МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное автономное
образовательное учреждение высшего образования
«Национальный исследовательский университет ИТМО»
(Университет ИТМО)

Международный научно-образовательный центр
Физики наноструктур

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА
GRADUATION THESIS

по теме:
ИССЛЕДОВАНИЕ АЛГОРИТМОВ КОМПЬЮТЕРНОГО ЗРЕНИЯ ДЛЯ
ОТСЛЕЖИВАНИЯ ПОДВОДНЫХ ОБЪЕКТОВ

Студент:
Группа № R34372   *Нгуен Тоан*

Руководитель:
доцент, кандидат технических наук   *Шаветов Сергей Васильевич*

Санкт-Петербург 2025

СОДЕРЖАНИЕ

INTRODUCTION

Computer vision is one of the most rapidly developing fields in modern artificial intelligence, with applications spanning a wide range of tasks, from autonomous driving to medical diagnostics and industrial automation systems. One of the critical and complex challenges researchers face is object tracking in underwater environments. Underwater conditions are characterized by complex optical distortions, low image contrast, uneven lighting, and various noise effects, which significantly complicate the process of video data processing and analysis.

The relevance of this topic is driven by the need to create reliable monitoring systems and manage underwater robotic complexes, as well as to improve methods for marine security and scientific research of ocean resources. The application of modern machine learning and neural network techniques to solve the problem of object tracking in underwater environments allows for enhanced accuracy and robustness of algorithms, which opens up new opportunities for the practical implementation of developed solutions.

The goal of this work is to investigate and develop computer vision algorithms capable of efficiently tracking underwater objects under significant visual distortions. To achieve this goal, several tasks need to be accomplished: conducting an analytical review of existing methods and algorithms used for tracking objects in complex visual conditions; preparing and analyzing the UOT100 dataset, which contains over 100 video sequences under various underwater conditions; evaluating the effectiveness of classical machine learning methods compared to modern neural network approaches, including algorithms based on correlation filters and Siamese networks; and developing software in Python using libraries such as OpenCV, PyTorch, and TensorFlow to implement and compare the selected algorithms.

The structure of the work includes an introduction that justifies the relevance of the topic and sets the research objectives, an overview of existing solutions, a theoretical part dedicated to studying the principles of computer vision algorithms, a description of the experimental setup and analysis conducted, as well as conclusions and recommendations for further development of the topic. The research carried out will not only identify the strengths and weaknesses of existing

approaches but also propose directions for their improvement in the context of underwater applications.

Thus, the work aims to address practical problems related to improving the efficiency of object tracking systems in extreme conditions, which is of great significance for the development of autonomous underwater technologies and the implementation of new solutions in marine safety and monitoring.

1   PROBLEM STATEMENT

Tracking underwater objects, which refers to the continuous and accurate estimation of a target's position within a video sequence captured underwater, is a critical task with significant implications for marine applications. These applications span across environmental monitoring, resource development, and ecological protection [1], where precise tracking can aid in assessing marine ecosystems, monitoring aquatic life, and ensuring the sustainability of ocean resources. However, the unique characteristics of the underwater environment introduce substantial challenges that complicate the task of object tracking in this setting.

One of the foremost challenges is the issue of **water turbidity and blurring effects**. Underwater optical images are frequently affected by turbidity, which results in significant blurring and fog. This occurs due to the absorption and scattering of light by seawater and various suspended particles, including minerals, salt, sand, and plankton [2][3][4]. The turbid nature of water severely limits visibility, introducing haze that further complicates the clarity of captured images. Additionally, turbulence in the water flow can cause motion blur, which further disrupts the tracking process and reduces the precision of object localization [4].

Another obstacle faced in underwater tracking is **color attenuation and light scattering**. As light travels through water, different wavelengths are absorbed at varying rates, with red light being absorbed more rapidly than other wavelengths [3]. This phenomenon leads to the attenuation of color, which distorts the natural hues of objects and affects the overall image quality. The scattering of light by suspended particles further diminishes light intensity and alters its direction, resulting in color distortion that can make it even more difficult to distinguish between different objects. This attenuation of color and reduction in light intensity contribute to the overall decolorization of underwater scenes, posing a significant challenge for accurate object tracking [3][4][5].

In addition to these optical challenges, **low-contrast targets and occlusion** are persistent problems in underwater environments. Low visibility, low contrast, and low light intensity are common characteristics of underwater images, making it particularly difficult to discern objects and extract meaningful features for tracking [2][4]. Moreover, many marine species exhibit camouflage, blending

seamlessly with their surroundings, which complicates the process of segmentation and accurate identification [3]. Furthermore, occlusion, where one object partially or fully obscures another, presents a substantial problem. This can significantly hinder the ability to maintain a consistent track of the target, especially in dynamic environments where objects frequently move in and out of view [2][3][6].

The underwater environment is also characterized by **dynamic backgrounds with similar-looking objects**, adding another layer of complexity. Marine organisms exhibit a wide range of sizes and shapes, often appearing in dense clusters or schools, such as schools of fish. The presence of these similar-looking objects in the background can lead to confusion, as tracking algorithms may mistakenly identify these objects as the target, causing errors and drift in the tracking process [7]. The complexity and dynamic nature of the underwater scene further exacerbate the challenges faced by tracking algorithms.

Traditional object tracking methods, primarily designed and optimized for open-air environments, struggle to perform effectively in underwater settings [3]. These conventional methods are not robust enough to handle the significant domain shifts introduced by the optical properties of water, such as color distortion, image blur, and light scattering [3]. Moreover, mainstream tracking models can be computationally expensive, often requiring high computational power that may not be available on resource-constrained underwater edge devices [1]. The inherent challenges of target feature distortion and contrast attenuation in underwater scenes can further undermine the ability of general tracking algorithms to differentiate between the target and its background, reducing their reliability. Therefore, specialized tracking models that take into account the unique characteristics of the underwater environment, along with underwater-specific datasets, are essential for improving the accuracy and reliability of object tracking in such challenging conditions [3].

2 OVERVIEW OF EXISTING SOLUTIONS

2.1 Correlation Filter-Based Tracking Methods

Correlation Filter (CF)-Based Tracking has emerged as a prominent approach in visual object tracking, known for achieving a good balance between tracking accuracy and speed [8]. These methods utilize a dynamic model to track the same target across consecutive video frames [9]. The core idea revolves around learning a correlation filter in the frequency domain to efficiently locate the target in subsequent frames [10].
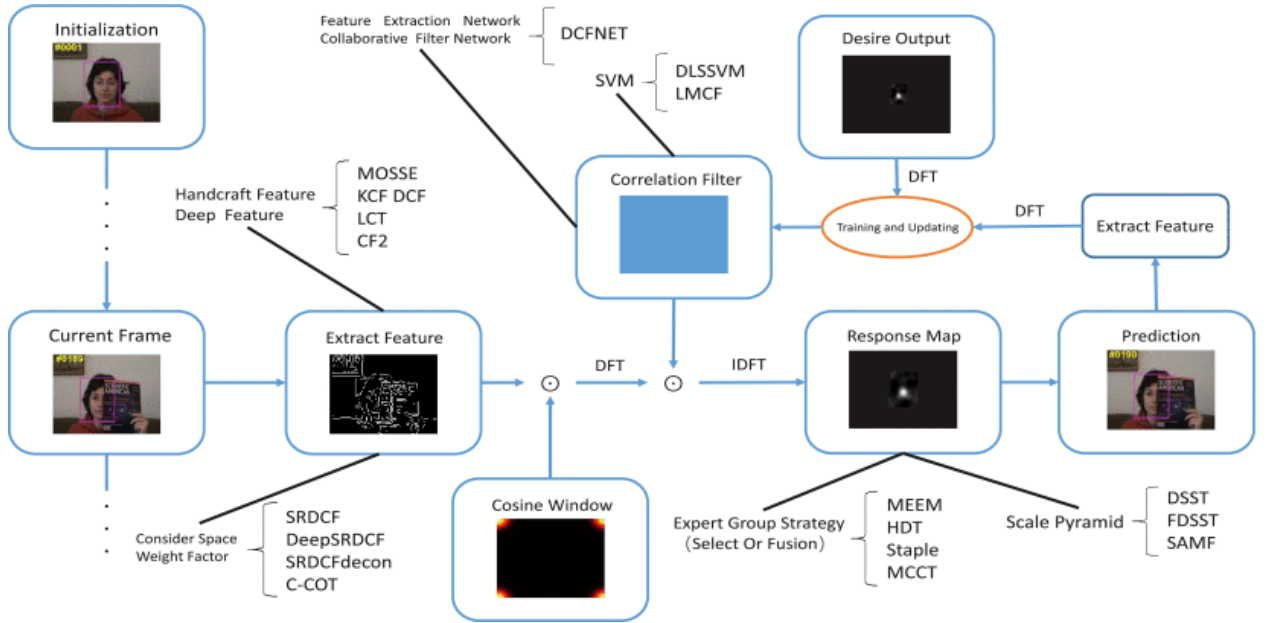

Рисунок 1 — General framework for correlation-filter-based object tracking [9]

The general framework for correlation filter-based object tracking typically involves the following steps [9][10]:

1. **Feature Extraction**: In the initial frame, features are extracted from an image block at the target's location [9] [10]. These features can include handcrafted features like grayscale, Histogram of Oriented Gradients (HOG) [10][9], and color names (CN) [11], or deep features extracted from Convolutional Neural Networks (CNNs) [9][10][11] . A cosine window is often applied to reduce boundary effects [9].

2. **Filter Training:** A correlation filter is learned based on the extracted features and a desired output, often a Gaussian-shaped response map

centered at the target [8][9] . This learning process aims to create a filter that produces a strong correlation response when convolved with the target and a weak response elsewhere [1][9]. Techniques like the Minimum Output Sum of Squared Error (MOSSE) algorithm represent early and straightforward approaches to filter learning [8] [9] [10]. Kernel Correlation Filters (KCF) further enhance this by utilizing a cyclic matrix to construct training samples and employing kernel functions to handle non-linear features [9][10][11]. Dual Correlation Filters (DCF) represent another early development [9].

3. **Target Localization:** In subsequent frames, a search window centered around the previous target location is cropped, and features are extracted [10]. The learned correlation filter is then convolved with these features to generate a response map [9][10]. The location of the maximum response in this map indicates the new position of the target [9][10]. The response map's characteristics, such as its peak value and shape, can also be used for tasks like failure detection [11].

4. **Filter Update:** To adapt to changes in the target's appearance and the environment, the correlation filter is typically updated online with new information from the tracked target in the current frame [1][9][10]. The learning rate of this update can be adaptively determined based on tracking reliability [9].

Correlation filter-based object tracking methods have undergone significant advancements and can be categorized based on several key characteristics [9]. Early approaches utilized categorized features such as grayscale information, exemplified by the Minimum Output Sum of Squared Error (MOSSE) tracker [8][10][11]. Subsequent developments incorporated more sophisticated handcrafted features like Histogram of Oriented Gradients (HOG) and color names [3][7][8][11]. The integration of deep convolutional neural network (CNN) features has further enhanced performance by providing robust and discriminative representations [8][9][11]. Some trackers leverage a combination of handcrafted and deep features to capitalize on their complementary strengths [7][9][11].

Another crucial categorization factor is the consideration of the spatial context through space weight factors. Spatially Regularized Discriminative Correlation Filters (SRDCF) were introduced to mitigate boundary effects

by penalizing filter coefficients outside the target region [9][8][11]. Variants like DeepSRDCF combine SRDCF with deep features [8][9], and CSRDCF incorporates channel and spatial reliability [8][9][10]. Dynamic Saliency-Aware Regularized CF Tracking (DSAR-CF) refines spatial regularization by integrating object saliency information, allowing the weight map to adapt to shape variations [8][9][11]. Scale factors represent another important category. Discriminative Scale Space Tracking (DSST) decouples translation and scale estimation to accurately determine the object's size [7][8][11].

Expert strategies involve utilizing multiple filters or cues and selecting the most reliable one for tracking. Multiexpert Entropy Minimization (MEEM) based on Support Vector Machines (SVM) is an example of this approach [9][12]. Methods employing decision-level fusion also fall under this category, combining the outputs of multiple experts based on different features or filter types [11][3]. Large Margin Object Tracking with Circulant Feature Maps (LMCF) combines correlation filters with structured SVM for robust tracking [8][9].

Recent advancements include the development of datasets tailored for specific domains, such as WebUOT-1M for underwater object tracking [7]. The introduction of Transformer-based architectures in tracking has also shown promising results [7]. Furthermore, the integration of motion estimation and failure correction modules has been explored to enhance tracking robustness in challenging scenarios, such as in satellite videos [11]. The Walsh-Hadamard transform (WHT) has also been investigated for feature extraction in underwater object tracking within a particle filter framework [5]. These categorizations and advancements highlight the continuous efforts to improve the accuracy, robustness, and efficiency of correlation filter-based object tracking algorithms for diverse applications [8][9][10][11].

Correlation filter-based object tracking has evolved significantly with the introduction of various methods that enhance accuracy, robustness, and efficiency. Below is a detailed categorization and description of specific CF-based trackers, highlighting their key features and advancements.

Specific CF-Based Trackers:

– MOSSE (Minimum Output Sum of Squared Error) [11]: An early, fast tracker using grayscale features.

– KCF (Kernelized Correlation Filters) [10]: Utilizes HOG features and kernel functions for improved accuracy and robustness.
– DCF (Dual Correlation Filters) [8]: An early approach based on correlation filters.
– SRDCF (Spatially Regularized Discriminative Correlation Filters) [8]: Addresses boundary effects using spatial regularization.
– DeepSRDCF [10]: Combines deep features with SRDCF
– C-COT (Continuous Convolution Operators for Tracking) [9]: Employs continuous convolution operators and deep features, achieving high accuracy.
– ECO (Efficient Convolution Operators) [8]: An efficient version of C-COT with a focus on speed and reduced overfitting.
– DSST (Discriminative Scale Space Tracking) [8]: Specifically designed for accurate scale estimation.
– SAMF (Scale Adaptive with Multiple Features) [11]: Integrates multiple features and scale estimation.
– MCCTH (Multicue Correlation Tracker-Based Handcrafted Feature) [9]: Uses multiple handcrafted feature experts.
– LMCF (Large Margin With Circulant Feature Maps Tracker) [9]: Combines correlation filters with structured SVM.
– DCFNET (Discriminant Correlation Filters Network) [9]: An end-to-end trainable network for CF tracking.
– STRCF (Spatial-Temporal Regularized Correlation Filter) [11]: Incorporates both temporal and spatial regularization.
– MACF (Motion-Aware Correlation Filter) [11]: Integrates motion estimation for satellite video tracking.
– CSR-DCF (Discriminative Correlation Filter with Channel and Spatial Reliability) [11]: Utilizes channel reliability and spatial confidence.
– BACF (Background-Aware Correlation Filters) [10]: Learns from real negative background examples.
– DSAR-CF (Dynamic Saliency-Aware Regularized CF Tracking) [8]: Uses dynamic saliency-aware regularization.

Correlation filter-based trackers exhibit several notable advantages and limitations. One of the primary strengths of these methods lies in their ability

to achieve a favorable balance between tracking accuracy and computational efficiency. This efficiency is largely attributed to the utilization of the Fast Fourier Transform (FFT), which enables rapid convolution operations in the frequency domain [8]. Additionally, these trackers are capable of learning discriminative filters that effectively distinguish the target from its surroundings [1].

However, traditional correlation filter-based trackers also face several challenges. Trackers relying on shallow features are particularly susceptible to environmental factors such as background clutter, occlusions, variations in illumination, and target deformations [8][9][10]. Furthermore, the inherent circular shift operation in these methods can lead to undesirable boundary effects, although spatial regularization techniques have been developed to address this issue [8]. Another limitation is the sensitivity of these trackers to hyperparameter settings, which often require careful tuning to achieve optimal performance [9].

While correlation filter trackers offer a balance of efficiency and robustness, ongoing research continues to explore avenues for improvement, including the design of more effective filters, the fusion of complementary features, the development of robust scale estimation and occlusion handling techniques, and the adaptation to specific application scenarios [9][10]. The integration of advanced deep learning architectures and attention mechanisms also represents a promising direction for future work [1].

2.2 Deep Learning-Based Tracking Methods

Object tracking based on object detection leverages the advancements in object detection algorithms to identify and track objects across video frames. By detecting objects in each frame and associating them over time, these methods provide a robust framework for tracking. Common deep learning-based object detection algorithms can be categorized into two-stage and single-stage algorithms based on convolutional neural networks (CNNs), as well as transformer-based object detection algorithms [2].

Two-stage object detection algorithms, such as Faster R-CNN, operate by first generating region proposals to identify potential object locations. These proposals are then classified and refined to achieve precise object detection and boundary estimation. This method is particularly effective in scenarios requiring

high accuracy, as it systematically narrows down the search space for object localization [2].

In contrast, **single-stage object detection algorithms**, such as YOLO and SSD, predict object classes and locations in a single step. This streamlined process offers a faster and more efficient approach, albeit sometimes at the cost of accuracy. These models, particularly YOLO, have demonstrated reliability in underwater object detection, where computational efficiency is often critical [13].
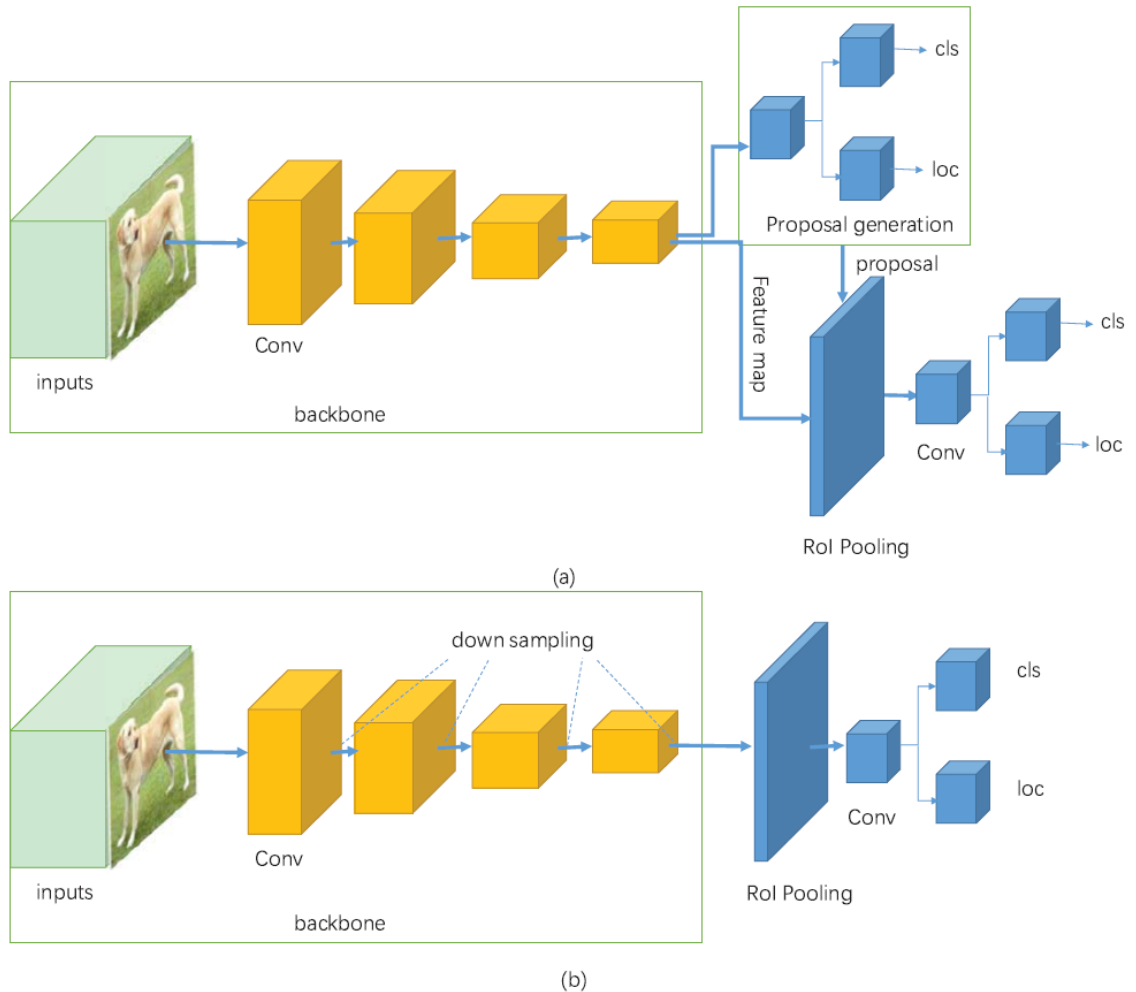

Рисунок 2 — (a) Exhibits the basic architecture of two-stage detectors, which consists of region proposal network to feed region proposals into classifier and regressor. (b) Shows the basic architecture of one-stage detectors, which predicts bounding boxes from input images directly [14].

Additionally, **transformer-based object detection algorithms**, such as DETR and RT-DETR, integrate transformers with convolutional neural networks (CNNs) to directly predict object categories and boundaries. This combination enhances efficiency and accuracy, making them a promising choice for modern object tracking applications [2].
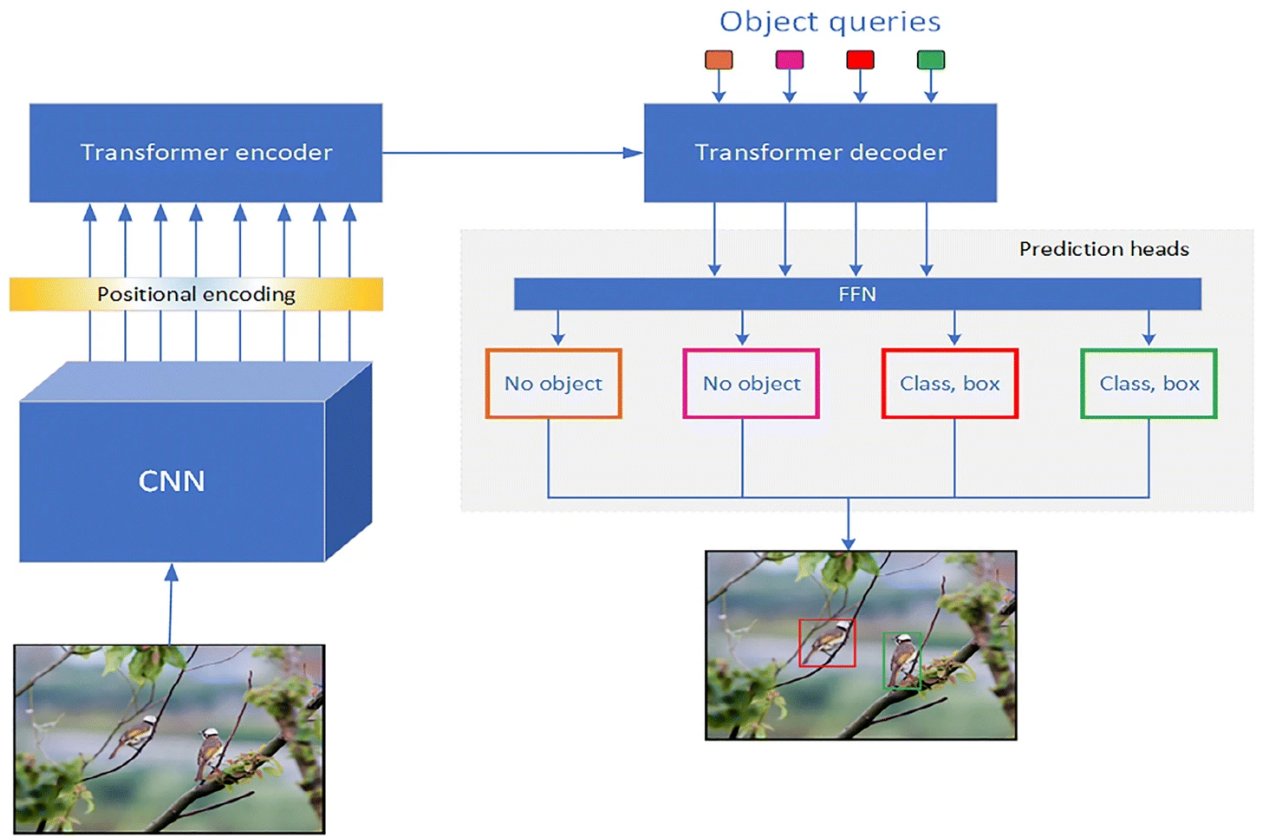
Рисунок 3 — The process of DETR and its structure [15].

Despite demonstrating strong performance in terrestrial environments, these algorithms encounter significant challenges in underwater object detection due to factors such as low light conditions and turbid backgrounds [1][2][6]. Underwater object tracking plays a critical role in applications such as marine resource exploration and military security [1]. Recent advancements in deep learning have garnered considerable attention for their potential in data-driven underwater image enhancement and object recognition [16].

Advanced Deep Learning Techniques Tailored for Underwater Object Tracking (UOT):

– **Knowledge Distillation (KD):** This involves efficiently training a smaller "student"network by learning from a larger, pre-trained "teacher"network. This approach has been applied to UOT, where a teacher model trained on massive open-air data can enhance the tracking performance of a student model dedicated to underwater frames [7]. Omni-knowledge distillation combines different types of distillation losses, such as token contrastive representation, similarity matrix, feature embeddings, and response maps distillation losses [7].

- **Motion-Aware Target Prediction (MATP):** To address model drift caused by similar distractors in underwater environments, a training-free MATP module based on Kalman filtering can be used [7]. This involves prediction and correction stages to estimate and refine the target's state. UOSTrack also uses motion-based post-processing to mitigate the influence of similar targets [1][7].

- **Boundary Attention and Sparse Feature Learning:** The DBSF network utilizes a differential boundary attention distribution model to accurately perceive the underwater object edge structure even in noisy and low-resolution images [1]. It then learns to perceive highly discriminative sparse features on the object structure, reducing the computational demands for edge devices [1].

- **Hybrid Training:** Some approaches, like UOSTrack, use hybrid training with both underwater images and open-air sequences to address the sample imbalance problem in underwater datasets [7].

- **Siamese network-based trackers**: Siamese networks employ identical neural networks with shared weights to process the target object's initial appearance (template) and subsequent frames (search region) [3]. By performing a convolutional feature cross-correlation between these two inputs, Siamese networks learn a similarity function to robustly identify the target across varying conditions and frames [3][10][17]. The tracking task is essentially formulated as a similarity-matching function [17].

- **Multi-modal Data:** Leveraging multi-modal data like depth and event information, alongside specialized architectures like multi-scale feature pyramids and transformer-based attention mechanisms, holds significant potential for advancing underwater detection and tracking [3].

Deep learning-based object detection models like YOLO, R-CNN, and SSD are used for underwater tracking [13][3]. Studies have shown YOLO to be reliable for underwater object detection [13]. To enhance temporal stability, YOLO can be merged with SORT [13]. Transformer-based architectures are also showing promising results in UOT [6].

This paradigm involves using object detection models (like YOLOv3) to detect objects in each frame and then linking these detections across frames using tracking algorithms such as Deep SORT, which can be enhanced with LSTM
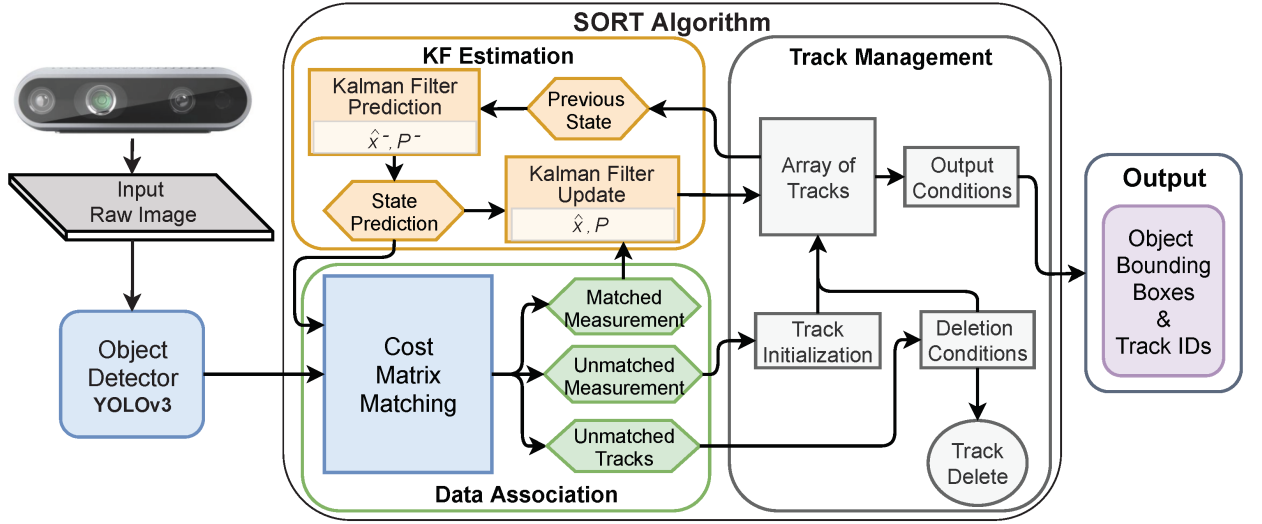
Рисунок 4 — Overview of SORT algorithm based YOLOv3 and Kalman filter [18].

for handling occlusions, as proposed in the HADSYv3 approach [6][3]. This method can be effective in dynamic underwater environments, although its efficacy depends on the quality of the detection algorithm [6].

Several datasets have been established to promote research in UOT, including UOT32, UOT100, UTB180, VMAT, and UVOT400 [7][1][5]. However, many lack training sets or have limitations in size and scenario coverage. The WebUOT-1M dataset has been introduced as a larger benchmark to facilitate the development of more powerful deep UOT algorithms [7]. FishTrack23 is another large-scale dataset specifically for fish tracking [3].

Evaluations on datasets like WebUOT-1M show that Transformer-based trackers (e.g., OKTrack, UOSTrack, All-in-One, GRM, OSTrack) often perform well [7]. UOT-specific trackers like OKTrack and UOSTrack, even with plain ViT backbones, can surpass state-of-the-art open-air trackers, highlighting the domain gap between underwater and open-air environments [7]. Retraining open-air trackers on underwater datasets like WebUOT-1M can effectively reduce this domain gap and improve performance. The DBSF tracker has also demonstrated optimal tracking results on UOT32 and UOT100 benchmarks by focusing on boundary information and sparse confidence features, particularly for edge computing devices [1].

Correlation filter (CF) based tracking offers real-time capabilities, and combining it with deep convolutional features has significantly improved performance [1][9][10][11]. Algorithms like **DeepSRDCF**, which uses CNN features, and **SiamFC**, a fully-convolutional Siamese network, have shown

excellent results [10]. ECO is another efficient correlation filter-based tracker utilizing deep features [10][8][11]. **MACF** is a motion-aware correlation filter algorithm developed for challenging scenarios like satellite videos with small objects and similar distractors, and it has achieved superior accuracy compared to state-of-the-art trackers [11].

Despite the advancements, challenges remain in UOT, including handling occlusion, small or camouflaged objects, low visibility, optimizing computational efficiency for real-time processing on autonomous systems, and mitigating model drift [1][6][7][11]. Future work includes introducing spatial-temporal features and semi-supervised learning to improve tracking in complex underwater scenes [1][6]. Enhancing the detection of small or camouflaged objects, optimizing computational efficiency, and improving real-time processing capabilities are also crucial for deployment in autonomous systems [3]. For correlation filter-based tracking, future research can focus on designing better filters by considering categorized features, space weight factors, scale factors, and expert strategies, as well as combining manual and deep features [9].

2.3 Sensor Fusion-Based Tracking Methods

Sensor fusion-based approaches in underwater environments leverage data from multiple sensors to enhance surveillance and tracking capabilities [19][20][21]. In contrast to traditional methods that rely on single platforms like submarines or frigates, distributed networks of stationary and mobile sensors, such as autonomous underwater vehicles (AUVs), offer advantages in scalability, robustness, and reliability through intelligent networking [19]. **Distributed information fusion (DIFFUSION) strategies**, where local information is shared among sensors, are a key aspect of these intelligent networks [19]. This allows for the combination of data from spatially separated sensors to achieve higher performance [19].

Two primary DIFFUSION schemes are proposed for underwater surveillance: one based on the sharing of local contacts generated by the detection stage, and another based on the sharing of tracks produced by the local tracking stage [19]. In the contact-sharing scheme, contacts are combined at each node using optimal Bayesian tracking based on the random finite set (RFS) formulation [19]. The track-sharing scheme employs a track-to-track (T2T) association and

16

fusion procedure [19]. A local tracker on each AUV provides a set of tracks and their associated covariance matrices, which are then associated and fused with tracks from other AUVs to obtain more accurate state estimates [19]. Unassociated tracks are treated as originating from a single sensor [19].

The problem of sensor selection within a large network is also critical for optimizing tracking performance under constraints like communication bandwidth [20]. Efficient search techniques, such as convex optimization followed by greedy local search, can be employed to determine near-optimal sensor utilization strategies in real-time for multitarget tracking [20]. Different approaches to sensor selection include "closest-sensor"strategies, which select sensors nearest to estimated target positions, and more sophisticated methods like "coarse-step"and "fine-step"planning that consider tracking performance over time [20]. The posterior Cramér–Rao lower bound (PCRLB) can serve as a basis for network management and as a cost function for optimization [20].

In the context of passive sensor fusion for underwater surveillance, the **Generalized Labeled Multi-Bernoulli (GLMB)** filter can be used for joint filtering of measurements from multiple sensors, enabling multi-target tracking [21]. While detection sets from passive sensors may be staggered in time, the iterated multi-sensor update approach in the GLMB filter has proven sufficiently accurate for simultaneous detections [21]. This Bayesian recursion over time allows for long-term integration of sensor data, even when dealing with the non-linearities inherent in bearings-only tracking, often tackled using sequential Monte Carlo methods [21].

3 THEORETICAL STUDY OF ALGORITHMS

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. *Qiu H.*, *Li N.*, *Li P.*, [et al.]. Boundary attention guided sparse feature learning for underwater object tracking in edge computing // ACM Transactions on Multimedia Computing, Communications and Applications. — 2024.

2. *Zhou H.*, *Kong M.*, *Yuan H.*, [et al.]. Real-time underwater object detection technology for complex underwater environments based on deep learning // Ecological Informatics. — 2024. — Vol. 82. — P. 102680.

3. *Elmezain M.*, *Saoud L.S.*, *Sultan A.*, [et al.]. Advancing Underwater Vision: A Survey of Deep Learning Models for Underwater Object Recognition and Tracking // IEEE Access. — 2025.

4. *Bhadouria A.S.* Underwater image enhancement techniques: an exhaustive study // International journal for research in applied science & engineering technology (IJRASET), ISSN. —. — P. 2321–9653.

5. *Rout D.K.*, *Subudhi B.N.*, *Veerakumar T.*, *Chaudhury S.* Walsh–Hadamard-kernel-based features in particle filter framework for underwater object tracking // IEEE Transactions on Industrial Informatics. — 2019. — Vol. 16, no. 9. — P. 5712–5722.

6. *Mathias A.*, *Dhanalakshmi S.*, *Kumar R.* Occlusion aware underwater object tracking using hybrid adaptive deep SORT-YOLOv3 approach // Multimedia Tools and Applications. — 2022. — Vol. 81, no. 30. — P. 44109–44121.

7. *Zhang C.*, *Liu L.*, *Huang G.*, [et al.]. Webuot-1m: Advancing deep underwater object tracking with a million-scale benchmark // arXiv preprint arXiv:2405.19818. — 2024.

8. *Feng W.*, *Han R.*, *Guo Q.*, *Zhu J.*, *Wang S.* Dynamic saliency-aware regularization for correlation filter-based object tracking // IEEE Transactions on Image Processing. — 2019. — Vol. 28, no. 7. — P. 3232–3245.

9. *Du S.*, *Wang S.* An overview of correlation-filter-based object tracking // IEEE Transactions on Computational Social Systems. — 2021. — Vol. 9, no. 1. — P. 18–31.

10. *Zhao S.*, *Sun K.*, *Ji Y.*, *Guo N.*, *Jia X.* Correlation filter-based object tracking algorithms // 2020 IEEE 3rd International Conference on Information Communication and Signal Processing (ICICSP). — IEEE. 2020. — P. 57–62.

11. *Lin B.*, *Zheng J.*, *Xue C.*, [et al.]. Motion-aware correlation filter-based object tracking in satellite videos // IEEE Transactions on Geoscience and Remote Sensing. — 2024. — Vol. 62. — P. 1–13.

12. *Srigowri M.* Enhancing unpaired underwater images with cycle consistent network // 2022 International Conference on Inventive Computation Technologies (ICICT). — IEEE. 2022. — P. 305–311.

13. *Lotfi F.*, *Virji K.*, *Dudek N.*, *Dudek G.* A comparison of RL-based and PID controllers for 6-DOF swimming robots: hybrid underwater object tracking // arXiv preprint arXiv:2401.16618. — 2024.

14. *Jiao L.*, *Zhang F.*, *Liu F.*, [et al.]. A Survey of Deep Learning-Based Object Detection // IEEE Access. — 2019. — Vol. 7. — P. 128837–128868.

15. *Arkin E.*, *Yadikar N.*, *Xu X.*, *Aysa A.*, *Ubul K.* A survey: object detection methods from CNN to transformer // Multimedia Tools and Applications. — 2023. — Vol. 82, no. 14. — P. 21353–21383.

16. *Yang H.*, *Xu J.*, *Lin Z.*, *He J.* LU2Net: a lightweight network for real-time underwater image enhancement // arXiv preprint arXiv:2406.14973. — 2024.

17. *Wu X.*, *Han X.*, *Zhang Z.*, [et al.]. A hybrid excitation model based lightweight siamese network for underwater vehicle object tracking missions // Journal of Marine Science and Engineering. — 2023. — Vol. 11, no. 6. — P. 1127.

18. *Pereira R.*, *Carvalho G.*, *Garrote L.*, *Nunes U.J.* Sort and deep-SORT based multi-object tracking for mobile robotics: Evaluation with new data association metrics // Applied Sciences. — 2022. — Vol. 12, no. 3. — P. 1319.

19. *Braca P.*, *Goldhahn R.*, *Ferri G.*, *LePage K.D.* Distributed information fusion in multistatic sensor networks for underwater surveillance // IEEE Sensors Journal. — 2015. — Vol. 16, no. 11. — P. 4003–4014.

20. *Tharmarasa R.*, *Kirubarajan T.*, *Hernandez M.* Large-scale optimal sensor array management for multitarget tracking // IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews). — 2007. — Vol. 37, no. 5. — P. 803–814.

21. *Uney M.*, *Stinco P.*, *Dreo R.*, [et al.]. Passive sensor fusion and tracking in underwater surveillance with the GLMB model // 2022 25th International Conference on Information Fusion (FUSION). — IEEE. 2022. — P. 1–8.