

TrackPFN: Multiple Object Tracking via (Tabular) Prior Data Fitted Networks

Max Kahl¹, Margret Keuper^{1,2}

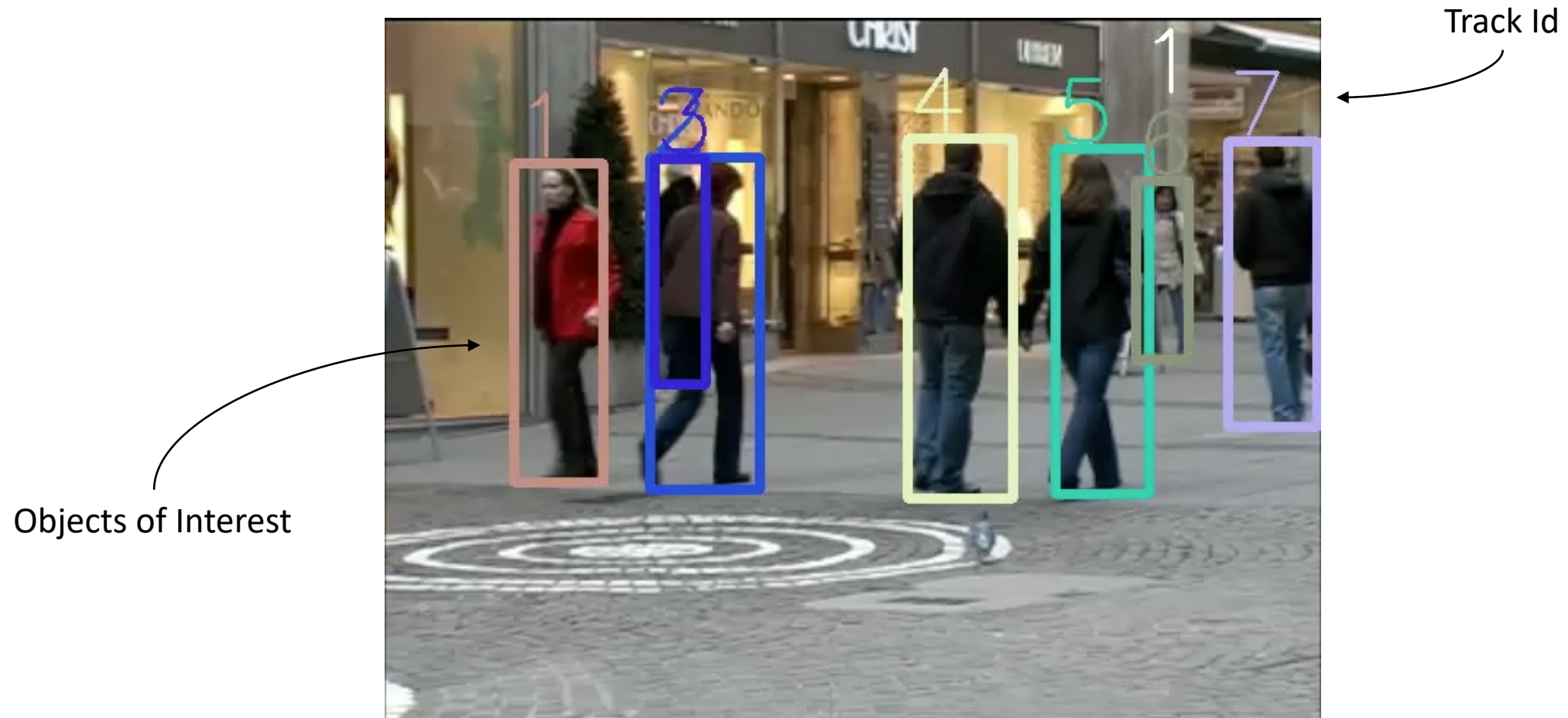
¹Max-Planck Institute for Informatics, ²University of Mannheim

MPI Winter Symposium 2026

Multiple Object Tracking (MOT)



Multiple Object Tracking (MOT)

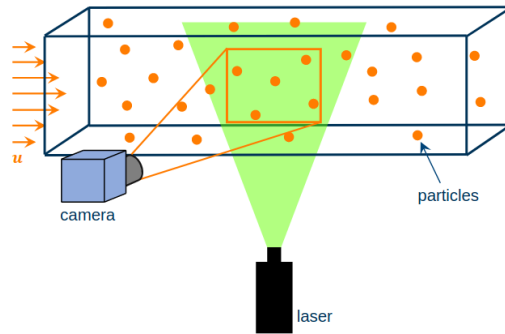


MOT Lacks a Generalist

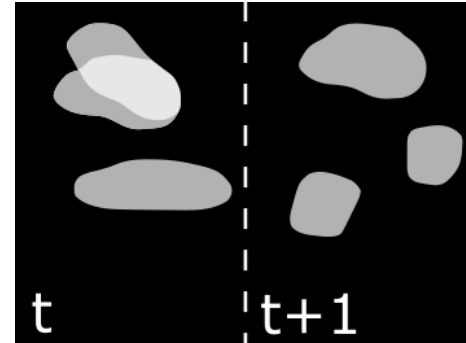
“Classical” Computer Vision



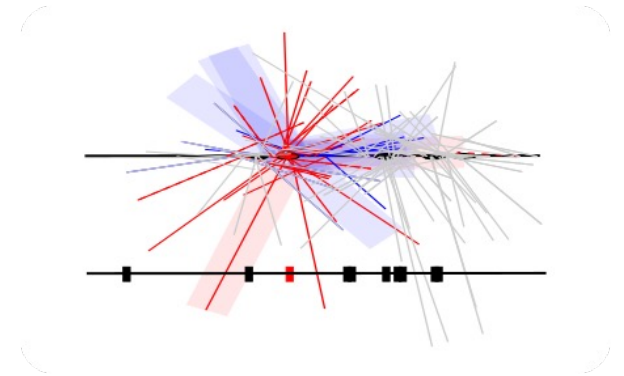
Microfluidics



Biology



Particle Physics



Data Challenges:

Scarcity

Uncertainty

Domain Shift

Lack of Expressive
Features

Our Goal: A Generalist Association Module for Tracking-by-Detection

How Could This Generalist Look Like?

Desiderata: • Feature-Agnostic • Differentiable • Out-Of-the-Box



Idea: In-Context Learning

How Could This Generalist Look Like?

Desiderata: • Feature-Agnostic • Differentiable • Out-Of-the-Box

→ Idea: In-Context Learning via PFNs

Prior Data Fitted Networks (PFNs)

Sample Prior Data Sets $D^{(i)} \sim p(\mathcal{D})$

$$\begin{aligned} D^{(1)} &= D_{\text{train}}^{(1)} \cup \{(x_{\text{test}}^{(1)}, y_{\text{test}}^{(1)})\} \\ &\vdots \\ D^{(K)} &= D_{\text{train}}^{(K)} \cup \{(x_{\text{test}}^{(K)}, y_{\text{test}}^{(K)})\} \end{aligned}$$

Train via PFN Loss

$$-\sum_{i=1}^K \log q_{\theta} \left(y_{\text{test}}^{(i)} \mid x_{\text{test}}^{(i)}, D_{\text{train}}^{(i)} \right)$$

PFN with parameters θ^*

Actual Dataset and Test Input

$$(D_{\text{train}}, x_{\text{test}})$$

Provable Bayesian Inference

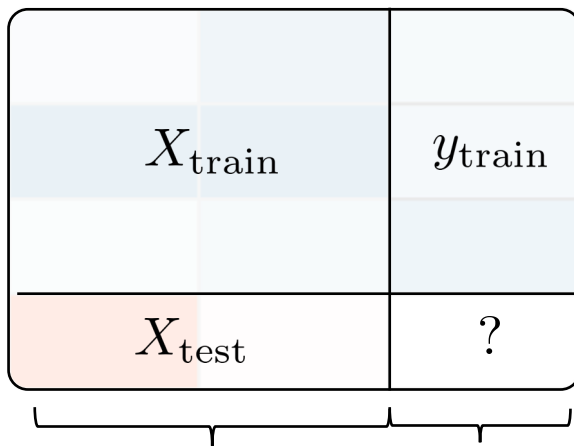
$$q_{\theta^*}(y_{\text{test}} \mid x_{\text{test}}, D_{\text{train}}) \approx p(y_{\text{test}} \mid x_{\text{test}}, D_{\text{train}})$$

(Müller et al., 2021)

Tabular Prior Data Fitted Network (TabPFN)

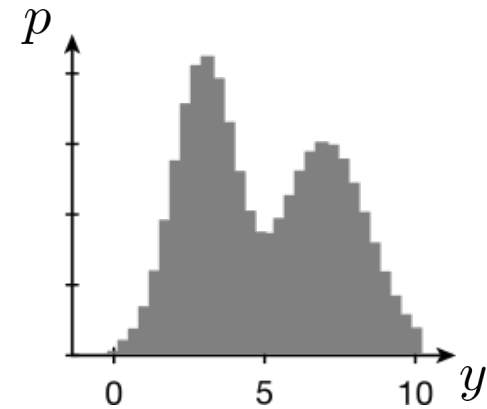
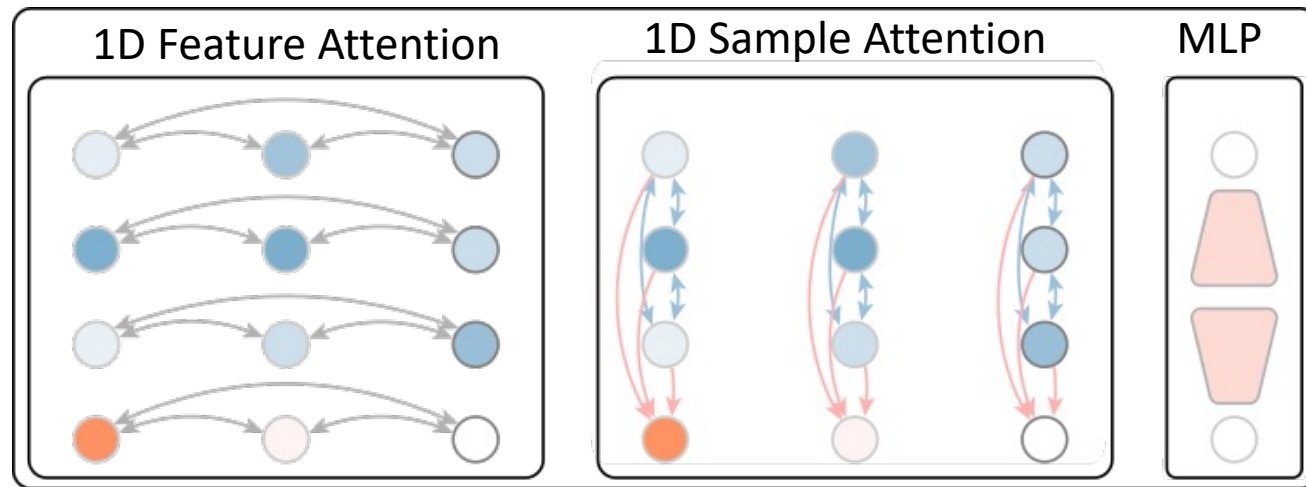
TabPFN: PFN for Tabular Data with Transformer Architecture

Input Data → 2D TabPFN Layer (12x) → Predictive Distribution



Features

Target



(Hollman et al., 2025)

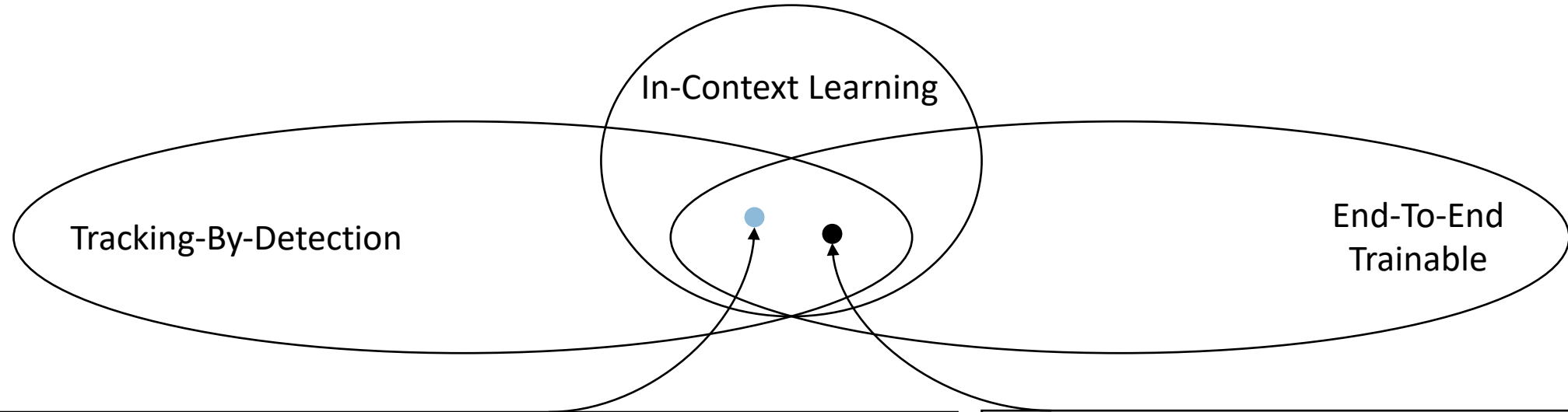
Why?

Classification &
Regression

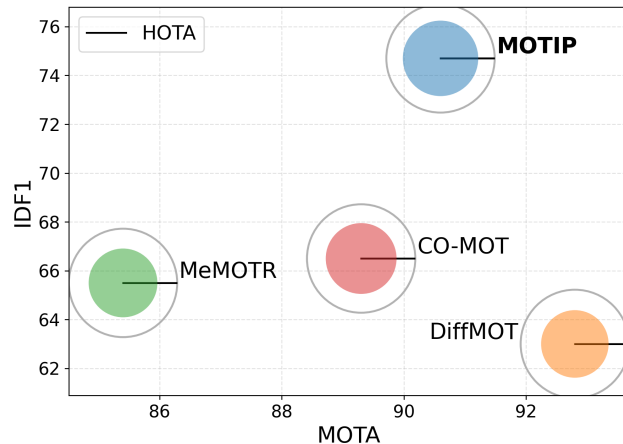
Powerful Prior Based On
Structural Causal Models

Impressive Results on
Time Series

What has been done so far?



MOTIP (Gao et al., 2025)



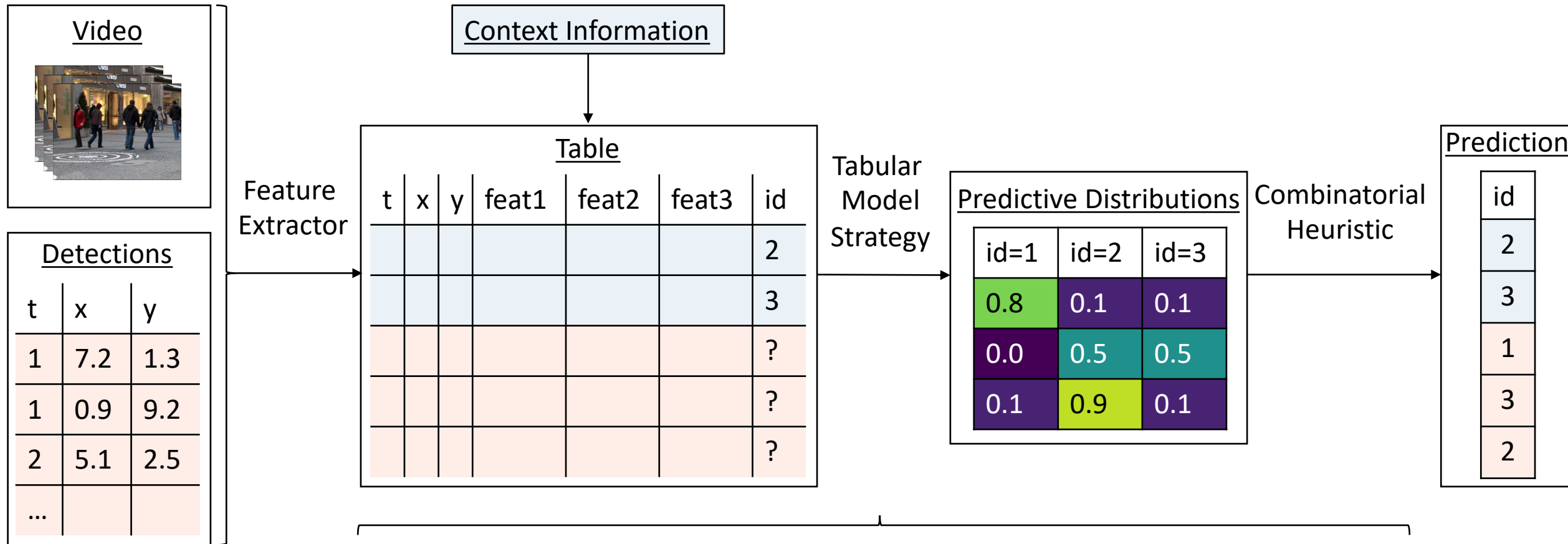
✗ Task-Specific Association Module

✗ Additional Tracking Cues
Require Retraining

TrackPFN (Ours)

Expect Better Generalization Due to
Structural Prior

MOT via Tabular Models



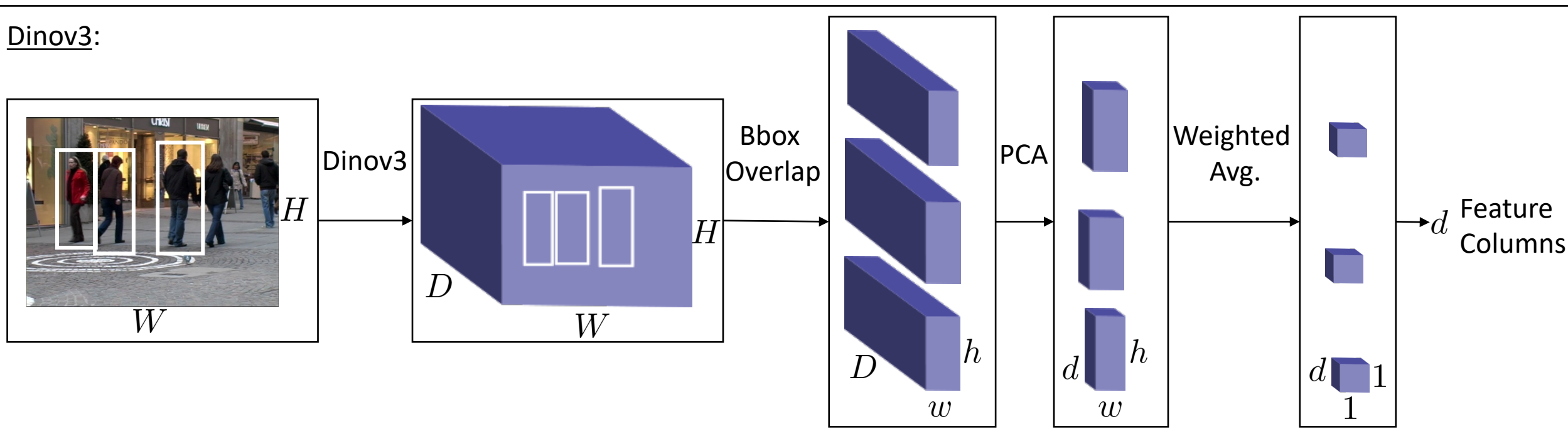
Limitations:

- #Samples $\approx$ 10.000
- #Features $\approx$ 512
- Univariate Prediction

Feature Extraction

Handcrafted: Area, Color Histograms, Lagging Gradient

Dinov3:



Next Steps: Reduction via Contrastive Learning, Person ReID, DETR, ...

Context Information

Dummy Frame: Expected Number of Tracks

t	x	y	...	id
-1	nan	nan	nan	1
-1	nan	nan	nan	2
-1	nan	nan	nan	3

Transfer-Learning: Other Annotated Scenes

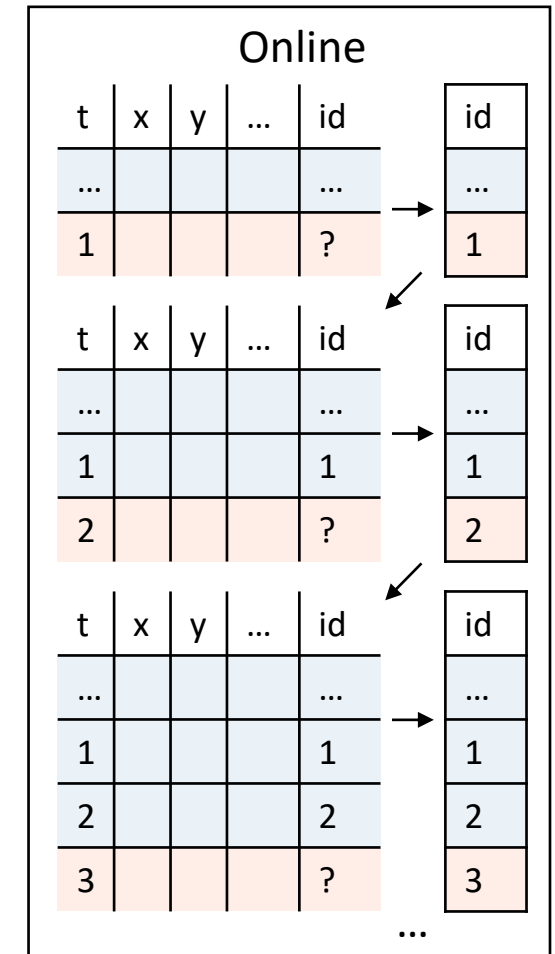
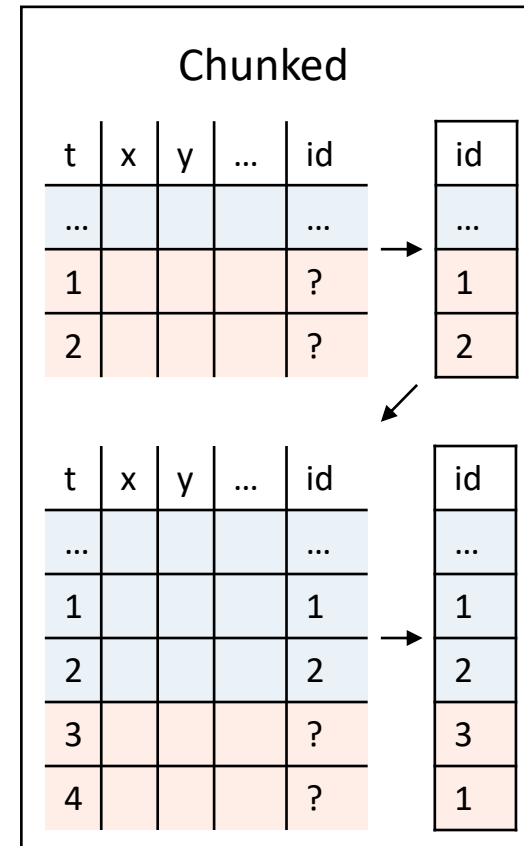
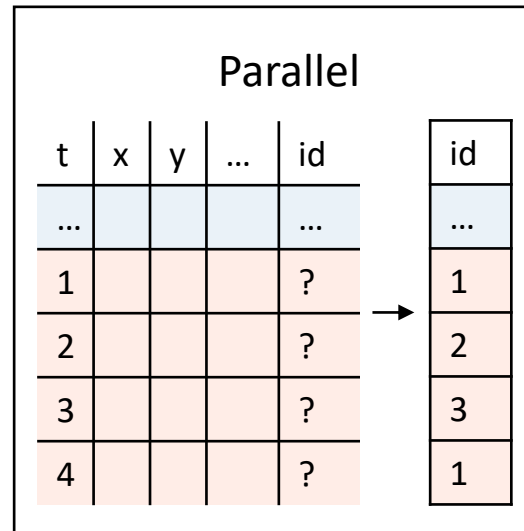
t	x	y	...	scene	id
1				A	1
2				A	3
1				B	1
2				B	1
2				B	3

Next Steps: Self-Supervised, e.g., via Kalman Filter and/or Others

t	x	y	...	alg1	alg2	...	id
1				1	1		nan
2				2	3		nan
2				1	1		nan

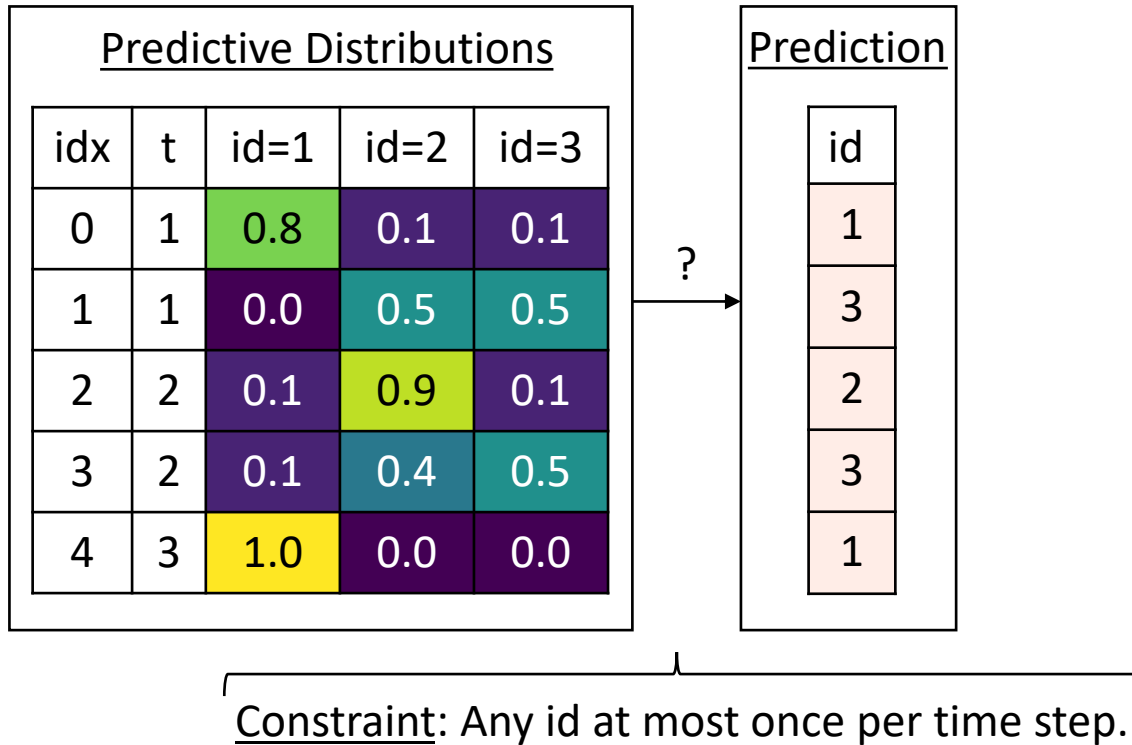
Tabular Model Strategy

Autoregression:



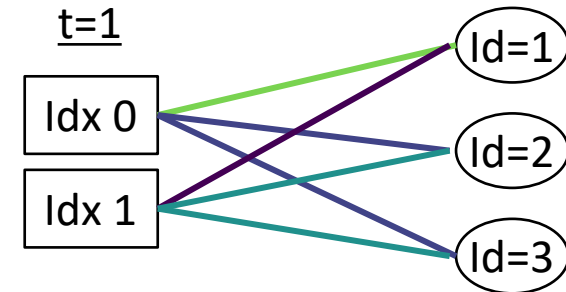
Next Steps: Beam Search, Memory, (Spatial) Reasoning, Self-Correction

Combinatorial Heuristic

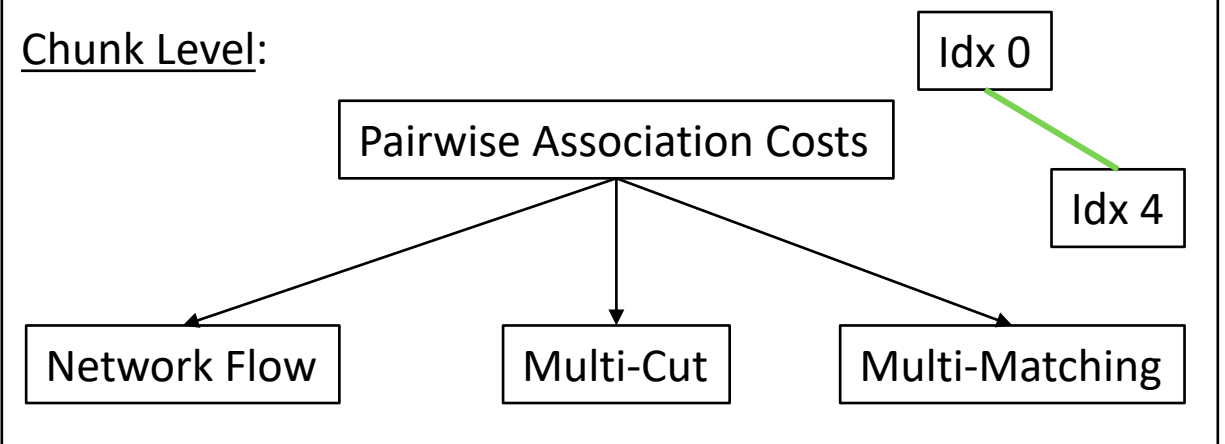


Next Steps: Regression -> SORT Like Cost Gating

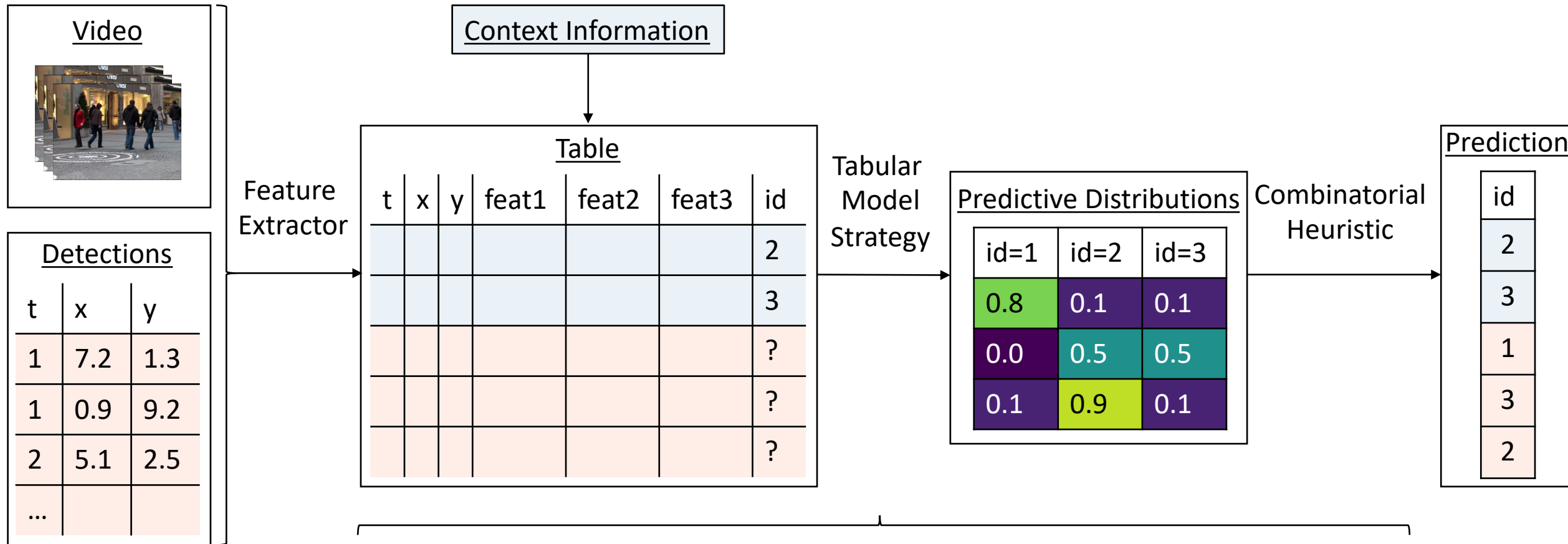
Frame Level: Linear Assignment Problem



Chunk Level:



MOT via Tabular Models



Limitations:

- #Samples $\approx$ 10.000
- #Features $\approx$ 512
- Univariate Prediction

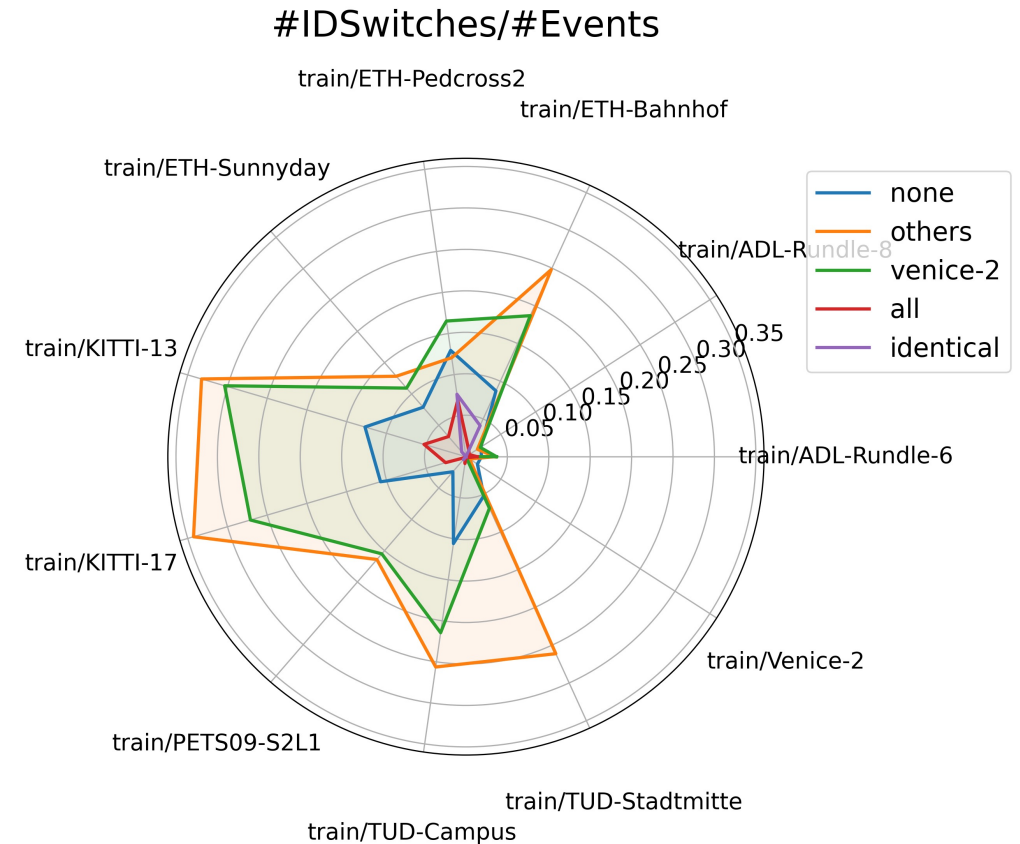
Results on MOT15 – Context Information

Setting:

- Ground Truth Bboxes as Detections
- 10 Frame Chunks

Findings:

- No Context performs best
- Problem too high dimensional for transfer learning
- Default = Look Up, if possible



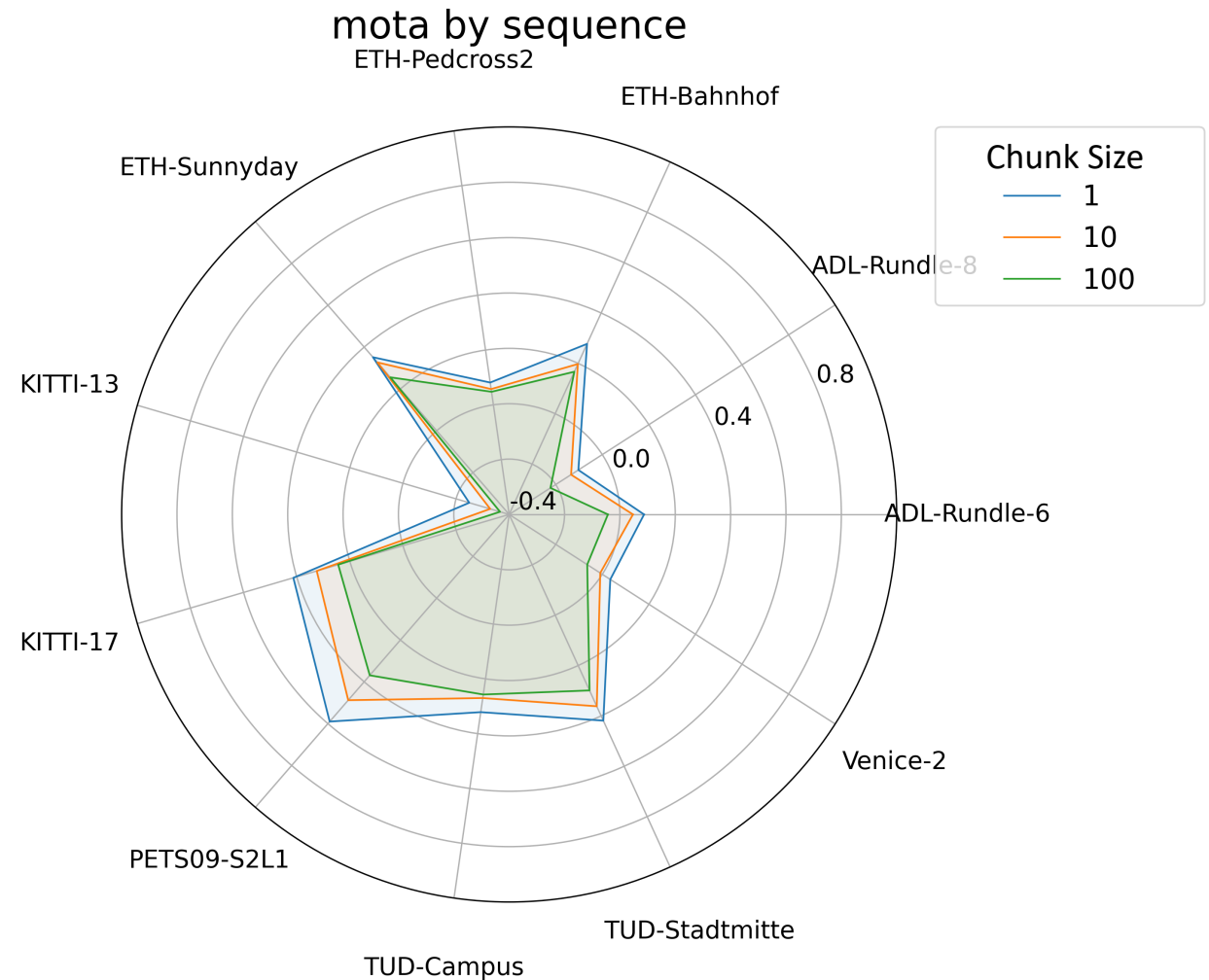
Results on MOT15 – Chunk Size

Setting:

- MOT15 Public Detections
- 10 Frame Chunks
- **First 10% Frames as Context Information**

Findings:

- Online outperforms all others



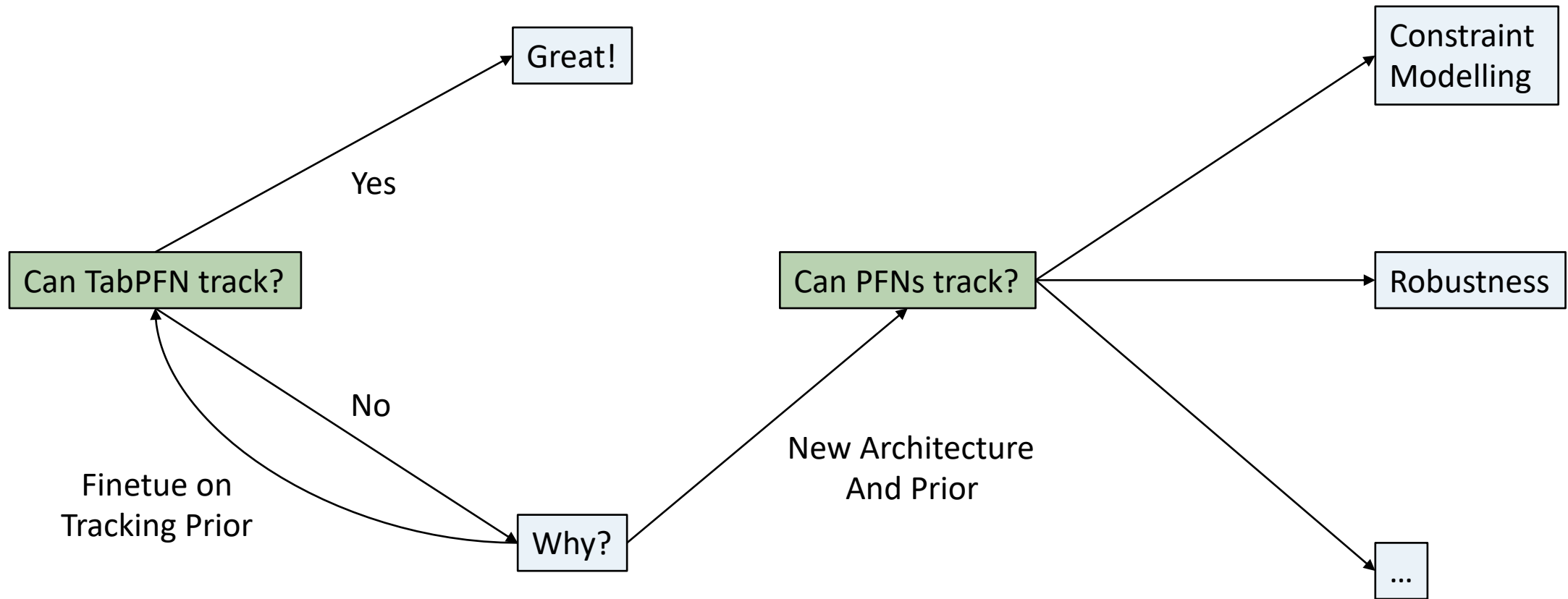
Qualitative Results on MOT15

Strong Positional Bias

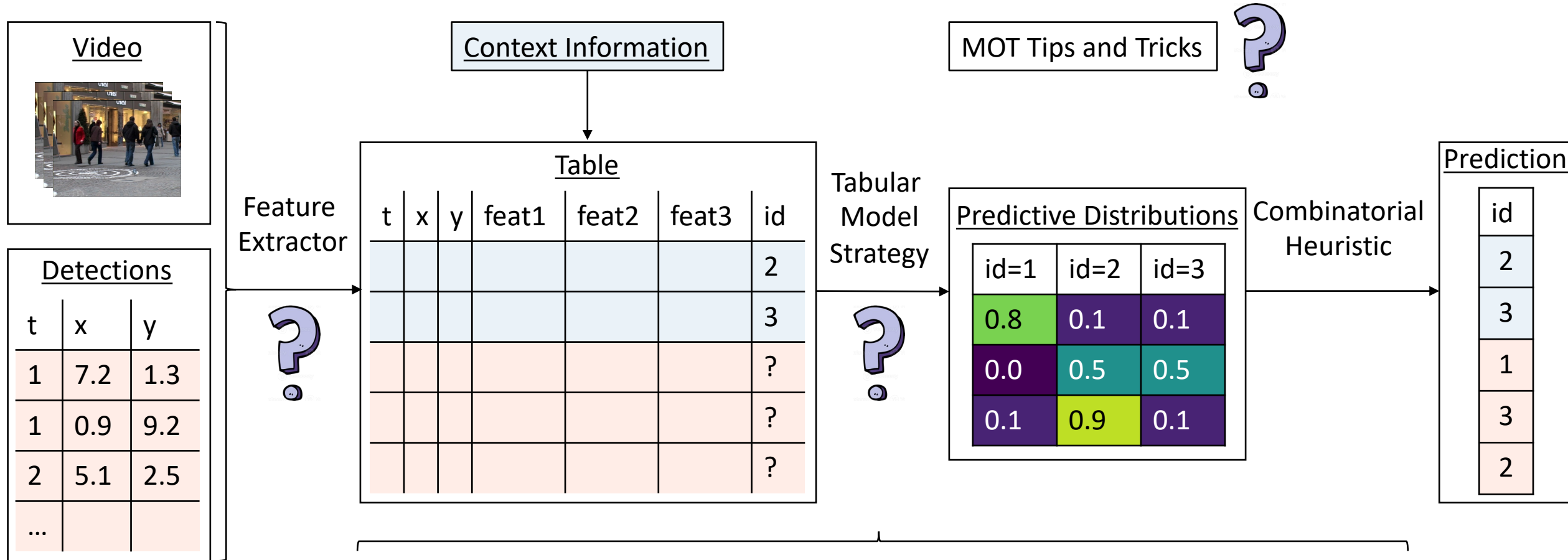


Ground Truth Bboxes as
Detections

Future Perspectives



What Should We Discuss?



Limitations:

- #Samples $\approx$ 10.000
- #Features $\approx$ 512
- Univariate Prediction

Qualitative Results on MOT15

Setting:

- 300 Expected Tracks
- Online
- MOT15 Public Detections

