# About Multiple Object Tracking (MOT) using deepSORT

ANNA AVANYAN

**ABSTRACT**

**Recently, multi-object tracking (MOT) has attracted increasing attention, and accordingly, remarkable progress has been achieved. However, existing methods tend to use various basic models (e.g., detector and embedding models) and different training or inference tricks. In this paper revisited the article "StrongSORT: Make DeepSORT Great Again" by Yunhao Du, Zhicheng Zhao, Yang Song, Yanyun Zhao, Fei Su, Tao Gong, and Hongying Meng. The algorithm proposed in the article was studied and applied to another dataset, with integrated code for the convenient use of the algorithm. Experiments were conducted on a dataset with poor quality and varying resolutions**

**Index Terms — Multi-Object Tracking, DeepSORT.**

**INTRODUCTION**

MULTI-OBJECT TRACKING (MOT) aims to detect and track all specific classes of objects frame by frame, which plays an essential role in video understanding. In the past few years, the MOT task has been dominated by the tracking-by-detection (TBD) paradigm [60, 3, 55, 4, 32], which performs per frame detection and formulates the MOT problem as a data association task. TBD methods tend to extract appearance and/or motion embeddings first and then perform bipartite graph matching. Benefiting from high-performing object detection models, TBD methods have gained favour due to their excellent performance. As MOT is a downstream task corresponding to object detection and object re-identification (ReID), recent works tend to use various detectors and ReID models to increase MOT performance [18, 39], which makes it difficult to construct a fair comparison between them. Another problem preventing fair comparison is the usage of various external datasets for training [64, 63]. Moreover, some training and inference tricks are also used to improve the tracking performance.

To solve the above problems, the paper "StrongSORT: Make DeepSORT Great Again" [1] presents a simple but effective MOT baseline called StrongSORT. In paper revisited the classic TBD tracker DeepSORT [55], which is among the earliest methods that apply a deep learning model to the MOT task. Is choosed DeepSORT because of its simplicity, expansibility and effectiveness. It is claimed that DeepSORT underperforms compared with state-of-the-art methods because of its outdated techniques, rather than its tracking paradigm. To be specific, first equiped DeepSORT with a strong detector [18] following [63] and embedding model [30]. Then, collected some inference tricks from recent works to further improve its performance. Simply equipping DeepSORT with these advanced components results in the proposed StrongSORT, and it is shown that it can achieve SOTA results on the popular benchmarks MOT17 [31] and MOT20 [9] but in other dataset which has no high qualitie it works not very good so is integrated some helpfull information about location of objects in each frame.. There are two "missing" problems in the MOT task, i.e., missing association and missing detection. Missing association means the same object is spread in more than one tracklet. This problem is particularly common in online trackers because they lack global information in association. Missing detection, also known as false negatives, refers to recognizing the object as background, which is usually caused by occlusion and low resolutions.

**RELATED WORK**

MOT methods can be classified into separate and joint trackers. Separate trackers [60, 3, 55, 4, 32, 21] follow the tracking-by-detection paradigm, which localizes targets first and then associates them with information on appearance, motion, etc. Benefiting from the rapid development of object detection [39, 38, 18], separate trackers have been widely applied in MOT tasks. Recently, several joint tracking methods [57, 59, 28, 51] have been proposed to jointly train detection and other components, such as motion, embedding and association models. The main advantages of these trackers are low computational cost and comparable performance.
Meanwhile, several recent studies [42, 43, 63, 7] have abandoned appearance information, and relied only on highperformance detectors and motion information, which achieve high running speed and state-of-the-art performance on MOTChallenge benchmarks [31, 9]. However, abandoning appearance features would lead to poor robustness in more complex scenes. In the paper [1], was adopt the DeepSORT-like [55] paradigm and equip it with advanced techniques from various aspects to confirm the effectiveness of this classic framework.

**METHOD**

In the paper [1] was presented and explained DeepSORT, StrongSORT, Advanced modules EMA , ECC, NSA Kalman, Vanilla Matching. Let's look in more detail ECC.

Camera movements exist in multiple benchmarks. In article [1] adopt the enhanced correlation coefficient maximization (ECC) [13] model for camera motion compensation. It is a technique for parametric image alignment that can estimate the global rotation and translation between adjacent frames. Specifically, it is based on the following criterion to quantify the performance of the warping transformation:

$$E_{ECC}(\mathbf{p}) = \left\| \frac{\bar{\mathbf{i}}_{\mathbf{r}}}{\|\bar{\mathbf{i}}_{\mathbf{r}}\|} - \frac{\bar{\mathbf{i}}_{\mathbf{w}}(\mathbf{p})}{\|\bar{\mathbf{i}}_{\mathbf{w}}(\mathbf{p})\|} \right\|^2,$$

where $\|\cdot\|$ denotes the Euclidean norm, $\mathbf{p}$ is the warping parameter, and $\bar{\mathbf{i}}_{\mathbf{r}}$ and $\bar{\mathbf{i}}_{\mathbf{w}}(\mathbf{p})$ are the zero-mean versions of the reference (template) image $\mathbf{i}_{\mathbf{r}}$ and warped image $\mathbf{i}_{\mathbf{w}}(\mathbf{p})$. Then, the image alignment problem is solved by minimizing $E_{ECC}(\mathbf{p})$, with the proposed forward additive iterative algorithm or inverse compositional iterative algorithm. Due to its efficiency and effectiveness, ECC is widely used to compensate for the motion noise caused by camera movement in MOT tasks.

*Theorem I:* *Consider the scalar function*

$$f(\mathbf{x}) = \frac{u + \mathbf{u}^t\mathbf{x}}{\sqrt{v + 2\mathbf{v}^t\mathbf{x} + \mathbf{x}^tQ\mathbf{x}}} \tag{10}$$

*where $u, v$ are scalars; $\mathbf{u}, \mathbf{v}$ are vectors of length $N$; $Q$ is a square, symmetric and positive definite matrix of size $N$ and $v, \mathbf{v}, Q$ are such that*

$$v > \mathbf{v}^tQ^{-1}\mathbf{v} \tag{11}$$

*then, as far as the maximal value of $f(\mathbf{x})$ is concerned, we distinguish the following two cases:*

*Case $u > \mathbf{u}^tQ^{-1}\mathbf{v}$: here we have a maximum, specifically*

$$\max_{\mathbf{x}} f(\mathbf{x}) = \sqrt{\frac{(u - \mathbf{u}^tQ^{-1}\mathbf{v})^2}{v - \mathbf{v}^tQ^{-1}\mathbf{v}} + \mathbf{u}^tQ^{-1}\mathbf{u}}. \tag{12}$$

*which is attainable for*

$$\mathbf{x} = Q^{-1}\left\{\frac{v - \mathbf{v}^tQ^{-1}\mathbf{v}}{u - \mathbf{u}^tQ^{-1}\mathbf{v}}\mathbf{u} - \mathbf{v}\right\}. \tag{13}$$

*Case $u \le \mathbf{u}^tQ^{-1}\mathbf{v}$: here we have a supremum which is equal to*

$$\sup_{\mathbf{x}} f(\mathbf{x}) = \sqrt{\mathbf{u}^tQ^{-1}\mathbf{u}} \tag{14}$$

*and can be approached arbitrarily close by selecting*

$$\mathbf{x} = Q^{-1}\{\lambda\mathbf{u} - \mathbf{v}\}, \tag{15}$$

*with $\lambda$ positive scalar and of sufficiently large value*[1].

*Proof:* The proof makes repeated use of the Schwartz inequality. All details are presented in the Appendix. ■

**EXPERIMENTS**

Experiments were conducted using a newly compiled dataset specifically designed to challenge current object detection algorithms. This dataset features scenarios with poor lighting, high occlusion, and rapid object movements which traditionally result in low detection rates

Due to the subpar performance of automated detection algorithms on new dataset, we introduced a manual labeling feature allowing users to select objects of interest directly on the initial video frame. This manual intervention aims to bypass the unreliable detection stage, ensuring that tracking processes commence with accurate data.

Is adapted the Deep SORT tracking algorithm to operate based on these manually selected bounding boxes without relying on subsequent automated detections. This modification tests the tracker's dependency on the initial user input and its subsequent ability to predict object trajectories based solely on motion analysis.

Implementation Challenges

Integrating manual selection posed significant technical challenges. The transition from manual selection to automatic tracking involved adjustments in how the tracking algorithm initializes and sustains object identities over time. We encountered issues where the tracker would either fail to continue tracking post-initialization or inaccurately track unrelated objects. These problems were indicative of the complexities involved in maintaining tracking continuity without continuous detection inputs.

Key issues included:

- Initialization Failures: Occasionally, the tracker did not successfully transition from the manual selection phase to the tracking phase, leading to a complete halt in tracking.
- False Tracking: In several instances, the tracker incorrectly associated the manually selected boxes with background elements or other non-target objects in subsequent frames, leading to erroneous track continuations.

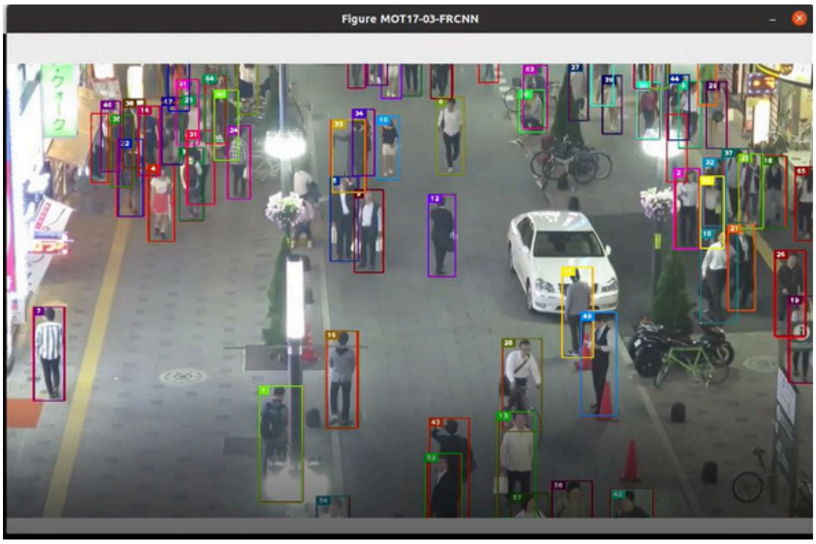



Image 1: Tracking on MOT17 Dataset

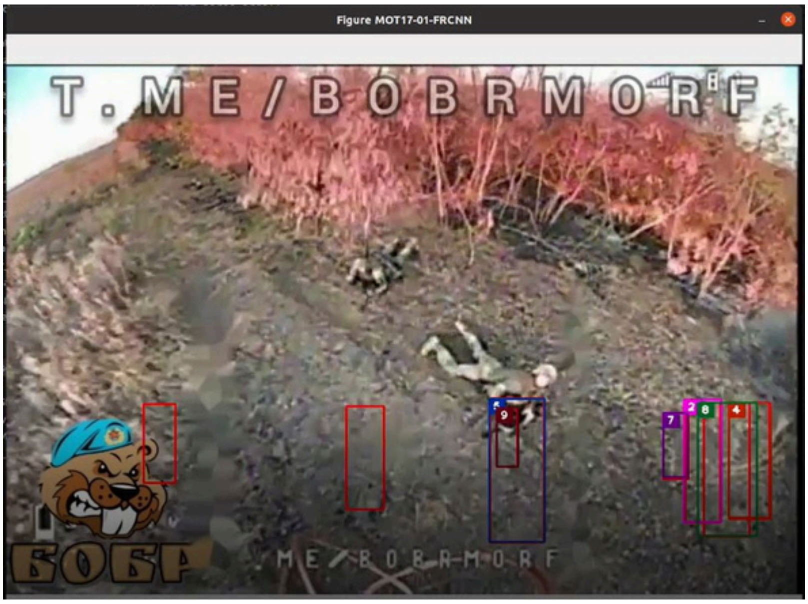



Image 2: Tracking on Integrated Dataset

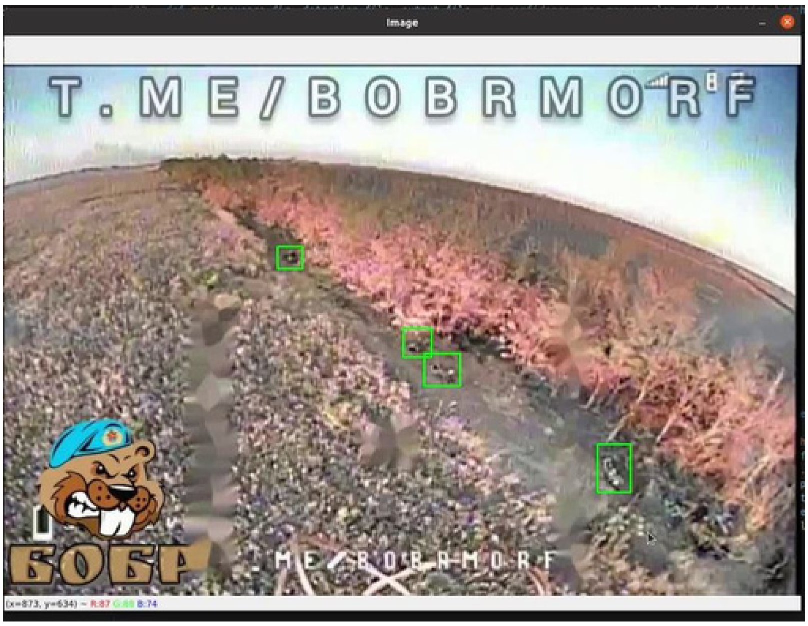
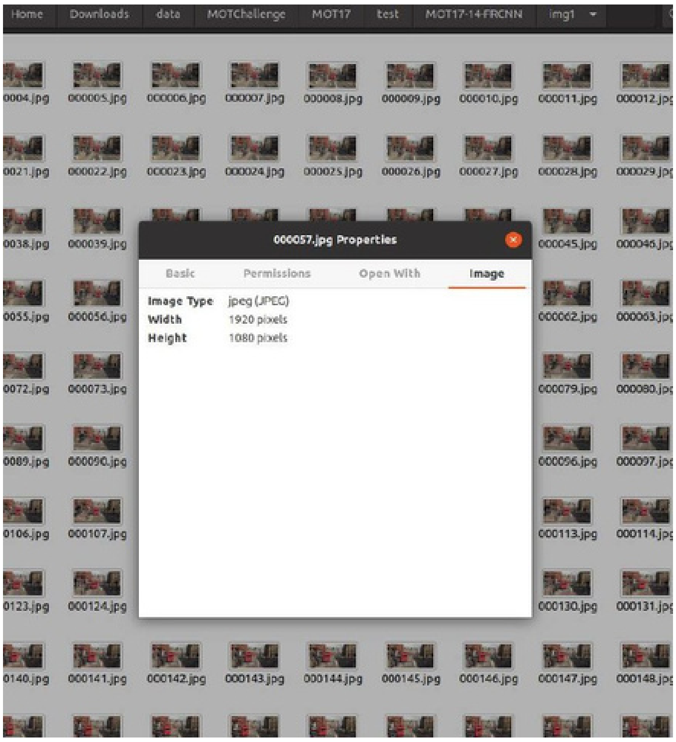


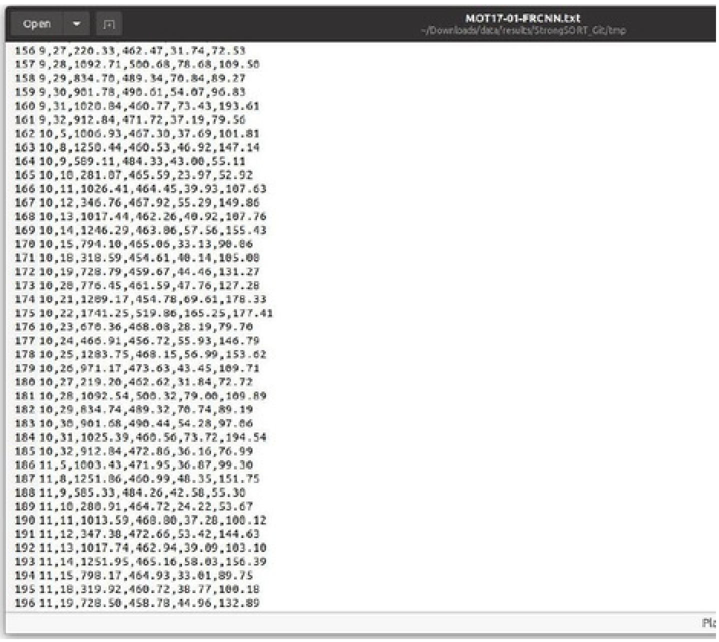

Image 3: Initial Labeling on Integrated Dataset

## CONCLUSION

Based on the detailed examination of the article [1] and extensive experimentation with the associated code, several critical insights and conclusions can be drawn regarding the applicability and limitations of the discussed detection algorithm. Firstly, the structure inherent to the algorithm significantly intertwines detection and tracking components, rendering it challenging to isolate and utilize these components separately. This tightly coupled architecture can limit flexibility in applications where modular adjustments are desired for enhanced detection or tracking accuracy.
Furthermore, the algorithm demonstrates a notable degree of error when applied to newly introduced datasets.

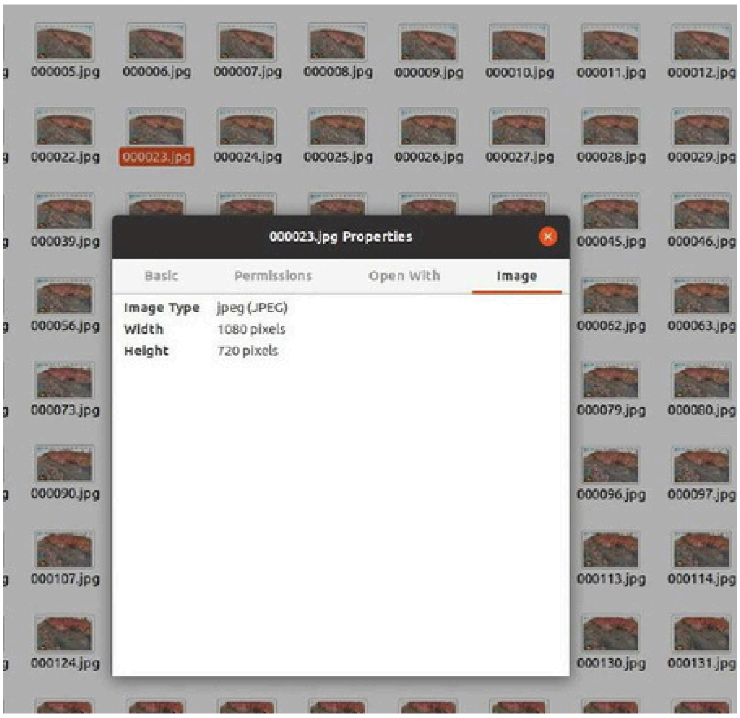



MOT17 dataset image properties.

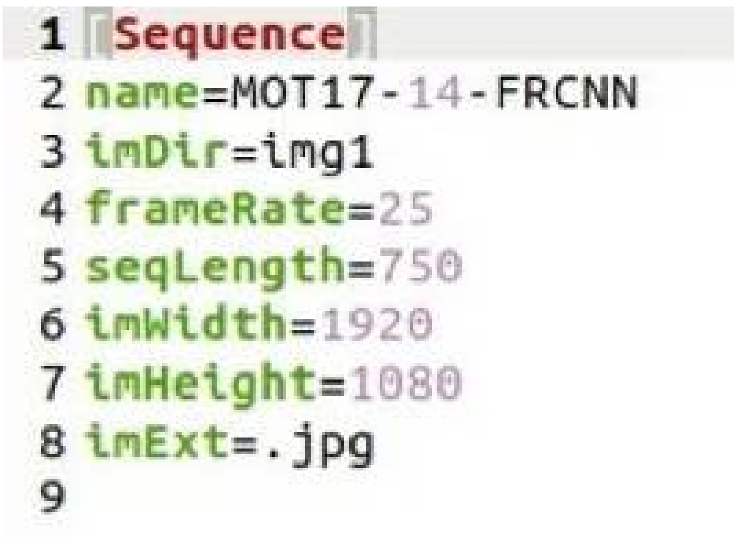


```
9,27,220.33,462.47,31.74,72.53
9,28,1092.71,500.68,78.68,109.50
9,29,834.70,489.34,70.84,89.27
9,30,901.78,490.61,54.07,96.83
9,31,1020.84,460.77,73.43,193.61
9,32,912.84,471.72,37.19,79.56
10,5,1006.93,467.30,37.69,101.81
10,8,1250.44,460.53,46.92,147.14
10,9,589.11,484.33,43.00,55.11
10,10,281.07,465.59,23.97,52.92
10,11,1026.41,464.45,39.93,107.63
10,12,346.76,467.92,55.29,149.86
10,13,1017.44,462.26,40.92,107.76
10,14,1246.29,463.06,57.56,155.43
10,15,794.10,465.06,33.13,90.06
10,18,318.59,454.61,40.14,105.08
10,19,728.79,459.67,44.46,131.27
10,20,776.45,461.59,47.76,127.28
10,21,1289.17,454.78,69.61,178.33
10,22,1741.25,519.86,165.25,177.41
10,23,670.36,468.08,28.19,79.70
10,24,466.91,456.72,55.93,146.79
10,25,1283.75,468.15,56.99,153.62
10,26,971.17,473.63,43.45,109.71
10,27,219.20,462.62,31.84,72.72
10,28,1092.54,500.32,79.00,109.89
10,29,834.74,489.32,70.74,89.19
10,30,901.68,490.44,54.28,97.06
10,31,1025.39,460.56,73.72,194.54
10,32,912.84,472.86,36.16,76.99
11,5,1003.43,471.95,36.87,99.30
11,8,1251.86,460.99,48.35,151.75
11,9,585.33,484.26,42.58,55.30
11,10,280.91,464.72,24.22,53.67
11,11,1013.59,468.80,37.28,100.12
11,12,347.38,472.66,53.42,144.63
11,13,1017.74,462.94,39.09,103.10
11,14,1251.95,465.16,58.03,156.39
11,15,798.17,464.93,33.01,89.75
11,18,319.92,460.72,38.77,100.18
11,19,728.50,458.78,44.96,132.89
```

The results are saved in the following format: (frame, x, y, w, h, conf).




new dataset image properties.



```
[Sequence]
name=MOT17-14-FRCNN
imDir=img1
frameRate=25
seqLength=750
imWidth=1920
imHeight=1080
imExt=.jpg
```

seqinfo.ini

## References

[1] Du, Y., Zhao, Z., Song, Y., Zhao, Y., Su, F., Gong, T., Meng, H. "StrongSORT: Make DeepSORT Great Again."

[2] Bernardin, K., Stiefelhagen, R. "Evaluating multiple object tracking performance: the clear MOT metrics." EURASIP Journal on Image and Video Processing, 2008, pp. 1–10.

[3] Bewley, A., Ge, Z., Ott, L., Ramos, F., Upcroft, B. "Simple online and realtime tracking." In: 2016 IEEE International Conference on Image Processing (ICIP), pp. 3464–3468.

[4] Bochinski, E., Eiselein, V., Sikora, T. "High-speed tracking-by-detection without using image information." In: 2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), pp. 1–6.

[5] Bochinski, E., Senst, T., Sikora, T. "Extending iou based multi-object tracking by visual information." In: 2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), pp. 1–6.

[6] Braso, G., Leal-Taixé, L. "Learning a neural solver for multiple object tracking." In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 6247–6257, 2020.

[7] Cao, J., Weng, X., Khirodkar, R., Pang, J., Kitani, K. "Observation-centric sort: Rethinking sort for robust multi-object tracking." arXiv preprint arXiv:2203.14360, 2022.

[8] Dai, P., Wang, X., Zhang, W., Chen, J. "Instance segmentation enabled hybrid data association and discriminative hashing for online multi-object tracking." IEEE Transactions on Multimedia, vol. 21, no. 7, pp. 1709–1723, 2018.

[9] Dendorfer, P., Rezatofighi, H., Milan, A., Shi, J., Cremers, D., Reid, I., Roth, S., Schindler, K., Leal-Taixé, L. "MOT20: A benchmark for multi-object tracking in crowded scenes." arXiv preprint arXiv:2003.09003, 2020.

[10] Du, Y., Tong, Z., Wan, J., Zhang, B., Zhao, Y. "Pami-ad: An activity detector exploiting part-attention and motion information in surveillance videos." In: 2022 IEEE International Conference on Multimedia and Expo Workshops (ICMEW), pp. 1–6.

[11] Du, Y., Wan, J., Zhao, Y., Zhang, B., Tong, Z., Dong, J. "Giaotracker: A comprehensive framework for mcmot with global information and optimizing strategies in visdrone 2021." In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 2809–2819, 2021.

[12] Ess, A., Leibe, B., Schindler, K., Van Gool, L. "A mobile vision system for robust multi-person tracking." In: 2008 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8.

[13] Evangelidis, G.D., Psarakis, E.Z. "Parametric image alignment using enhanced correlation coefficient maximization." IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 30, no. 10, pp. 1858–1865, 2008.

[14] Fernandez Llorca, D., Hernández Martínez, A., García Daza, I. "Vision-based vehicle speed estimation: A survey." IET Intelligent Transport Systems, vol. 15, no. 8, pp. 987–1005, 2021.

[15] Fu, Z., Angelini, F., Chambers, J., Naqvi, S.M. "Multilevel cooperative fusion of gm-phd filters for online multiple human tracking." IEEE Transactions on Multimedia, vol. 21, no. 9, pp. 2277–2291, 2019.

[16] Gao, J., Nevatia, R. "Revisiting temporal modeling for video-based person reid." arXiv preprint arXiv:1805.02104, 2018.

[17] Gao, T., Pan, H., Wang, Z., Gao, H. "A crf-based framework for tracklet inactivation in online multi-object tracking." IEEE Transactions on Multimedia, vol. 24, pp. 995–1007, 2021.

[18] Ge, Z., Liu, S., Wang, F., Li, Z., Sun, J. "Yolox: Exceeding yolo series in 2021." arXiv preprint arXiv:2107.08430, 2021.

[19] Geiger, A., Lenz, P., Stiller, C., Urtasun, R. "Vision meets robotics: The kitti dataset." The International Journal of Robotics Research, vol. 32, no. 11, pp. 1231–1237, 2013.

[20] Han, S., Huang, P., Wang, H., Yu, E., Liu, D., Pan, X. "Mat: Motion-aware multi-object tracking." Neurocomputing, 2022.

[21] He, J., Huang, Z., Wang, N., Zhang, Z. "Learnable graph matching: Incorporating graph partitioning with deep feature learning for multiple object tracking." In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5299–5309, 2021.

[22] Hofmann, M., Haag, M., Rigoll, G. "Unified hierarchical multi-object tracking using global data association." In: 2013 IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS), pp. 22–28.

[23] Hu, H.N., Yang, Y.H., Fischer, T., Darrell, T., Yu, F., Sun, M. "Monocular quasi-dense 3d object tracking." IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022.

[24] Kalman, R.E. "A new approach to linear filtering and prediction problems." Journal of Basic Engineering, vol. 82D, pp. 35–45, 1960.

[25] Khurana, T., Dave, A., Ramanan, D. "Detecting invisible people." In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 3174–3184, 2021.

[26] Kingma, D.P., Ba, J. "Adam: A method for stochastic optimization." In: International Conference on Learning Representations, 2014.

[27] Kuhn, H.W. "The Hungarian method for the assignment problem." Naval Research Logistics Quarterly, vol. 2, no. 1-2, pp. 83–97, 1955.

[28] Liang, C., Zhang, Z., Zhou, X., Li, B., Zhu, S., Hu, W. "Rethinking the competition between detection and reid in multi-object tracking." IEEE Transactions on Image Processing, vol. 31, pp. 3182–3196, 2022.

[29] Luiten, J., Osep, A., Dendorfer, P., Torr, P., Geiger, A., Leal-Taixé, L., Leibe, B. "Hota: A higher order metric for evaluating multi-object tracking." International Journal of Computer Vision, vol. 129, no. 2, pp. 548–578, 2021.

[30] Luo, H., Jiang, W., Gu, Y., Liu, F., Liao, X., Lai, S., Gu, J. "A strong baseline and batch normalization neck for deep person re-identification." IEEE Transactions on Multimedia, vol. 22, no. 10, pp. 2597–2609, 2019.

[31] Milan, A., Leal-Taixé, L., Reid, I., Roth, S., Schindler, K. "MOT16: A benchmark for multi-object tracking." arXiv preprint arXiv:1603.00831, 2016.

[32] Naiel, M.A., Ahmad, M.O., Swamy, M., Lim, J., Yang, M.H. "Online multi-object tracking via robust collaborative model and sample selection." Computer Vision and Image Understanding, vol. 154, pp. 94–107, 2017.

[33] Pang, B., Li, Y., Zhang, Y., Li, M., Lu, C. "Tubetk: Adopting tubes to track multi-object in a one-step training model." In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 6308–6318, 2020.

[34] Pang, J., Qiu, L., Li, X., Chen, H., Li, Q., Darrell, T., Yu, F. "Quasi-dense similarity learning for multiple object tracking." In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 164–173, 2021.

[35] Peng, J., Wang, T., Lin, W., Wang, J., See, J., Wen, S., Ding, E. "Tpm: Multiple object tracking with tracklet-plane matching." Pattern Recognition, vol. 107, 107480, 2020.

[36] Perera, A.A., Srinivas, C., Hoogs, A., Brooksby, G., Hu, W. "Multi-object tracking through simultaneous long occlusions and split-merge conditions." In: 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06), vol. 1, pp. 666–673.

[37] Possegger, H., Mauthner, T., Roth, P.M., Bischof, H. "Occlusion geodesics for online multi-object tracking." In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1306–1313, 2014.

[38] Redmon, J., Farhadi, A. "Yolov3: An incremental improvement." arXiv preprint arXiv:1804.02767, 2018.

[39] Ren, S., He, K., Girshick, R., Sun, J. "Faster R-CNN: Towards real-time object detection with region proposal networks." Advances in Neural Information Processing Systems 28, 2015.

[40] Ristani, E., Solera, F., Zou, R., Cucchiara, R., Tomasi, C. "Performance measures and a data set for multi-target, multi-camera tracking." In: European Conference on Computer Vision, pp. 17–35. Springer, 2016.

[41] Shao, S., Zhao, Z., Li, B., Xiao, T., Yu, G., Zhang, X., Sun, J. "Crowdhuman: A benchmark for detecting human in a crowd." arXiv preprint arXiv:1805.00123, 2018.

[42] Stadler, D., Beyerer, J. "On the performance of crowd-specific detectors in multi-pedestrian tracking." In: 2021 17th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), pp. 1–12.

[43] Stadler, D., Beyerer, J. "Modelling ambiguous assignments for multi-person tracking in crowds." In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 133–142, 2022.

[44] Sun, P., Cao, J., Jiang, Y., Yuan, Z., Bai, S., Kitani, K., Luo, P. "Dancetrack: Multi-object tracking in uniform appearance and diverse motion." In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 20993–21002, 2022.

[45] Sun, P., Cao, J., Jiang, Y., Zhang, R., Xie, E., Yuan, Z., Wang, C., Luo, P. "Transtrack: Multiple object tracking with transformer." arXiv preprint arXiv:2012.15460, 2020.

[46] Tokmakov, P., Li, J., Burgard, W., Gaidon, A. "Learning to track with object permanence." In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 10860–10869, 2021.

[47] Wang, B., Wang, G., Chan, K.L., Wang, L. "Tracklet association by online target-specific metric learning and coherent dynamics estimation." IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 3, pp. 589–602, 2016.

[48] Wang, G., Gu, R., Liu, Z., Hu, W., Song, M., Hwang, J.N. "Track without appearance: Learn box and tracklet embedding with local and global motion patterns for vehicle tracking." In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 9876–9886, 2021.

[49] Wang, G., Wang, Y., Gu, R., Hu, W., Hwang, J.N. "Split and connect: A universal tracklet booster for multi-object tracking." IEEE Transactions on Multimedia, pp. 1–1, 2022. fff

[50] Wang, G., Wang, Y., Zhang, H., Gu, R., Hwang, J.N. "Exploit the connectivity: Multi-object tracking with trackletnet." In: Proceedings of the 27th ACM International Conference on Multimedia, pp. 482–490, 2019.

[51] Wang, Q., Zheng, Y., Pan, P., Xu, Y. "Multiple object tracking with correlation learning." In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 3876–3886, 2021.

[52] Wang, Z., Zheng, L., Liu, Y., Li, Y., Wang, S. "Towards real-time multi-object tracking." In: European Conference on Computer Vision, pp. 107–122. Springer, 2020.

[53] Weng, X., Wang, J., Held, D., Kitani, K. "3d multi-object tracking: A baseline and new evaluation metrics." In: 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 10359–10366.

[54] Williams, C., Rasmussen, C. "Gaussian processes for regression." Advances in Neural Information Processing Systems 8, 1995.

[55] Wojke, N., Bewley, A., Paulus, D. "Simple online and realtime tracking with a deep association metric." In: 2017 IEEE International Conference on Image Processing (ICIP), pp. 3645–3649.

[56] Wu, J., Cao, J., Song, L., Wang, Y., Yang, M., Yuan, J. "Track to detect and segment: An online multi-object tracker." In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 12352–12361, 2021.

[57] Xu, Y., Osep, A., Ban, Y., Horaud, R., Leal-Taixé, L., Alameda-Pineda, X. "How to train your deep multi-object tracker." In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 6787–6796, 2020.

[58] Yang, F., Chang, X., Sakti, S., Wu, Y., Nakamura, S. "Remot: A model-agnostic refinement for multiple object tracking." Image and Vision Computing, vol. 106, 104091, 2021.

[59] Yu, E., Li, Z., Han, S., Wang, H. "Relationtrack: Relation-aware multiple object tracking with decoupled representation." IEEE Transactions on Multimedia, pp. 1–1, 2022. fff

[60] Yu, F., Li, W., Li, Q., Liu, Y., Shi, X., Yan, J. "Poi: Multiple object tracking with high performance detection and appearance feature." In: European Conference on Computer Vision, pp. 36–42. Springer, 2016.

[61] Zeng, F., Dong, B., Zhang, Y., Wang, T., Zhang, X., Wei, Y. "Motr: End-to-end multiple-object tracking with transformer." In: Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXVII, pp. 659–675. Springer, 2022.

[62] Zhang, S., Benenson, R., Schiele, B. "Citypersons: A diverse dataset for pedestrian detection." In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3213–3221, 2017.

[63] Zhang, Y., Sun, P., Jiang, Y., Yu, D., Weng, F., Yuan, Z., Luo, P., Liu, W., Wang, X. "Bytetrack: Multi-object tracking by associating every detection box." In: Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXII, pp. 1–21. Springer, 2022.

[64] Zhang, Y., Wang, C., Wang, X., Zeng, W., Liu, W. "Fairmot: On the fairness of detection and re-identification in multiple object tracking." International Journal of Computer Vision, vol. 129, no. 11, pp. 3069–3087, 2021.

[65] Zheng, L., Bie, Z., Sun, Y., Wang, J., Su, C., Wang, S., Tian, Q. "Mars: A video benchmark for large-scale person re-identification." In: European Conference on Computer Vision, pp. 868–884. Springer, 2016.

[66] Zhou, X., Koltun, V., Krähenbühl, P. "Tracking objects as points." In: European Conference on Computer Vision, pp. 474–490. Springer, 2020.

[67] Zhu, Y., Zhou, K., Wang, M., Zhao, Y., Zhao, Z. "A comprehensive solution for detecting events in complex surveillance videos." Multimedia Tools and Applications, vol. 78, no. 1, pp. 817–838, 2019.

**ACKNOWLEDGMENT**

This work was done within the framework of the Computer Vision course, thanks to the expert guidance and steadfast support of the lecturer Vardan Sahakyan and the seminarist Armen Sardaryan.