

DeMeTer: Dense Memory-based Point Tracker

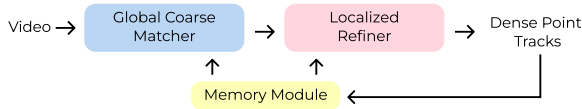
Visual Point Tracking

Track all points selected in the first frame throughout a video:



- Existing trackers **lack long-term robustness** under occlusions and large appearance changes.
- We address this using **global matching with a “memory” of encoded point appearances**.

DeMeTer decouples robustness and accuracy:



- Global Coarse Matcher:** uses appearance memories to robustly re-identify points and initialize dense tracks.
- Localized Refiner:** relies on localized spatio-temporal features to achieve precise localization.
- Memory Module:** dynamically encodes appearance & motion.

Training Procedure

DeMeTer is trained on *synthetic point-tracking* datasets:

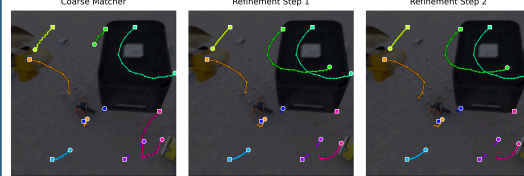
- Global Coarse Matcher:** uses a Gaussian target and dustbin token for robust point localization and presence detection:

$$\mathcal{L}_G = \text{CE}(\mathbf{q}_t, \mathbf{\hat{q}}_t)$$

- Localized Refiner:** uses L_1 loss for accurate localization ($\mathcal{L}_{track}$), BCE for visibility ($\mathcal{L}_{vis}$) and confidence ($\mathcal{L}_{conf}$).
- The total loss is a weighted sum of all components:

$$\mathcal{L}_{total} = 0.1 \cdot \mathcal{L}_G + 0.05 \cdot \mathcal{L}_{track} + \mathcal{L}_{vis} + \mathcal{L}_{conf}$$

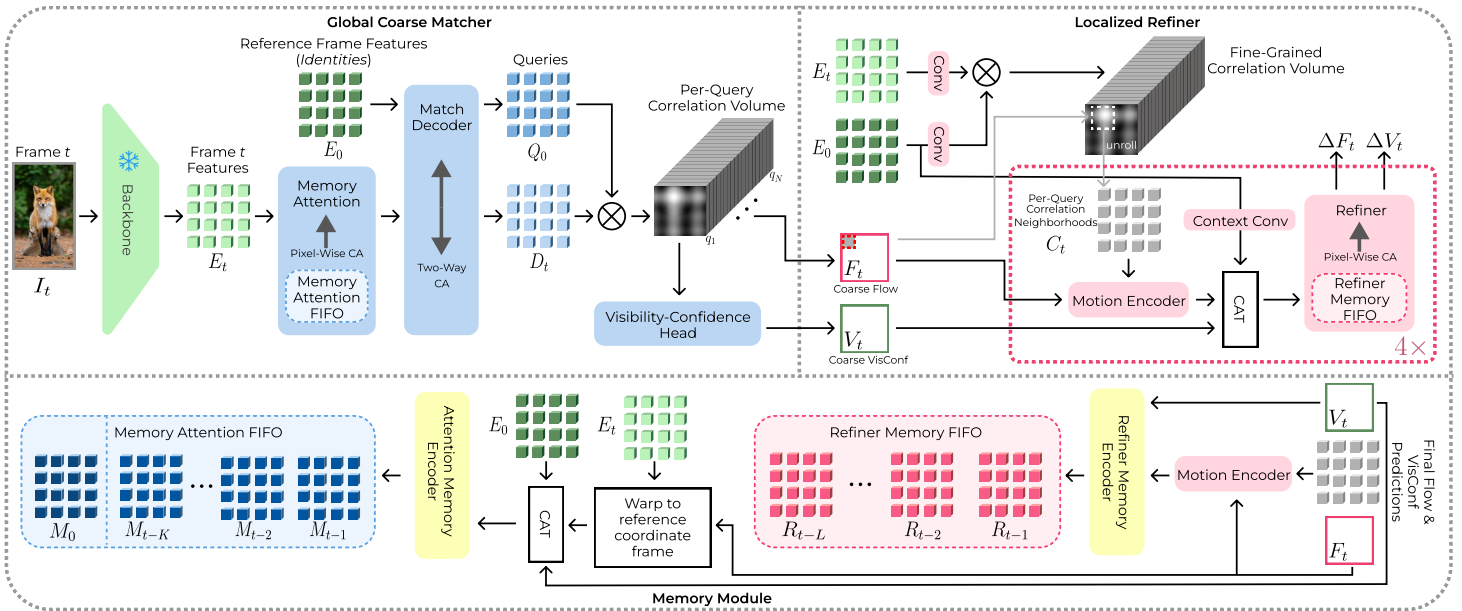
Take-away Message



Step	DAVIS $< \delta x_{avg} \uparrow$	
	AllTracker	DeMeTer
Coarse Matcher	-	66.5
Refiner Step 1	69.0	75.0
Refiner Step 2	74.5	75.3
Refiner Step 3	75.1	75.3
Refiner Step 4	75.2	75.3

- Global matching **robustly localizes points**.
- Refinement **improves localization** and **resolves ambiguities**.
- DeMeTer requires a smaller number of refinement steps than other trackers.

DeMeTer Architecture



Comparison with State-of-the-Art

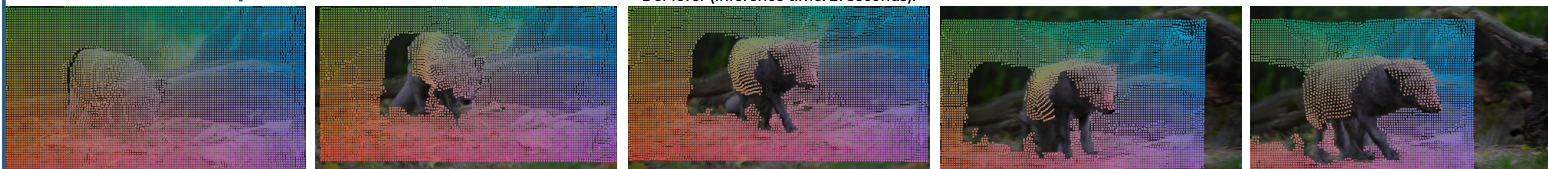
Online Method	DAVIS			RGB-Stacking			Kinetics			RoboTAP			Average		
	AJ $\uparrow$	$< \delta x_{avg} \uparrow$	OA $\uparrow$	AJ $\uparrow$	$< \delta x_{avg} \uparrow$	OA $\uparrow$	AJ $\uparrow$	$< \delta x_{avg} \uparrow$	OA $\uparrow$	AJ $\uparrow$	$< \delta x_{avg} \uparrow$	OA $\uparrow$	AJ $\uparrow$	$< \delta x_{avg} \uparrow$	OA $\uparrow$
TAP-Net	33.0	48.6	78.8	53.5	68.1	86.3	38.5	54.4	80.6	45.1	62.1	82.9	42.5	58.3	82.2
MFT	47.3	66.8	77.8	-	-	-	39.6	60.4	72.7	-	-	-	43.5	63.6	75.3
DOT	53.3	67.8	85.4	61.3	75.3	86.5	45.3	58.0	81.4	51.9	62.9	79.9	53.0	66.0	83.3
CoTracker3 (Stride=1)	56.0	70.8	87.0	64.3	80.9	88.6	-	-	-	52.9	70.1	82.6	57.7	73.9	86.1
Online TAPIR	56.7	70.2	85.7	67.7	78.6	89.5	51.5	64.4	85.2	59.1	69.9	86.1	58.8	70.8	86.6
AllTracker (Stride=1)	58.5	71.1	85.9	77.6	88.3	91.4	55.3	68.4	87.0	64.5	77.9	89.0	64.0	76.4	88.3
SPOT	61.5	75.0	88.9	73.3	86.4	90.3	50.2	62.7	85.5	60.7	73.6	87.5	61.4	74.4	88.1
ReTracker	63.7	77.5	88.2	68.2	80.7	85.8	54.5	67.0	86.4	65.5	78.1	89.9	63.0	75.8	87.6
DeMeTer	62.2	75.3	89.5	78.3	88.0	92.5	58.9	72.0	90.3	71.2	82.4	93.4	67.7	79.4	91.4

DeMeTer sets a new state-of-the-art among **online** as well as **offline** trackers, while delivering dense output.

Method	Params.	Training	Cro.	Dav.	Dri.	Ego.	Kin.	Rgb.	Rob.	Avg.
AllTracker (offline)	16.48M	Kubric	42.3	75.2	66.1	60.3	71.3	90.1	82.2	69.6
DeMeTer (online)	23.33M (10.51M Trained)	Kubric	44.1	75.3	67.6	60.6	72.0	88.0	82.4	70.0

Visual Comparison

DeMeTer (Inference time: 21 seconds):



ReTracker (Inference time: 40 minutes)

