International Journal on Advanced Science Engineering Information Technology

Vol.14 (2024) No. 5
ISSN: 2088-5334

Advanced Object Tracking through Conditional Online Updates and Noise Suppression

Suchang Lim [a], Jongchan Kim [a,*]

[a] *Department of Computer Engineering, Sunchon National University, 235, Jungang-ro, Suncheon-si, Jeollanam-do, Republic of Korea*
Corresponding author: [*]seaghost@scnu.ac.kr

***Abstract*— Object tracking under complex environmental conditions, such as background clutter, occlusion, and rapid motion, presents significant challenges. This paper addresses these issues by proposing a tracking algorithm integrating background suppression, target region enhancement, and an adaptive online template update mechanism to improve tracking accuracy. The proposed method uses the initial bounding box of the target object as a reference template and selectively updates specific regions online to suppress noise and retain critical features. We evaluated the proposed method using the OTB dataset to validate it. The baseline model without the proposed method showed a success rate of 0.417 and a precision of 0.586, while the algorithm with the proposed method achieved improved values of 0.524 and 0.728, respectively. Qualitative evaluations further confirmed the robustness of the proposed method, demonstrating high performance in scenarios with occlusion and complex backgrounds. Rather than updating all regions indiscriminately, the proposed method selectively updates the template using representative values from the target object's information. This selective update mechanism ensures the incorporation of the most relevant and accurate features, enabling the algorithm to adapt to changes in the target's appearance while minimizing noise integration. Emphasizing the feature regions and suppressing noise are also critical for maintaining a clear and precise representation of the target object, reducing the likelihood of confusion by irrelevant background information. Future research will focus on developing balanced update strategies that integrate new information while maintaining stable and reliable target characteristics.**

***Keywords*— Computer vision; data analysis; object tracking; deep learning; image similarity; online learning.**

Manuscript received 5 Apr. 2024; revised 19 Jun. 2024; accepted 12 Sep. 2024. Date of publication 31 Oct. 2024.

I. INTRODUCTION

In computer vision, object tracking involves continuously monitoring the location and trajectory of a specific object within an image sequence, making it a crucial technology in various vision-based applications and domains [1]. Object tracking plays a key role in areas such as security and surveillance systems, autonomous vehicles, and AR/VR applications [2], [3], [4]. Object tracking is challenging within computer vision due to several complex factors influencing the tracking process [5]. Objects within a video frame can move in various angles and directions over time, causing changes in their appearance. Concurrently, as the object moves, the background also changes, often leading to scenarios where distinguishing between the object and the background becomes difficult, particularly in complex environments [6]. Objects rarely move in ideal, predictable patterns. When another object partially or entirely occludes an object, the tracking algorithm may experience a drift phenomenon, where it recognizes only a portion of the object or fails to track it altogether. Thus, object tracking in computer vision is complicated by various factors such as changes in the object's appearance, background complexity, lighting variations, rapid movements, occlusions, and noise. Extensive research is being conducted to address these challenges.

Feature extraction techniques are pivotal in object tracking, as they are essential for clearly identifying objects and accurately tracking their location and spatial changes over time. Notably, the extracted features must be highly discriminable to distinguish the object from the background or other objects [7]. Additionally, these features must provide consistent performance despite various real-world environmental changes. Features can be categorized into generic features and semantic features. Generic features consider the overall pattern of the entire image, enabling the identification of the object's overall shape and structure. Semantic features, containing unique semantic information

about the object, are beneficial for accurately identifying the type or state of the object.

The field of object tracking can be categorized into classical manual methods and deep learning-based methods according to the type of feature handling [8] [21], [26]. Classical object tracking methods primarily utilize hand-crafted feature extraction techniques. Notable examples include the Kalman filter, mean shift, particle filter, correlation filter, and optical flow-based methods [9], [10]. These manual feature extraction techniques recognize and track objects by extracting specific patterns or features from images. Representative manual feature extraction methods include SIFT, SURF, and HOG [11], [25]. Recently, deep learning-based feature extraction techniques have garnered significant attention. These methods utilize Deep Neural Networks to automatically learn kernels for extracting object features[31], [32], offering powerful expressiveness to process more complex and non-linear data compared to classical methods. This results in robust tracking performance under various conditions and variations [12], [13], [14].

In object tracking, filter updating refers to the continuous adaptation of the filter to changes in the object's appearance and environmental variations [15-17]. This process primarily involves updating the filter using features of the object extracted from the current frame [30]. A filter generally learns a specific pattern of an object and subsequently recognizes and tracks that pattern in future frames [22], [23], [24]. This process is termed online update because it is performed concurrently with the tracking process. However, if the background area is incorrectly incorporated during this process, the feature may include noise from the background, thereby deteriorating the ability to distinguish the object accurately. When updating a filter in object tracking, various issues arise if the filter learns not only the characteristics of the object but also those of the surrounding background. Firstly, if the filter includes background information, its ability to differentiate between the object and the background diminishes. The filter may be confounded by the background information, which typically exhibits more variability and diversity than the target information. Additionally, if background information is repeatedly included in the learning process, the filter tends to learn more about the background [27], [28], [29]. This results in the maximum peak point of the filter's response being distorted by the background, reducing tracking reliability. When these errors accumulate, the filter deviates from the state in which it learned the original object pattern, resulting in a decreased ability to distinguish objects.

To address this problem, it is crucial to emphasize the object region, update only that area, and disregard the background. This paper proposes a selective partial region filter update method to resolve the issue above and enhance object tracking accuracy. Masking techniques are applied to highlight the object regions and ignore the background. The mask enhances only the object's features and attenuates background information during the filter update process. The mask's shape is dynamically generated according to the object's external size and location. The proposed filter introduces a conditional filter update procedure. If the change in the object's appearance is insignificant, the filter update frequency is reduced to minimize the inclusion of background information. Conversely, if the object's appearance changes significantly or the environment substantially changes, the filter update frequency is increased to reflect the object's latest characteristics. Through this process, the object's features can be learned more accurately, and the impact of background noise can be minimized.

II. Materials and Method

A. Foreground Attention Method based on Feature Extractor

Object tracking presents a variety of challenging problems. First, in environments where objects and backgrounds are intermixed, tracking performance deteriorates when background information is included in the target template update. Second, if the filter does not accurately reflect changes in the appearance or location of an object, the likelihood of tracking failure increases. Third, when an object is partially occluded, the filter learns incorrect target information, raising the possibility of degraded tracking performance. To address these issues, a filter update technique is required to extract object features, minimize the background's influence, and is robust to changes and occlusions.

This paper proposes a filter update method to resolve these challenges and improve object tracking accuracy. First, masking and attention techniques emphasize the foreground and the object area while ignoring the background. This approach aims to ensure accurate tracking even in complex backgrounds by focusing on the foreground, the area with high similarity. The mask enhances only the features of the object and attenuates background information during the filter update process. The mask's shape is dynamically created according to the external size and position of the object. Notably, in addition to binary masks, cosine windows, and Hann windows are applied to smooth discontinuities occurring at the boundaries of the signal, preventing spectral leakage and maintaining robust tracking performance. Fig. 1 shows the basic structure of the network model used for object tracking.

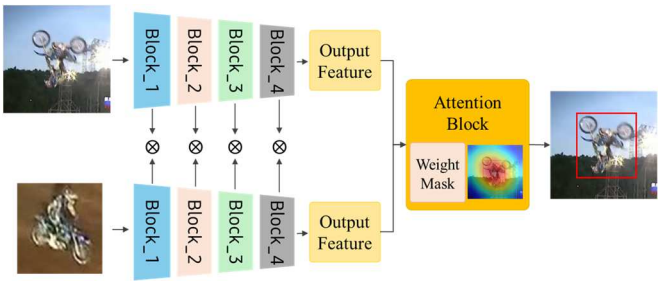

Fig. 1 The overall structure of the proposed tracking algorithm

The target object and search image are used as inputs to the network. Each image passes through the backbone network and is converted into a feature map. The feature maps are output at different sizes in each backbone block. Each output feature map is resized and converted to a uniform size of 127x127. These feature maps are merged into a single feature map and passed through the attention module. In object tracking, the attention mechanism is applied to distinguish the target object area from the background area. By introducing an attention mechanism, the accuracy and efficiency of object

tracking can be improved by directing the network to focus more on specific areas of interest.

An attention map is created to infer the location and area of the target object within the search image. An attention map assigns importance to each pixel or region in the image. A feature map is extracted using a CNN as the backbone network, followed by an attention layer to highlight the target's features. The attention map emphasizes the target object area while reducing the importance of the background area. The attention map then passes through linear and ReLU layers, and a normalization function is applied to convert the significance of each pixel into a value between 0 and 1. This enhanced feature map of the input image is used to highlight the object's features. The target object's area is dynamically selected using the attention map, and a filter is trained in the online learning process using only the features from this area.

B. Partial Response Map Filter for Online Learning

If the model is updated, every frame, drift, and overfitting problem is likely to occur. Drift refers to the phenomenon where a model gradually deviates from the original object it tries to track because it incorporates background or other object characteristics during the update process. The overfitting problem occurs when the model, updated every frame, becomes overly sensitive to small changes or noise, reacting to these transient changes more than to general features. This likely reduces the consistency and accuracy of tracking.

The similarity map reflects consistency with the previous frame, enhancing object features and minimizing background noise when updating the filter through the combined map. The area of the target object is dynamically selected using the attention map, and the correlation filter is updated using only the features from this area. After confirming the target object's location through similarity analysis, only areas with high importance in the attention map are incorporated into the update process. This minimizes the influence of the background area.

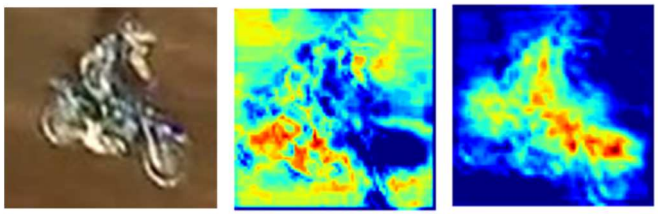

Fig. 2 Overall structure of the proposed tracking algorithm

Fig. 2 shows the visualization results for each block of the feature map of the target area in the first frame after passing through the backbone network. The images in the second and third columns distinctly represent the characteristics of the background and foreground areas separately. The applied filter introduces a conditional filter update procedure. Updating the model every frame increases the likelihood of drift and overfitting problems. Drift refers to the phenomenon where a model gradually deviates from the original object it tries to track due to the inclusion of background or other object characteristics during the update process. Overfitting occurs when the model, updated every frame, becomes overly sensitive to small changes or noise, reacting to these transient changes more than to general features. This increases the likelihood of deteriorating tracking consistency and accuracy.

If the change in the object's appearance is insignificant, reducing the filter update frequency can mitigate the inclusion of background information. Conversely, when the object's appearance changes significantly, or the environment undergoes substantial changes, the filter update frequency is increased to reflect the latest characteristics of the object. Through similarity analysis, the similarity between the current and previous frames is calculated, and the filter is updated only when the similarity exceeds a threshold, thereby preventing filter contamination due to background or incorrect location information.

Similarity and reliability estimation are used to determine occlusion. The similarity between the target area in the current frame and the previous frame is calculated using cosine similarity or correlation coefficient metrics. A significant drop in similarity indicates the possibility of occlusion. Confidence scores are also employed. A high confidence score indicates a strong focus on the target area, significantly impacting accuracy. Convolution-based similarity measurement can effectively capture detailed characteristics of an object. Remarkably, the attention module helps suppress background information and emphasize the foreground, the target area. This is achieved by weighting important features, reducing unnecessary background information, and highlighting relevant features.

III. Result and Discussion

A. Implementation Details

The proposed tracking algorithm was evaluated on hardware consisting of the PyTorch deep learning library, an Intel i7-8700k CPU, an NVIDIA GeForce RTX 3080 GPU with 8GB VRAM, and 32GB of RAM. The dataset used in the experiment is the ImageNet VID dataset, commonly used in object tracking and object detection research [18]. This dataset comprises over 3,000 video sequences and approximately 1,000,000 annotated video frames. During the training phase, the model was trained using the training dataset. The data used for training underwent a series of preprocessing operations. First, two images were randomly selected from the video sequences and paired, one to be used as the target object and the other as the search image. The sizes of the target object and search image inputs to the network were defined as 127 × 127 and 255 × 255, respectively.

We used ResNet-52, employed for object classification tasks in ImageNet, as the feature extractor for the tracking algorithm [19]. The dataset was used to train the Attention module, and the convolutional layer was added to the backbone network components. To enhance the discriminative power of the dataset, the amount of image data was augmented by applying random transformations, random rotations, random resizing, and random cropping to each image. The weights of the Attention module and the parameters of the added convolutional layer were initialized using Xavier initialization. The network parameter optimizer was set to AdamW, with a learning rate 0.0001.

In the object tracking stage, performance verification is conducted using the Object Tracking Benchmark (OTB)

dataset [20]. The initial frame of the video sequence is used as a template image, and the search area is expanded by padding it to more than 20% of the center coordinates of the target object derived from the previous frame. In typical Siamese-based tracking algorithms, these outputs require extensive post-processing to obtain the best-predicted bounding box, including resizing, bounding box smoothing, and cosine window penalties. However, these operations are susceptible to hyperparameters, where small changes can significantly impact the final results. Since the test dataset consists of sequences containing various environmental changes, considerable effort is required to derive optimal parameters during inference. To address this issue, a window mask is used to obtain the final region score. The window penalty operation can reduce the reliability of the target feature point from the previous frame, enabling the selection of the ideal bounding box.

B. Evaluation

OTB is a data set derived from pioneering work on visual tracking algorithms and is widely used to evaluate the performance of tracking algorithms. This dataset contains 100 environments with 12 different challenging environment properties, including low resolution, in-plane rotation, fast motion, motion blur, deformation, occlusion, scale deformation, out-of-plane rotation, illumination deformation, background clutter, and out-of-view. It consists of a video sequence. The number of frames that make up each sequence is different. Each video sequence contains one or more environmental properties. Precision and success indicators are used to measure the performance of the tracking algorithm with OTB. Each indicator is an evaluation indicator uniquely assigned to OTB. Success is an indicator that indicates the degree of overlap between the area manually extracted from the sequence using the coordinates of the annotation and the area derived through the algorithm. Precision indicates the drift rate by measuring the distance the bounding box is from the object center. In the experiment, the impact of the annotation module on tracking accuracy is analyzed and verified. For this purpose, evaluation is conducted using the tracking accuracy before the module is applied and the tracking algorithm after the module is applied. Fig. 3 compares the performance of the proposed algorithm with an existing general Siamese network-based tracking algorithm using the OTB50 dataset.

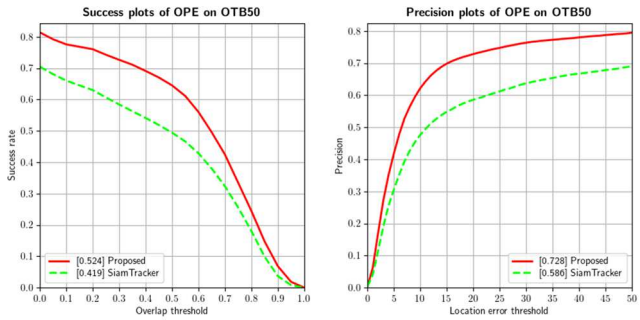

Fig. 3 Experimental results for algorithms with and without the proposed method applied using the OTB benchmark dataset Overall structure of the proposed tracking

The graph on the left is the Success plot, and the graph on the right is the Precision plot. By applying the attention module and the conditional dynamic online update method, we aim to improve tracking performance through background suppression and foreground emphasis. In the Success plot, the proposed algorithm achieved a score of 0.524, which is 0.105 higher than SiamTracker. In the Precision plot, the proposed algorithm achieved a score of 0.728, which is 0.142 higher than SiamTracker.

In the Success plot, both algorithms show a tendency for the success rate to decrease as the threshold increases, but the proposed algorithm maintains a high success rate in the low threshold range. This demonstrates that the proposed module is practical for accurate object tracking in the early stages. The precision plot shows higher precision for the proposed algorithm than SiamTracker. In particular, there is a noticeable difference in the section where the position error is small, indicating that the applied online update method maintains the target template robustly, thus aiding in accurately tracking the target's location in the search image.

Fig. 4 Qualitative results with and without applying the attention module on the OTB dataset (from top to bottom: Basketball, Soccer, and DragonBaby)

For further comparison, we present the qualitative results of the proposed algorithm and SiamTrakcer on the OTB dataset in Fig. 4. This image shows the qualitative tracking results. The blue box represents the proposed algorithm, and the red box represents SiamTracker, a comparison algorithm. The performance of the two algorithms is compared in terms of various environmental properties.

Fig. 5 shows an example of the tracking failure of the proposed algorithm. The green bounding box represents the area of the actual object, while the blue bounding box represents the result of the proposed algorithm tracking the target area. The causes of object tracking failure can be analyzed as follows. Due to changes in brightness and lower background contrast within the video, when the tracking target entered the tree area, it was confused with the background, making clear distinction difficult. The failure in the first column is due to the reduced contrast between the

object and the background caused by bright sunlight and the reflection of snow, making it difficult to distinguish between the background and the object clearly.

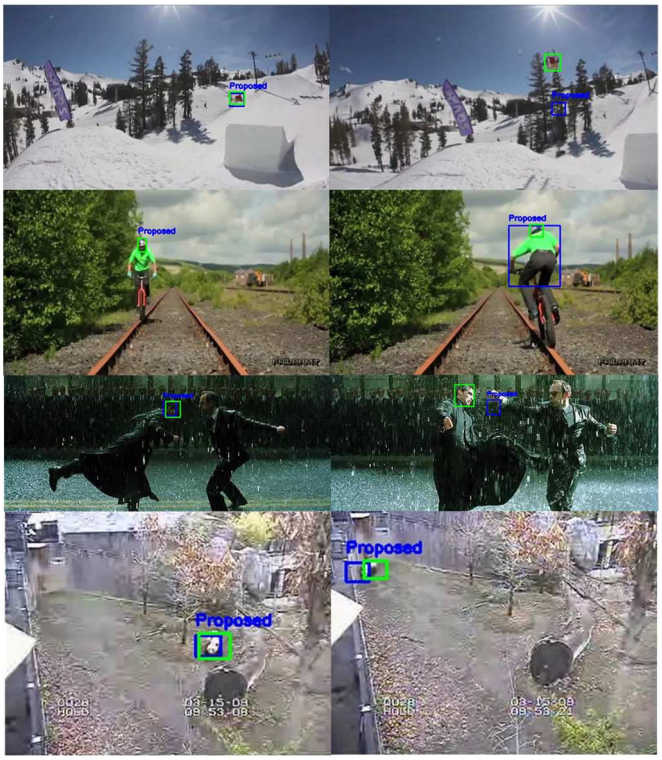

Fig. 5 Tracker failure cases in the OTB data set (Green and blue bounding boxes indicate ground-truth and inference results)

Environmental factors also need to be more clear about the object's characteristics, confusing with the background. The failure in the second column occurs because the object moves quickly, and the location of the tracking target cannot be accurately determined. Specifically, when the target moves forward and then rotates, there are instances where the rear features are missing, leading to tracking failure. The failure in the third column is caused by rain and rapid changes in brightness, which obscure the object's outline and reduce the contrast with the background. This results in an unclear outline and failure to track the object. In the fourth column, the low resolution makes it difficult to extract detailed features of the object, and the information is insufficient to construct a comprehensive template for the target object. Consequently, tracking is carried out using only partial features, failing. Additionally, because the object's movement radius between frames is narrow, tracking is performed by bounding the surrounding area rather than the precise area of the object.

IV. Conclusion

The comparative analysis between the proposed tracking algorithm and the existing Siamese tracker shows that the background suppression, target area highlighting, and dynamic conditional template online update modules positively influenced the tracking results by improving performance. As illustrated in the first uploaded image, the proposed algorithm achieved higher success rates and precision scores at various overlap and position error thresholds. Notably, the success plot of the OTB dataset used in the experiment shows an increase from 0.419, as seen in the comparison algorithm, to 0.524 with the proposed algorithm. Similarly, the precision plot demonstrates that the proposed method achieved a precision score of 0.728, compared to 0.586 by the comparison algorithm, validating the effectiveness of the proposed methodology.

The proposed tracking algorithm improved estimation accuracy by consistently maintaining accurate target tracking in scenarios involving complex environments, occlusion, and fast movement. It provides adaptability and reliability under various tracking conditions, such as scale changes, occlusion, and motion blur. Online updates play a crucial role in maintaining the accuracy and robustness of tracking algorithms. Instead of indiscriminately updating all regions in the template, the proposed method selectively updates the template using representative values from the target object's information. This selective update mechanism ensures that the most relevant and accurate features are incorporated into the template, enabling the algorithm to adapt to changes in target shape while minimizing noise integration.

However, it has been shown that the tracking algorithm can be confused when sufficient information is not provided or when encountering situations that deviate from the template format. This indicates that confusion and performance degradation of the tracking model may occur over time due to continuous updates. Based on these conclusions, future research will focus on developing and applying a balanced update strategy that integrates new information while maintaining stable and reliable target characteristics.

Acknowledgment

This paper was supported by Sunchon National University Research Fund in 2024. (Grant number: 2024-0396)

References

[1] H. Cai, L. Lan, J. Zhang, X. Zhang, C. Xiao, and Z. Luo, “Online intervention siamese tracking,” *Information Sciences*, vol. 637, p. 118954, Aug. 2023, doi:10.1016/j.ins.2023.118954.

[2] H. Lu, Y. Zhang, Y. Li, C. Jiang, and H. Abbas, “User-oriented virtual mobile network resource management for vehicle communications,” *IEEE Trans. Intell. Transp Syst*., vol. 22, no. 6, pp. 3521–3532, May. 2021, doi:10.1109/tits.2020.2991766.

[3] H. Lu, Y. Li, S. Mu, D. Wang, H. Kim, and S. Serikawa, “Motor anomaly detection for unmanned aerial vehicles using reinforcement learning,” *IEEE Internet ThingsJ*., vol. 5, no. 6, pp.2315–2322, Aug. 2018, doi: 10.1109/jiot.2017.2737479.

[4] C. Li, X. Liang, Y. Lu, N. Zhao, and J. Tang, “RGB-T object tracking:benchmark and baseline,” *Pattern Recognition*, vol. 96, p. 106977, Aug. 2019, doi:10.1016/j.patcog.2019.106977.

[5] P. Li, D. Wang, L. Wang, and H. Lu, “Deep visual tracking: review and experimental comparison,” *Pattern Recognition*, vol. 76, pp. 323–338, Apr. 2018, doi:10.1016/j.patcog.2017.11.007.

[6] A. W. Smeulders, D. M. Chu, R. Cucchiara, S. Calderara, A. Dehghan, and M. Shah, “Visual tracking: an experimental survey,” *IEEE Trans. Pattern Anal. Mach.Intell.*, vol. 36, no. 7, pp. 1442–1468, 2013, doi:10.1109/tpami.2013.230.

[7] L. Zhou, Y. Jin, H. Wang, Z. Hu, and S. Zhao, “Robust DCF object tracking with adaptive spatial and temporal regularization based on target appearance variation,” *Signal Processing*, vol. 195, p. 108463, Jun. 2022, doi:10.1016/j.sigpro.2022.108463.

[8] S. Javed, M. Danelljan, F. S. Khan, M. H. Khan, M. Felsberg, and J. Matas, "Visual Object Tracking With Discriminative Filters and Siamese Networks: A Survey and Outlook," *in IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 5, pp. 6552-6574, May 2023, doi: 10.1109/tpami.2022.3212594.

[9] T. Zhang, C. Xu, and M. H. Yang, "Learning Multi-Task Correlation Particle Filters for Visual Tracking," *in IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 2, pp. 365-378, Feb. 2019, doi:10.1109/tpami.2018.2797062.

[10] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "High-Speed Tracking with Kernelized Correlation Filters," *in IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 3, pp. 583-596, Mar. 2015, doi:10.1109/tpami.2014.2345390.

[11] N. Dalal, and B. Triggs, "Histograms of oriented gradients for human detection," *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Diego, CA, USA, vol. 1, pp. 886-893, 2005, doi:10.1109/cvpr.2005.177.

[12] A. Berthelier, T. Chateau, S. Duffner, C. Garcia, and C. Blanc, "Deep model compression and architecture optimization for embedded systems: A survey," *Journal of Signal Process. System*, vol. 93, pp. 863–878, Oct. 2021, doi:10.1007/s11265-020-01596-1.

[13] M. Danelljan, L. Van Gool, and R. Timofte, "Probabilistic Regression for Visual Tracking," *In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, Seattle, WA, USA, pp. 7181-7190, 2020, doi:10.1109/cvpr42600.2020.00721.

[14] K. Nai, Z. Li, and H. Wang, "Dynamic feature fusion with spatial-temporal context for robust object tracking," *Pattern Recognition*, vol. 130, p. 108775, Oct. 2022, doi:10.1016/j.patcog.2022.108775.

[15] D. Elayaperumal, and Y. H. Joo, "Robust visual object tracking using context-based spatial variation via multi-feature fusion," *Information Sciences*, vol. 577, pp. 467-482, Oct. 2021, doi:10.1016/j.ins.2021.06.084.

[16] M. Mueller, N. Smith, and B. Ghanem, "Context-Aware Correlation Filter Tracking," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, pp. 1387-1395, 2017, doi: 10.1109/cvpr.2017.152.

[17] J. Zhang, M. Miao, H. Zhang, J. Wang, Y. Zhao, Z. Chen, and J. Qiao, "Object semantic-guided graph attention feature fusion network for Siamese visual tracking," *Journal of Visual Communication and Image* Representation, vol. 90, p. 103705, Feb. 2023, doi:10.1016/j.jvcir.2022.103705.

[18] O. Russakovsky, J. Deng, H. Su et al., "ImageNet Large Scale Visual Recognition Challenge," *International journal of computer vision*, vol. 115, pp. 211–252, Apr. 2015, doi:10.1007/s11263-015-0816-y.

[19] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, pp. 770-778, Jun. 2016, doi: 10.1109/cvpr.2016.90.

[20] Y. Wu, J. Lim, and M. H. Yang, "Object Tracking Benchmark," *in IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 9, pp. 1834-1848, Sep. 2015, doi:10.1109/tpami.2014.2388226.

[21] S. Xuan, S. Li, Z. Zhao, L. Kou, Z. Zhou, and G. S. Xia, "Siamese Networks with Distractor-Reduction Method for Long-term Visual Object Tracking," *Pattern Recognition*, vol. 112, p. 107698, Apr. 2021, doi:10.1016/j.patcog.2020.107698.

[22] Y. Zeng, B. Zeng, H. Hu, and H. Zhang, "PRAT: Accurate Object Tracking Based on Progressive Attention," *Engineering Applications of Artificial Intelligence*, vol. 126, p. 106988, Nov. 2023, doi:10.1016/j.engappai.2023.106988.

[23] H. Wang, and F. Guo, "Online Object Tracking Based Interactive Attention," *Computer Vision and Image Understanding*, vol. 236, p. 103809, Nov. 2023, doi:10.1016/j.cviu.2023.103809.

[24] J. Lu, S. Li, W. Guo, M. Zhao, J. Yang, Y. Liu, and Z. Zhou, "Siamese Graph Attention Networks for Robust Visual Object Tracking," *Computer Vision and Image Understanding*, vol. 229, p. 103634, Mar. 2023, doi:10.1016/j.cviu.2023.103634.

[25] Y. Zhang, T. Wang, K. Liu, B. Zhang, and L. Chen, "Recent advances of single-object tracking methods: A brief survey," *Neurocomputing*, vol. 455, pp. 1-11, Sep. 2021, doi:10.1016/j.neucom.2021.05.011.

[26] F. Chen, X. Wang, Y. Zhao, S. Lv, and X. Niu, "Visual Object Tracking: A Survey," *Computer Vision and Image Understanding*, vol. 222, p. 103508, Sep. 2022, doi:10.1016/j.cviu.2022.103508.

[27] J. Zhang, J. Sun, J. Wang, Z. Li, and X. Chen. "An Object Tracking Framework with Recapture Based on Correlation Filters and Siamese Networks," *Computers and Electrical Engineering*, vol. 98, p. 107730, Mar. 2022, doi:10.1016/j.compeleceng.2022.107730.

[28] X. Gao, Y. Zhou, S. Huo, Z. Li, and K. Li, "Robust Object Tracking via Deformation Samples Generator," *Journal of Visual Communication and Image Representation*, vol. 83, p. 103446, Feb. 2022, doi: 10.1016/j.jvcir.2022.103446.

[29] S. Gao, C. Zhou, and J. Zhang, "Generalized Relation Modeling for Transformer Tracking," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Vancouver, BC, Canada, pp. 18686-18695, Jun. 2023, doi:10.1109/CVPR52729.2023.01792.

[30] H. Zhao, G. Yang, D. Wang, and H. Lu, "Deep Mutual Learning for Visual Object Tracking," *Pattern Recognition*, vol. 112, p. 107796, Apr. 2021, doi:10.1016/j.patcog.2020.107796.

[31] H. Rumapea, M. Zarlis, S. Efendy, and P. Sihombing, "Improving Convective Cloud Classification with Deep Learning: The CC-Unet Model", *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 14, no. 1, pp. 28–36, Feb. 2024, doi: 10.18517/ijaseit.14.1.18658.

[32] X. Y. Chan, T. Connie, and M. K. O. Goh, "Facial and Body Gesture Recognition for Determining Student Concentration Level", Int. J. Adv. Sci. Eng. Inf. Technol., vol. 13, no. 5, pp. 1693–1702, Oct. 2023, doi: 10.18517/ijaseit.13.5.19035