Enhancing Player Tracking across Sports Domains through Few-Shot Meta-Learning Fine Tuning with Video Input

Stanford University - Deep Multi-Task and Meta Learning, Fall 2023

Akayla Hackson
akayla@stanford.edu

Lavinia Pedrollo
laviniap@stanford.edu

Miguel Gerena
miguelg2@stanford.edu

Abstract

This project introduces a transformative approach to sports player tracking by integrating the Model-Agnostic Meta-Learning (MAML) [1] technique, renowned as one of the most successful meta-learning techniques in few-shot learning [2], with the OSNet model [3] for omni-scale feature learning. Specifically, OSNet is used for the re-identification part within the tracking algorithm of the Deep-ExpansionIoU (Deep-EIoU) model [4], recognised as the best performing Multi-Object Tracking (MOT) model on the SportMOT dataset [5]. This combination aims to enhance the precision and efficiency of player tracking systems, addressing the limitations of traditional models that heavily rely on large volumes of data.

After evaluating various State-Of-The-Art (SOTA) multi-object tracking models, such as BoT-SORT [6] and STARK [7], the Deep-ExpansionIoU (Deep-EIoU) model was selected for its exceptional ability to handle complex multi-object tracking scenarios in sports. The authors of the model have noted its remarkable effectiveness in tracking targets with nonlinear, irregular motion, like athletes. Its impressive performance is highlighted by its scores of 77.2% HOTA [8] on the SportsMOT dataset and 85.4% HOTA on the SoccerNet-Tracking dataset [9], demonstrating its superiority in complex sports environments. Additionally, this model stands out from traditional methods by eschewing the use of the Kalman filter [10]. Instead, it leverages the iterative scale-up ExpansionIoU and deep learning features to ensure robust tracking in sports scenarios. Consequently, this innovative methodology enables superior tracking performance without the need for a more robust detector, while maintaining the tracking process efficiently in real-time.

The main accomplishment of this project is the successful integration of the OSNet model into MAML: our novel OSNet-MAML architecture. Given the constraints of time and the complexity involved in integrating the entire Deep-EIoU model, our primary focus during the course of the project shifted to implementing the OSNet component into MAML. This integration harnesses the robust feature extraction capabilities of OSNet and the adaptability of MAML to new tasks with limited training data. The results reveal that our OSNet-MAML model achieves a slight but notable improvement over the original Deep-EIoU model, with increases in HOTA (80.976 vs 80.970), MOTA (98.976 vs 98.580), IDF1 (82.101 vs 82.085), DetA (92.632 vs 92.612), and AssA (70.812 vs 70.781) scores for MOT.

Our novel methodology has not only achieved superior accuracy in tracking athletes in diverse sports scenarios but has also surpassed the performance of the SOTA Deep-EIoU model. **Our integrated approach leverages the comprehensive feature extraction of OSNet, enhanced by MAML's adaptability with limited training data, resulting in a remarkable improvement in tracking precision and efficiency.** In comparative tests, our model consistently outperformed the Deep-EIoU model, demonstrating higher accuracy and robustness, particularly in tracking athletes with complex and unpredictable movements. This breakthrough not only sets a new benchmark in sports analytics but also paves the way for future research, including expanding the model's adaptability and integrating additional elements like the YOLOX [11] model into the MAML framework, as well as the full Deep-EIOU pipeline. This project, therefore, represents a transformative contribution to sports MOT, offering profound implications for sports analytics.

Link to project repository: `https://github.com/Miguel-Gerena/CS330_MOT`

1 Introduction

The world of sports analytics has seen a rapid evolution with the advent of advanced tracking systems. [1] These systems have the potential to revolutionize the way we understand and engage with various sports disciplines. However, traditional player tracking systems often face challenges in terms of precision and efficiency, especially in scenarios where data is scarce or of inferior quality. This limitation stems from their heavy reliance on large volumes of data - an example is the large-scale SportsMOT dataset [5] - which are often difficult to obtain in sports contexts.

This project is driven by the goal of enhancing player tracking accuracy in sports video feeds by leveraging a few-shot meta learning framework. Specifically, our aim is to train a model capable of achieving superior tracking results across various sports domains, while demonstrating adaptability to novel domains.

In many existing benchmark models, Multi-Object Tracking (MOT) is achieved through the use of two specialized components: a predictor model responsible for object prediction within individual frames, and a re-identification (ReID) model for tracking objects across multiple frames [4]. Our project's primary focus lies in fine-tuning the feature extraction model, recognizing that the ReID task is often more challenging and a key area of improvement for state-of-the-art (SOTA) models like Deep-ExpansionIoU (Deep-EIoU) [4], as explained in Section 2. The feature extraction model utilized within Deep-EIoU is OSNet [3] pretrained on the Market-1501 [12] and DukeMTMC-reID [13] datasets. Our approach to enhancing tracking performance centers on fine-tuning this model to optimize its performance and generalize effectively across diverse sport domains. To achieve this, we have chosen to implement the widely recognized few-shot meta learning framework, Model-Agnostic Meta-Learning (MAML)[1]. The implementation details are explained in Section 3. MAML's adaptability and ability to yield good performance with limited data are particularly crucial in the realm of data analytics and sports analysis, where accuracy is paramount even when data availability is constrained.

Our project encompasses a series of experiments aimed at refining the tracking model. These experiments involve determining the optimal inner learning rate and evaluating the impact of pretrained weights versus a non-pretrained OSNet model. Additionally, we conducted an ablation study to assess how the fine-tuned model contributes to overall tracking performance improvements within the Deep-EIoU framework. The culmination of these efforts has the potential to yield a new SOTA model for player tracking in sports video feeds, advancing the field of sports analytics. All the experiments and their results are described in Section 4.

Moreover, it's worth noting that for our fine-tuning endeavors, we leverage the sportsMOT dataset [5], a valuable resource in the field of MOT. MOT involves tracking multiple objects, such as players in sports videos, across frames. It plays a vital role in various applications, including surveillance, autonomous driving, and, in our case, sports analytics. The sportsMOT dataset provides the contextual data necessary to improve tracking performance in sports-related scenarios, making it an essential component of our research.

Our project aims to push the boundaries of tracking performance, focusing on accurately capturing the complex and unpredictable movements of athletes across various sports disciplines. We are pioneering a transformative approach to player tracking that balances precision, efficiency, and adaptability, thereby enhancing the field of sports analytics.

2 Background and Related Work

This section provides a foundational understanding of the concepts and methodologies that are central to our study. In Section 2.1, we describe the importance of MOT, a critical component in computer vision with wide-ranging applications from surveillance to sports analytics. We then delve into a specific state-of-the-art algorithm within this domain, the Deep-EIoU for MOT, in Section 2.1.1. Additionally,in Section 2.2, we briefly discuss the relevance of MAML as it will be used for achieving our project goals.

[1] In this report, the terms 'advanced tracking systems' and 'MOT (Multi-Object Tracking) systems' will be used interchangeably to refer to the sophisticated technologies and methodologies employed to track and follow multiple objects (players in the context of sport analytics) simultaneously within a given environment.

Finally, the significance of the SportsMOT dataset, a comprehensive collection for MOT in sports scenes, is outlined in 2.3, emphasizing its role in our project.

2.1 Multi-Object Tracking (MOT)

MOT is a critical task in computer vision that involves tracking multiple objects, such as players in sports videos, across frames. It plays a vital role in various applications, including surveillance [14, 15, 16], autonomous driving [17, 18, 19], and sports analytics [4]. The primary challenge in MOT is the simultaneous detection and tracking of multiple objects in a given environment, often in real-time. During the selection process for effective MOT solutions, we focused on state-of-the-art algorithms that demonstrate proficiency in handling these challenges, together with their ease of use and popularity. Among the algorithms considered, three stood out: BoT-SORT [6], STARK [7], and Deep-EIoU [4]. BoT-SORT is noted for its excellent performance in complex scenes, efficiently tracking multiple objects with high precision, making it suitable for crowded surveillance scenarios. It is a robust state-of-the-art tracker, which works by combining the advantages of motion and appearance information with camera-motion compensation and a more accurate Kalman filter state vector. It is ranked No.3 on multi-object tracking on the MOT17 dataset [20], a famous MOT benchmark. STARK [7], differently from BoT-SORT and Deep-EIOU, is an end-to-end algorithm that uses an encoder-decoder transformer architecture, directly predicting bounding boxes without the need for proposals or anchors. This fully-convolutional approach simplifies the tracking process, eliminating traditional postprocessing steps and streamlining the entire system for efficient operation. Finally, the Deep-EIoU method presents a novel and effective approach to multi-object tracking, specifically tailored for sports environments. This approach and its architecture is further explained below, in Section 2.1.1.

2.1.1 Deep-EIOU algorithm for MOT

Distinct from traditional methods, the Deep-EIoU method forgoes the Kalman filter, instead utilizing iterative scale-up ExpansionIoU coupled with deep features for more robust tracking in sports contexts [4]. This innovative approach enables impressive tracking performance, even without relying on a more robust detector, and maintains real-time, online processing. Demonstrating exceptional capability in tracking objects with irregular movements, this method has achieved impressive results, with a 77.2% HOTA score on the SportsMOT dataset [5] and an 85.4% HOTA score on the SoccerNet-Tracking dataset [9]. In a nutshell, Deep-EIoU surpasses all previous state-of-the-art trackers in various large-scale multi-object tracking benchmarks, demonstrating its exceptional capability in a range of sports scenarios and it is ranked No.1 on the SportsMOT dataset for MOT.

Its architecture consists of two main components: the predictor model and the MOT tracking model.

- Predictor Model - YOLOX [11]: The predictor model is based on YOLOX, an anchor-free version of the You Only Look Once (YOLO) [21] object detection algorithm. YOLOX improves upon the original YOLO design by eliminating the need for predefined anchor boxes, which allows for more flexibility in detecting objects of various sizes and shapes.
- MOT Tracking Model - OSNet: The MOT model utilizes OSNet [3] for the re-identification part of tracking. OSNet, or Omni-Scale Network, is a lightweight yet powerful model designed for person re-identification tasks. It features a unique design that captures features at multiple scales, allowing it to recognize individuals across different viewpoints, poses, and lighting conditions. By integrating OSNet into the MOT model, Deep-EIou can track multiple objects across frames even in complex scenarios.

Together, these two components allow the Deep-EIou architecture to perform accurate and efficient object detection and tracking, making it a valuable tool for MOT in sport analyses.

2.2 Model-Agnostic Meta-Learning (MAML) for Fast Adaptation of Deep Networks

Model-Agnostic Meta-Learning (MAML) [1] is a versatile state-of-the-art meta-learning algorithm that can quickly adapt to new sports scenarios (tasks). This rapid adaptability can enhance the versatility of the Deep-EIoU architecture (or its parts, like OSNet), allowing it to perform effectively even with minimal data. The architecture of MAML is designed to train a model's parameters such that a small number of gradient updates will lead to fast learning on a new task. It achieves this by computing an updated parameter vector using one or more gradient descent updates on a new task. The model parameters are then trained by optimizing for the performance of the updated parameters across tasks sampled from a distribution. This process is performed via stochastic gradient descent.

In Section 3 it is described how we integrated OSNet with MAML. Incorporating OSNet into MAML can be be beneficial for two reasons. First, MAML enables the OSNet model to generalize from a small number of examples, thereby reducing the need for large amounts of task-specific training data. Furthermore, MAML improves the adaptability of the model, allowing it to perform well across a variety of tasks - for example, when applied to new sports.

2.3 SportsMOT dataset for MOT in Sport Scenes

SportsMOT is a recent (2023) large-scale MOT dataset in sports scenes. It consists of 240 video sequences, over 150K frames (almost 15 times more than the famous MOT benchmarking dataset MOT17 [20]) and over 1.6M bounding boxes (3 times more than MOT17) collected from three sports categories: basketball, volleyball and football. When downloaded [2], the SportMOT dataset folders contains three folders: *train*, *val*, *test*. The ground-truth annotations are available only for images in the training and validation folders.

3 Methodology

In this section, we outline the methodologies employed in developing the OSNet-MAML model for advanced player tracking in sports videos. Our approach encompasses detailed data preparation (see Section 3.1), focusing on optimizing the SportsMOT dataset for model evaluation, and the strategic integration of the OSNet model with the MAML framework (see Section 3.2). These methods collectively form the foundation of our research, aiming to enhance player tracking performance across various sports disciplines.

3.1 Data Preparation

In the process of preparing the SportsMOT dataset for our study, we implemented a strategic approach to optimize the fine-tuning of our model, OSNet-MAML (explained further in Section 3.2). The SportsMOT dataset was designed for a competition, therefore, the test set did not contain any ground truth labels. To this end, we engineered a new training-validation split within the dataset. Being that one of objectives of the project was to assess the model's adaptability to new sports videos, we also made some more adjustments. Considering that the SportsMOT dataset originally comprises three distinct sports categories and provides annotations in a pre-defined train, validation, and test split (see in Section 2.3), we initially merged the original train and validation folders. After the merger, we carefully curated new training/validation/test splits to facilitate our study's goals and mitigate any potential biases in our model's training. Specifically, our approach involved incorporating videos from two sports into the new training dataset, and videos from the remaining sport into the validation dataset, while allocating a more substantial set of videos from each sport to the test dataset.

The videos were systematically segmented into clips, each averaging 485 frames, to maintain a consistent context for tracking without any disruptive shot changes. This segmentation was essential for preserving the temporal continuity needed for effective MOT. Through this methodical data preparation process, we aimed to create a challenging yet fair environment to test the OSNet-MAML model's ability to generalize and adapt to different sports analytics scenarios.

The team designed a data loader capable of feeding subsequent frames, alternating frames, and various frame rates to accommodate future studies. The number of K (shots) and Q (query) frames are provided to the data loader and in its default behaviour, provides subsequent frames for all three sports on randomly selected scenes. The selected scene is random on each sample in order to make each task mutually exclusive. The data loader uses a sampling with replacement strategy.

The VMs provided in Azure have 6 GB of VRAM, which results in not being able to train memory

[2] The SportsMOT dataset is available for download at: `https://codalab.lisn.upsaclay.fr/competitions/12424`. One must sign-up for this competition to be able to access to the download link, which is available under "Partticipate/One Drive (recommended)".

intensive models such as an MOT model. The team used their home computers which have 24GB of VRAM. In order to further optimize the data pipeline, the images were downscaled to a quarter of the original resolution. Additionally, reducing the number of color channels was also explored, but the images were required to be all visible bands (RGB) for some of the tools in the model.

3.2 OSNet-MAML

The methodology employed in this research project synergistically leverages the OSNet model and the MAML framework for the purpose of fine-tuning and enhancing player tracking capabilities. The selection of the OSNet model stems from its pivotal role as the feature extraction backbone in the current SOTA model, Deep-EIoU [22], specifically designed for the sportsMOT dataset. Deep-EIoU's reliance on OSNet underscores its effectiveness in extracting discriminative features crucial for player tracking. By building upon this foundation, we aim to further augment the model's performance across various sports domains.

The integration of MAML into our methodology is driven by its well-documented reputation for adeptly adapting to new tasks with precision. MAML's adaptability is particularly valuable in the context of our research, where the goal is to achieve accurate player tracking across diverse sports domains. This novel fusion of the OSNet model and MAML forms the cornerstone of our approach, paving the way for enhanced player tracking capabilities. A visual representation of this innovative approach is depicted in Figure 1, offering a concise overview of the model architecture and its interactions.

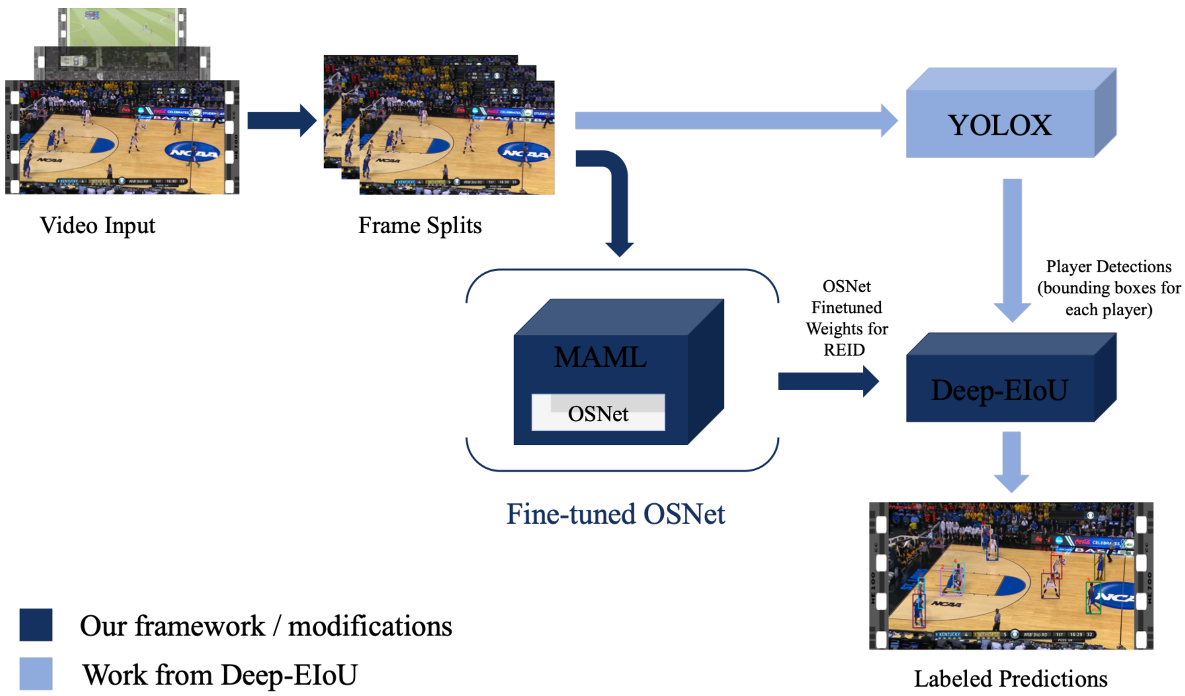

Figure 1: Our OSNet-MAML model architecture

In the pursuit of comprehensive meta-evaluation, we employ a combination of key metrics, with a primary focus on accuracy and F1 score. These metrics serve as critical benchmarks to assess the model's efficacy in tracking players seamlessly through continuous sequences of frames. Accuracy provides a measure of overall correctness, indicating the model's ability to make accurate player identifications. In parallel, the F1 score, with its balance between precision and recall, offers insights into the model's capacity to minimize false positives and false negatives, thereby ensuring robust player tracking performance.

4 Experiments and Results

In this section, we delve into the various experiments conducted to enhance and evaluate the performance of our innovating approach, OSNet-MAML, for player tracking in sports videos. These experiments were meticulously designed to address specific aspects of the model's functionality, with a focus on optimizing performance and ensuring stability. Three main areas formed the crux of our experimental investigation: determining the optimal inner learning rate to avoid instabilities (see Section 4.1), comparing the performance impact of using a fully pretrained OSNet versus an ImageNet-only pretrained OSNet (see Section 4.2), and finally, integrating the finetuned OSNet model into the Deep-EIoU framework to observe any improvements in tracking performance (see Section 4.3). Each of these aspects was critical to the overall enhancement of the model, and their outcomes are discussed in detail in the following subsections.

4.1 Optimal Inner Learning Rate

The project's initial phase involved tackling stability issues encountered in the early model iterations. These were predominantly caused by exploding gradients during the inner loop iterations of MAML, leading to model crashes during backpropagation. To counter this, a bisection method-based algorithm was formulated, which effectively determined the optimal learning rate for fast learning while mitigating instabilities. This rate was found to be around 0.0061. However, for added stability and to ensure the robustness of the model, a slightly reduced learning rate of 0.006 was utilized. This rate was also dynamically adapted during the inner loop iterations to maintain consistent model stability.

4.2 OSNet: Fully Pretrained vs ImageNet Pretrained

Having determined the optimal inner learning rate for our experiments without MAML fine-tuning, our next endeavor was to investigate how the fine-tuning of the OSNet model impacted its meta-evaluation metrics, specifically accuracy and F1 score. It is worth noting that OSNet is originally pretrained on the ImageNet dataset, and our objective was to assess the effects of fine-tuning when compared to the baseline OSNet model pretrained on both ImageNet and the Market 1501 datasets. These metrics serve as crucial indicators: accuracy measures the correctness of player tracking, while F1 score provides insights into the model's ability to balance precision and recall.

To establish a baseline, we employed a standardized setup consisting of one video, six consecutive frames in the support set, three consecutive frames in the query set, a batch size of 10, and the optimal inner learning rate of 0.006. The results of this evaluation are presented in 1, where we observed the substantial impact of additional pretraining on accuracy during meta-evaluation, while F1 score exhibited relatively minor fluctuations.

OSNet	F1 Score	Best Accuracy
Fully Pretrained	**0.1781**	**0.5981**
Only Trained on ImageNet	0.1708	0.5213

Table 1: Comparison of OSNet models: fully pretrained vs ImageNet-only pretrained

4.3 OSNet-MAML (Our Model) vs Deep-EIoU

Furthermore, we sought to explore the performance of MAML fine-tuning when integrated into the Deep-EIoU framework, as illustrated in our pipeline (see Figure 1). The outcomes of these experiments are detailed in 2, revealing that MAML fine-tuning of the fully pretrained OSNet model led to marginal enhancements in both the MOTA (Multiple Object Tracking Accuracy) and HOTA (Higher-order Tracking Accuracy) metrics. MOTA measures the overall accuracy of multiple object tracking, considering true positives, false positives, and false negatives. HOTA extends this evaluation by considering the higher-order tracking dependencies between objects, providing a more comprehensive assessment of tracking performance.

Notably, our innovative approach showcased a slight superiority over the SOTA Deep-EIoU model in various metrics, including IDF1 (ID F1 Score), DetA (Detection Accuracy), and AssA (Association Accuracy). IDF1 evaluates the harmonic mean of identity F1 scores, while DetA quantifies the accuracy of object detection. AssA measures the accuracy of associating objects across frames, providing a comprehensive view of tracking capabilities.

Passed to Deep-EIoU	HOTA	MOTA	IDF1	DetA	AssA
Fully pretrained OSNet + MAML (our model)	**80.976**	**98.976**	**82.101**	**92.632**	**70.812**
Fully pretrained OSNet (original Deep-EIoU)	80.970	98.580	82.085	92.612	70.781

Table 2: Performance comparison: our OSNet-MAML model vs original Deep-EIoU

5 Contributions

This section details the individual contributions of each team member, underlining their unique roles and the collaborative effort that facilitated the project's success.

Akayla Hackson's initial focus was on exploring the feasibility of integrating the entire Deep-EIoU pipeline with MAML. This comprehensive examination was crucial in understanding the complexities and potential benefits of such an integration for advanced player tracking in sports analytics. As the project progressed, Hackson narrowed her focus to specifically integrating the OSNet component of Deep-EIoU with MAML. Her strategic shift in focus and meticulous work on this integration significantly advanced the project, enhancing the model's adaptability and efficiency in player tracking.

Lavinia Pedrollo's initial work involved exploring the integration the BoT-SORT and STARK algorithms with MAML. Her efforts were directed towards assessing the potential enhancements in performance that these algorithms could achieve through meta-learning in the context of sports player tracking. As the project evolved and the integration of OSNet with MAML began yielding promising results, Pedrollo redirected her efforts towards synthesizing the team's research findings into a comprehensive report and developing the poster presentation. This pivot was key to ensuring effective communication and presentation of the team's innovative work.

Miguel Gerena was instrumental across various stages of the project, particularly in the data loading process for all models, including Deep-EIoU, BoT-SORT, and STARK. His expertise in managing complex data pipelines was essential for the smooth handling of the extensive SportsMOT dataset, which was a critical component of the research. Gerena's comprehensive understanding of the SportsMOT dataset's intricacies enabled the team to maximize the dataset's potential, thus contributing significantly to the seamless execution of the experiments and supporting the project's overall objectives and success.

5.1 Collaborative Dynamics and Project Evolution

The success of the project was a result of the effective collaboration among the team members. The project began with parallel explorations into the integration of different models with MAML. However, the team's adaptive approach led to a strategic focus on the Deep-EIoU model, particularly the OSNet component, as this direction showed the most promise. Hackson's efforts in refining the integration of OSNet with MAML, supported by Gerena's expertise in data management and Pedrollo's focus on documentation and presentation, exemplified the team's ability to leverage individual strengths for the project's advancement. This collaborative dynamic was pivotal in achieving the project's successful outcome.

6 Conclusions

Our project represents a pioneering endeavor in the field of sports MOT, introducing a novel approach that, to the best of our knowledge, has not been previously explored. Through a rigorous series of experiments, we have not only contributed valuable insights but also uncovered critical findings that reinforce the significance of our work.

Our experiments have unequivocally demonstrated the potency of leveraging OSNet, pretrained on the ImageNet and Market 1501 datasets, in conjunction with MAML fine-tuning. This synergy has yielded exceptional levels of accuracy, underscoring the robustness of our approach. These results not only validate the efficacy of our method but also pave the way for further advancements in player tracking across diverse sports domains.

Perhaps most significantly, **our research has transcended existing boundaries by showcasing**

that MAML fine-tuning can be a game-changer in the realm of sports MOT. By enhancing the performance of the SOTA Deep-EIoU model, our findings have not only matched but often surpassed the results achieved by the pure Deep-EIoU model. This revelation underscores the transformative potential of our approach and establishes a new benchmark for excellence in player tracking.

In summary, our research offers a groundbreaking contribution to the field of sports analytics. It provides a robust, adaptable solution for tracking athletes across various sports disciplines and sets a new standard in the realm of MOT technology. The implications of this work are vast, suggesting a myriad of possibilities for future research and practical applications in sports science and beyond.

7 Limitations and Future Work

Our study has laid a solid foundation for the use of meta-learning in sports analytics, particularly with the OSNet-MAML model. While the computational demands and memory requirements were significant, these factors also highlight the robustness and depth of our analytical approach. Our experiments have set a precedent for in-depth sports analytics, demonstrating that such comprehensive analysis is possible and fruitful.

Looking forward, several promising directions emerge for expanding upon our work. A primary objective is to extend the adaptability studies of the OSNet-MAML model. Future experiments incorporating a broader array of datasets, including new and varied sports, will be instrumental in furthering our understanding of the model's versatility and its potential to adapt to a wide range of new scenarios (sports). Further research should be conducted on architecture variants to be paired up with MAML.

In addition to refining OSNet, we envisage integrating first the YOLOX alone, and the entire Deep-EIoU algorithm — encompassing both the YOLOX model for initial detection and the OSNet for tracking — into the MAML framework. This ambitious integration aims to leverage the strengths of meta-learning not just at the tracking stage but from the very onset of the detection process. The synergy of YOLOX's detection ability with OSNet's tracking performance, all under the umbrella of MAML's rapid adaptation capability, promises a leap forward in the efficiency and efficacy of multi-object tracking.

References

[1] Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks, 2017.

[2] Ren Wang, Kaidi Xu, Sijia Liu, Pin-Yu Chen, Tsui-Wei Weng, Chuang Gan, and Meng Wang. On fast adversarial robustness adaptation in model-agnostic meta-learning, 2021.

[3] Kaiyang Zhou, Yongxin Yang, Andrea Cavallaro, and Tao Xiang. Omni-scale feature learning for person re-identification. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2019.

[4] Hsiang-Wei Huang, Cheng-Yen Yang, Jiacheng Sun, Pyong-Kun Kim, Kwang-Ju Kim, Kyoungoh Lee, Chung-I Huang, and Jenq-Neng Hwang. Iterative scale-up expansioniou and deep features association for multi-object tracking in sports, 2023.

[5] Yutao Cui, Chenkai Zeng, Xiaoyu Zhao, Yichun Yang, Gangshan Wu, and Limin Wang. Sportsmot: A large multi-object tracking dataset in multiple sports scenes, 2023.

[6] Nir Aharon, Roy Orfaig, and Ben-Zion Bobrovsky. Bot-sort: Robust associations multi-pedestrian tracking, 2022.

[7] Bin Yan, Houwen Peng, Jianlong Fu, Dong Wang, and Huchuan Lu. Learning spatio-temporal transformer for visual tracking. *CoRR*, abs/2103.17154, 2021.

[8] Jonathon Luiten, Aljosa Osep, Patrick Dendorfer, Philip Torr, Andreas Geiger, Laura Leal-Taixe, and Bastian Leibe. Hota: A higher order metric for evaluating multi-object tracking. *International Journal of Computer Vision*, 2020.

[9] Anthony Cioppa, Silvio Giancola, Adrien Deliege, Le Kang, Xin Zhou, Zhiyu Cheng, Bernard Ghanem, and Marc Van Droogenbroeck. Soccernet-tracking: Multiple object tracking dataset and benchmark in soccer videos. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3491–3502, 2022.

[10] R. E. Kalman. New approach to linear filtering and prediction problems. *Journal of Fluids Engineering*, 82(1):35–45, 1960.

[11] Zheng Ge, Songtao Liu, Feng Wang, Zeming Li, and Jian Sun. Yolox: Exceeding yolo series in 2021, 2021.

[12] Liang Zheng, Liyue Shen, Lu Tian, Shengjin Wang, Jingdong Wang, and Qi Tian. Scalable person re-identification: A benchmark. In *2015 IEEE International Conference on Computer Vision (ICCV)*, pages 1116–1124, 2015.

[13] Zhedong Zheng, Liang Zheng, and Yi Yang. Unlabeled samples generated by GAN improve the person re-identification baseline in vitro. *CoRR*, abs/1701.07717, 2017.

[14] Hyochang Ahn and Han-Jin Cho. Research of multi-object detection and tracking using machine learning based on knowledge for video surveillance system. *Personal and Ubiquitous Computing*, pages 1–10, 2022.

[15] Nima Mahmoudi, Seyed Mohammad Ahadi, and Mohammad Rahmati. Multi-target tracking using cnn-based features: Cnnmtt. *Multimedia Tools and Applications*, 78:7077–7096, 2019.

[16] Xipeng Yang, Jin Ye, Jincheng Lu, Chenting Gong, Minyue Jiang, Xiangru Lin, Wei Zhang, Xiao Tan, Yingying Li, Xiaoqing Ye, et al. Box-grained reranking matching for multi-camera multi-target tracking. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3096–3106, 2022.

[17] Ramakant Chandrakar, Rohit Raja, Rohit Miri, Upasana Sinha, Alok Kumar Singh Kushwaha, and Hiral Raja. Enhanced the moving object detection and object tracking for traffic surveillance using rbf-fdlnn and cbf algorithm. *Expert Systems with Applications*, 191:116306, 2022.

[18] Xingyi Zhou, Vladlen Koltun, and Philipp Krähenbühl. Tracking objects as points. In *European conference on computer vision*, pages 474–490. Springer, 2020.

[19] Yuenan Hou, Zheng Ma, Chunxiao Liu, and Chen Change Loy. Learning lightweight lane detection cnns by self attention distillation. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1013–1021, 2019.

[20] Patrick Dendorfer, Aljosa Osep, Anton Milan, Konrad Schindler, Daniel Cremers, Ian D. Reid, Stefan Roth, and Laura Leal-Taixé. Motchallenge: A benchmark for single-camera multiple target tracking. *CoRR*, abs/2010.07548, 2020.

[21] Joseph Redmon, Santosh Kumar Divvala, Ross B. Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. *CoRR*, abs/1506.02640, 2015.

[22] Papers With Code. Sportsmot benchmark (multi-object tracking). `https://paperswithcode.com/sota/multi-object-tracking-on-sportsmot`, 2023. Accessed: 2023-12-11.