Chair of Connected Mobility
TUM School of Computation, Information and Technology
Technical University of Munich

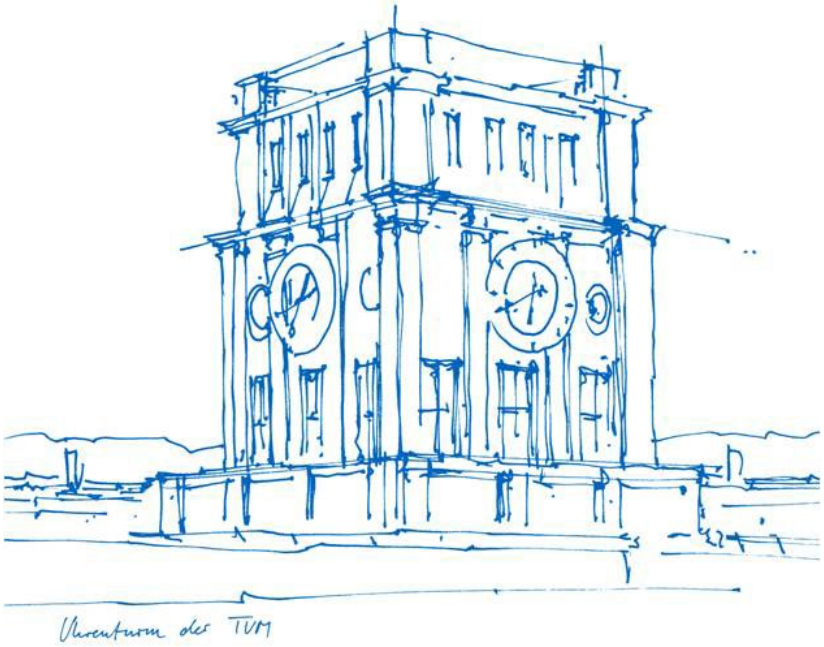
Test-Time Accuracy Indicators for Object Detection

Jul 08, 2026

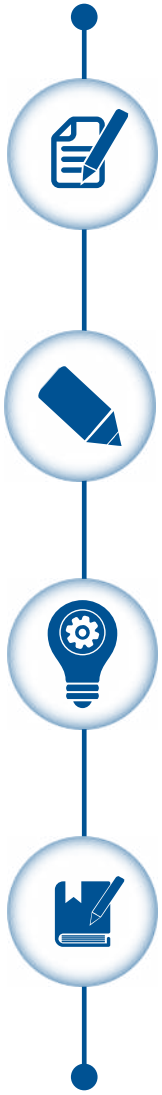
Bachelor Thesis

Student
Zhenghao Lu

Supervisor
Wei Geng, M.Phil., M.Eng



Outline



Part 1 Introduction and Related Work

Part 2 Proposed PTC-IoU Method

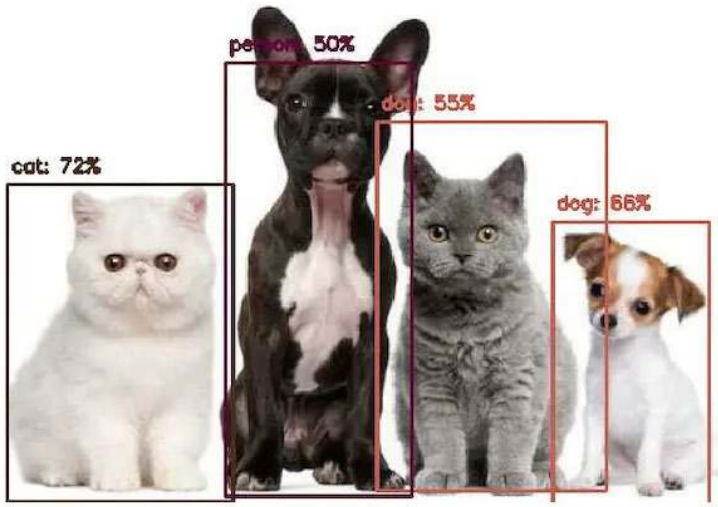
Part 3 Experimental Setup & Evaluation

Part 4 Conclusion & Future Work

Part 1 Introduction and Related Work

Background & Motivation — Why adaptive video object detection?

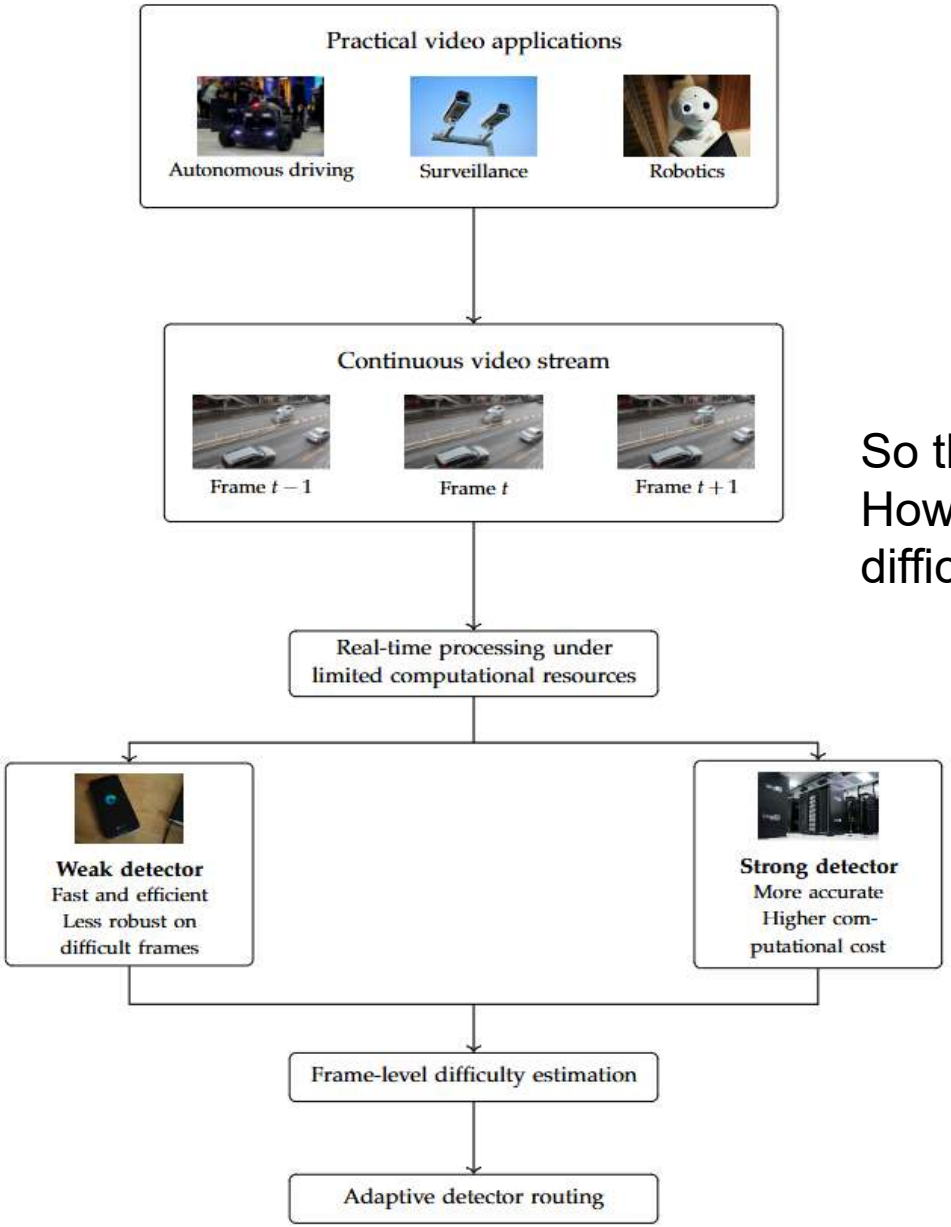
- Object detection has achieved strong performance with CNN-based and Transformer-based detectors.



- But in practical applications:
  - frames arrive continuously
  - computation is limited

- Trade-off:
  - Strong detector → accurate and reliable but expensive
  - Weak detector → fast and efficient but less robust

Motivation Example: Adaptive Detector Routing



So the central problem becomes: How can we estimate frame-level difficulty during testing?

Key Observation — Temporal consistency in videos

- Adjacent video frames are temporally related.

- **Main assumption:**

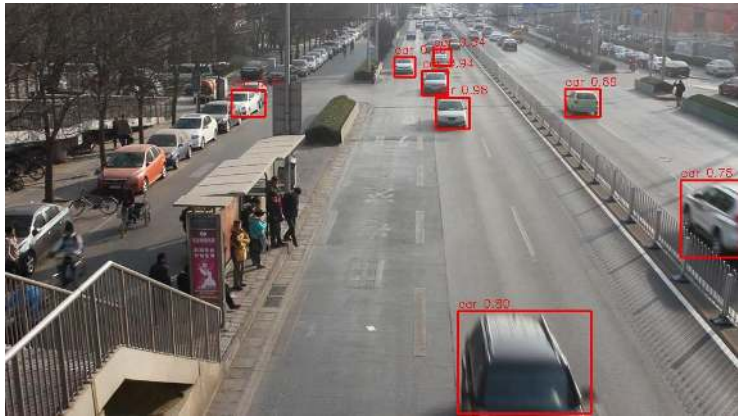
  Easy frames should produce temporally continuous and stable detections.

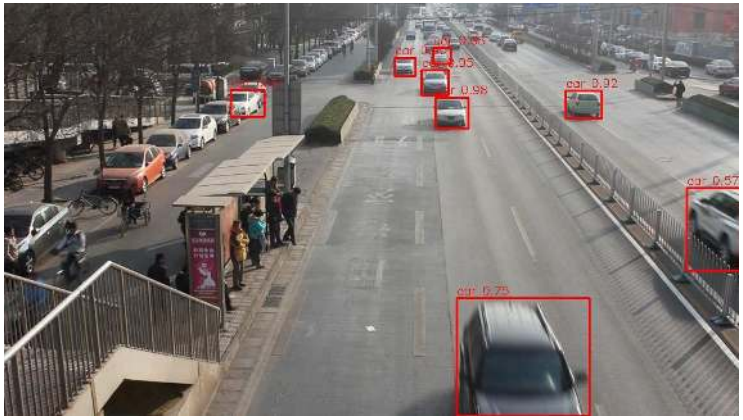
  If detections become discontinuous or unstable across adjacent frames, the current frame is more likely to be difficult for the detector.

Part 1 Introduction and Related Work

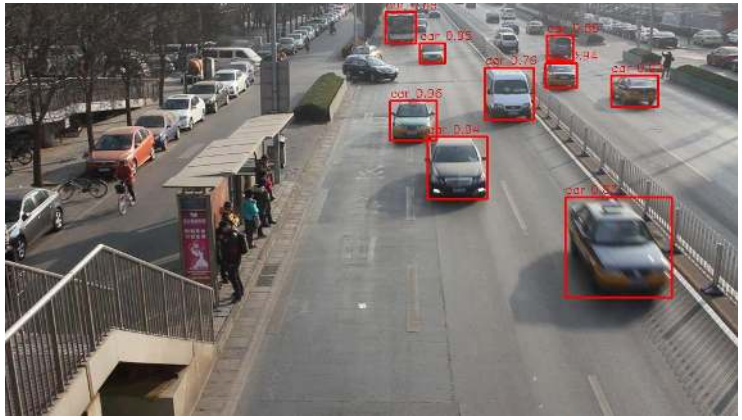
Key Observation — Temporal consistency in videos

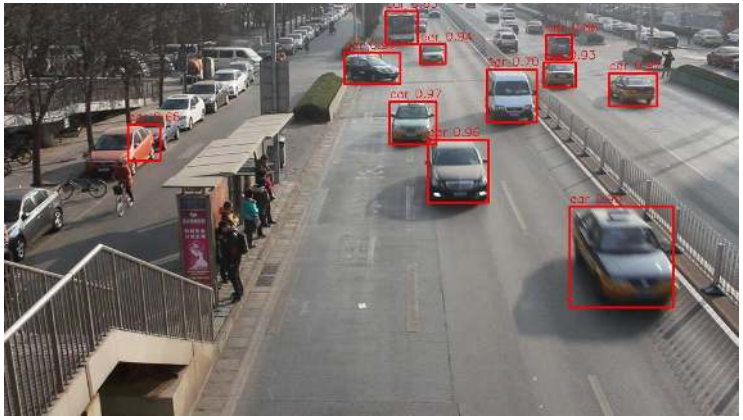
temporally stable detection pair:





temporally unstable detection pair:





Part 1 Introduction and Related Work

Related Work

- Temporal consistency:

  useful in videos, but rarely used directly as a test-time difficulty proxy.

- Quality estimation:

  IoU-Net replaces classification scores with localization confidence. (predicted IoU between a detected bbox and the corresponding gt box.)

  Generalized Focal Loss (GFL) jointly models classification and localization quality.

  Require additional learned quality components and need extra inference operations.

- Adaptive routing:

  improves efficiency, but requires reliable frame-level difficulty signals.

- Evaluation metrics:

  mAP / HOTA inspire evaluation and decomposition, but are not test-time indicators.

- Gap:

  lightweight, training-free and test-time frame-level difficulty indicator for adaptive video object detection.

Part 2 Proposed PTC-IoU Method

Method Overview — HOTA-inspired Design of PTC-IoU (Proxy-based Temporal Consistency IoU)

Inspiration from HOTA:

HOTA evaluates tracking quality from three aspects:

- Detection: are objects correctly detected?
- Association: are identities consistently associated?
- Localization: are boxes spatially accurate?

Mapping:

HOTA Detection → PTC-Det: object presence consistency

HOTA Association → PTC-Ass: matching ambiguity between frames

HOTA Localization → PTC-Loc: motion-aware box stability

Final score:

PTC-IoU combines these three consistency sub-dimensions into one frame-level difficulty indicator.

Part 2 Proposed PTC-IoU Method

Temporal Matching Between Adjacent Frames

Given adjacent frames:

$$D_{t-1} = \{d_i^{t-1}\}_{i=1}^{N_{t-1}} \qquad D_t = \{d_j^t\}_{j=1}^{N_t}$$

Each detection:

d = (b, s, c)

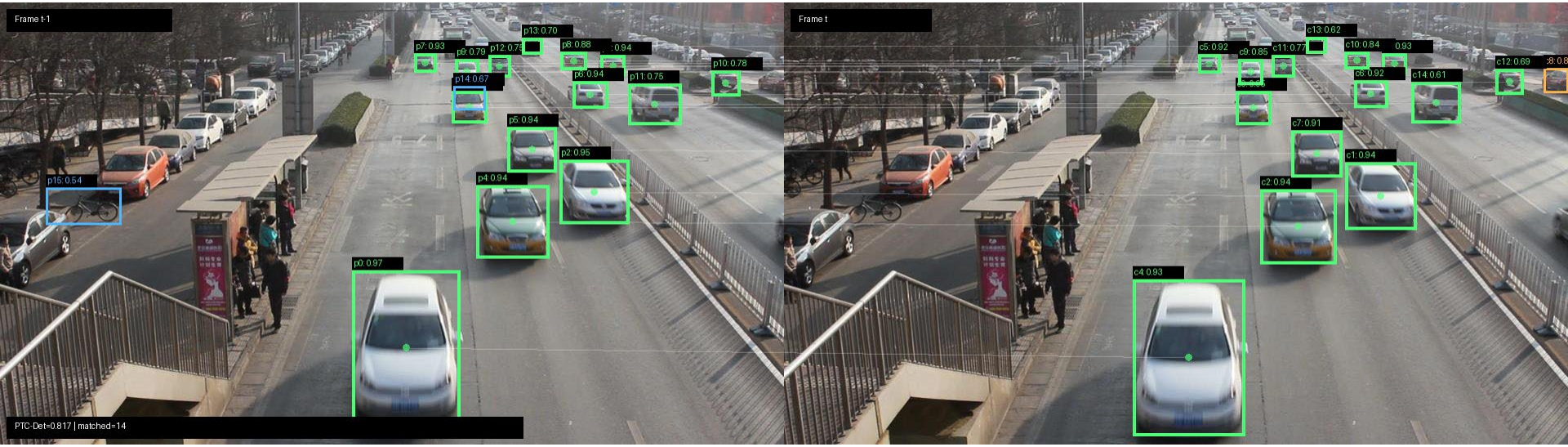
- b → bounding box
- s → confidence score
- c → class

$$\mathrm{IoU}(d_i^{t-1}, d_j^t) = \frac{\mathrm{Area}(b_i^{t-1} \cap b_j^t)}{\mathrm{Area}(b_i^{t-1} \cup b_j^t)}$$

$$m(j) = \arg\max_i \mathrm{IoU}(d_i^{t-1}, d_j^t)$$

Matching:

- class-aware IoU matching
- one-to-one matched detection pairs
- used to measure temporal consistency

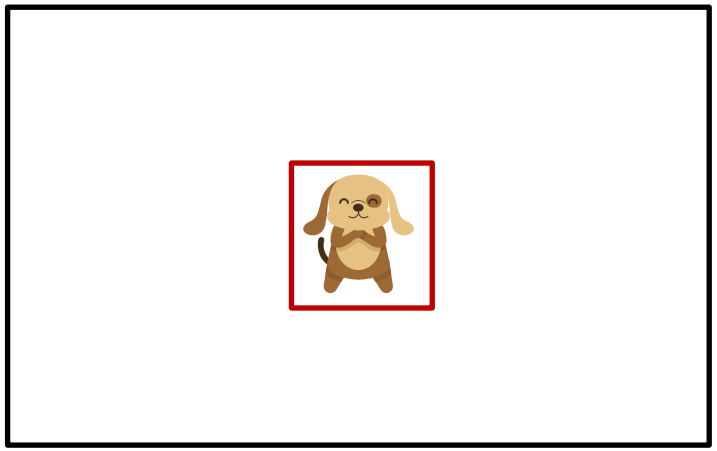


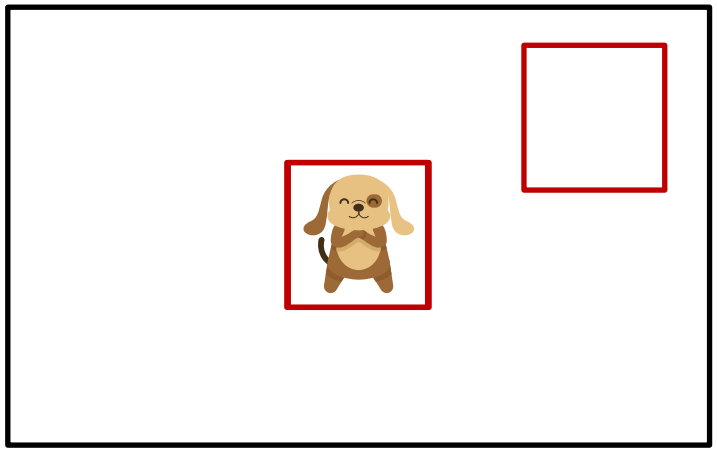
Temporal Matching tries to find the same object in the previous frame

Part 2 Proposed PTC-IoU Method

Three Temporal Consistency Components - Intuition

PTC-Det **Detection Consistency**	PTC-Ass Association Consistency	PTC-Loc Localization Consistency
Does the object remain detected?	Is the temporal matching clear?	Is the bounding box motion smooth?
Missing or new detections indicate temporal instability.	Ambiguous matches indicate crowded or occluded scenes.	Deviation from predicted motion indicates unstable localization.
Confidence-weighted Jaccard similarity	Gap between best and second-best match	Overlap + center consistency





Part 2 Proposed PTC-IoU Method

Three Temporal Consistency Components – PTC-Det Formulation

PTC-Det **Detection Consistency**	PTC-Ass Association Consistency	PTC-Loc Localization Consistency
Does the object remain detected?	Is the temporal matching clear?	Is the bounding box motion smooth?
Missing or new detections indicate temporal instability.	Ambiguous matches indicate crowded or occluded scenes.	Deviation from predicted motion indicates unstable localization.
Confidence-weighted Jaccard similarity	Gap between best and second-best match	Overlap + center consistency

number of matched detection pairs

$$\text{PTC-Det}_t^{\text{count}} = \frac{|\mathcal{M}_t|}{N_{t-1} + N_t - |\mathcal{M}_t|}.$$

union of detections from two frames

confidence-weighted Jaccard similarity

$$w_{ij} = \min(s_i^{t-1}, s_j^t).$$

$$W_t^{\text{match}} = \sum_{(i,j)\in\mathcal{M}_t} w_{ij}. \qquad W_{t-1} = \sum_{i=1}^{N_{t-1}} s_i^{t-1}, \qquad W_t = \sum_{j=1}^{N_t} s_j^t.$$

$$\text{PTC-Det}_t = \frac{W_t^{\text{match}}}{W_{t-1} + W_t - W_t^{\text{match}}}.$$

Part 2 Proposed PTC-IoU Method

Three Temporal Consistency Components - Intuition

PTC-Det
Detection Consistency

Does the object remain detected?

Missing or new detections indicate temporal instability.

Confidence-weighted Jaccard similarity

PTC-Ass
Association Consistency

Is the temporal matching clear?

Ambiguous matches indicate crowded or occluded scenes.

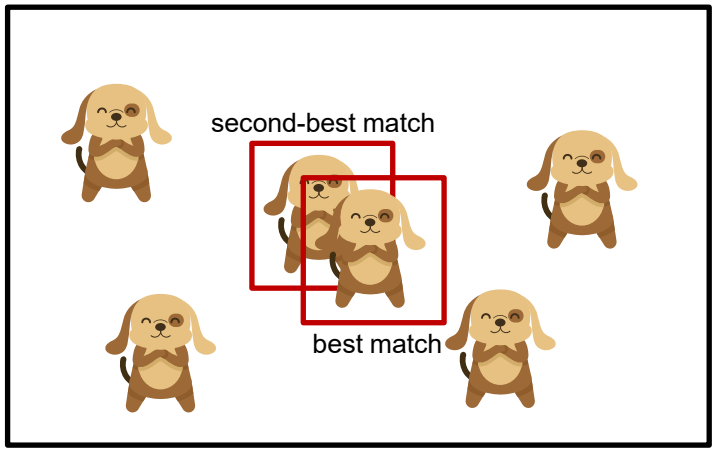
Gap between best and second-best match

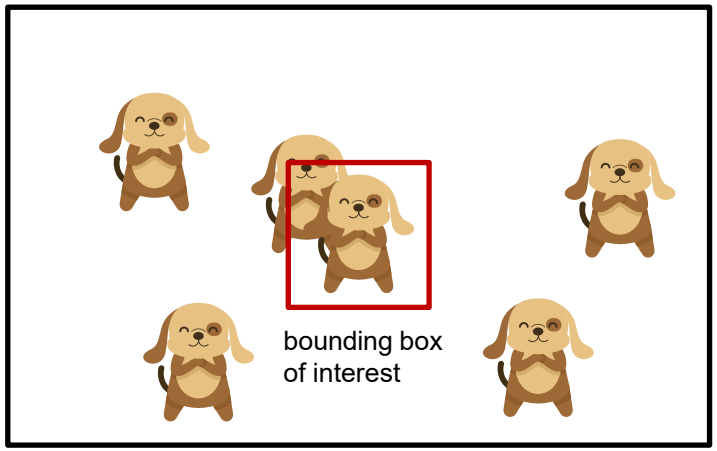
PTC-Loc
Localization Consistency

Is the bounding box motion smooth?

Deviation from predicted motion indicates unstable localization.

Overlap + center consistency





Part 2 Proposed PTC-IoU Method

Three Temporal Consistency Components – PTC-Ass Formulation

PTC-Det Detection Consistency	**PTC-Ass Association Consistency**	PTC-Loc Localization Consistency
Does the object remain detected?	**Is the temporal matching clear?**	Is the bounding box motion smooth?
Missing or new detections indicate temporal instability.	**Ambiguous matches indicate crowded or occluded scenes.**	Deviation from predicted motion indicates unstable localization.
Confidence-weighted Jaccard similarity	**Gap between best and second-best match**	Overlap + center consistency

large best–second gap → clear association
small gap → ambiguous association

best matching score:

$$s_1(j) = \max_i \mathrm{IoU}(d_i^{t-1}, d_j^t),$$

second-best matching score:

$$s_2(j) = \max_{i \neq m(j)} \mathrm{IoU}(d_i^{t-1}, d_j^t),$$

$$a_j = \frac{s_1(j) - s_2(j)}{s_1(j) + \epsilon},$$

$$\mathrm{PTC\text{-}Ass}_t = \frac{\sum_{j=1}^{M_t} s_j^t a_j}{\sum_{j=1}^{M_t} s_j^t},$$

Part 2 Proposed PTC-IoU Method

Three Temporal Consistency Components - Intuition

PTC-Det
Detection Consistency

Does the object remain detected?

Missing or new detections indicate temporal instability.

Confidence-weighted Jaccard similarity

PTC-Ass
Association Consistency

Is the temporal matching clear?

Ambiguous matches indicate crowded or occluded scenes.

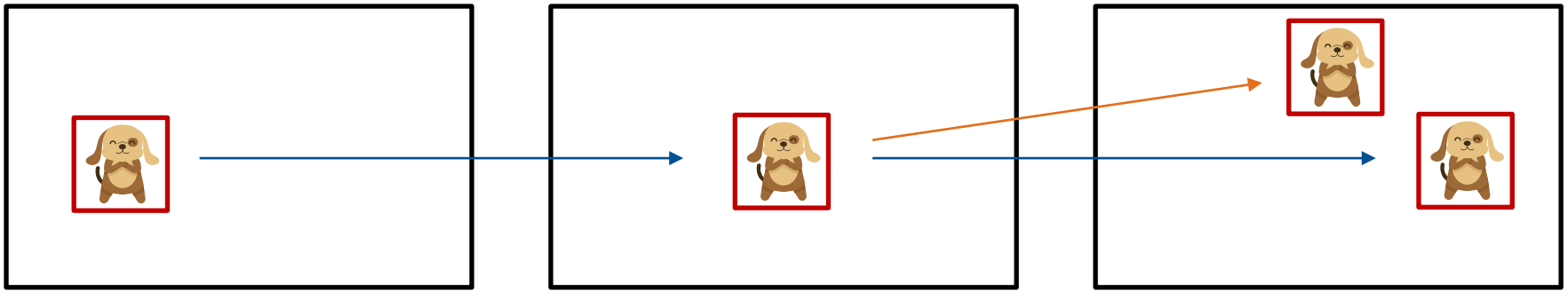
Gap between best and second-best match

PTC-Loc
Localization Consistency

Is the bounding box motion smooth?

Deviation from predicted motion indicates unstable localization.

Overlap + center consistency



Part 2 Proposed PTC-IoU Method

Three Temporal Consistency Components – PTC-Loc Formulation

PTC-Det Detection Consistency	PTC-Ass Association Consistency	**PTC-Loc Localization Consistency**
Does the object remain detected?	Is the temporal matching clear?	**Is the bounding box motion smooth?**
Missing or new detections indicate temporal instability.	Ambiguous matches indicate crowded or occluded scenes.	**Deviation from predicted motion indicates unstable localization.**
Confidence-weighted Jaccard similarity	Gap between best and second-best match	**Overlap + center consistency**

overlap with predicted box + center consistency

center coordinates of a bbox $c = (x, y).$

center similarity $$\mathrm{CS}(i,j) = \exp\left(-\frac{(\hat{c}_{i,x}^t - c_{j,x}^t)^2 + (\hat{c}_{i,y}^t - c_{j,y}^t)^2}{C^2(\hat{b}_i^t, b_j^t) + \epsilon}\right),$$

motion vector $\Delta c_i^{t-1} = c_i^{t-1} - c_i^{t-2}.$

combination $$l_j = \left(\mathrm{IoU}(\hat{b}_i^t, b_j^t)\right)^{\alpha} \cdot \left(\mathrm{CS}(i,j)\right)^{1-\alpha},$$

the predicted center position $\hat{c}_i^t = c_i^{t-1} + \Delta c_i^{t-1}.$

$$\text{PTC-Loc}_t = \frac{\sum_{j=1}^{M_t} s_j^t l_j}{\sum_{j=1}^{M_t} s_j^t},$$

localization overlap consistency $\mathrm{IoU}(\hat{b}_i^t, b_j^t).$

Part 2 Proposed PTC-IoU Method

Composite PTC-IoU

- Combine complementary consistency sub-dimensions:

- $\text{PTC-IoU}_t = \mathcal{F}\left(\text{PTC-Det}_t, \text{PTC-Ass}_t, \text{PTC-Loc}_t\right),$

  Aggregation:

  • weighted combination

  • parameters selected from training split

  • lower PTC-IoU → higher estimated difficulty

- Example:

  weighted arithmetic mean $\text{PTC-IoU}_t = w_d \cdot \text{PTC-Det}_t + w_a \cdot \text{PTC-Ass}_t + w_l \cdot \text{PTC-Loc}_t,$

  weighted geometric mean $\text{PTC-IoU}_t = \left(\text{PTC-Det}_t^{w_d} \cdot \text{PTC-Ass}_t^{w_a} \cdot \text{PTC-Loc}_t^{w_l}\right).$

Part 3 Experimental Setup & Evaluation

Experimental Setup

Dataset	MOT17 pedestrian sequences
Split	Video-level train / validation split
Weak detector	Faster R-CNN MobileNet V3 Large 320 FPN
Strong detector	Faster R-CNN ResNet50 FPN V2
Method	PTC-IoU
Baselines	IoU-Net, GFL
Routing references	Random routing, Oracle routing
Metrics	Ranking: Pairwise, Kendall, Pearson, Spearman; Routing: AP, Precision, Recall
Hardware	NVIDIA A40 GPU

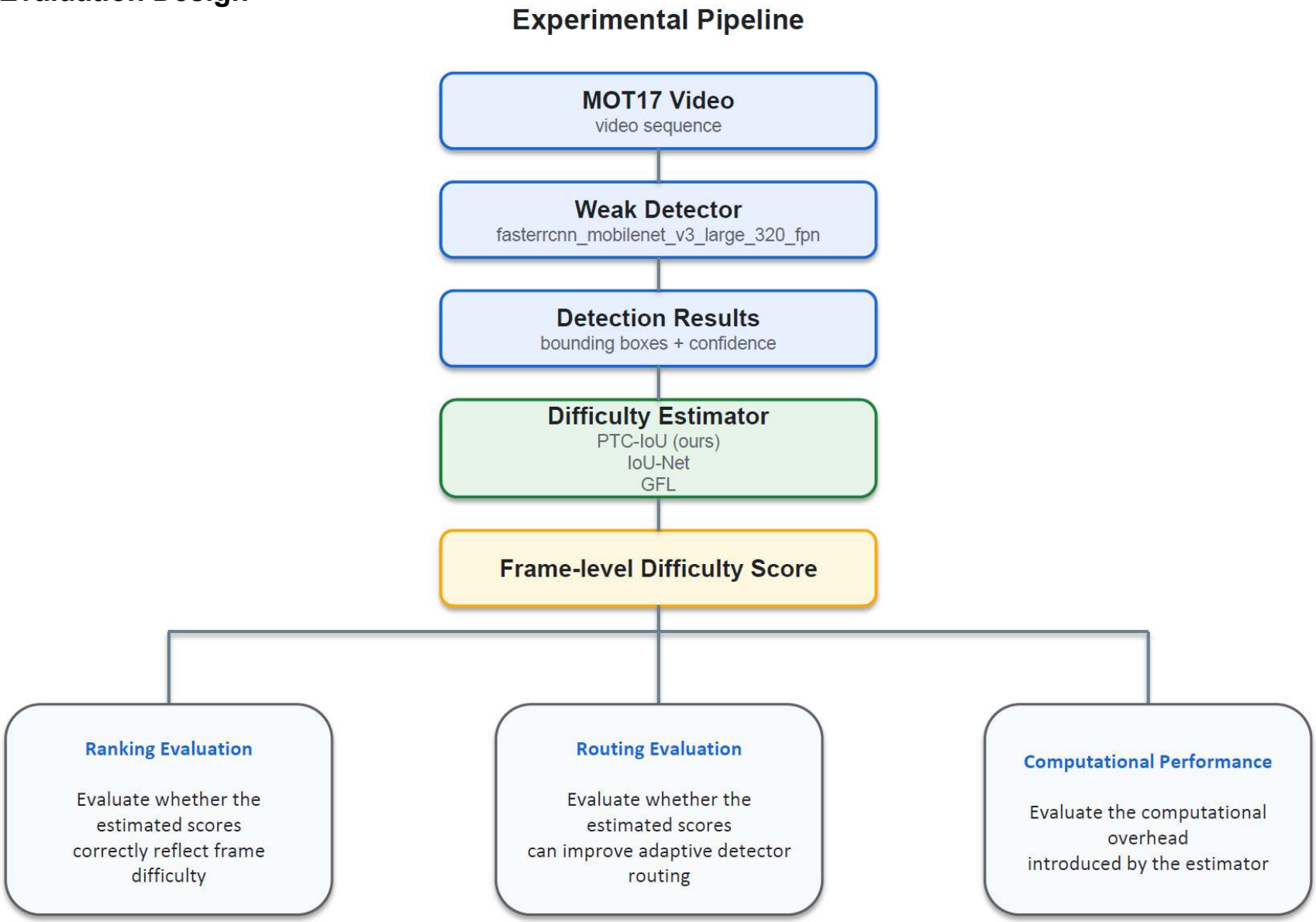
MOT17-02

MOT17-09

MOT17-11

Part 3 Experimental Setup & Evaluation

Evaluation Design



Part 3 Experimental Setup & Evaluation

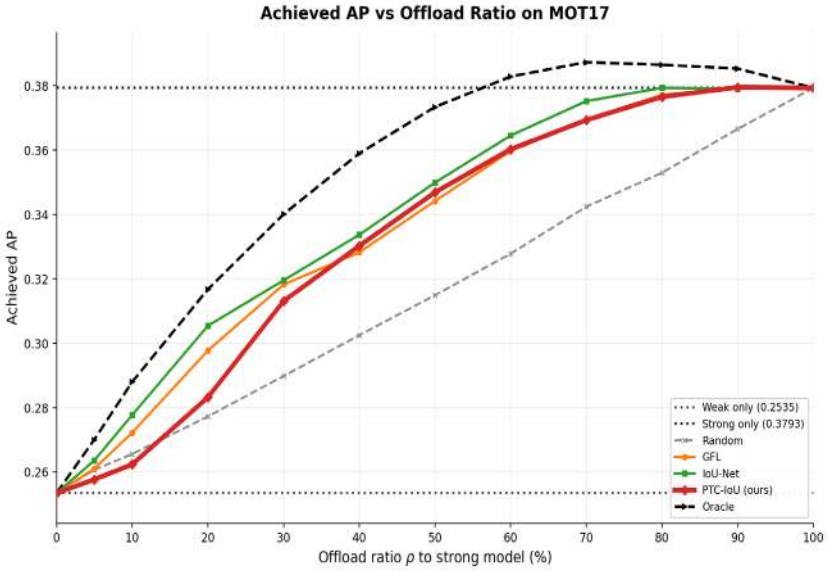
Ranking Evaluation Results

Target	Split	Method	Pairwise ↑	Kendall τ ↑	Spearman ρ ↑	Pearson r ↑
AP	Split 1	GFL	0.748	0.487	0.679	0.667
		IoU-Net	**0.757**	**0.509**	**0.708**	**0.697**
		PTC-IoU (Ours)	0.709	0.406	0.568	0.396
	Split 2	GFL	0.657	0.315	0.464	0.451
		IoU-Net	0.706	0.397	0.581	0.554
		PTC-IoU (Ours)	**0.768**	**0.531**	**0.738**	**0.694**
Precision	Split 1	GFL	**0.827**	**0.643**	**0.837**	**0.821**
		IoU-Net	0.813	0.621	0.817	0.810
		PTC-IoU (Ours)	0.772	0.526	0.696	0.468
	Split 2	GFL	0.734	0.458	0.638	0.669
		IoU-Net	0.746	0.485	0.677	**0.680**
		PTC-IoU (Ours)	**0.771**	**0.524**	**0.719**	0.679
Recall	Split 1	GFL	0.736	0.464	0.655	0.660
		IoU-Net	**0.740**	**0.475**	**0.668**	**0.671**
		PTC-IoU (Ours)	0.682	0.354	0.508	0.353
	Split 2	GFL	0.662	0.315	0.457	0.449
		IoU-Net	0.709	0.399	0.576	0.553
		PTC-IoU (Ours)	**0.771**	**0.535**	**0.736**	**0.708**

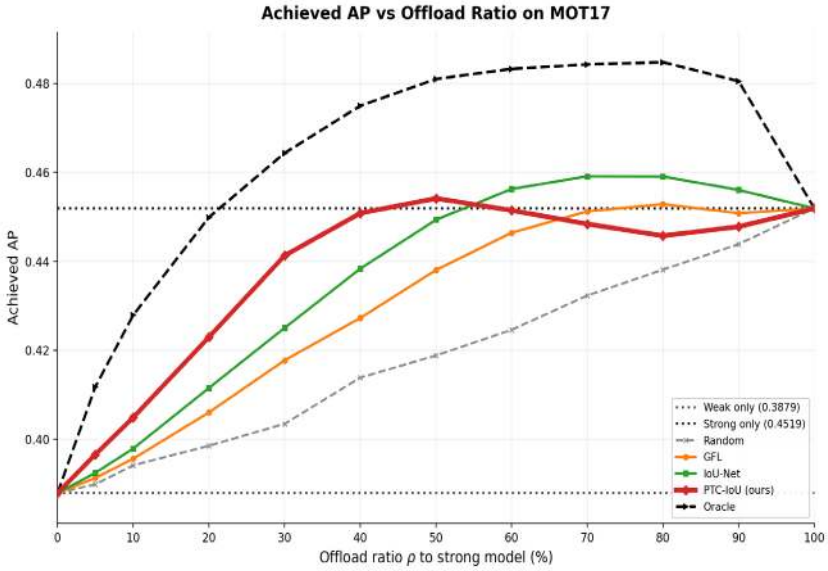
Part 3 Experimental Setup & Evaluation

Routing Evaluation Results

Split 1



Split 2

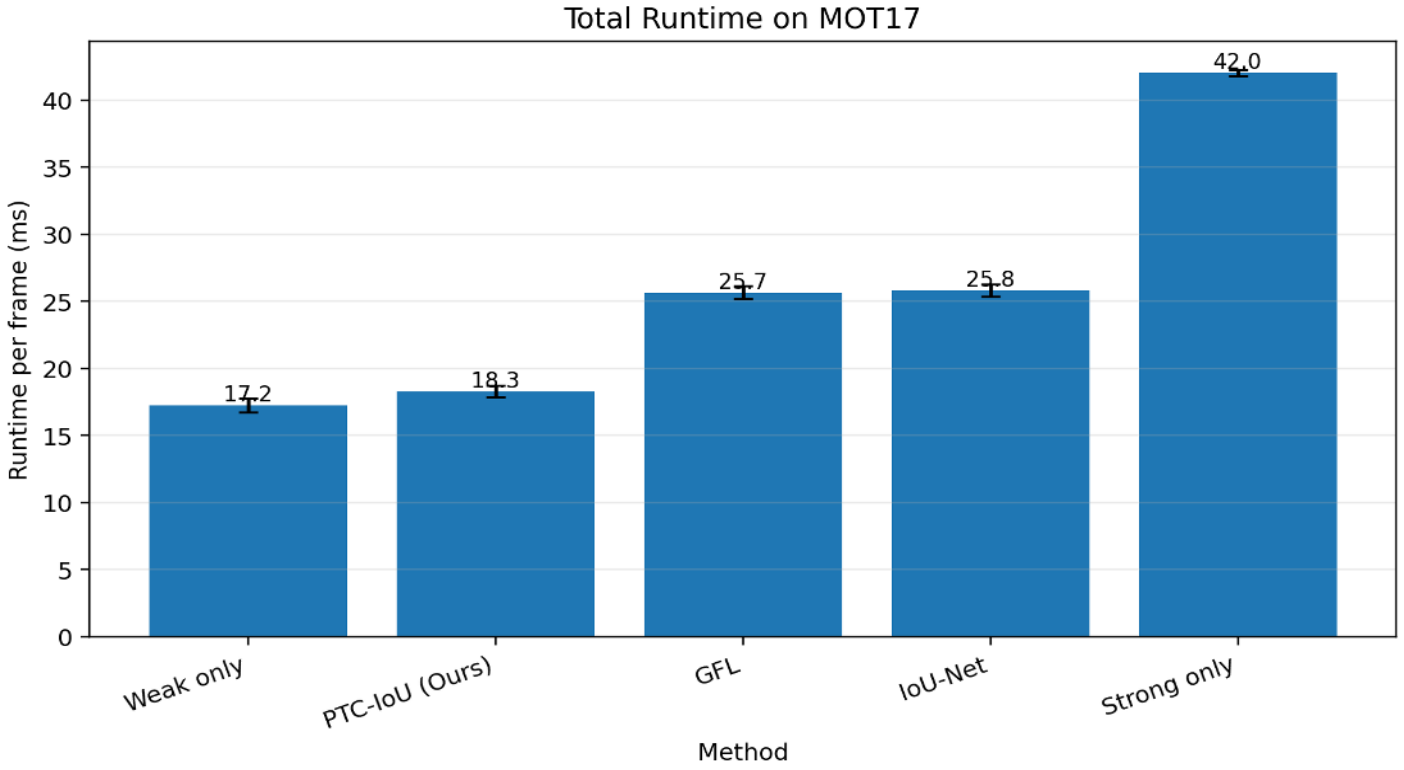


Routing performance summary at routing budget ρ = 0.3

Method	Split 1			Split 2		
	AP ↑	Precision ↑	Recall ↑	AP ↑	Precision ↑	Recall ↑
Weak only	0.253	0.165	0.309	0.388	0.210	0.447
Random routing	0.290	0.193	0.356	0.403	0.227	0.467
GFL	0.318	0.227	0.383	0.418	0.237	0.487
IoU-Net	**0.320**	**0.227**	**0.384**	0.425	0.240	0.497
PTC-IoU (Ours)	0.313	0.222	0.376	**0.441**	**0.252**	**0.514**
Oracle	0.340	0.232	0.402	0.464	0.257	0.529
Strong only	0.379	0.265	0.468	0.452	0.270	0.519

Part 3 Experimental Setup & Evaluation

Computational Performance Results



Method	ms/frame ↓	FPS ↑	Overhead ↓	Slowdown ↓
Weak only	17.25 ± 0.53	58.05 ± 1.86	0.00 ± 0.00	1.00x
PTC-IoU (Ours)	18.29 ± 0.44	54.70 ± 1.33	1.04 ± 0.24	1.06x
GFL	25.65 ± 0.48	39.00 ± 0.75	8.40 ± 0.10	1.49x
IoU-Net	25.82 ± 0.46	38.75 ± 0.70	8.57 ± 0.10	1.50x
Strong only	42.04 ± 0.21	23.79 ± 0.12	24.79 ± 0.41	2.44x

Part 4 Conclusion & Future Work

Conclusion

• PTC-IoU estimates frame-level difficulty using temporal consistency.

• It combines three sub-dimensions:
Detection consistency
Association consistency
Localization consistency

• Results show:
meaningful difficulty ranking
improved adaptive routing compared with random routing
low additional runtime overhead

Limitations

• Dataset sensitivity
• Limited parameter / detector generalization
• Gap to oracle routing
• Simple aggregation strategy

Future Work

• Evaluate on more diverse datasets
• Study thresholds, smoothing, and detector pairs
• Improve components and aggregation strategy
• Extend to multi-model routing and edge deployment

Thank you for your attention

Questions?

Zhenghao LU

Munich, 08.07.2026