

ЛИТЕРАТУРЕН ОБЗОР НА ДЪЛБОКИ НЕВРОННИ МРЕЖИ ЗА ПРОСЛЕДЯВАНЕ НА ОБЕКТИ ВЪВ ВИДЕОИЗОБРАЖЕНИЯ

Цветислава Лавчиева

Технически университет – Габрово, ул. „Хаджи Димитър“ 4, Габрово, България
*кореспондиращ автор: tvetislava.lavchieva@edu.mon.bg

LITERATURE REVIEW OF DEEP NEURAL NETWORKS FOR OBJECT TRACKING IN VIDEO FRAMES

Tsvetislava Lavchieva

Technical University of Gabrovo, 4 Hadji Dimitar Street, Gabrovo, Bulgaria
*Corresponding author: tvetislava.lavchieva@edu.mon.bg

Abstract

This research explores the application of deep neural networks for object tracking in video frames. The main objective is to analyze and compare state-of-the-art multi-object tracking algorithms such as DeepSORT, ByteTrack, and FairMOT in terms of their accuracy, speed, and efficiency across various video scenarios. The study involves key stages including object detection, trajectory association, and evaluation of tracking stability under challenging conditions such as illumination changes, camera motion, and object occlusions. The study highlights the potential of deep learning-based tracking methods for practical applications in video surveillance, intelligent transportation systems, autonomous systems, and behavioral analysis.

Keywords: deep neural networks, object tracking algorithms, artificial intelligence

ВЪВЕДЕНИЕ

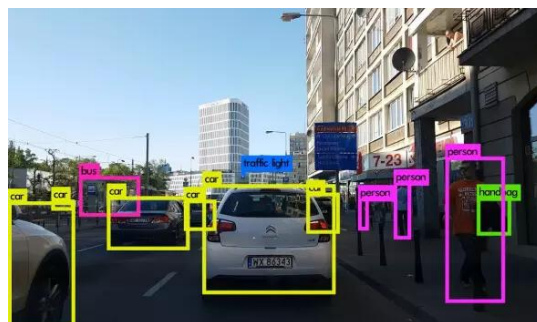
Проследяването на обекти във видео-изображения е една от основните задачи в компютърното зрение и намира приложение в разнообразни области – от системи за видеонаблюдение и интелигентен транспорт до автономни превозни средства и медицинска диагностика.

Целта не е просто да бъдат открити обектите във всеки кадър, а да се поддържа връзката между тях във времето, така че системата да може да разпознава движението им и да следи техните промени.

Съвременните методи за проследяване на обекти комбинират алгоритми за детекция и за асоцииране на траектории.

През последните години водеща роля

в този процес имат дълбоките невронни мрежи, които позволяват да се постигне висока точност и стабилност на системите дори в сложни ситуации – например при движение на камерата, частично закриване на обекти или промяна в осветеността (Фиг. 1).



Фиг. 1. Детекция на множество обекти в реална улична сцена [1].

Сред най-популярните алгоритми, използвани на практика, се открояват DeepSORT, ByteTrack и FairMOT, които успяват да намерят баланс между скорост и точност и са подходящи за реализация в реално време.

Въпреки това остава открит въпросът за тяхната оптимизация и сравнение при различни сценарии – именно върху това е насочен настоящият труд.

Целта на изследването е да се анализират и сравнят съвременни алгоритми за проследяване на обекти във видеоизображения, базирани на дълбоки невронни мрежи.

За постигане на тази цел са поставени следните задачи:

- Да се направи обзор на принципите и особеностите на алгоритмите DeepSORT, ByteTrack и FairMOT;
- Да се изследва тяхната ефективност при работа в реално време;
- Да се обсъдят техните предимства и ограничения в различни практически приложения.

Настоящото изследване има за цел да представи аналитичен поглед върху съвременните подходи за проследяване на обекти чрез дълбоки невронни мрежи, като подчертае техните силни страни и възможности за приложение в реални условия.

ИЗЛОЖЕНИЕ

Проследяването на обекти във видеоизображения може да се разглежда като логично продължение на задачата по откриване и разпознаване на обекти.

Докато детекцията се концентрира върху идентифицирането на обектите във всеки отделен кадър, проследяването има за цел да поддържа връзката между тях във времето и да определи движението им в последователността от кадри.

През последните години тази задача добива все по-голямо значение благодарение на развитието на дълбоките невронни мрежи.

Тяхното използване позволява да се съчетаят висока точност и стабилност на резултатите дори при реални, динамични условия.

Благодарение на това съвременните системи за проследяване вече могат да работят в ситуации, в които по-старите методи са имали сериозни затруднения – например при бързо движение, закриване или промяна на осветлението.

Днес алгоритмите за проследяване намират приложение в различни направления – от интелигентни системи за видеонаблюдение и управление на трафика до автономни превозни средства и медицински анализ на изображения.

В следващите раздели ще бъдат разглеждани както традиционните методи за проследяване, които поставят основите на областта, така и някои от най-използваните съвременни алгоритми, базирани на дълбоко обучение и пригодени за работа в реално време.

ПРОСЛЕДЯВАНЕ НА ОБЕКТИ – СЪЩНОСТ И ПРЕДИЗВИКАТЕЛСТВА

Проследяването на обекти във видеоизображения е една от основните задачи в компютърното зрение.

То е насочено към откриване, локализиране и проследяване на обекти през последователни кадри, така че да се формират коректни и непрекъснати траектории на движение.

За разлика от детекцията, която идентифицира обектите само в отделни кадри, проследяването има за цел да осигури тяхното разпознаване и съответствие във времето, дори когато сцената се променя.

Основните предизвикателства при изпълнение на тази задача [2] са свързани с влиянието на редица фактори, които могат да намалят точността на системата:

- Промени в осветеността – видео сцените често са динамични, а различните нива на светлина и сенки могат да променят видимостта на обектите;

- Закриване на обекти – когато един обект частично или напълно се скрие зад друг или напусне кадъра;
- Бързо движение – при висока скорост обектите могат да се размазват или изкривяват, което затруднява откриването им;
- Голям брой обекти – при сцени с много участници асоциирането на траектории става по-сложно и податливо на грешки;
- Движение на камерата – при динамично заснемане се появяват фонове измествания и шум, които влияят на стабилността на проследяването.

Всички тези фактори налагат използването на надеждни алгоритми, които съчетават висока точност, бърза реакция и устойчивост в реални условия на работа.

ТРАДИЦИОННИ МЕТОДИ ЗА ПРОСЛЕДЯВАНЕ НА ОБЕКТИ

Преди навлизането на дълбоките невронни мрежи проследяването на обекти се е извършвало чрез класически алгоритми от областта на компютърното зрение.

Сред най-използваните методи е оптичният поток (Optical Flow) [3], който оценява движението на обектите, като анализира промените в яркостта между последователни кадри.

Друг широко приложим подход е филтърът на Калман (Kalman Filter) [4], използван за предсказване на бъдещото положение на обектите чрез вероятностен модел, който отчита шумове и несигурности в измерванията.

Популярен и сравнително прост за реализация е алгоритъмът Mean-Shift [5], както и неговото адаптивно разширение CamShift [6].

Те работят чрез намиране на област с най-голяма плътност на пиксели по даден признак (например цвят), като позицията на търсещия прозорец се актуализира многократно в последователни итерации, докато се достигне най-доброто съвпадение.

Тези методи поставят основите на днешните решения и са важна стъпка в развитието на проследяването на обекти.

Въпреки това те показват ограничена устойчивост при по-динамични сцени – например при закриване на обекти, промени в осветеността или движение на камерата.

Поради тези ограничения през последните години вниманието постепенно се насочва към подходи, които използват дълбоки невронни мрежи.

Те комбинират процесите на детекция и проследяване, осигурявайки значително по-висока точност и гъвкавост в реални условия.

В настоящото изследване акцентът е поставен именно върху съвременните алгоритми, които използват дълбоко обучение и демонстрират отчетлив напредък спрямо класическите подходи.

СЪВРЕМЕННИ АЛГОРИТМИ ЗА ПРОСЛЕДЯВАНЕ НА ОБЕКТИ

Подходите, базирани на дълбоки невронни мрежи, обединяват процесите на детекция и асоцииране на траектории, което им позволява да постигат висока точност и стабилност дори в сложни сцени с множество обекти, закриване или движение на камерата.

Благодарение на тези предимства те се превръщат в основен инструмент при разработването на съвременни системи за автоматично видеоанализиране.

В тази част са разгледани три от най-популярните съвременни алгоритъма за многократно проследяване на обекти, изградени на базата на дълбоки невронни мрежи.

Алгоритъм DeepSORT за проследяване на обекти с невронни мрежи

Алгоритъмът DeepSORT (Deep Simple Online and Realtime Tracking) [7] е сред най-разпространените методи за многократно проследяване на обекти, използващи дълбоко обучение.

Той комбинира няколко ключови етапа: първо, обектите се откриват във все-

ки кадър чрез детектор, след това филтър на Калман прогнозира тяхното бъдещо местоположение, и накрая асоциирането на траекториите се извършва чрез допълнителна невронна мрежа за повторна идентификация (ReID).

Тази допълнителна мрежа извлича уникални характеристики за всеки обект, което позволява системата да разпознава кой е кой, дори при частично закриване или когато множество обекти се движат близо един до друг.

Благодарение на тази комбинация DeepSORT постига значително по-висока устойчивост в сравнение с традиционните методи и намира приложение в разнообразни области – от видеонаблюдение и интелигентен транспорт до автономни превозни средства и спортен анализ.

Сред основните предимства на алгоритъма се открояват високата точност и надеждност при реални условия, но тези качества идват за сметка на по-големи изчислителни изисквания, породени от ReID модула, както и зависимостта му от качеството на използвания детектор за обекти.

Алгоритъм ByteTrack за ефективно многократно проследяване на обекти

ByteTrack е един от най-съвременните и ефективни подходи за многократно проследяване на обекти.

Той се отличава с отличен баланс между точност и изчислителна ефективност [8], което го прави особено подходящ за системи, работещи в реално време.

Основната идея зад алгоритъма е проста, но изключително резултатна – вместо да използва само детекциите с висока достоверност (както правят повечето класически методи), ByteTrack взема предвид и тези с по-ниска такава, които други системи обикновено отхвърлят.

Чрез внимателно асоцииране на всички налични обекти алгоритъмът успява да запази по-пълна информация за дви-

жението и да намали случаите на изгубени траектории.

Този подход значително повишава устойчивостта на системата при ситуации като частично закриване на обекти, промяна в осветлението или рязко движение на камерата.

Благодарение на това ByteTrack демонстрира отлична стабилност в реални приложения – например при видеонаблюдение или в автономни системи, където дори краткото „изчезване“ на обект може да доведе до грешки в анализа.

Комбинацията между висока скорост, точност и ниски хардуерни изисквания прави ByteTrack (Фиг.2) предпочитан избор за мащабни решения за проследяване в реално време.



Фиг. 2. Последователност от три кадъра (t_1 , t_2 , t_3), при която ByteTrack използва както високодостоверни, така и нискодостоверни детекции за непрекъснато проследяване на обекта [8].

Алгоритъм FairMOT за обединяване на детекция и повторна идентификация

FairMOT представлява съвременен подход, който се отличава със своята уникална архитектура за едновременно извършване на детекция и повторна идентификация на обекти в рамките на една обща невронна мрежа [9].

За разлика от повечето методи, които разглеждат тези две задачи поотделно и ги изпълняват последователно, FairMOT ги обединява в споделено представяне.

Това позволява да се постигне подобър баланс между точност и скорост при работа в реални условия.

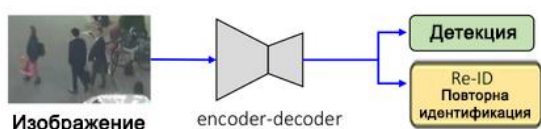
Ключова особеност на алгоритъма е неговият двоен изход – единият служи за локализация и класификация на обектите, а другият генерира уникални вектор-

ни характеристики, използвани за повторна идентификация.

Благодарение на това решение моделът преодолява проблема с неравномерното представяне на обектите и осигурява по-“справедливо” асоцииране на траекториите, дори когато обектите се застъпват или временно се закриват.

Тази архитектура прави FairMOT изключително ефективен при сцени с множество движещи се обекти, като съчетава висока точност и необходимата скорост за приложения в реално време (Фиг. 3).

Алгоритъмът намира приложение в широк кръг от области – от интелигентни системи за видеонаблюдение и транспортно управление до анализ на човешко поведение в обществени пространства.



Фиг. 3. Схематично представяне на алгоритъма FairMOT [8].

За по-ясна систематизация на разгледаните съвременни алгоритми е представен обобщен сравнителен анализ, включващ техните основни характеристики, особености и приложения (Табл. 1).

Табл. 1. Обобщено сравнение на съвременни алгоритми за проследяване на обекти

Алгоритъм	Основни характеристики	Особености и приложение
DeepSORT	Комбиниращ детекция, филтър на Калман и мрежа за повторна идентификация (Re-ID)	Висока точност при закриване на обекти; подходящ за видеонаблюдение и транспортни системи
ByteTrack	Използва както високо-, така и ниско-достовърни детекции за по-пълно асоцииране	Осигурява стабилно проследяване при динамични сцени; ефективен за работа в реално време
FairMOT	Обединява детекция и Re-ID в една невронна мрежа	Добър баланс между точност и скорост; приложим при множество обекти и сложни среди

През последните години се появяват усъвършенствани варианти на вече съществуващите алгоритми, които надграждат основните принципи на DeepSORT, ByteTrack и FairMOT.

Сред тях се откроява OC-SORT (Observation-Centric SORT) [10], който подобрява процеса на асоцииране на обекти, като използва по-надеждни наблюдения вместо изцяло предсказани позиции.

Друг популярен метод е BoT-SORT (Bag of Tricks for SORT) [11], който включва набор от оптимизации, насочени към повишаване на точността при идентификацията на обектите.

Съвременни решения като StrongSORT++ [12] и ByteTrack++ [13] комбинират предимствата на предходните методи, като добавят по-добро управление на визуалните характеристики и по-висока адаптивност към различни сцени и среди.

Тези подходи бележат нов етап в развитието на системите за многократно проследяване, демонстрирайки още по-добър баланс между скорост и точност при работа в реално време.

За оценка на ефективността на алгоритмите обикновено се използват публично достъпни набори от данни, като MOT16, MOT17 и MOT20, които съдържат реални видеопоследователности с множество обекти и техните анотирани траектории.

В някои случаи се прилагат и допълнителни набори, например KITTI Tracking, предназначен за автомобилни сцени, и VisDrone, включващ въздушни видеа с голямо разнообразие от движения и мащаби.

Изпълнението на алгоритмите обикновено се извършва в среди, базирани на Python, с използване на библиотеки като PyTorch, TensorFlow и OpenCV, които предоставят необходимите инструменти за обработка на видео, обучение на модели и визуализация на резултатите.

За обективна оценка на производителността на алгоритмите се използват няколко стандартни метрики.

Най-разпространената сред тях е MOTA (Multiple Object Tracking Accuracy), която отчита общата точност на проследяването, като съчетава броя на пропуснатите, неправилно асоциираните и излишно създадените траектории.

Показателят IDF1 (Identification F1 Score) измерва степента, до която проследените траектории съответстват на реалните идентичности на обектите,

а FPS (Frames per Second) отразява скоростта на обработка и позволява да се оцени възможността за работа в реално време.

Тези метрики осигуряват унифициран подход за сравнение на различните алгоритми и се използват в повечето съвременни изследвания и състезания по многократно проследяване на обекти.

Проведеният сравнителен анализ в рамките на настоящото изследване цели да послужи като основа за избора на оптимален алгоритъм, който ще бъде внедрен в разработвания софтуерен прототип на следващ етап от проекта.

ЗАКЛЮЧЕНИЕ

В настоящия доклад беше разгледан проблемът с проследяването на обекти във видеоизображения и основните предизвикателства, които го съпътстват – закриване на обекти, промени в осветеността и движение на камерата.

Бяха анализирани класическите методи Optical Flow, филтърът на Калман, както и алгоритмите Mean-Shift и CamShift, които имат съществен принос за развитието на тази област.

Въпреки това те показват ограничения при работа в динамични и комплексни среди, където сцената и условията се променят бързо.

Развитието на методите за дълбоко обучение открива нови възможности за по-ефективно многократно проследяване.

Алгоритмите DeepSORT, ByteTrack и FairMOT комбинират детекция и асоци-

иране на траектории с помощта на невронни мрежи и постигат значително по-висока точност и устойчивост в реални условия.

Анализът на техните характеристики показва, че всеки от разгледаните подходи има свои силни страни - DeepSORT се отличава със стабилност при закриване на обекти, ByteTrack е особено ефективен при детекции с по-ниска достоверност, а FairMOT осигурява едновременно детекция и повторна идентификация в рамките на една мрежа.

Следващият етап от изследването предвижда избор на най-подходящ алгоритъм и изграждане на софтуерен прототип, който да демонстрира приложимостта на подхода в реални условия.

Това ще позволи по-задълбочена експериментална оценка и ще подпомогне избора на най-ефективните решения за практическа реализация.

Извършеният анализ представлява подготвителен етап от дисертационната работа и ще послужи като основа за създаване на реална система за проследяване на обекти във видеоизображения.

Източник на финансиране: Настоящият документ е изготвен с финансовата помощ на договор № НИП2025-9 за провеждане на научни изследвания по проект на тема: „Изследване и изграждане на софтуерно дефинирани мрежи” към Технически университет – Габрово.

ЛИТЕРАТУРА

- [1] Kathuria, A. (2025, October 24). *How to implement a YOLO (v3) object detector from scratch in PyTorch: Part 5*. Paperspace Blog. <https://blog.paperspace.com/how-to-implement-a-yolo-v3-object-detector-from-scratch-in-pytorch-part-5/>
- [2] Smeulders, A., Chu, D., Cucchiara, R., Calderara, S., Dehghan, A., & Shah, M. (2014). Visual tracking: An experimental survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(4), 685–707.

- [3] Horn, B. K., & Schunck, B. G. (1981). Determining optical flow. *Artificial Intelligence*, 17(1–3), 185–203.
- [4] Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Transactions of the ASME – Journal of Basic Engineering*, 82(1), 35–45.
- [5] Comaniciu, D., & Meer, P. (2002). Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5), 603–619.
- [6] Bradski, G. R. (1998). Computer vision face tracking for use in a perceptual user interface. *Intel Technology Journal*, Q2, 1–15.
- [7] Wojke, N., Bewley, A., & Paulus, D. (2017). Simple online and real-time tracking with a deep association metric (DeepSORT). *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, 3645–3649.
- [8] Zhang, Y., Sun, P., Jiang, Y., Yu, D., Weng, X., Yuan, Z., Luo, P., Liu, W., & Wang, X. (2021). ByteTrack: Multi-object tracking by associating every detection box. *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 1–9.
- [9] Zhang, Y., Wang, C., Wang, X., Zeng, W., & Liu, W. (2020). FairMOT: On the fairness of detection and re-identification in multiple object tracking. *International Journal of Computer Vision*, 129(11), 3069–3087.
- [10] Cao, J., Weng, X., & Kim, K. (2023). OC-SORT: Observation-centric SORT for robust multi-object tracking. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 1–10.
- [11] Aharon, N., Orfaig, R., & Bobrovsky, B. (2022). BoT-SORT: Robust associations for multi-pedestrian tracking. *arXiv preprint arXiv:2206.14651*.
- [12] Du, Y., Chen, K., & Sun, Y. (2023). StrongSORT++: Make DeepSORT great again. *arXiv preprint arXiv:2302.05046*.
- [13] Zhang, Y., Weng, X., & Wang, X. (2023). ByteTrack++: Better, faster, stronger multi-object tracking. *arXiv preprint arXiv:2305.14572*.