# 3D-ZeF: A 3D Zebrafish Tracking Benchmark Dataset
## Supplementary Material

Malte Pedersen*, Joakim Bruslund Haurum*, Stefan Hein Bengtson, Thomas B. Moeslund
Visual Analysis of People (VAP) Laboratory, Aalborg University, Denmark
mape@create.aau.dk, joha@create.aau.dk, shbe@create.aau.dk, tbm@create.aau.dk

## A. Content

In these supplementary materials we provide: examples of the visual difference between the fish in the three splits, a more detailed explanation of the detectors, details on the benchmark pipeline parameters, a more detailed presentation of the relation between the proposed dataset complexity measures and tracking metrics, and the full set of tracking metrics for each step in the proposed benchmark pipeline.

## B. Fish Examples

The visual appearance of the fish varies both within and between the splits as unique groups of fish have been used for each split. However, the two groups of fish used in the train and validation sets are from the same cohort, whereas the zebrafish in the test split are from a cohort of younger and smaller fish. We present a fish from each of the three splits in Figure 1 captured at the approximately same position in the water tank. The resolution of the zebrafish from the test split is significantly smaller compared to the fish in the training and validation splits, mainly due to the physical size of the fish.

## C. Naive Object Detection

The Naive object detection algorithms presented briefly in the paper are described in more details in this section. The methods are inspired by Qian *et al*. [1] and their work on tracking zebrafish in 2D.

### C.1. Pre-Processing

Initially the background image, without any fish, is estimated by taking the median of $N$ images sampled uniformly across the video. In the tests presented in the paper a set of 80 images were used to create the background image for each recording. All further processing is restricted within a defined region of interest based on the boundaries of the water tank and the level of water, in order to limit the processing time.

*Equal contribution

The video is downsampled by a factor of 2, in order to further decrease the processing time. Background subtraction is applied by calculating the absolute difference image, $|im - bg|$, choosing the max value across all color channels, and normalizing the image into the range $[0; 255]$. The resulting image is filtered with a $5 \times 5$ median filter to reduce noise.

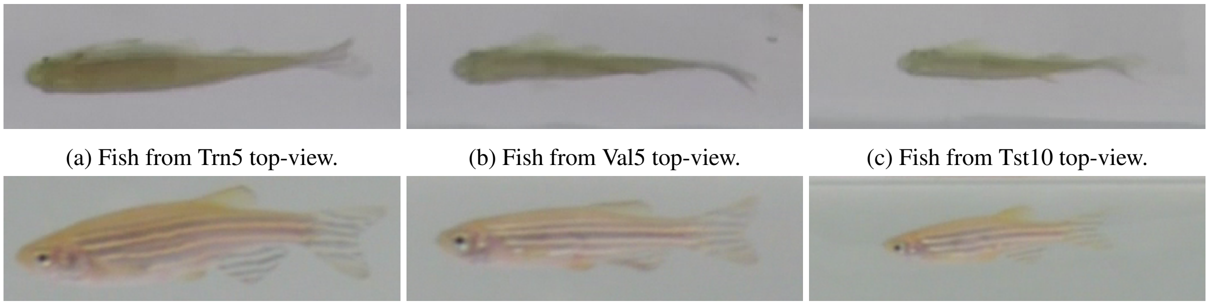
### C.2. Top-View Detection

The top-view is thresholded based on the assumption of the image histogram being bimodal, due to a near uniform bright background, and darker zebrafish. Therefore the intermodal approach of Prewitt *et al*. [2] can be utilized. The threshold is set to the middle point between the two modes in the image histogram. If the histogram is not bimodal, the frame is filtered with a $3 \times 3$ mean filter, until its histogram is bimodal.

The fish are detected by applying a skeletonization based approach. BLOBs are initially detected and filled, and the skeletonization approach of Zhang and Suen [3] is applied. The skeleton is analyzed in order to find the skeleton keypoints: head, tail, and junctions. This is done by convolving the frame with the kernel in Equation (1). All end points of the skeleton have a value of 116, 117, 118, or 131, while junctions have a value of 148, 149, 150, or 151. Values, like 132, that can represent both an end point or an arbitrary point on the skeleton are not considered.

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 15 & 15 & 15 & 1 \\ 1 & 15 & 100 & 15 & 1 \\ 1 & 15 & 15 & 15 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} \tag{1}$$

Each keypoint is assigned a weight, $w$, consisting of the smallest eigenvalue of the covariance matrix of the BLOB coordinates extracted from a $20 \times 20$ window around the keypoint. The smallest eigenvalue is used, as it is assumed the variance will be smallest along the width of the head. This way the head keypoints will be assigned a larger $w$ than the tail keypoints, as the width of the zebrafish head is



(a) Fish from Trn5 top-view. (b) Fish from Val5 top-view. (c) Fish from Tst10 top-view.

(d) Fish from Trn5 front-view. (e) Fish from Val5 front-view. (f) Fish from Tst10 front-view.

Figure 1: Examples of zebrafish from the train, validation, and test split captured from the approximately same position in the water tank. The top- and front-view pairs show the same fish. Notice the variation in size of the zebrafish.

| Parameter | $c$ | $\delta_T$ | $\delta_F$ | $\tau_k$ | $\alpha$ | $\tau_p$ | $\beta$ |
|---|---|---|---|---|---|---|---|
| Value | 95 | 15 | 0.5/15 | 10 | 10 | 25 | 0.02 |

Table 1: Pipeline parameters used for testing.

usually larger than the width of the zebrafish tail. The $w$ of the junction points are reduced by a factor of $2.5$, as they are usually larger than endpoint keypoints.

Keypoints too close to each other are removed by applying non-maximal suppression (NMS). Each keypoint is assigned a bounding box of size $w \times w$, and kept if the bounding box does not overlap with any other keypoint by more than $\text{NMS}_{\text{thresh}}\%$ of the area of the keypoint in focus. In case there is an overlap, only the keypoint with the largest $w$ is kept. $\text{NMS}_{\text{thresh}}$ was set to 50 in the conducted tests.

Finally, all keypoints with $w < 1$ are discarded. Per skeleton, the two keypoints furthest apart are determined, and the keypoint with largest $w$ is kept. Furthermore, half of the additional extra keypoints with the largest $w$ values are also kept in order to handle crossings and occlusions. Every keypoint ideally represent a zebrafish head and they are all kept as individual detections.

### C.3. Front-View Detection

The front frame cannot be thresholded based on the assumption of a bimodal distribution of the image histogram, as the stripes of the zebrafish results in a non-uniform appearance. Instead the frame is thresholded by finding the point maximizing the total entropy of the image [4]. The entropy is modeled as *background* and *foreground* entropy for each non-zero bin.

For bin $k$ the background entropy, $b_k$, is calculated by

$$b_k = |\sum_{i=0}^{k} \frac{h(i)}{h_c(i)} \log(\frac{h(i)}{h_c(i)})|, \tag{2}$$

where $h$ is the normalized image histogram and $h_c$ is the cumulative histogram of $h$. The foreground entropy, $w_k$, is calculated by

$$w_k = |\sum_{i=k+1}^{255} \frac{h(i)}{1-h_c(i)} \log(\frac{h(i)}{1-h_c(i)})|, \tag{3}$$

resulting in the two entropy sets $B = \{b_0, b_1, .., b_{255}\}$ and $W = \{w_0, w_1, .., w_{255}\}$ which are combined into

$$E = B + W, \tag{4}$$

and the threshold, $t$, is then finally determined by

$$t = \operatorname*{argmax}_x E(x). \tag{5}$$

The BLOBs in the thresholded image are found through simple Connected Component Analysis (CCA). Only the $2N_{\text{fish}}$ largest BLOBs are kept, as long as the area of the BLOBs are larger than a predefined threshold. This allows for potential reflections being detected alongside the true detection of the fish. Otherwise, if only the $N_{\text{fish}}$ largest BLOBs are considered for further analysis, it is possible that reflections are considered at the expense of real fish.

As it is not known where the head of the fish is, two points are saved as proxies for the head, together with the center-of-mass. If the width of the BLOB bounding box is larger than the height, the proxy points are $(\min(x), \mu(y))$ and $(\max(x), \mu(y))$, where x and y are the coordinates of the BLOB pixels, and $\mu$ is the arithmetic mean function. Otherwise, the proxy points are $(\mu(x), \min(y))$ and $(\mu(x), \max(y))$. The center-of-mass is further used for constructing 2D tracklets, as it is the most stable of the three proxy points.

## D. FRCNN-H Object Detection

Two Faster R-CNN networks were trained to detect the zebrafish heads in each view, respectively. The official PyTorch implementation of Faster R-CNN with a ResNet50

| | Property | Mean | Std. Dev. | Median |
|---|---|---|---|---|
| Trn | Reproj. Error $[px]$ | 8.03 | 5.26 | 7.59 |
| | Speed $[cm/s^2]$ | 2.13 | 2.32 | 1.54 |
| Val | Reproj. Error $[px]$ | 6.22 | 4.50 | 5.43 |
| | Speed $[cm/s^2]$ | 2.02 | 2.37 | 1.41 |
| Tst | Reproj. Error $[px]$ | 4.36 | 3.27 | 3.69 |
| | Speed $[cm/s^2]$ | 2.11 | 1.93 | 1.58 |

Table 2: Statistics of the reprojection error and speed of the zebrafish, for each split, based on the ground truth annotations.

based Feature Pyramid Network [5] backbone was utilized in both cases. The model was pretrained on the COCO train2017 dataset. The network was fine-tuned for 30 epochs, using stochastic gradient descent with momentum. A learning rate of 0.005, momentum of 0.9, weight decay of 0.0005, and batch size of 8 was used. The learning rate was "warmed up" during the first epoch, linearly interpolating the learning rate from 0.001 to 0.005. Each model was trained on an RTX 2080TI.

## E. Pipeline Parameters

The full system pipeline has a set of parameters, which needs to be manually set. The parameters were chosen based on empirical investigation on the training data and they are shown in Table 1. The front-view distance threshold, $\delta_F$, has two different values, depending on the method applied. For the Naive method it is set to $0.5$, as the distance is measured in standard deviations, whereas for the FRCNN-H method it is set to $15$, where the distance is measured in pixels.

Furthermore, during the association of 2D tracklets between views, a set of exponential distributions are utilized. These distributions are parameterized using the mean of the reprojection error of the ground truth training annotations and the average speed of the zebrafish in the training split. This builds on the assumption that these parameters generalize across the different splits. As shown in Table 2, the average speed of the zebrafish for each split are within $0.1\frac{cm}{s^2}$, which can be seen as a negligible difference, whereas the mean reprojection error is $2-4$ pixels larger in the training and validation splits than in the testing split. The cause of this difference is hypothesized to be the larger amount and duration of occlusions in the training and validation splits, which can introduce some uncertainty during the annotation phase. The training split has the largest reprojection error, which means the utilized exponential distribution penalize larger reprojection errors less harshly. However, as the mean training reprojection error is still low, it is still deemed fitting for the task.

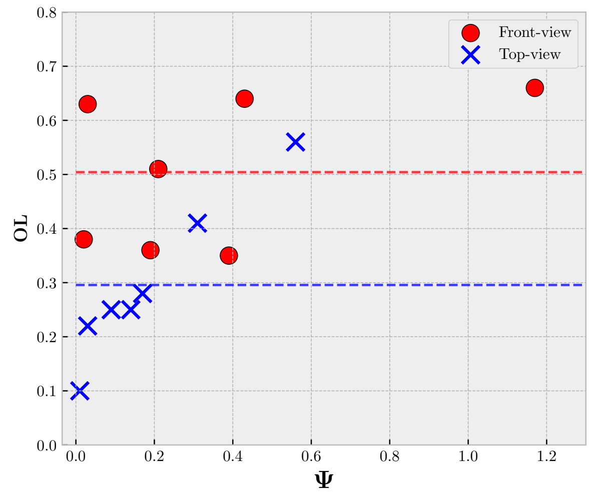



Figure 2: $\Psi$ is the proposed complexity measure based on the front- and top-view recordings. OL is the average length of occlusion events and the dashed lines show the mean OL for the two views. Tst1 is not included.

## F. Tracking Metrics and Results

When evaluating tracking tasks, a wide suite of metrics are often used, each expressing different properties of the task. Commonly, the MOTA metric is used as a representative metric for the overall performance of the tracker, as it incorporates false positives and negatives, as well as the amount of identity swaps. However, as shown by Carr and Collins [6] the MOTA metric is not guaranteed to be robust at all times. We evaluate the generated tracks using the collection of metrics from the MOT challenge, consisting of the CLEAR MOT [7], mostly tracked/lost metric [8], and identity based metrics of Ristani *et al.* [9], as well as using the MTBF metric proposed by Carr and Collins [6]. A description of each of the used metrics are presented in Table 3. Each step of the pipeline has been evaluated according to the previous mentioned metrics. Specifically the following steps have been evaluated:

- 2D tracklets after 2D Tracklet Construction (for top-view see Table 4 and for front-view see Table 5)
- 2D tracklets after 2D Tracklet Association Between Views (for top-view see Table 6 and for front-view see Table 7)
- 3D tracklets after 2D Tracklet Association Between Views (see Table 8)
- 3D tracks after 3D Tracklet Association (see Table 9)

For the 2D tracklets, a distance threshold of 20 px is used.

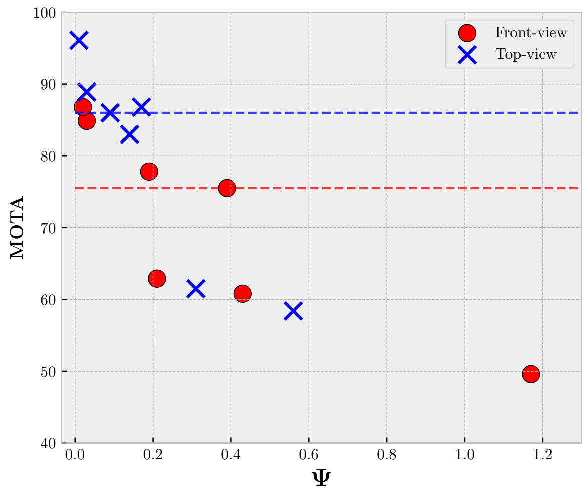



Figure 3: $\Psi$ is the proposed complexity measure based on the front- and top-view recordings. MOTA is based on the oracle tracker and the dashed lines show the mean MOTA for the two views. Tst1 is not included.

## G. Occlusions

The two views are not equal when it comes to occlusions. Due to the social behavior of zebrafish they tend to swim in the same horizontal layer of the water column. This is indicated by the graph in Figure 2, where the occlusion lengths are displayed with respect to the proposed complexity measure, $\Psi$. It should be noted that $\Psi$ is calculated per view and not as a mean between the two, as in the main article. Data points are presented for all the recordings from both views, except Tst1 as it only contains a single fish and therefore has a complexity measure of zero. The dashed lines illustrate the mean occlusion lengths and shows that there is a significant difference between the two views. In general, the occlusion events seems longer in the front-view recordings.

The MOTA of the hypothetical oracle tracker is plotted against the proposed complexity measure in Figure 3. MOTA is calculated for the top- and front-view 2D tracklets as presented in Table 4 and Table 5, respectively. The tracking performance of the oracle tracker correlates well with the complexity in both views. Notice the significant difference in performance of the tracker in the two views; the mean scores are $\text{MOTA}_{\text{top}} = 86.0\%$ and $\text{MOTA}_{\text{front}} = 75.5\%$ as illustrated by the two dashed lines.

## References

[1] Z.-M. Qian, S. H. Wang, X. E. Cheng, and Y. Q. Chen, "An effective and robust method for tracking multiple fish in video image based on fish head detection," *BMC Bioinformatics*, vol. 17, p. 251, June 2016.

[2] J. M. S. Prewitt and M. L. Mendelsohn, "The analysis of cell images*," *Annals of the New York Academy of Sciences*, vol. 128, no. 3, pp. 1035–1053, 1966.

[3] T. Y. Zhang and C. Y. Suen, "A Fast Parallel Algorithm for Thinning Digital Patterns," *Commun. ACM*, vol. 27, pp. 236–239, Mar. 1984.

[4] P. Sahoo, S. Soltani, and A. Wong, "A survey of thresholding techniques," *Computer Vision, Graphics, and Image Processing*, vol. 41, no. 2, pp. 233 – 260, 1988.

[5] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.

[6] P. Carr and R. T. Collins, "Assessing tracking performance in complex scenarios using mean time between failures," in *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 1–10, March 2016.

[7] K. Bernardin and R. Stiefelhagen, "Evaluating multiple object tracking performance: The CLEAR MOT metrics," *EURASIP Journal on Image and Video Processing*, vol. 2008, p. 246309, May 2008.

[8] Bo Wu and R. Nevatia, "Tracking of multiple, partially occluded humans based on static body part detection," in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, vol. 1, pp. 951–958, June 2006.

[9] E. Ristani, F. Solera, R. Zou, R. Cucchiara, and C. Tomasi, "Performance measures and a data set for multi-target, multi-camera tracking," in *Computer Vision – ECCV 2016 Workshops* (G. Hua and H. Jégou, eds.), (Cham), pp. 17–35, Springer International Publishing, 2016.

[10] Y. Li, C. Huang, and R. Nevatia, "Learning to associate: HybridBoosted multi-target tracker for crowded scene," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2953–2960, June 2009.

| Metric | Better | Range | Description |
|---|---|---|---|
| MOTA | Higher | $]-\infty, 100]$ | Overall tracking accuracy based on ID-swaps, and false negatives/positives [7] |
| MOTP | Lower | $[0, \infty[$ | Distance between predicted position and ground truth [7] |
| Prc. | Higher | $[0, 100]$ | Precision: The percentage of correctly predicted positions |
| Rcll. | Higher | $[0, 100]$ | Recall: The percentage of ground truth positions detected |
| ID-Prc. | Higher | $[0, 100]$ | ID-Precision: The percentage of correctly identified predicted positions [9] |
| ID-Rcll. | Higher | $[0, 100]$ | ID-Recall: The percentage of correctly identified ground truth positions [9] |
| ID-F1 | Higher | $[0, 100]$ | F1-score of the predicted identities [9] |
| FP | Lower | $[0, \infty[$ | Amount of incorrectly predicted positions |
| FN | Lower | $[0, \infty[$ | Amount of ground truth positions missed |
| MT | Higher | $[0, 100]$ | Amount of ground truth tracks which are 80% or more correctly tracked [8] |
| ML | Lower | $[0, 100]$ | Amount of ground truth tracks which are 20% or less correctly tracked [8] |
| ID Sw. | Lower | $[0, \infty[$ | Total amount of identity switches [10] |
| Frag. | Lower | $[0, \infty[$ | Total amount of track fragmentations |
| $\text{MTBF}_s$ | Higher | $[0, \infty[$ | Mean time between ID-switches or false negatives/positives [6] |
| $\text{MTBF}_m$ | Higher | $[0, \infty[$ | Monotonic version of $\text{MTBF}_s$ [6] |

Table 3: Description of the full collection of utilized tracking metrics.

| | Method | MOTA ↑ | MOTP ↓ | Prc. ↑ | Rcll. ↑ | ID-Rcll. ↑ | ID-Prc. ↑ | ID-F1 ↑ | FP ↓ | FN ↓ | MT ↑ | ML ↓ | ID Sw. ↓ | Frag. ↓ | $\text{MTBF}_s$ ↑ | $\text{MTBF}_m$ ↑ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Trn2 | Naive | 21.9% | 3.309 | 57.2 | 91.0 | 10.7 | 6.7 | 8.2 | 9798 | 1301 | 2 | 0 | 136 | 428 | 26.316 | 14.109 |
| | FRCNN-H | 91.3% | 2.128 | 98.8 | 92.8 | 12.4 | 13.2 | 12.8 | 156 | 1030 | 2 | 0 | 72 | 140 | 87.829 | 45.408 |
| | Oracle | 61.5% | 0.000 | 100.0 | 63.0 | 6.7 | 10.6 | 8.2 | 0 | 5314 | 0 | 0 | 216 | 216 | 41.587 | 20.699 |
| Trn5 | Naive | 38.2% | 3.196 | 64.1 | 90.1 | 31.5 | 22.4 | 26.2 | 2274 | 447 | 5 | 0 | 58 | 124 | 27.020 | 14.792 |
| | FRCNN-H | 90.2% | 2.038 | 99.1 | 91.6 | 31.9 | 34.5 | 33.2 | 39 | 378 | 5 | 0 | 24 | 50 | 73.607 | 38.523 |
| | Oracle | 58.4% | 0.000 | 100.0 | 59.6 | 19.1 | 32.0 | 23.9 | 0 | 1819 | 0 | 0 | 52 | 52 | 47.035 | 23.726 |
| Val2 | Naive | 59.8% | 3.050 | 72.6 | 96.7 | 57.1 | 42.8 | 48.9 | 1315 | 118 | 2 | 0 | 13 | 44 | 68.275 | 36.653 |
| | FRCNN-H | 96.8% | 2.165 | 99.2 | 97.9 | 75.9 | 76.8 | 76.3 | 30 | 76 | 2 | 0 | 10 | 17 | 167.810 | 92.737 |
| | Oracle | 88.9% | 0.000 | 100.0 | 89.7 | 25.8 | 28.8 | 27.2 | 0 | 372 | 2 | 0 | 28 | 28 | 107.600 | 55.655 |
| Val5 | Naive | 89.4% | 3.552 | 93.1 | 97.0 | 51.7 | 49.6 | 50.6 | 330 | 135 | 5 | 0 | 19 | 36 | 92.083 | 52.619 |
| | FRCNN-H | 96.4% | 2.782 | 99.5 | 97.2 | 48.3 | 49.4 | 48.8 | 22 | 128 | 5 | 0 | 15 | 23 | 147.567 | 83.528 |
| | Oracle | 86.0% | 0.000 | 100.0 | 86.9 | 37.2 | 42.8 | 39.8 | 0 | 598 | 4 | 0 | 40 | 40 | 87.933 | 46.553 |
| Tst1 | Naive | 83.1% | 4.095 | 85.6 | 100.0 | 100.0 | 85.6 | 92.2 | 152 | 0 | 1 | 0 | 0 | 0 | 900.000 | 900.000 |
| | FRCNN-H | 85.1% | 2.513 | 100.0 | 85.7 | 49.7 | 58.0 | 53.5 | 0 | 129 | 1 | 0 | 5 | 34 | 22.029 | 11.174 |
| | Oracle | 100.0% | 0.000 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 0 | 0 | 1 | 0 | 0 | 0 | 900.000 | 900.000 |
| Tst2 | Naive | 98.6% | 3.771 | 99.6 | 99.1 | 99.1 | 99.6 | 99.3 | 8 | 17 | 2 | 0 | 0 | 10 | 148.583 | 81.045 |
| | FRCNN-H | 72.6% | 2.443 | 100.0 | 74.1 | 39.0 | 52.7 | 44.8 | 0 | 467 | 1 | 0 | 26 | 93 | 14.032 | 7.053 |
| | Oracle | 96.1% | 0.000 | 100.0 | 96.7 | 32.6 | 33.7 | 33.1 | 0 | 60 | 2 | 0 | 10 | 10 | 145.000 | 79.091 |
| Tst5 | Naive | 82.8% | 3.801 | 89.3 | 94.7 | 49.7 | 46.8 | 48.2 | 512 | 237 | 5 | 0 | 24 | 82 | 45.839 | 24.222 |
| | FRCNN-H | 70.5% | 2.685 | 99.9 | 71.8 | 25.6 | 35.6 | 29.7 | 4 | 1271 | 0 | 0 | 52 | 145 | 21.384 | 10.836 |
| | Oracle | 83.0% | 0.000 | 100.0 | 84.0 | 44.7 | 53.2 | 48.6 | 0 | 721 | 3 | 0 | 44 | 44 | 77.122 | 39.779 |
| Tst10 | Naive | 83.4% | 3.752 | 89.3 | 95.4 | 59.0 | 55.2 | 57.0 | 1026 | 418 | 10 | 0 | 49 | 138 | 52.329 | 28.323 |
| | FRCNN-H | 70.8% | 2.803 | 99.8 | 72.4 | 26.4 | 36.4 | 30.6 | 13 | 2483 | 4 | 0 | 130 | 417 | 14.879 | 7.543 |
| | Oracle | 86.8% | 0.000 | 100.0 | 87.5 | 42.1 | 48.1 | 44.9 | 0 | 1128 | 10 | 0 | 64 | 64 | 106.378 | 56.229 |

Table 4: Full metric evaluation of the 2D tracking in the top-view, on all sequences.

| | Method | MOTA ↑ | MOTP ↓ | Prc. ↑ | Rcll. ↑ | ID-Rcll. ↑ | ID-Prc. ↑ | ID-F1 ↑ | FP ↓ | FN ↓ | MT ↑ | ML ↓ | ID Sw. ↓ | Frag. ↓ | MTBF$_s$ ↑ | MTBF$_m$ ↑ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Trn2 | Naive | -69.5% | 12.666 | 7.6 | 6.2 | 0.6 | 0.7 | 0.6 | 10756 | 13490 | 0 | 2 | 134 | 159 | 4.709 | 2.528 |
| | FRCNN-H | 92.8% | 3.521 | 99.8 | 95.0 | 6.4 | 6.8 | 6.6 | 26 | 723 | 2 | 0 | 288 | 126 | 43.218 | 30.898 |
| | Oracle | 62.9% | 0.000 | 100.0 | 64.1 | 2.4 | 3.8 | 3.0 | 0 | 5168 | 0 | 0 | 170 | 170 | 53.558 | 26.936 |
| Trn5 | Naive | -73.5% | 12.768 | 5.2 | 4.2 | 3.1 | 3.8 | 3.4 | 3478 | 4310 | 0 | 5 | 19 | 22 | 5.758 | 2.923 |
| | FRCNN-H | 90.7% | 3.634 | 99.9 | 92.4 | 23.2 | 25.1 | 24.1 | 6 | 341 | 5 | 0 | 72 | 47 | 45.207 | 29.496 |
| | Oracle | 60.8% | 0.000 | 100.0 | 61.7 | 20.1 | 32.6 | 24.9 | 0 | 1724 | 1 | 0 | 41 | 41 | 60.348 | 29.849 |
| Val2 | Naive | -93.1% | 12.918 | 3.7 | 3.6 | 0.8 | 0.8 | 0.8 | 3444 | 3469 | 0 | 2 | 37 | 28 | 3.119 | 1.795 |
| | FRCNN-H | 93.6% | 4.113 | 100.0 | 97.4 | 12.9 | 13.3 | 13.1 | 0 | 95 | 2 | 0 | 136 | 24 | 25.036 | 21.372 |
| | Oracle | 84.9% | 0.000 | 100.0 | 85.3 | 16.3 | 19.1 | 17.6 | 0 | 530 | 2 | 0 | 14 | 14 | 191.875 | 102.333 |
| Val5 | Naive | -68.8% | 12.419 | 10.0 | 8.5 | 2.9 | 3.4 | 3.1 | 3480 | 4162 | 0 | 5 | 37 | 44 | 6.690 | 3.464 |
| | FRCNN-H | 79.0% | 4.224 | 99.6 | 81.8 | 16.2 | 19.7 | 17.8 | 14 | 826 | 3 | 0 | 117 | 72 | 26.791 | 17.402 |
| | Oracle | 49.6% | 0.000 | 100.0 | 50.8 | 16.5 | 32.4 | 21.8 | 0 | 2240 | 0 | 0 | 54 | 54 | 39.153 | 19.914 |
| Tst1 | Naive | -44.8% | 14.682 | 27.5 | 26.2 | 5.2 | 5.5 | 5.4 | 621 | 664 | 0 | 0 | 18 | 28 | 6.378 | 3.522 |
| | FRCNN-H | 41.4% | 7.031 | 99.0 | 44.4 | 15.0 | 33.4 | 20.7 | 4 | 500 | 0 | 0 | 23 | 17 | 13.333 | 8.163 |
| | Oracle | 100.0% | 0.000 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 0 | 0 | 1 | 0 | 0 | 0 | 900.000 | 900.000 |
| Tst2 | Naive | -38.3% | 13.201 | 30.0 | 27.0 | 5.1 | 5.6 | 5.3 | 1135 | 1314 | 0 | 1 | 41 | 41 | 7.967 | 4.673 |
| | FRCNN-H | 16.4% | 7.828 | 100.0 | 18.7 | 6.6 | 35.4 | 11.1 | 0 | 1464 | 0 | 1 | 40 | 57 | 5.695 | 2.824 |
| | Oracle | 86.8% | 0.000 | 100.0 | 87.3 | 45.8 | 52.4 | 48.9 | 0 | 228 | 2 | 0 | 10 | 10 | 131.000 | 71.455 |
| Tst5 | Naive | -39.2% | 12.703 | 28.5 | 24.3 | 4.6 | 5.4 | 4.9 | 2734 | 3408 | 0 | 3 | 124 | 138 | 5.717 | 3.250 |
| | FRCNN-H | 53.5% | 5.601 | 98.7 | 56.4 | 15.4 | 27.0 | 19.7 | 33 | 1962 | 0 | 0 | 99 | 147 | 15.962 | 8.161 |
| | Oracle | 77.8% | 0.000 | 100.0 | 78.7 | 32.6 | 41.4 | 36.5 | 0 | 957 | 2 | 0 | 42 | 42 | 75.383 | 38.934 |
| Tst10 | Naive | -54.1% | 13.667 | 19.6 | 16.9 | 3.8 | 4.4 | 4.1 | 6219 | 7482 | 0 | 5 | 171 | 195 | 5.816 | 3.209 |
| | FRCNN-H | 47.4% | 5.604 | 99.4 | 50.3 | 13.0 | 25.6 | 17.2 | 27 | 4474 | 0 | 0 | 229 | 277 | 14.368 | 7.518 |
| | Oracle | 75.5% | 0.000 | 100.0 | 76.5 | 25.7 | 33.6 | 29.2 | 0 | 2115 | 4 | 0 | 94 | 94 | 66.202 | 33.261 |

Table 5: Full metric evaluation of the 2D tracking in the front-view, on all sequences.

| | Method | MOTA ↑ | MOTP ↓ | Prc. ↑ | Rcll. ↑ | ID-Rcll. ↑ | ID-Prc. ↑ | ID-F1 ↑ | FP ↓ | FN ↓ | MT ↑ | ML ↓ | ID Sw. ↓ | Frag. ↓ | MTBF$_s$ ↑ | MTBF$_m$ ↑ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Trn2 | Naive | 82.6% | 3.247 | 96.2 | 86.5 | 13.2 | 14.6 | 13.9 | 497 | 1939 | 2 | 0 | 73 | 403 | 28.732 | 14.846 |
| | FRCNN-H | 90.9% | 2.113 | 100.0 | 91.1 | 22.1 | 24.2 | 23.1 | 1 | 1273 | 2 | 0 | 41 | 144 | 85.667 | 43.836 |
| | Oracle | 44.8% | 0.000 | 100.0 | 46.2 | 2.4 | 5.3 | 3.3 | 0 | 7742 | 0 | 0 | 202 | 202 | 32.539 | 16.190 |
| Trn5 | Naive | 79.3% | 3.151 | 95.9 | 83.4 | 37.5 | 43.2 | 40.1 | 159 | 748 | 3 | 0 | 23 | 119 | 29.543 | 15.129 |
| | FRCNN-H | 90.2% | 2.028 | 100.0 | 90.5 | 45.6 | 50.4 | 47.8 | 0 | 427 | 5 | 0 | 13 | 44 | 81.460 | 42.874 |
| | Oracle | 36.6% | 0.000 | 100.0 | 37.8 | 10.9 | 28.9 | 15.9 | 0 | 2799 | 0 | 0 | 53 | 53 | 29.328 | 14.175 |
| Val2 | Naive | 89.0% | 3.047 | 99.4 | 89.8 | 63.3 | 70.1 | 66.6 | 20 | 369 | 2 | 0 | 8 | 52 | 57.696 | 29.642 |
| | FRCNN-H | 96.1% | 2.163 | 100.0 | 96.3 | 78.4 | 81.4 | 79.8 | 0 | 133 | 2 | 0 | 7 | 20 | 157.591 | 82.548 |
| | Oracle | 73.9% | 0.000 | 100.0 | 74.9 | 12.7 | 17.0 | 14.5 | 0 | 902 | 0 | 0 | 36 | 36 | 71.000 | 35.500 |
| Val5 | Naive | 92.1% | 3.524 | 99.8 | 92.5 | 53.3 | 57.4 | 55.3 | 9 | 341 | 4 | 0 | 9 | 57 | 67.968 | 34.826 |
| | FRCNN-H | 90.9% | 2.765 | 100.0 | 91.1 | 51.8 | 56.8 | 54.2 | 0 | 404 | 4 | 0 | 10 | 24 | 143.138 | 76.870 |
| | Oracle | 42.2% | 0.000 | 100.0 | 43.6 | 13.2 | 30.3 | 18.4 | 0 | 2567 | 0 | 0 | 64 | 64 | 28.812 | 14.406 |
| Tst1 | Naive | 99.6% | 4.102 | 100.0 | 99.6 | 99.6 | 100.0 | 99.8 | 0 | 4 | 1 | 0 | 0 | 2 | 298.667 | 179.200 |
| | FRCNN-H | 69.8% | 2.477 | 100.0 | 69.8 | 69.8 | 100.0 | 82.2 | 0 | 272 | 0 | 0 | 0 | 31 | 19.625 | 9.812 |
| | Oracle | 100.0% | 0.000 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 0 | 0 | 1 | 0 | 0 | 0 | 900.000 | 900.000 |
| Tst2 | Naive | 98.1% | 3.782 | 99.9 | 98.1 | 98.1 | 99.9 | 99.0 | 1 | 34 | 2 | 0 | 0 | 18 | 88.300 | 46.474 |
| | FRCNN-H | 40.8% | 2.514 | 100.0 | 40.9 | 32.8 | 80.1 | 46.5 | 0 | 1063 | 0 | 1 | 2 | 35 | 19.919 | 9.827 |
| | Oracle | 82.2% | 0.000 | 100.0 | 83.2 | 24.3 | 29.2 | 26.6 | 0 | 302 | 2 | 0 | 18 | 18 | 74.900 | 37.450 |
| Tst5 | Naive | 86.1% | 3.737 | 98.3 | 87.9 | 52.8 | 59.1 | 55.8 | 67 | 546 | 4 | 0 | 13 | 69 | 51.351 | 26.716 |
| | FRCNN-H | 65.8% | 2.666 | 100.0 | 66.2 | 34.0 | 51.3 | 40.9 | 0 | 1519 | 0 | 0 | 21 | 110 | 25.698 | 13.075 |
| | Oracle | 65.7% | 0.000 | 100.0 | 66.8 | 26.8 | 40.1 | 32.2 | 0 | 1492 | 1 | 0 | 50 | 50 | 54.691 | 26.857 |
| Tst10 | Naive | 87.0% | 3.757 | 98.4 | 88.6 | 68.1 | 75.7 | 71.7 | 126 | 1023 | 9 | 0 | 23 | 151 | 48.054 | 24.697 |
| | FRCNN-H | 62.4% | 2.797 | 100.0 | 62.9 | 31.4 | 49.9 | 38.6 | 2 | 3341 | 2 | 0 | 39 | 293 | 18.677 | 9.323 |
| | Oracle | 64.7% | 0.000 | 100.0 | 66.0 | 20.0 | 30.3 | 24.1 | 0 | 3059 | 1 | 0 | 119 | 119 | 46.054 | 22.676 |

Table 6: Full metric evaluation of the 2D tracking in the top-view, after 3D tracklet association, on all sequences.

| | Method | MOTA ↑ | MOTP ↓ | Prc. ↑ | Rcll. ↑ | ID-Rcll. ↑ | ID-Prc. ↑ | ID-F1 ↑ | FP ↓ | FN ↓ | MT ↑ | ML ↓ | ID Sw. ↓ | Frag. ↓ | MTBF$_s$ ↑ | MTBF$_m$ ↑ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Trn2 | Naive | -5.5% | 14.688 | 46.6 | 33.4 | 6.6 | 9.2 | 7.7 | 5513 | 9577 | 0 | 0 | 82 | 772 | 6.174 | 3.091 |
| | FRCNN-H | 85.3% | 3.482 | 99.9 | 85.6 | 19.5 | 22.8 | 21.0 | 10 | 2074 | 2 | 0 | 37 | 141 | 82.040 | 42.289 |
| | Oracle | 44.8% | 0.000 | 100.0 | 46.2 | 2.4 | 5.3 | 3.3 | 0 | 7742 | 0 | 0 | 202 | 202 | 32.539 | 16.190 |
| Trn5 | Naive | 2.2% | 13.648 | 51.9 | 37.8 | 19.1 | 26.2 | 22.1 | 1577 | 2800 | 0 | 2 | 22 | 168 | 9.770 | 4.843 |
| | FRCNN-H | 81.5% | 3.540 | 99.9 | 82.0 | 41.4 | 50.5 | 45.5 | 3 | 812 | 3 | 0 | 16 | 47 | 68.296 | 35.124 |
| | Oracle | 36.6% | 0.000 | 100.0 | 37.8 | 10.9 | 28.9 | 15.9 | 0 | 2799 | 0 | 0 | 53 | 53 | 29.328 | 14.058 |
| Val2 | Naive | -7.9% | 14.061 | 45.5 | 39.0 | 30.7 | 35.8 | 33.0 | 1678 | 2197 | 0 | 0 | 11 | 165 | 8.401 | 4.201 |
| | FRCNN-H | 82.1% | 4.095 | 100.0 | 82.3 | 63.6 | 77.2 | 69.7 | 0 | 638 | 1 | 0 | 7 | 36 | 75.949 | 39.493 |
| | Oracle | 73.9% | 0.000 | 100.0 | 74.9 | 12.7 | 17.0 | 14.5 | 0 | 902 | 0 | 0 | 36 | 36 | 71.000 | 35.500 |
| Val5 | Naive | -6.9% | 11.925 | 43.9 | 24.0 | 13.9 | 25.5 | 18.0 | 1391 | 3460 | 0 | 2 | 15 | 108 | 9.646 | 4.781 |
| | FRCNN-H | 69.3% | 4.068 | 99.9 | 69.7 | 35.0 | 50.1 | 41.2 | 4 | 1379 | 0 | 0 | 15 | 55 | 51.984 | 26.207 |
| | Oracle | 42.3% | 0.000 | 100.0 | 43.7 | 13.3 | 30.3 | 18.4 | 0 | 2562 | 0 | 0 | 64 | 64 | 28.812 | 14.099 |
| Tst1 | Naive | 72.1% | 8.828 | 92.8 | 78.2 | 78.2 | 92.8 | 84.9 | 55 | 196 | 0 | 0 | 0 | 35 | 19.556 | 9.778 |
| | FRCNN-H | 37.8% | 6.992 | 100.0 | 37.8 | 37.8 | 100.0 | 54.8 | 0 | 560 | 0 | 0 | 0 | 7 | 42.500 | 20.000 |
| | Oracle | 100.0% | 0.000 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 0 | 0 | 1 | 0 | 0 | 0 | 900.000 | 900.000 |
| Tst2 | Naive | 75.7% | 6.560 | 95.6 | 79.4 | 79.4 | 95.6 | 86.7 | 66 | 371 | 1 | 0 | 0 | 38 | 35.725 | 18.089 |
| | FRCNN-H | 7.8% | 6.187 | 100.0 | 7.9 | 4.8 | 61.3 | 9.0 | 0 | 1658 | 0 | 2 | 2 | 10 | 11.833 | 5.462 |
| | Oracle | 82.2% | 0.000 | 100.0 | 83.2 | 24.3 | 29.2 | 26.6 | 0 | 302 | 2 | 0 | 18 | 18 | 74.900 | 37.450 |
| Tst5 | Naive | 39.7% | 9.665 | 77.9 | 56.1 | 38.6 | 53.6 | 44.9 | 717 | 1977 | 0 | 0 | 21 | 229 | 10.736 | 5.403 |
| | FRCNN-H | 50.1% | 5.460 | 99.4 | 50.9 | 27.2 | 53.2 | 36.0 | 13 | 2209 | 0 | 0 | 23 | 91 | 23.619 | 11.749 |
| | Oracle | 65.7% | 0.000 | 100.0 | 66.8 | 26.8 | 40.1 | 32.2 | 0 | 1492 | 1 | 0 | 50 | 50 | 54.691 | 27.345 |
| Tst10 | Naive | 42.1% | 10.218 | 78.3 | 59.0 | 45.4 | 60.3 | 51.8 | 1467 | 3693 | 0 | 0 | 47 | 339 | 15.034 | 7.506 |
| | FRCNN-H | 39.8% | 5.488 | 99.9 | 40.3 | 19.6 | 48.8 | 28.0 | 3 | 5377 | 0 | 0 | 37 | 120 | 27.656 | 13.620 |
| | Oracle | 64.7% | 0.000 | 100.0 | 66.0 | 20.0 | 30.3 | 24.1 | 0 | 3059 | 1 | 0 | 119 | 119 | 46.054 | 22.589 |

Table 7: Full metric evaluation of the 2D tracking in the front-view, after 3D tracklet association, on all sequences.

| | Method | MOTA ↑ | MOTP ↓ | Prc. ↑ | Rcll. ↑ | ID-Rcll. ↑ | ID-Prc. ↑ | ID-F1 ↑ | FP ↓ | FN ↓ | MT ↑ | ML ↓ | ID Sw. ↓ | Frag. ↓ | MTBF$_s$ ↑ | MTBF$_m$ ↑ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Trn2 | Naive | 41.1% | 0.198 | 85.0 | 50.5 | 9.1 | 15.3 | 11.4 | 1285 | 7118 | 0 | 0 | 60 | 586 | 12.344 | 6.172 |
| | FRCNN-H | 73.7% | 0.066 | 96.9 | 76.3 | 18.8 | 23.8 | 21.0 | 352 | 3407 | 0 | 0 | 29 | 215 | 50.548 | 25.333 |
| | Oracle | 44.8% | 0.000 | 100.0 | 46.2 | 2.4 | 5.3 | 3.4 | 0 | 7738 | 0 | 0 | 202 | 202 | 32.539 | 16.190 |
| Trn5 | Naive | 40.3% | 0.169 | 83.5 | 50.5 | 25.5 | 42.1 | 31.8 | 448 | 2218 | 0 | 0 | 12 | 134 | 16.309 | 8.039 |
| | FRCNN-H | 63.5% | 0.064 | 92.1 | 69.7 | 37.1 | 49.1 | 42.3 | 269 | 1358 | 1 | 0 | 9 | 67 | 43.431 | 21.867 |
| | Oracle | 36.7% | 0.000 | 100.0 | 37.9 | 11.0 | 28.9 | 15.9 | 0 | 2784 | 0 | 0 | 53 | 53 | 29.328 | 14.175 |
| Val2 | Naive | 59.7% | 0.163 | 92.2 | 65.4 | 48.5 | 68.3 | 56.7 | 200 | 1244 | 0 | 0 | 5 | 86 | 26.682 | 13.341 |
| | FRCNN-H | 71.4% | 0.062 | 96.3 | 74.4 | 61.4 | 79.5 | 69.3 | 103 | 919 | 0 | 0 | 4 | 36 | 70.342 | 36.122 |
| | Oracle | 74.1% | 0.000 | 100.0 | 75.1 | 12.8 | 17.0 | 14.6 | 0 | 894 | 0 | 0 | 36 | 36 | 71.000 | 36.459 |
| Val5 | Naive | 28.4% | 0.204 | 78.8 | 39.1 | 23.2 | 46.8 | 31.1 | 477 | 2764 | 0 | 0 | 8 | 93 | 18.122 | 9.015 |
| | FRCNN-H | 61.3% | 0.070 | 95.4 | 64.6 | 33.7 | 49.7 | 40.2 | 143 | 1607 | 1 | 0 | 9 | 60 | 45.123 | 22.562 |
| | Oracle | 42.4% | 0.000 | 100.0 | 43.8 | 13.3 | 30.3 | 18.5 | 0 | 2552 | 0 | 0 | 64 | 64 | 28.812 | 14.618 |
| Tst1 | Naive | 77.6% | 0.152 | 96.0 | 80.9 | 80.9 | 96.0 | 87.8 | 30 | 171 | 1 | 0 | 0 | 28 | 25.000 | 12.500 |
| | FRCNN-H | 30.2% | 0.105 | 100.0 | 30.2 | 30.2 | 100.0 | 46.4 | 0 | 625 | 0 | 0 | 0 | 15 | 16.938 | 8.212 |
| | Oracle | 100.0% | 0.000 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 0 | 0 | 1 | 0 | 0 | 0 | 900.000 | 900.000 |
| Tst2 | Naive | 77.6% | 0.138 | 96.9 | 80.2 | 80.2 | 96.9 | 87.7 | 46 | 353 | 1 | 0 | 0 | 44 | 31.022 | 15.856 |
| | FRCNN-H | 5.7% | 0.133 | 100.0 | 5.8 | 3.5 | 60.2 | 6.6 | 0 | 1677 | 0 | 2 | 2 | 17 | 5.421 | 2.641 |
| | Oracle | 80.6% | 0.000 | 100.0 | 81.6 | 23.9 | 29.3 | 26.3 | 0 | 328 | 2 | 0 | 18 | 25 | 53.778 | 27.396 |
| Tst5 | Naive | 39.7% | 0.168 | 80.0 | 53.1 | 38.0 | 57.3 | 45.7 | 589 | 2078 | 0 | 0 | 9 | 186 | 12.340 | 6.219 |
| | FRCNN-H | 40.0% | 0.099 | 98.0 | 41.3 | 23.2 | 55.1 | 32.6 | 37 | 2605 | 0 | 0 | 17 | 117 | 15.000 | 7.469 |
| | Oracle | 66.7% | 0.000 | 100.0 | 67.8 | 27.2 | 40.1 | 32.4 | 0 | 1427 | 1 | 0 | 50 | 50 | 54.691 | 28.112 |
| Tst10 | Naive | 48.2% | 0.153 | 86.7 | 57.1 | 46.3 | 70.3 | 55.8 | 780 | 3824 | 0 | 0 | 16 | 273 | 18.007 | 8.925 |
| | FRCNN-H | 25.2% | 0.099 | 97.0 | 26.3 | 14.1 | 52.0 | 22.2 | 72 | 6571 | 0 | 3 | 32 | 225 | 9.996 | 4.904 |
| | Oracle | 65.2% | 0.000 | 100.0 | 66.6 | 20.2 | 30.3 | 24.3 | 0 | 2982 | 1 | 0 | 119 | 119 | 46.031 | 23.105 |

Table 8: Full metric evaluation of the 3D tracklets generated from the 3D tracklet association, on all sequences.

| | Method | MOTA ↑ | MOTP ↓ | Prc. ↑ | Rcll. ↑ | ID-Rcll. ↑ | ID-Prc. ↑ | ID-F1 ↑ | FP ↓ | FN ↓ | MT ↑ | ML ↓ | ID Sw. ↓ | Frag. ↓ | MTBF$_s$ ↑ | MTBF$_m$ ↑ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Trn2 | Naive | 41.4% | 0.198 | 85.3 | 50.4 | 28.5 | 48.3 | 35.9 | 1250 | 7133 | 0 | 0 | 40 | 573 | 12.597 | 6.298 |
| | FRCNN-H | 73.8% | 0.066 | 96.9 | 76.3 | 44.0 | 55.9 | 49.2 | 352 | 3407 | 0 | 0 | 14 | 216 | 50.317 | 25.216 |
| | Oracle | 46.2% | 0.000 | 100.0 | 46.2 | 46.2 | 100.0 | 63.2 | 0 | 7738 | 0 | 0 | 0 | 202 | 32.539 | 16.190 |
| Trn5 | Naive | 40.4% | 0.170 | 83.8 | 50.3 | 32.5 | 54.2 | 40.7 | 436 | 2228 | 0 | 0 | 7 | 135 | 16.121 | 7.947 |
| | FRCNN-H | 63.5% | 0.064 | 92.1 | 69.7 | 40.9 | 54.1 | 46.6 | 269 | 1359 | 1 | 0 | 7 | 66 | 44.028 | 22.170 |
| | Oracle | 37.9% | 0.000 | 100.0 | 37.9 | 37.9 | 100.0 | 55.0 | 0 | 2784 | 0 | 0 | 0 | 53 | 29.328 | 14.175 |
| Val2 | Naive | 59.6% | 0.163 | 92.2 | 65.3 | 50.8 | 71.7 | 59.4 | 199 | 1248 | 0 | 0 | 3 | 82 | 27.905 | 13.870 |
| | FRCNN-H | 71.5% | 0.062 | 96.3 | 74.4 | 71.7 | 92.8 | 80.9 | 103 | 920 | 0 | 0 | 2 | 35 | 72.216 | 37.111 |
| | Oracle | 75.1% | 0.000 | 100.0 | 75.1 | 75.1 | 100.0 | 85.8 | 0 | 894 | 0 | 0 | 0 | 36 | 71.000 | 36.459 |
| Val5 | Naive | 28.5% | 0.204 | 78.9 | 39.1 | 23.2 | 46.8 | 31.1 | 476 | 2764 | 0 | 0 | 7 | 93 | 18.122 | 9.015 |
| | FRCNN-H | 61.3% | 0.070 | 95.4 | 64.6 | 40.2 | 59.4 | 48.0 | 143 | 1607 | 1 | 0 | 5 | 60 | 45.123 | 22.562 |
| | Oracle | 43.8% | 0.000 | 100.0 | 43.8 | 43.8 | 100.0 | 60.9 | 0 | 2552 | 0 | 0 | 0 | 64 | 28.812 | 14.618 |
| Tst1 | Naive | 77.6% | 0.152 | 96.0 | 80.9 | 80.9 | 96.0 | 87.8 | 30 | 171 | 1 | 0 | 0 | 28 | 25.000 | 12.500 |
| | FRCNN-H | 30.2% | 0.105 | 100.0 | 30.2 | 30.2 | 100.0 | 46.4 | 0 | 625 | 0 | 0 | 0 | 15 | 16.938 | 8.212 |
| | Oracle | 100.0% | 0.000 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 0 | 0 | 1 | 0 | 0 | 0 | 900.000 | 900.000 |
| Tst2 | Naive | 77.6% | 0.138 | 96.9 | 80.2 | 80.2 | 96.9 | 87.7 | 46 | 353 | 1 | 0 | 0 | 44 | 31.022 | 15.856 |
| | FRCNN-H* | 5.7% | 0.133 | 100.0 | 5.8 | 3.5 | 60.2 | 6.6 | 0 | 1677 | 0 | 2 | 2 | 17 | 5.421 | 2.641 |
| | Oracle | 81.6% | 0.000 | 100.0 | 81.6 | 81.6 | 100.0 | 89.9 | 0 | 328 | 2 | 0 | 0 | 25 | 53.778 | 27.396 |
| Tst5 | Naive | 39.7% | 0.168 | 80.0 | 53.1 | 43.4 | 65.4 | 52.2 | 588 | 2079 | 0 | 0 | 7 | 185 | 12.400 | 6.249 |
| | FRCNN-H | 40.2% | 0.099 | 98.0 | 41.2 | 29.8 | 70.9 | 41.9 | 37 | 2609 | 0 | 0 | 7 | 115 | 15.217 | 7.577 |
| | Oracle | 67.8% | 0.000 | 100.0 | 67.8 | 67.8 | 100.0 | 80.8 | 0 | 1427 | 1 | 0 | 0 | 50 | 54.691 | 28.112 |
| Tst10 | Naive | 48.3% | 0.153 | 86.9 | 57.1 | 48.5 | 73.8 | 58.5 | 768 | 3829 | 0 | 0 | 11 | 268 | 18.313 | 9.075 |
| | FRCNN-H* | 25.2% | 0.099 | 97.0 | 26.3 | 14.1 | 52.0 | 22.2 | 72 | 6571 | 0 | 3 | 32 | 225 | 9.996 | 4.904 |
| | Oracle | 66.6% | 0.000 | 100.0 | 66.6 | 66.6 | 100.0 | 79.9 | 0 | 2982 | 1 | 0 | 0 | 119 | 46.031 | 23.105 |

Table 9: Full metric evaluation of the final 3D tracks generated from the 3D track association, on all sequences. An * indicates that the method did not complete the 3D Tracklet Association step for the sequence, as there were not enough concurrent 3D tracklets at any one point in the sequence. In this case, the 3D tracklet results are reported.