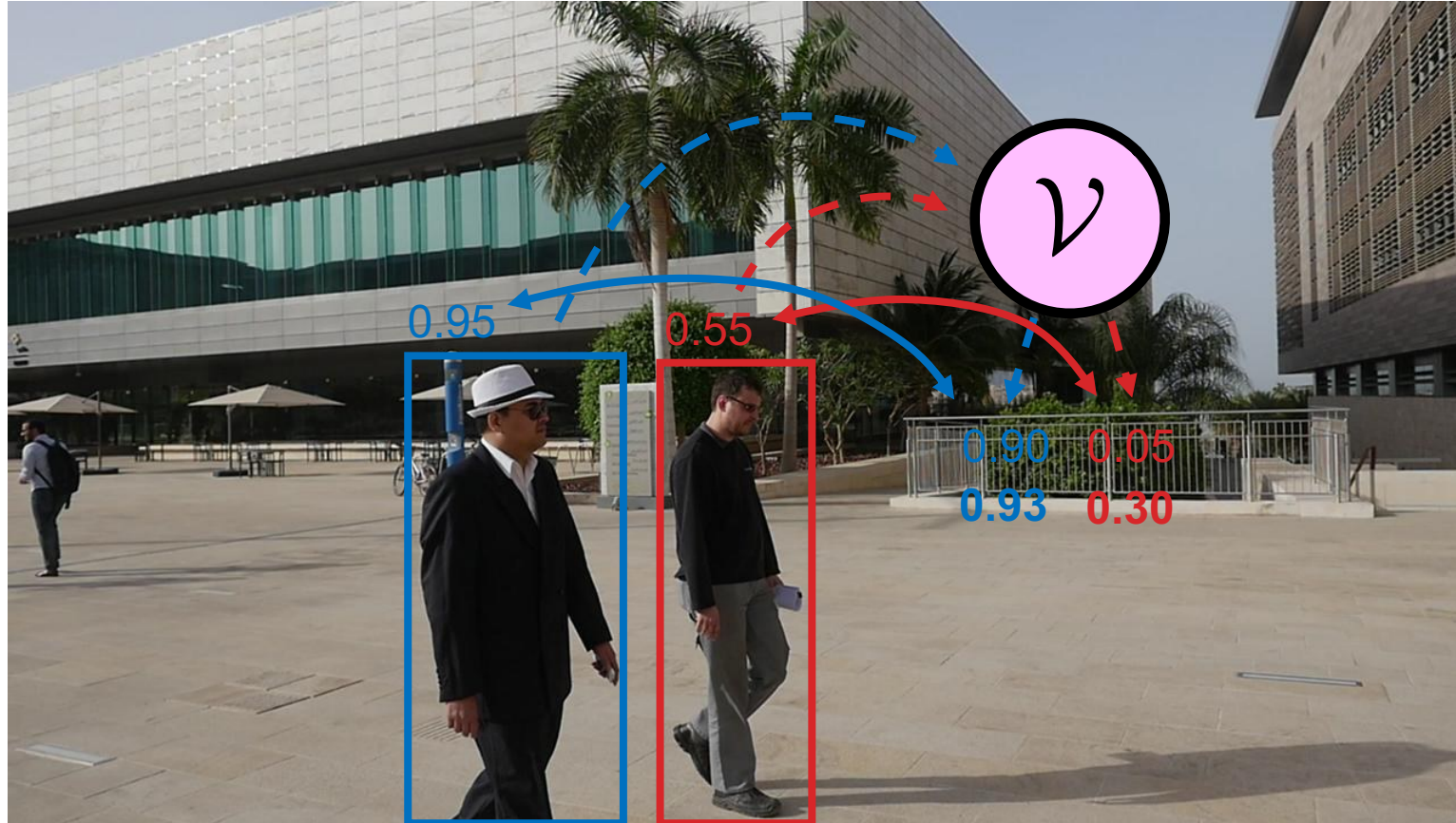
mipLT – Target Verification



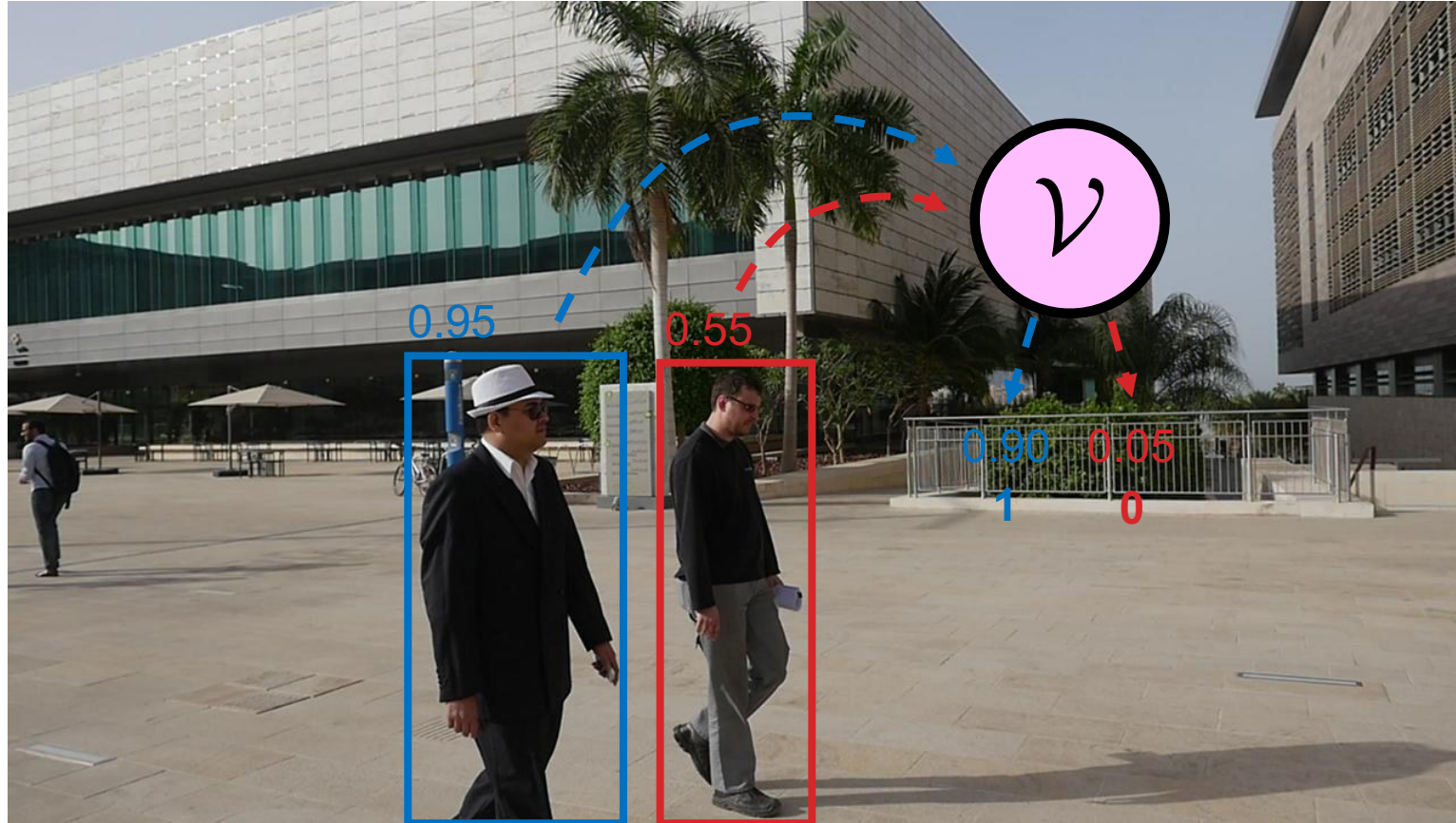
Results



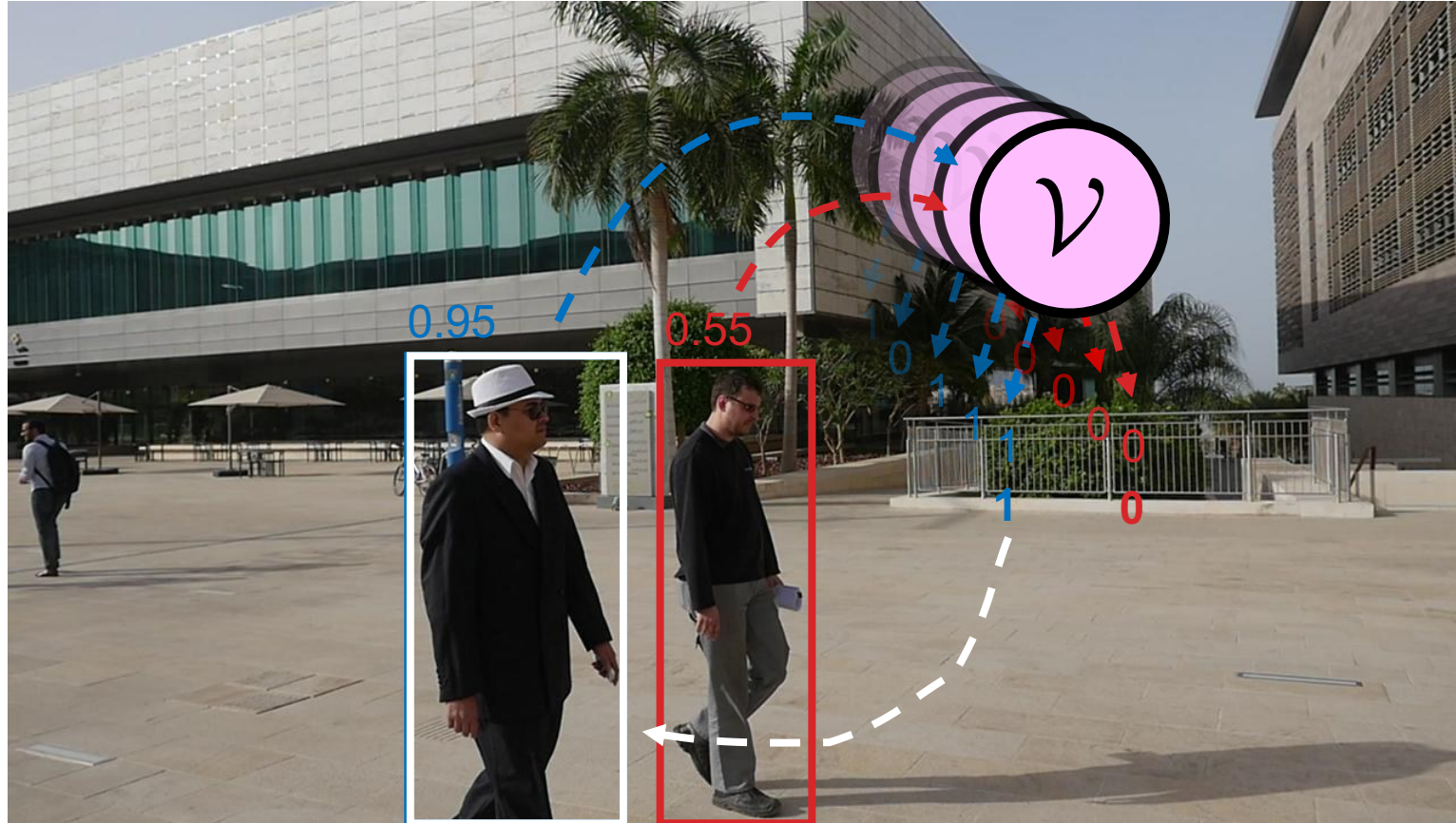
mipLT – Confidence Combination



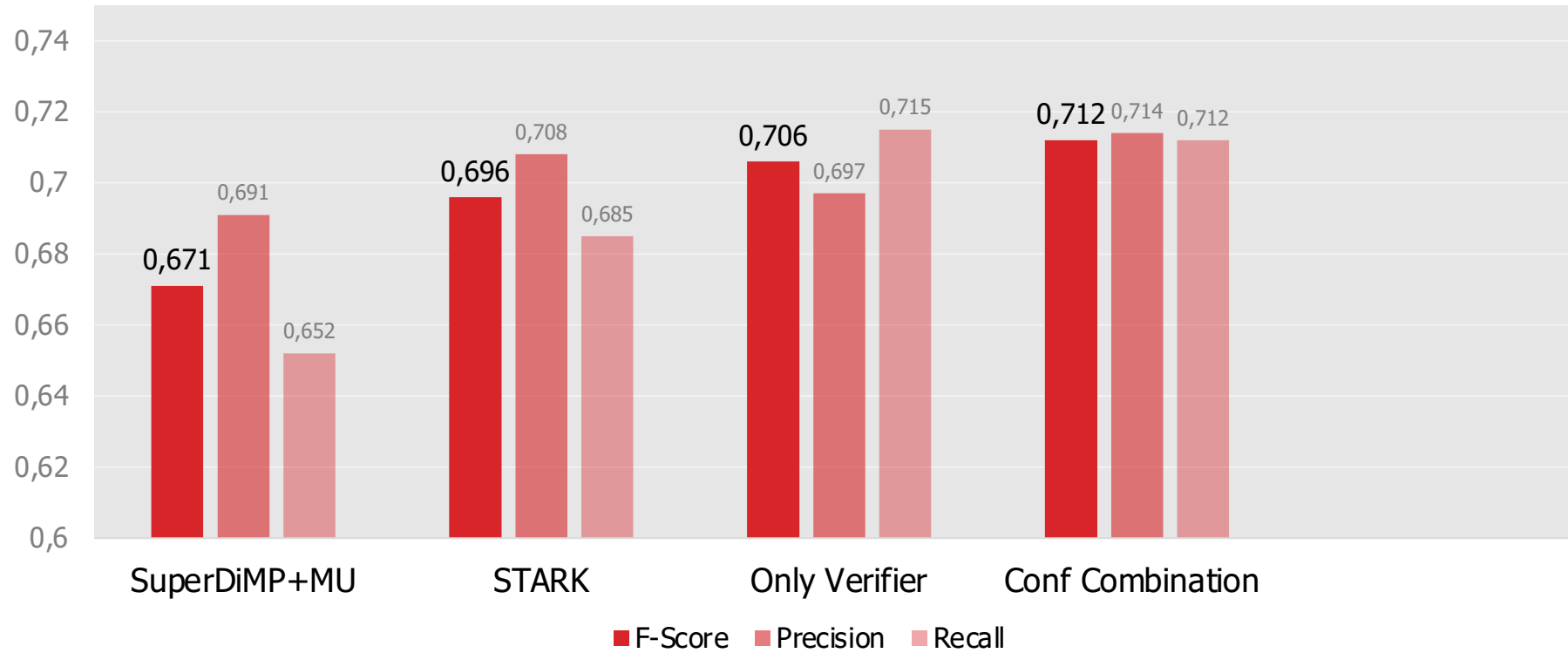
mipLT - Confidence Thresholding



mipLT - Confidence Over Frames



Results



mipLT – Tracker Correction



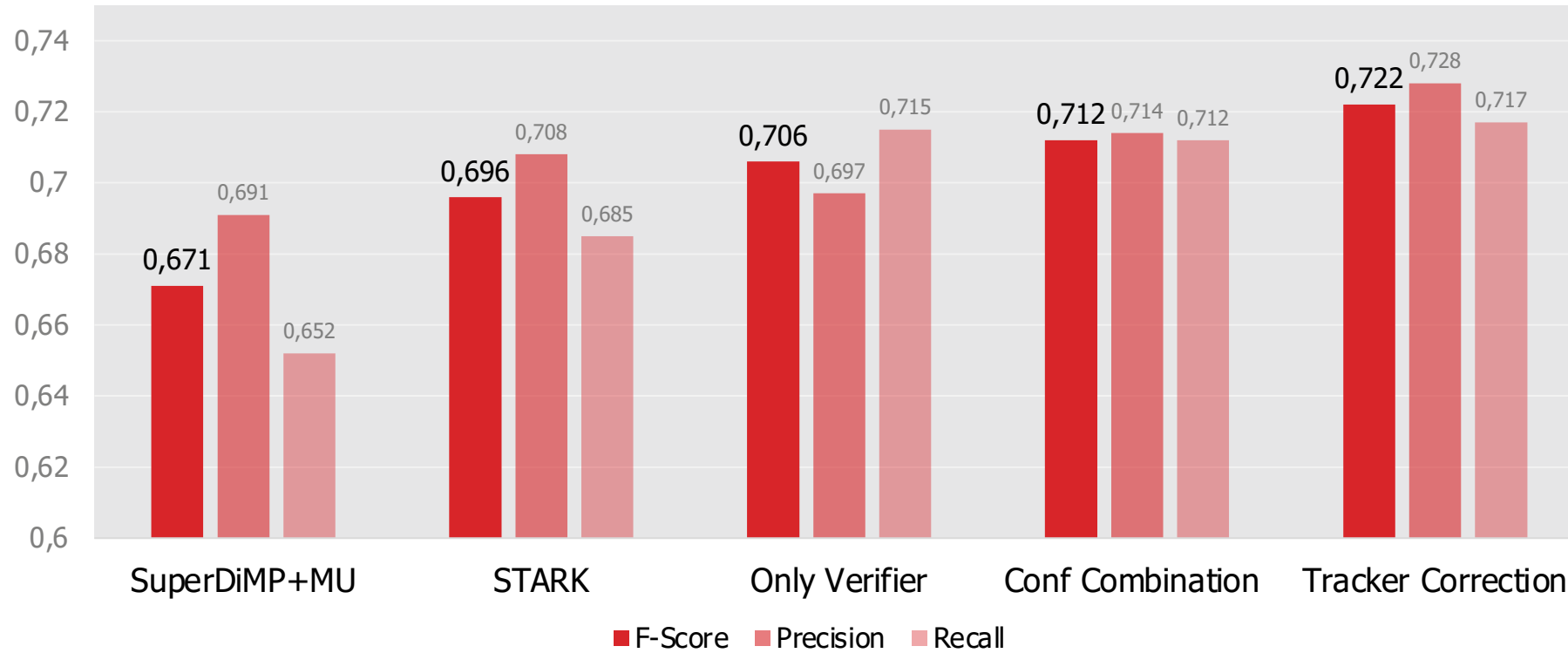
mipLT – Tracker Correction



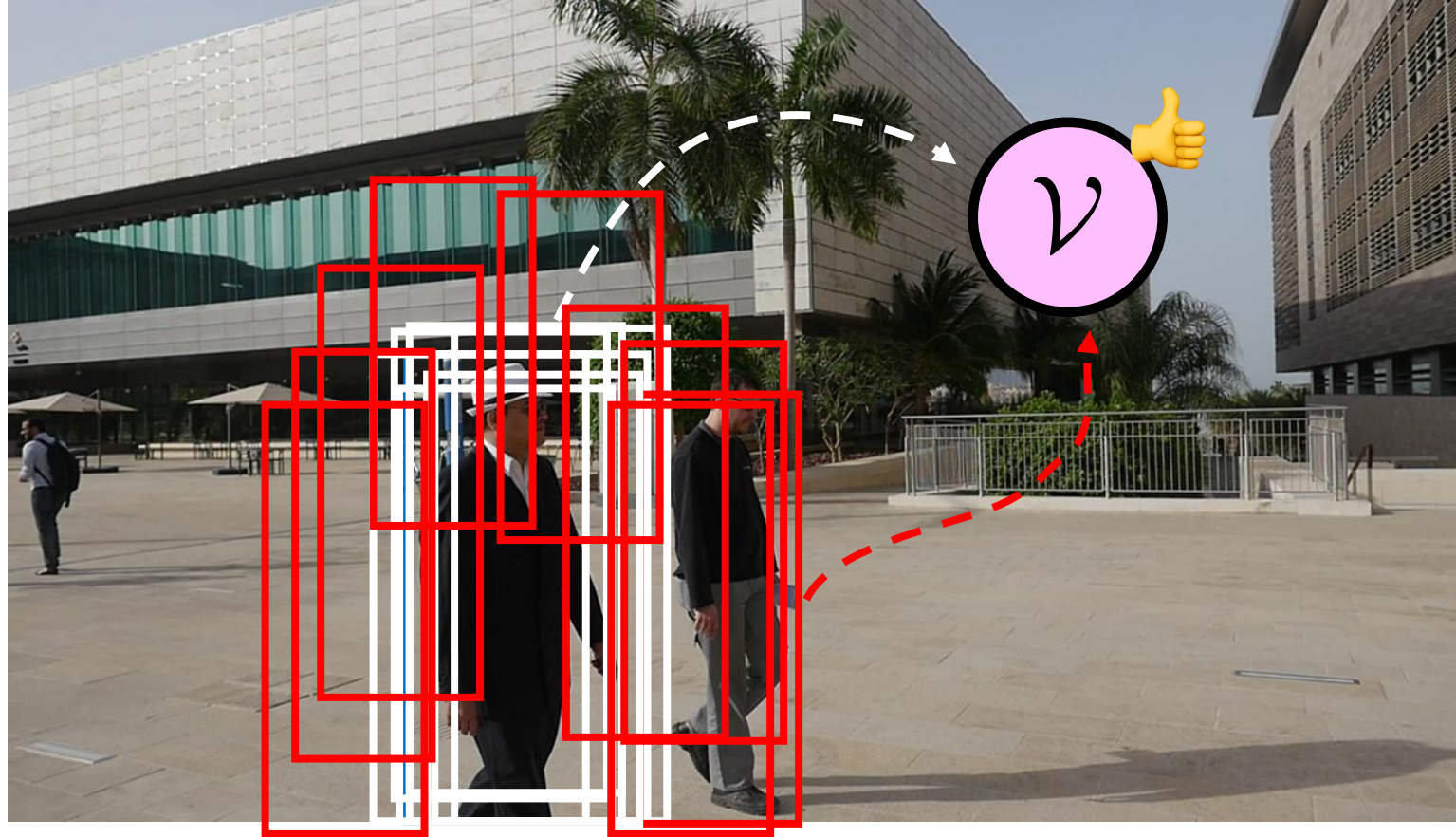
mipLT – Tracker Correction



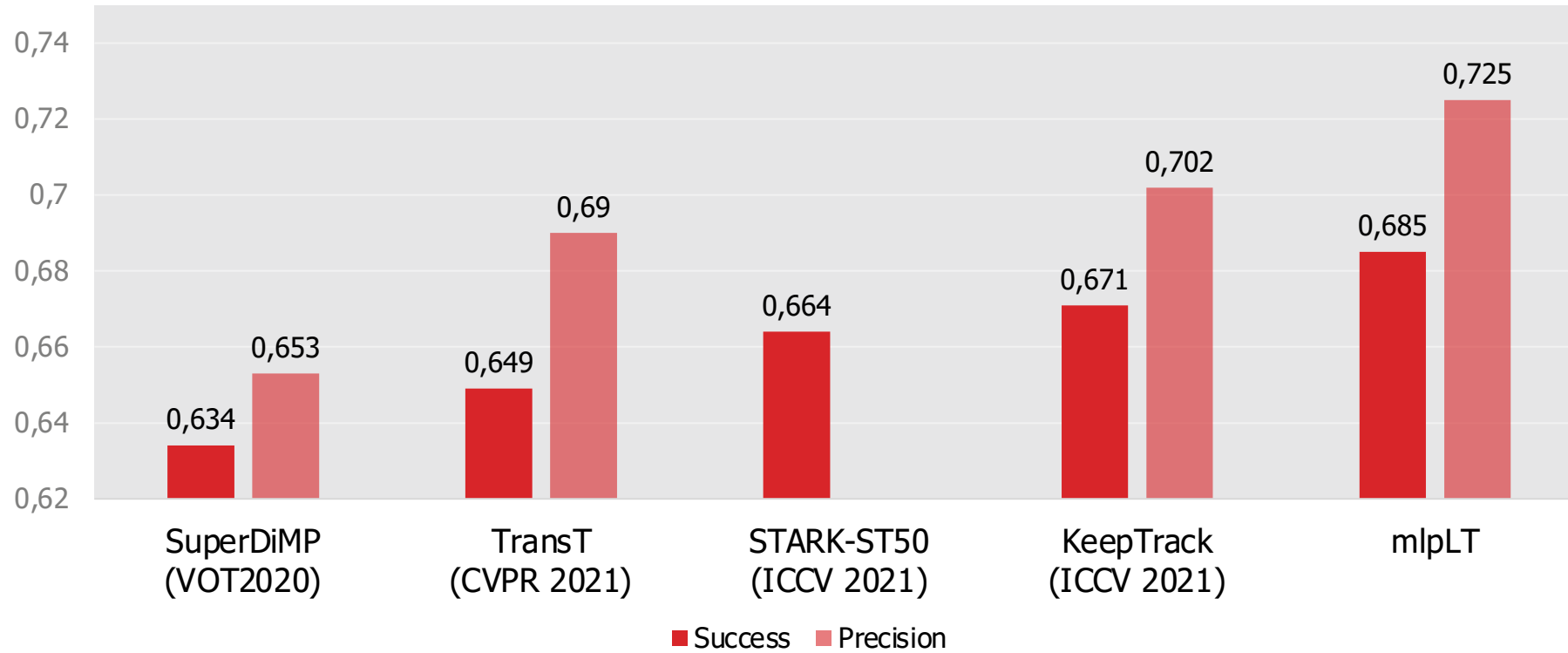
Results



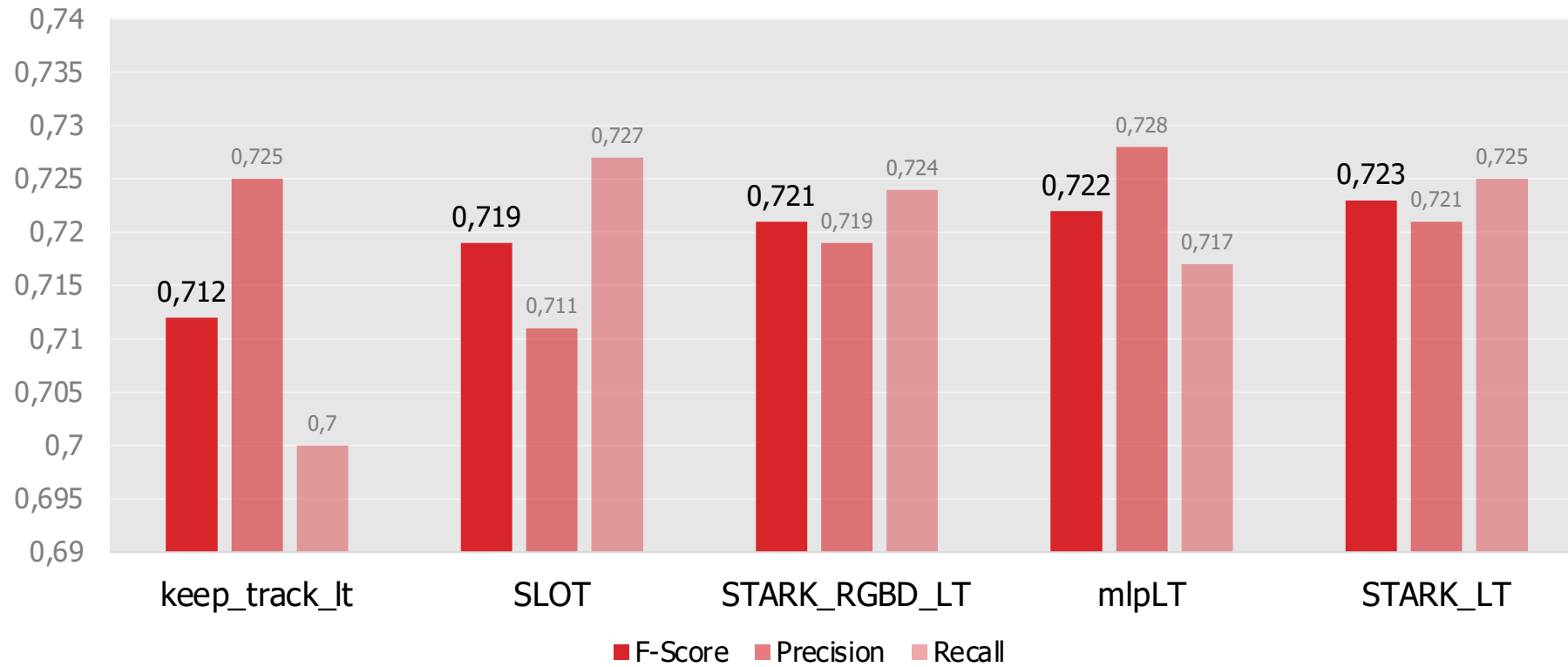
mipLT – Online Update



Results - LaSOT



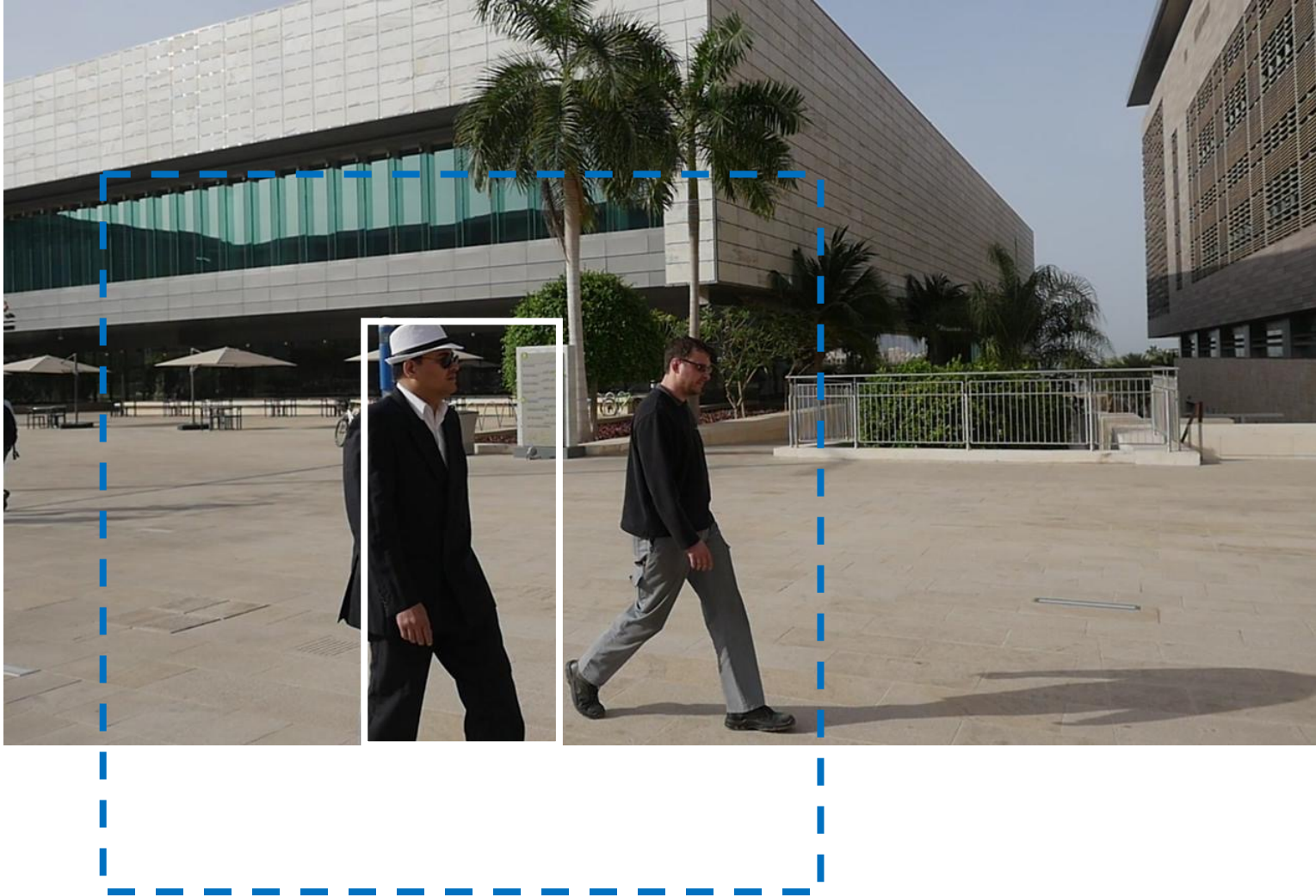
Results – VOTLT2021



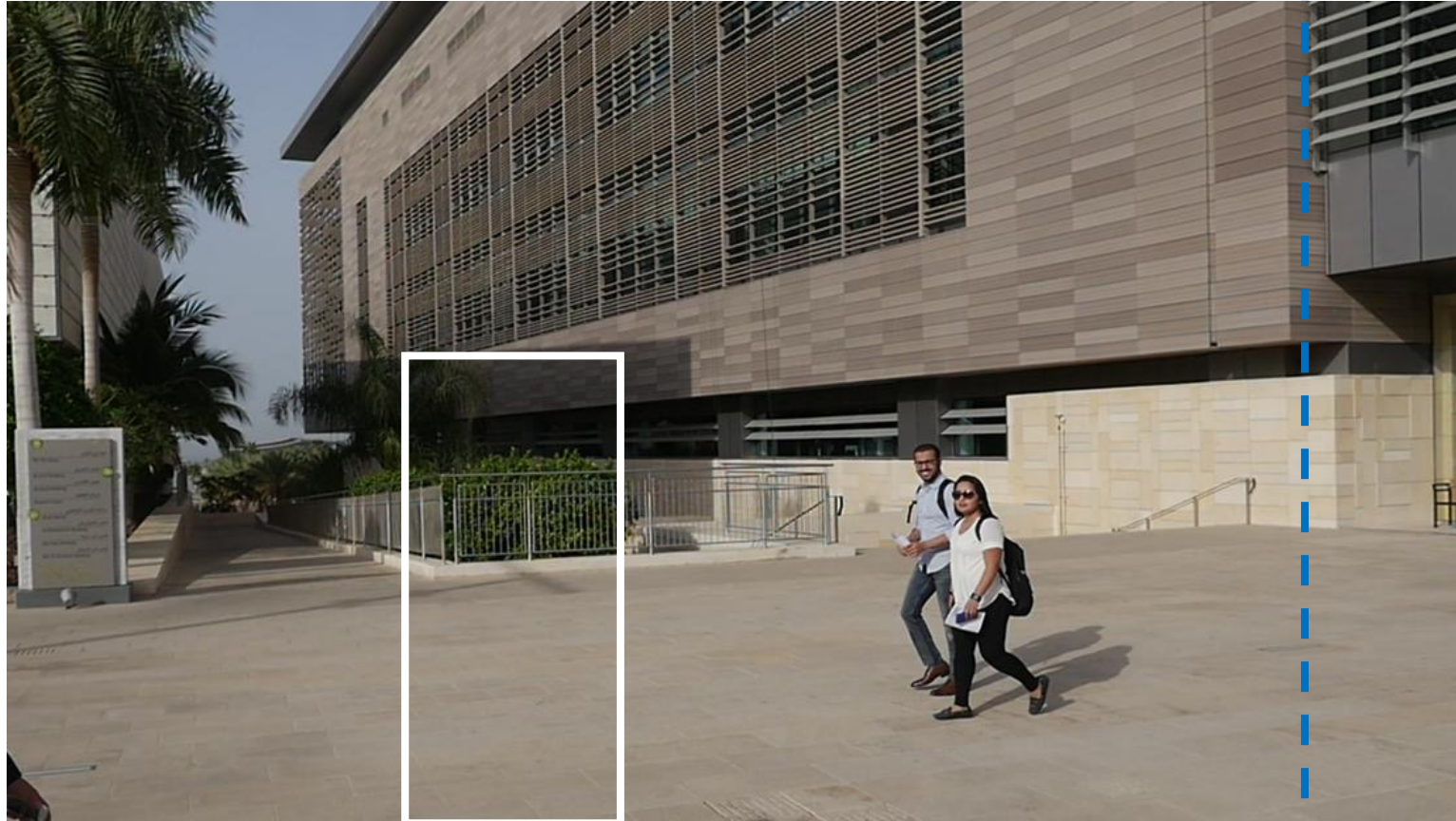
mipLT – Improvements to STARK



mlpLT – Improvements to STARK



mipLT – Improvements to STARK



mipLT – Improvements to STARK



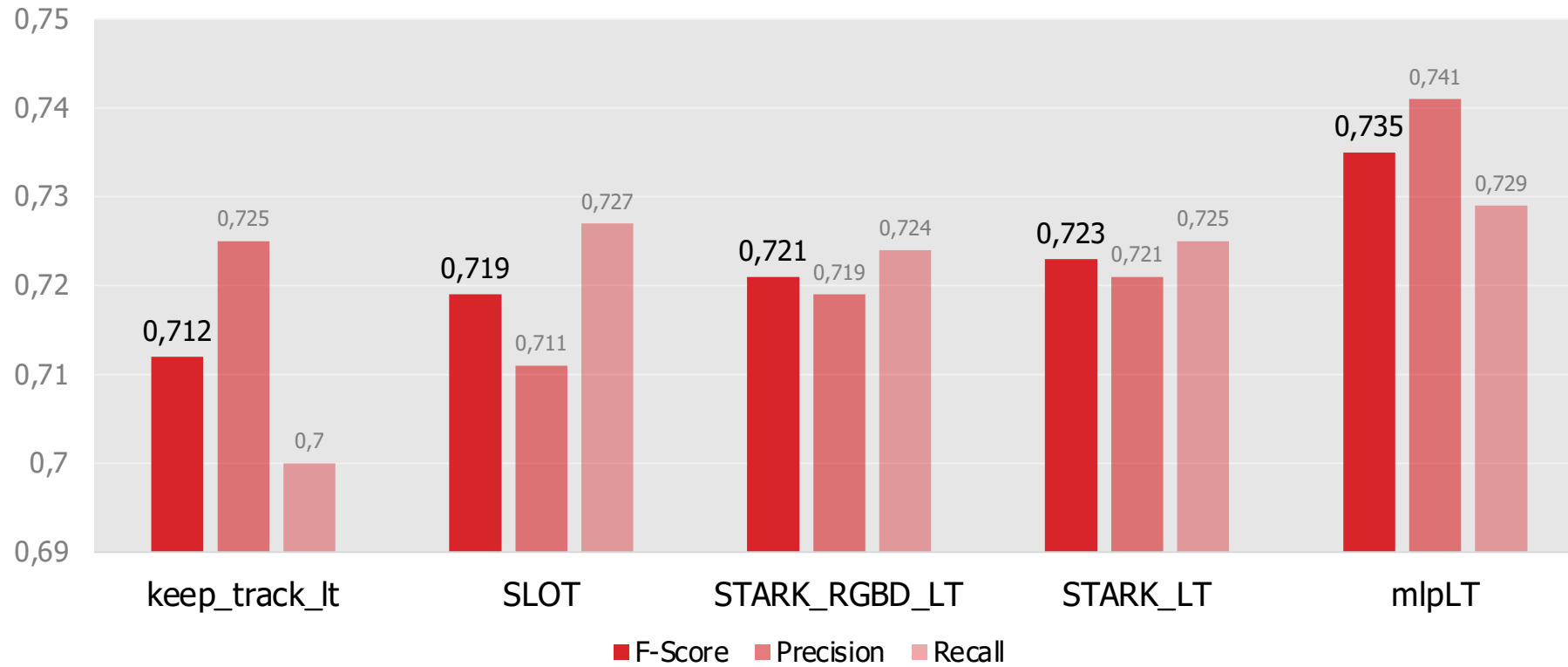
mipLT – Improvements to STARK



mipLT – Improvements to STARK



Results – VOTLT2021



Thank you!