
DOVTrack: Data-Efficient Open-Vocabulary Tracking

Supplementary Materials

Anonymous Author(s)

Affiliation

Address

email

1 Variance Implications of the Grouping Strategy

1.1 Minimizing Variance Sum through Affinity Group Construction

Problem Statement. Given a set of numbers x_1, x_2, \dots, x_n , we want to construct k groups, denoted as G_1, G_2, \dots, G_k , such that the sum of each group is T_k . The mean of each group is defined as $\mu_k = \frac{T_k}{m_k}$, where m_k is the number of elements in group k . We aim to minimize the overall variance S^2 of the sample set can be expressed as:

$$S^2 = \frac{1}{k} \sum_{i=1}^k \left[\frac{1}{m_k - 1} \sum_{j=1}^{m_k} (x_j - \mu_i)^2 \right].$$

Strategy for Minimizing S^2 . To minimize S^2 , we adopt a sequential grouping strategy as follows: First, sort the data values x_1, x_2, \dots, x_n in ascending order by the similarity from the F_{text} . Then divide the sorted data into k contiguous groups. The first group contains the smallest values, while the last group contains the largest values.

Proof. Next, we will provide a detailed proof of why this affinity group construction strategy can effectively achieve the minimization of S^2 as follows:

- **Definitions and Setup:** Let the mean of each group be $\mu_k = \frac{T_k}{m_k}$, where m_k is the number of elements in each group.
- **Exchange Argument:** Suppose there are two groups i and j , with μ_i and μ_j as their respective means. Let there exist an element $x < y$ such that $x \in G_j$ and $y \in G_i$. Based on the grouping method, we can conclude that $T_i < T_j$, meaning that the sum of group i is less than the sum of group j . We consider swapping these two elements:

$$T'_i = T_i - x + y,$$

$$T'_j = T_j - y + x.$$

The new group means after the swap will be:

$$\mu'_i = \frac{T'_i}{m_i} = \frac{T_i - x + y}{m_i},$$

$$\mu'_j = \frac{T'_j}{m_j} = \frac{T_j - y + x}{m_j}.$$

Next, we need to calculate the updated variances for groups i and j :

$$S_{i'}^2 = \frac{1}{m_i - 1} \sum_{j=1}^{m_i} (x_j - \mu'_i)^2,$$

$$S_{j'}^2 = \frac{1}{m_j - 1} \sum_{j=1}^{m_j} (y_j - \mu'_j)^2.$$

- **Calculating the Change in Variance:** The change in the overall variance ΔS^2 due to the swap is given by:

$$\Delta S^2 = S_{i'}^2 + S_{j'}^2 - (S_i^2 + S_j^2).$$

Calculating ΔS^2 directly can be complex; however, by swapping the larger value y from group G_j with the smaller value x from group G_i , we significantly alter the internal distribution of the elements within each group. The introduction of the larger value y in group G_i increases its variance because it increases the deviation from the new mean. Similarly, swapping the smaller value x into group G_j will also increase the variance of that group. Therefore, as a result of these changes, it can be inferred that:

$$\Delta S^2 > 0.$$

This indicates that the overall variance increases after any swap, which suggests that the process of selective grouping can minimize the spread of variance across the groups.

- **Conclusion:**

Therefore, using a sequential grouping method (i.e., partitioning the data into contiguous segments) will minimize the overall variance S^2 .

1.2 Maximizing Variance Sum through Dispersion Group Construction

Problem Statement. Given a set of numbers x_1, x_2, \dots, x_n , we want to construct k groups, denoted as G_1, G_2, \dots, G_k , such that the sum of each group is T_k . The mean of each group is defined as $\mu_k = \frac{T_k}{m_k}$, where m_k is the number of elements in group k . The overall variance S^2 of the sample set can be expressed as:

$$S^2 = \frac{1}{k} \sum_{i=1}^k \left[\frac{1}{m_k - 1} \sum_{j=1}^{m_k} (x_j - \mu_i)^2 \right].$$

Expanding this yields:

$$S^2 = \frac{1}{k} \sum_{i=1}^k \left[\frac{1}{m_k - 1} \left(\sum_{j=1}^{m_k} x_j^2 - m_k \mu_i^2 \right) \right].$$

This can be rearranged to show a fixed term:

$$S^2 = \frac{1}{k} \sum_{i=1}^k \left[\frac{1}{m_k - 1} \sum_{j=1}^{m_k} x_j^2 - \frac{m_k}{m_k - 1} \mu_i^2 \right].$$

The term $\sum_{j=1}^n x_j^2$ is a fixed quantity determined by the sample set. Therefore, to maximize S^2 , we need to minimize the term: $\sum_{i=1}^k \mu_i^2$.

Strategy for Minimizing $\sum_{i=1}^k \mu_i^2$. To minimize $\sum_{i=1}^k \mu_i^2$, we adopt a two-end grouping strategy as follows: First, sort the data values x_1, x_2, \dots, x_n in ascending order by the similarity from the F_{text} . Then, each group should be formed by selecting elements such that each group contains the largest

46 and smallest values available. Specifically, we can construct each group G_i by taking the maximum
 47 and minimum values from the remaining elements.

48 **Proof.** To prove that these strategies effectively minimize $\sum_{i=1}^k \mu_i^2$, consider the following:

- 49 • **Assuming Constant Total:** Let’s assume the overall sum of group means is constant, i.e.,
 50 $\mu_1 + \mu_2 + \dots + \mu_k = T$. The goal is to minimize $\sum_{i=1}^k \mu_i^2$ under this constraint.
- 51 • **Applying Cauchy-Schwarz Inequality:** By applying the Cauchy-Schwarz inequality in the
 52 context of these means:

$$k(\mu_1^2 + \mu_2^2 + \dots + \mu_k^2) \geq (\mu_1 + \mu_2 + \dots + \mu_k)^2 = T^2.$$

53 This implies that:

$$\mu_1^2 + \mu_2^2 + \dots + \mu_k^2 \geq \frac{T^2}{k}.$$

54 Therefore, minimizing $\sum_{i=1}^k \mu_i^2$ occurs under the condition that the means are as equal as
 55 possible.

- 56 • **Validating the Construction Method:** Our group construction method ensures that all
 57 μ_i values are as equal as possible because we are taking elements from both ends of the
 58 distribution. This approach ensures that the means converge to the overall mean of the
 59 sample set, thereby fulfilling the necessary condition for minimizing $\sum_{i=1}^k \mu_i^2$.
- 60 • **Conclusion:** By focusing on strategies that leverage the largest and smallest available
 61 values for group construction and maintaining the overall sum of means as constant, we can
 62 effectively minimize the sum of squares of the means $\mu_1^2, \mu_2^2, \dots, \mu_k^2$, thus maximizing the
 63 overall variance.

64 2 Qualitative Analysis

65 We compare our method with the baseline method OVTrack across several challenging scenarios
 66 involving novel object classes. As shown in Figure 1, in the first construction site scene, our approach
 67 effectively and accurately tracks fast-moving drones, whereas OVTrack fails to detect the drones
 68 at all. Moreover, our method precisely classifies the bulldozer as a novel object category, while
 69 OVTrack misclassifies it as a truck and initially fails to detect the object entirely. Our method also
 70 demonstrates superior detection and tracking capabilities for base object classes (persons).

71 In the second racing scenario, characterized by high-speed vehicles and occlusion, our method
 72 successfully tracks and correctly classifies the race cars. In contrast, OVTrack struggles to detect
 73 occluded targets and exhibits incorrect ID switching. Its classification is also less precise, categorizing
 74 the vehicles under the broader “car” class instead of the specific “race car” category.

75 Figure 2 illustrates the tracking performance in a field scenario, featuring a fast-moving dragonfly
 76 belonging to a novel category. Compared to OVTrack, our proposed method demonstrates superior
 77 detection and tracking capabilities, successfully identifying the dragonfly. In contrast, OVTrack fails
 78 to detect the dragonfly in most frames and cannot accurately classify it.

79 Figure 3 depicts a scene from the African savanna, featuring a novel category hippopotamus and two
 80 lions chasing it. Our method successfully classifies the hippopotamus correctly, whereas OVTrack
 81 misidentifies it as an “elephant”. Additionally, OVTrack incorrectly labels the lion as a “horse” and
 82 “cow”. Moreover, our approach demonstrates superior detection and tracking accuracy compared to
 83 OVTrack.

84 These results demonstrate that our method, through efficient training, significantly enhances localiza-
 85 tion, classification, and association capabilities across diverse and challenging tracking scenarios.

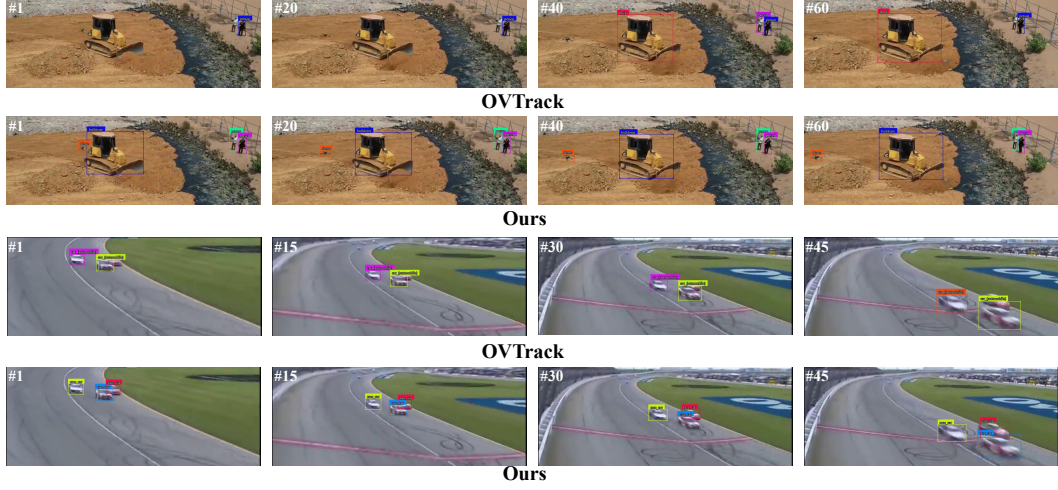


Figure 1: Visualization results from construction sites and race tracks with novel object categories, including drones and bulldozers in the construction scene, along with race cars in the racing track.

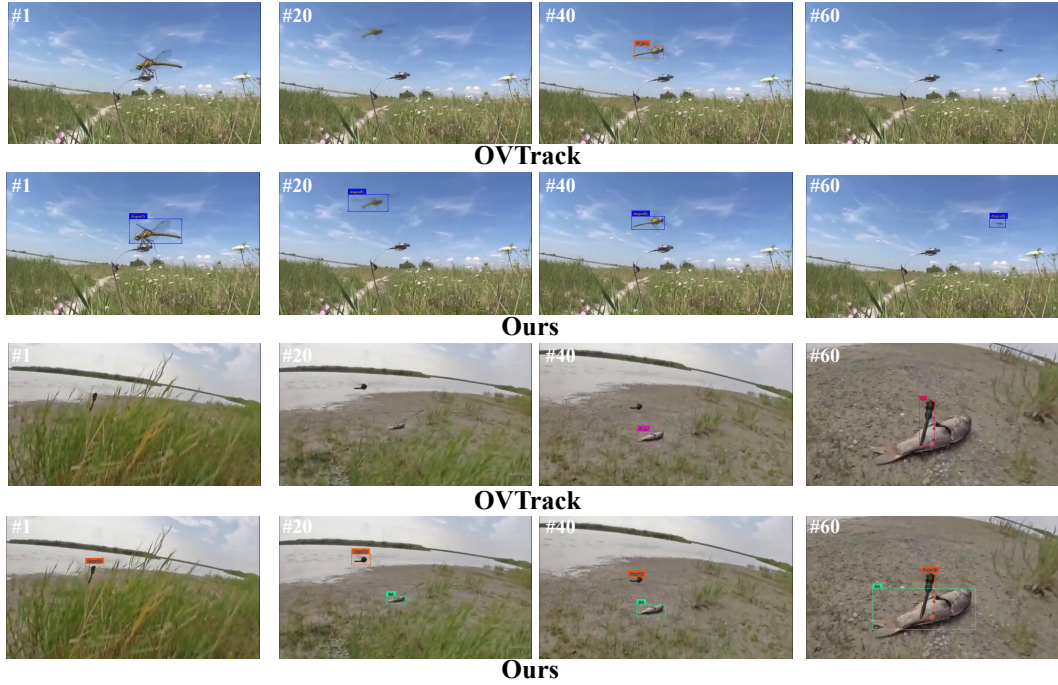


Figure 2: Visualization results in the field with the novel object category, dragonfly.

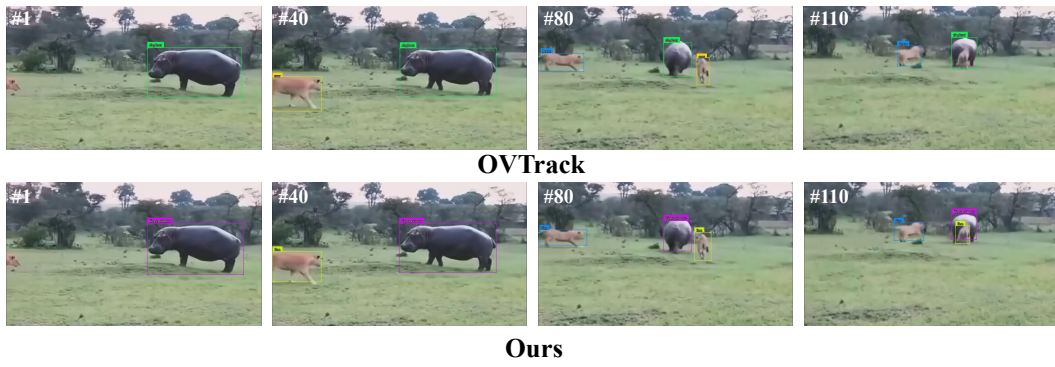


Figure 3: Visualization results in the African savanna with the novel object category, hippopotamus.