HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

Bidirectional Trackformer (BiTrackformer)

Novel approach to transformer-based multiple object tracking algorithm

Adrian Kosmala
Prof. Dr. Paul Swoboda

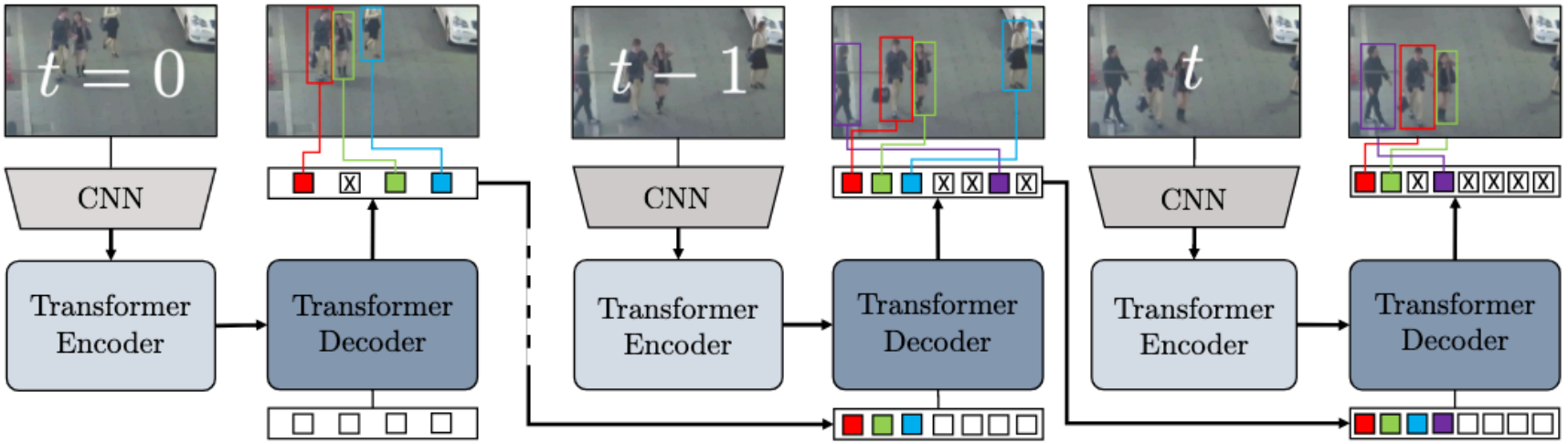
Siegen 26.03.2025

Vanilla Trackformer

HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

TrackFormer: Multi-Object Tracking with Transformers

(T. Meinhardt et al.)



Vanilla Trackformer

HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

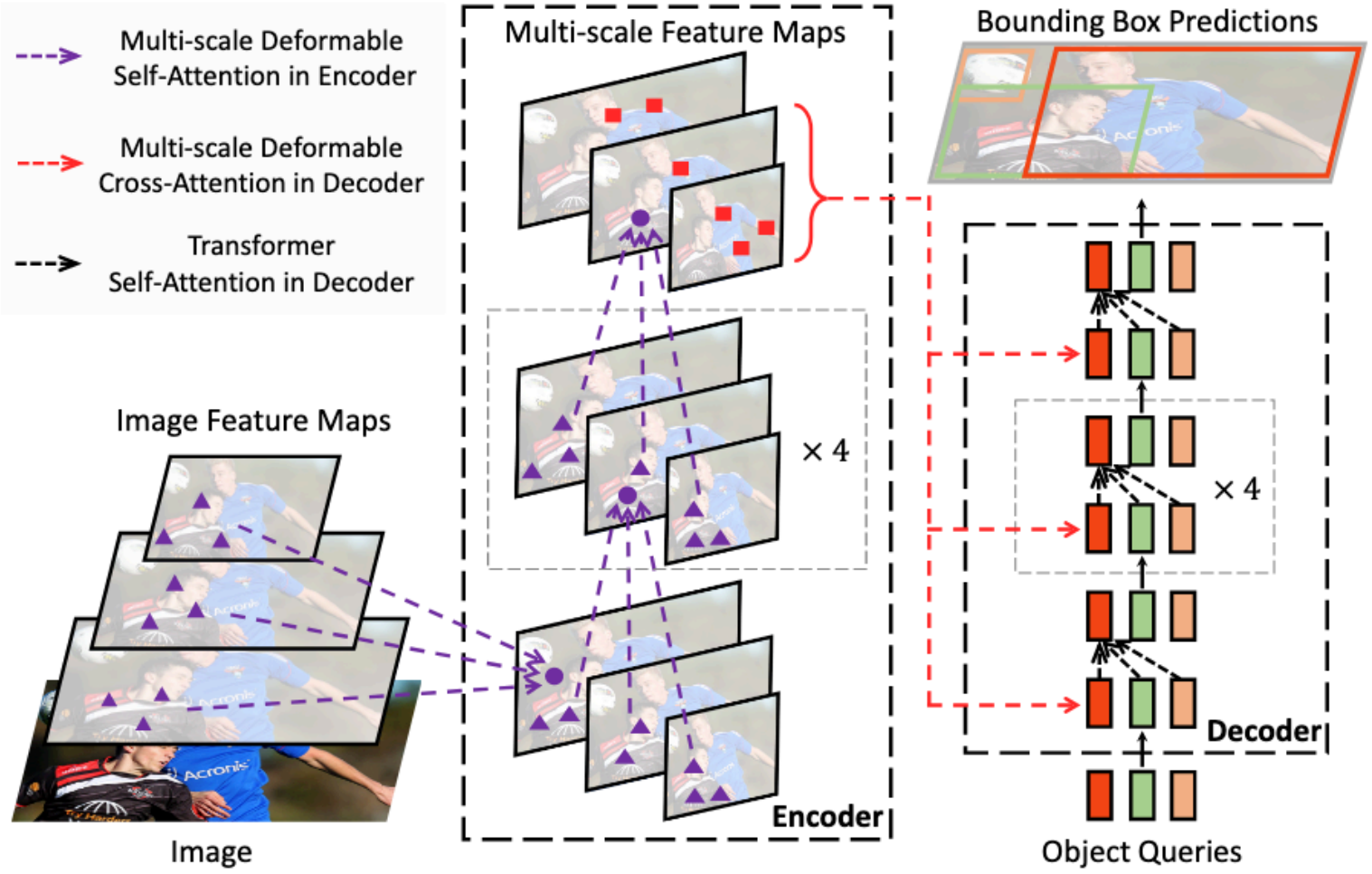
Training

- Short sequences of only 2 or 3 frames sampled from close interval
- Single images modified to imitate motion are also used as train sequences
- **Data augmentation:** false positives/negatives
- **Loss:** Linear combination of L1, GIoU and class error
- **Datasets:** CrowdHuman, MOT17

Inference

- Entire sequence from first to last frame
- **Postprocessing:**
  - non-maximum suppression (NMS)
  - Re-identification

Deformable DETR



Multiple Object Tracking (MOT)

HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

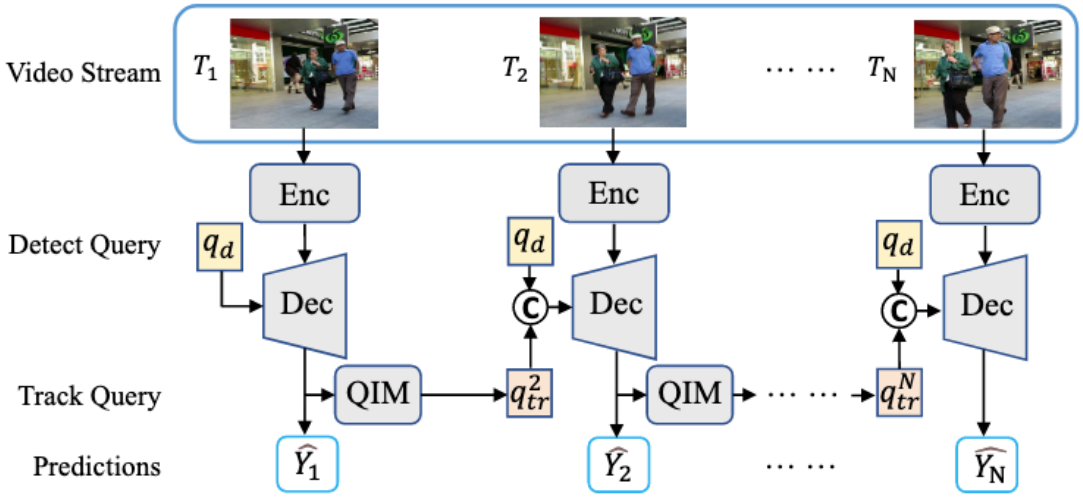
SOTA and the most important Transformer algorithms

Algorithm name	Description	MOTA in MOT17
Unified Multiple Object Tracking Model (UTM)	Composed of couple of modules which, based on previous frame in the previous frame, boost features of detected objects and suppress background in the current frame.	81,8
MotionTrack	Introduces novel Interaction Module and Refind Module which focuses on object movement estimation to help tracking with dense crowds and extreme occlusions.	81,1
SUSHI	Initially tracks on short sequences which are joined hierarchically by graph neural networks.	81,1
MOTR (T)	Similar to Trackformer, the main differences are longer training sequence and Query Interaction Module which additionally transforms track queries before passing into the next step.	78,6
TransCenter (T)	Employs dense detection queries to locate targets and efficient sparse tracking queries for object association across time.	75,8
TransTrack (T)	Similar to Trackformer, the main differences are separate decoders for object/track queries and two step tracking: association of 2-frame tracklets into longer trajectories.	74,5
MeMOT (T)	Utilizes information from both current frame detections and a large spatio-temporal memory storing identity embeddings of tracked objects.	72,5
Trackformer (T)		**74,1**

Multiple Object Tracking (MOT)

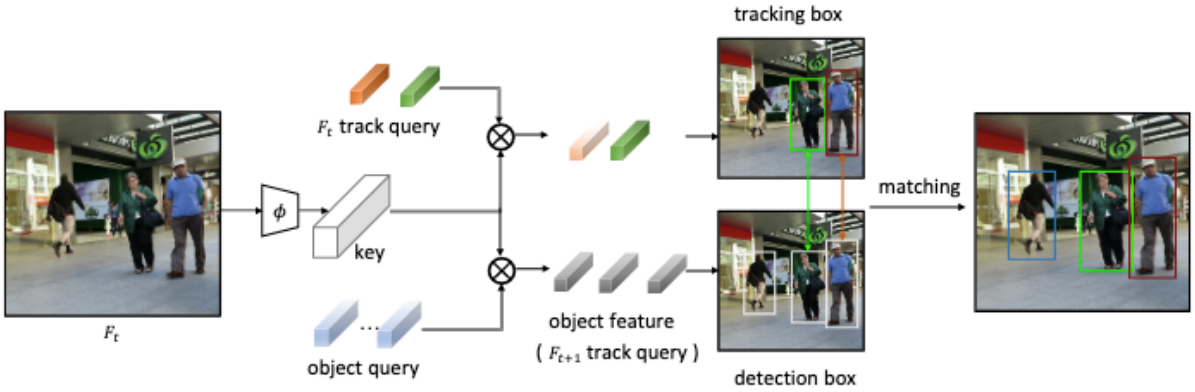
MOTR

- **Query Interaction Module** - transforms track queries between frames
- **longer training sequence**



TransTrack

- **Separate decoders** for object/ track queries
- **2 step tracking:**
  - Associate objects between 2 frames
  - Associate 2-frames tracklets into longer trajectories



Bidirectional Trackformer

HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

Key ideas

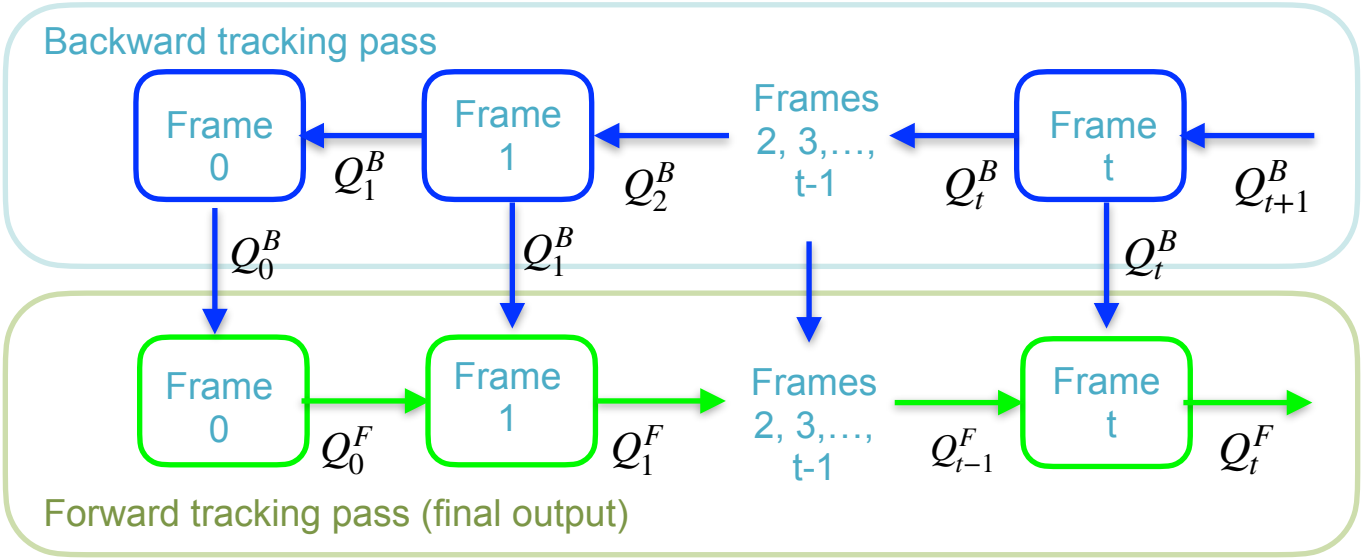
- **Change approach:**
  online -> offline

- **Utilization of both past and future frames:**
  backward and forward tracking passes

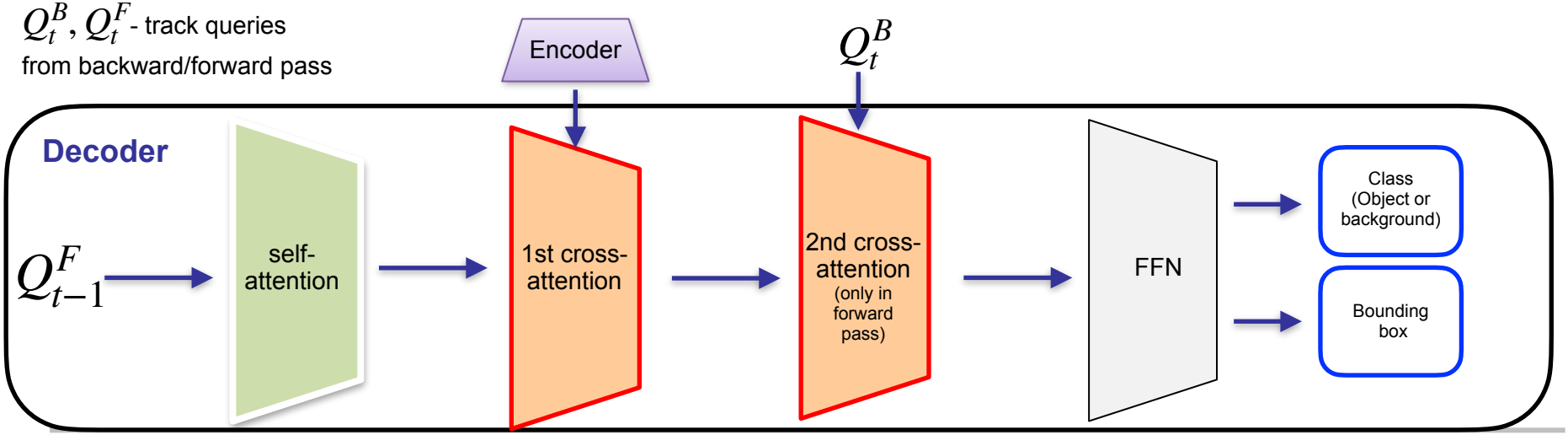
- **Minimal changes in the Trackformer architecture:**
  only one additional cross-attention module

- **General solution:**
  can be applied in other transformer trackers

Bidirectional Trackformer



Q_t^B, Q_t^F - track queries from backward/forward pass



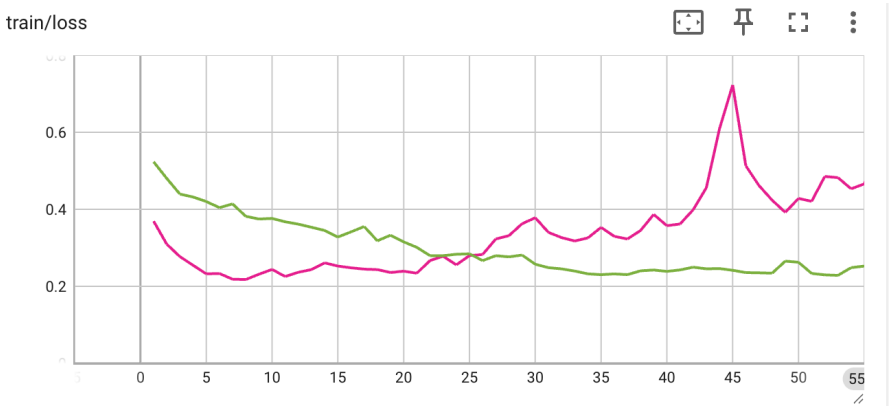
Bidirectional Trackformer

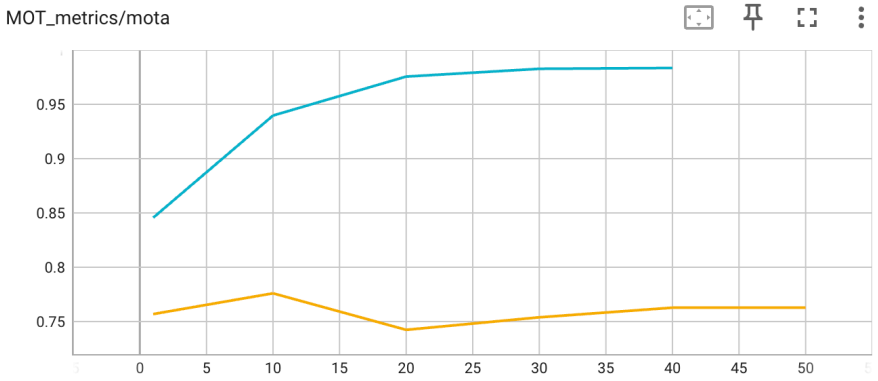
Challenges

- Applying loss only from last decoder layer instead of sum of losses from all decoder layers was found to increase loss

- Not including of false positives/ negatives in training was found to decrease MOTA

- A lot of small technical details (e.g. track ids propagation) which makes debugging and implementing new solutions sophisticated





Bidirectional Trackformer

HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

Initial results*

Algorithm	MOTA↑	IDF1↑	Mostly tracked↑	Mostly lost↓	ID Switches ↓
Vanilla Trackformer	72.8	**74.8**	176	41	**211**
BiTrackformer	**74.2**	73.1	**184**	**33**	327

* Trained and evaluated on entire MOT17 training set - further experiments are required

Bidirectional Trackformer

HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

Next steps

- Checking other training configurations, e.g. fine-tuning on MOT17+CrowdHuman
- Decision if we update all weights in the model or only in 2nd cross-attention module
- Decision if models for backward and forward tracking pass should share weights or be trained separately
- Applying better detector, e.g. RT-DETR
- Improving loss function, e.g. applying metric learning
- Improving the object queries initialization
- Applying solutions from other Transformer-based tracking algorithms, e.g. processing track queries before passing to the next frame