

# Dual-Path Temporal Decoder for End-to-End Multi-Object Tracking

Hyunseop Kim<sup>1\*</sup>, Juheon Jeong<sup>1\*</sup>, Hanul Kim<sup>2</sup>, Yeongjun Koh<sup>1†</sup>

<sup>1</sup>Chungnam National University, Korea

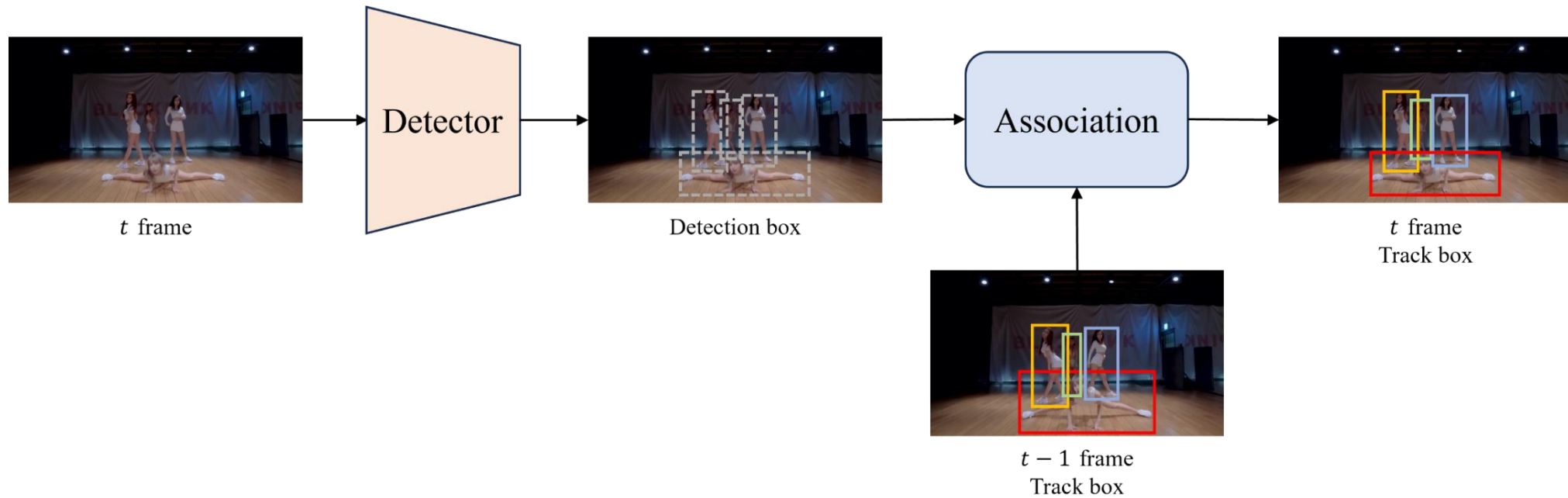
<sup>2</sup>Seoul National University of Science and Technology

NeurIPS 2025

# Background

- **Tracking by detection (TBD)**

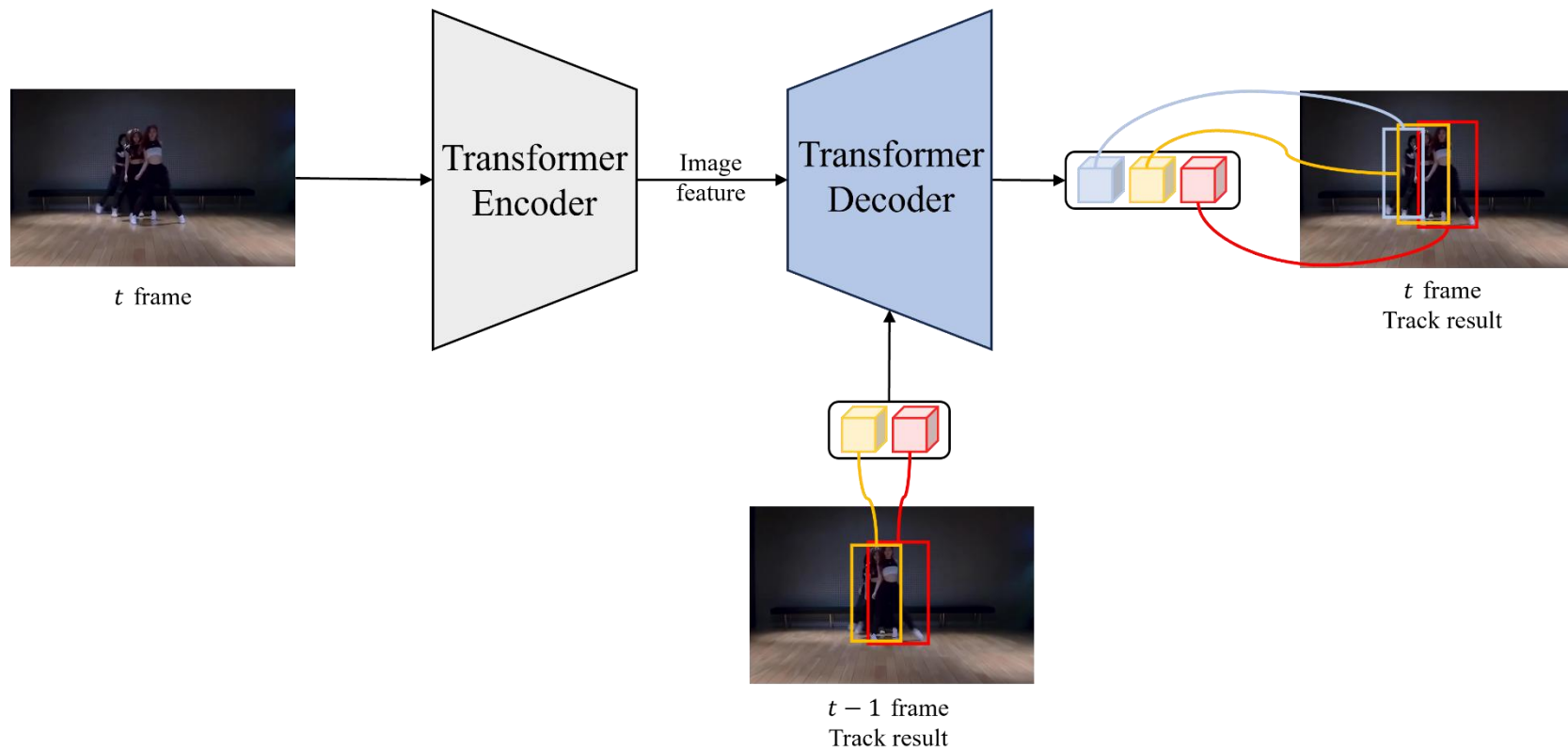
- Tracking-by-detection tracks objects by associating detections across frames using spatial and appearance similarities.



# Background

- **Transformer-based MOT**

- Transformer-based MOT tracks multiple objects by leveraging attention mechanisms to associate features across frames instead of relying solely on detections.

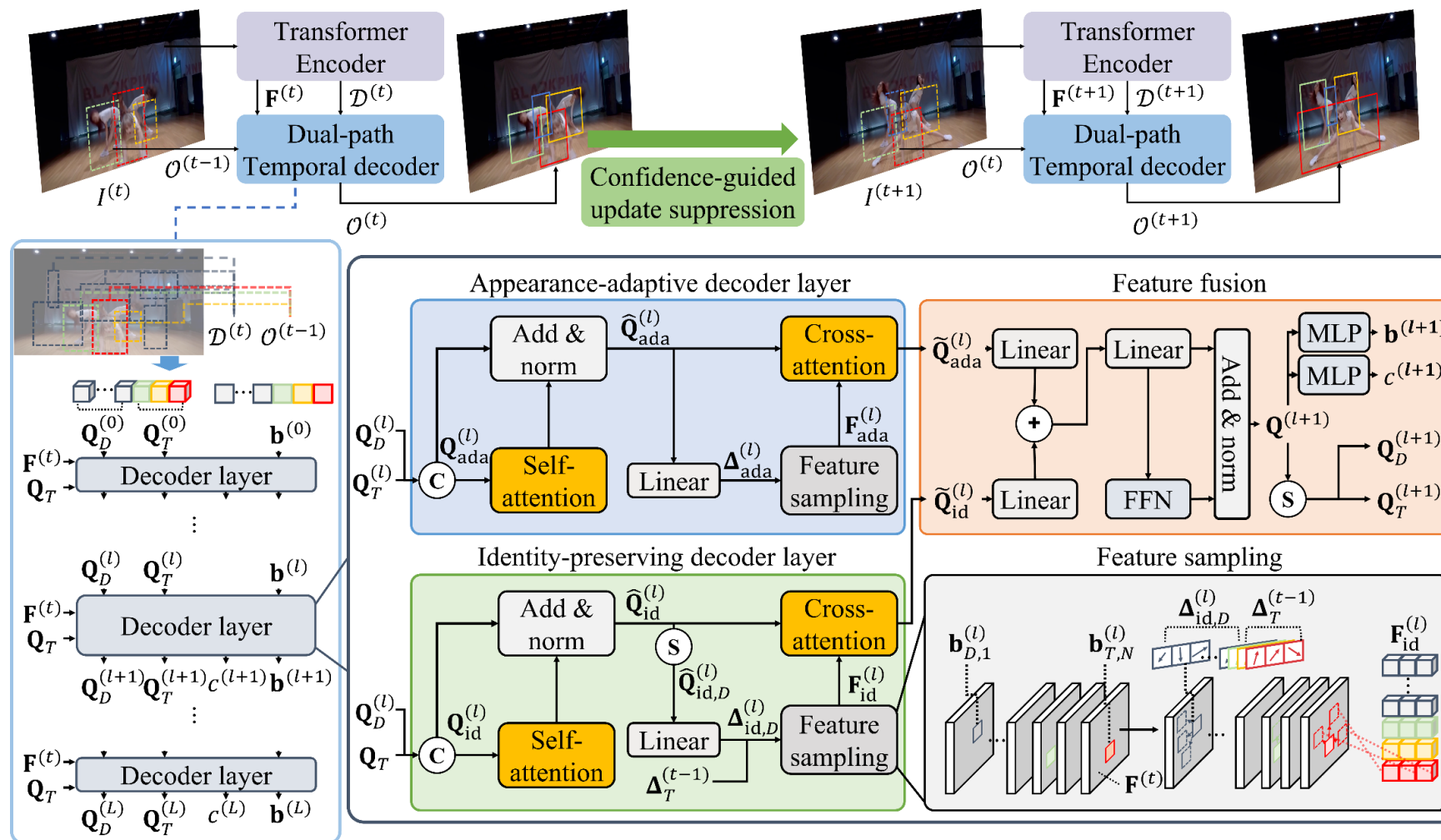


# Motivation

- **Limitation of tracking by detection (TBD)**
  - Not end-to-end : detection and tracking are separate.
  - Tracker performance is highly sensitive to detector errors
- **Limitation of transformer-based MOT**
  - Query drift : gradual degradation of per-object query embeddings due to accumulated erroneous updates, causing identity instability
- We proposed transformer-based **dual-path temporal decoder**
  - To reduce identity switches by preventing query drift
  - Appearance-adaptive decoder layer
  - Identity-preserving decoder layer

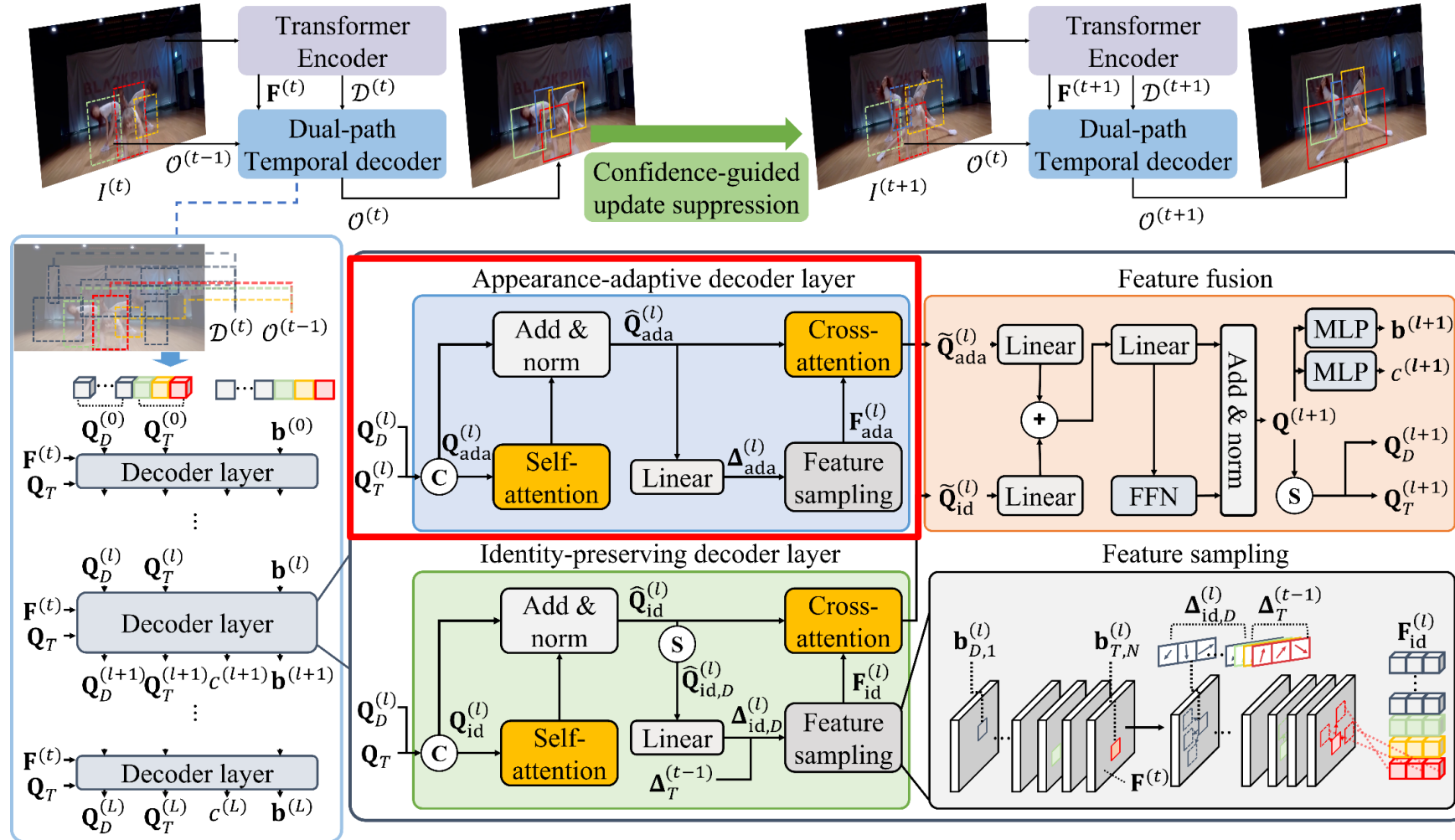
# Method

## • Overview



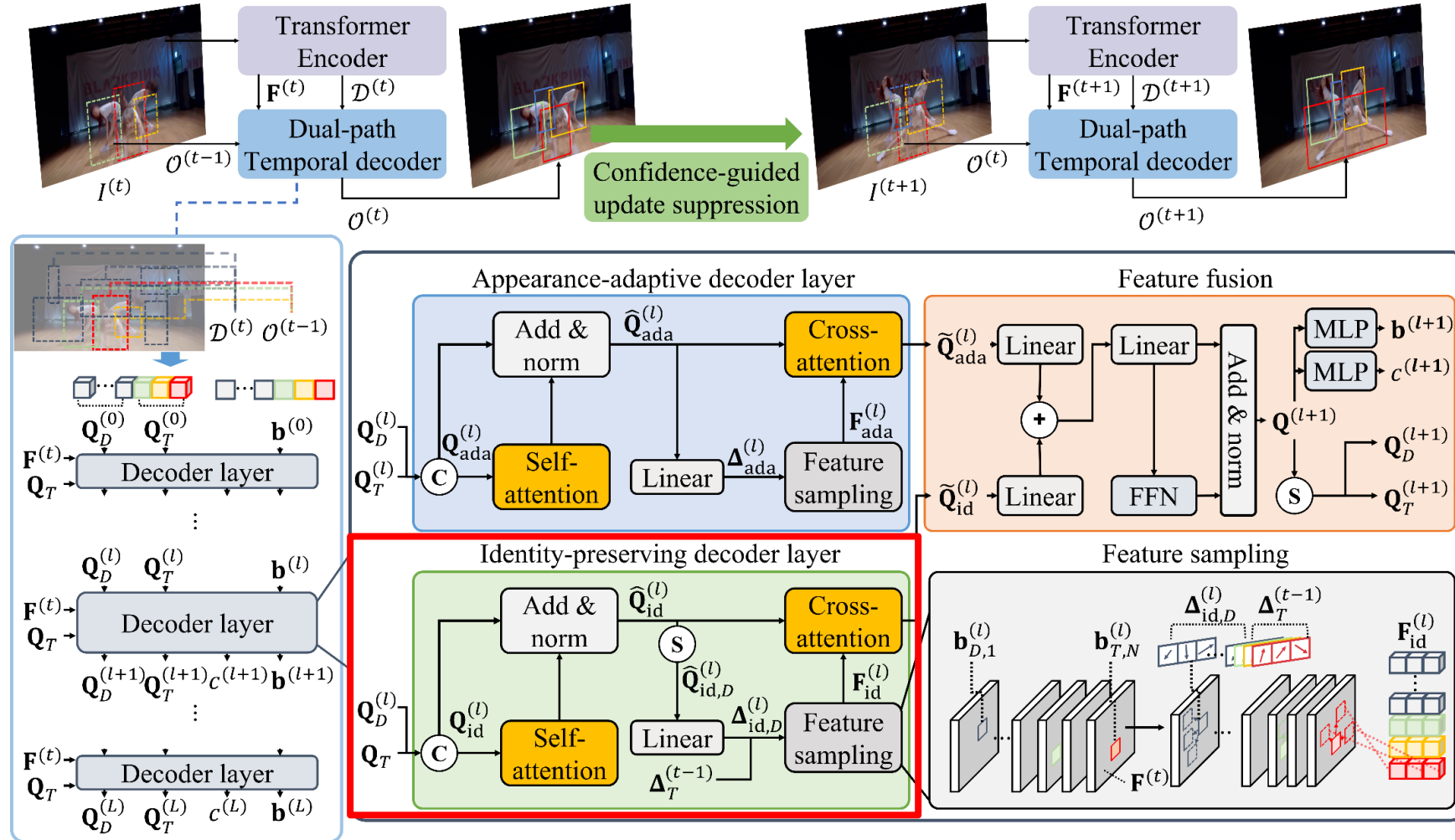
# Method

- Appearance-adaptive decoder layer (ADL)



# Method

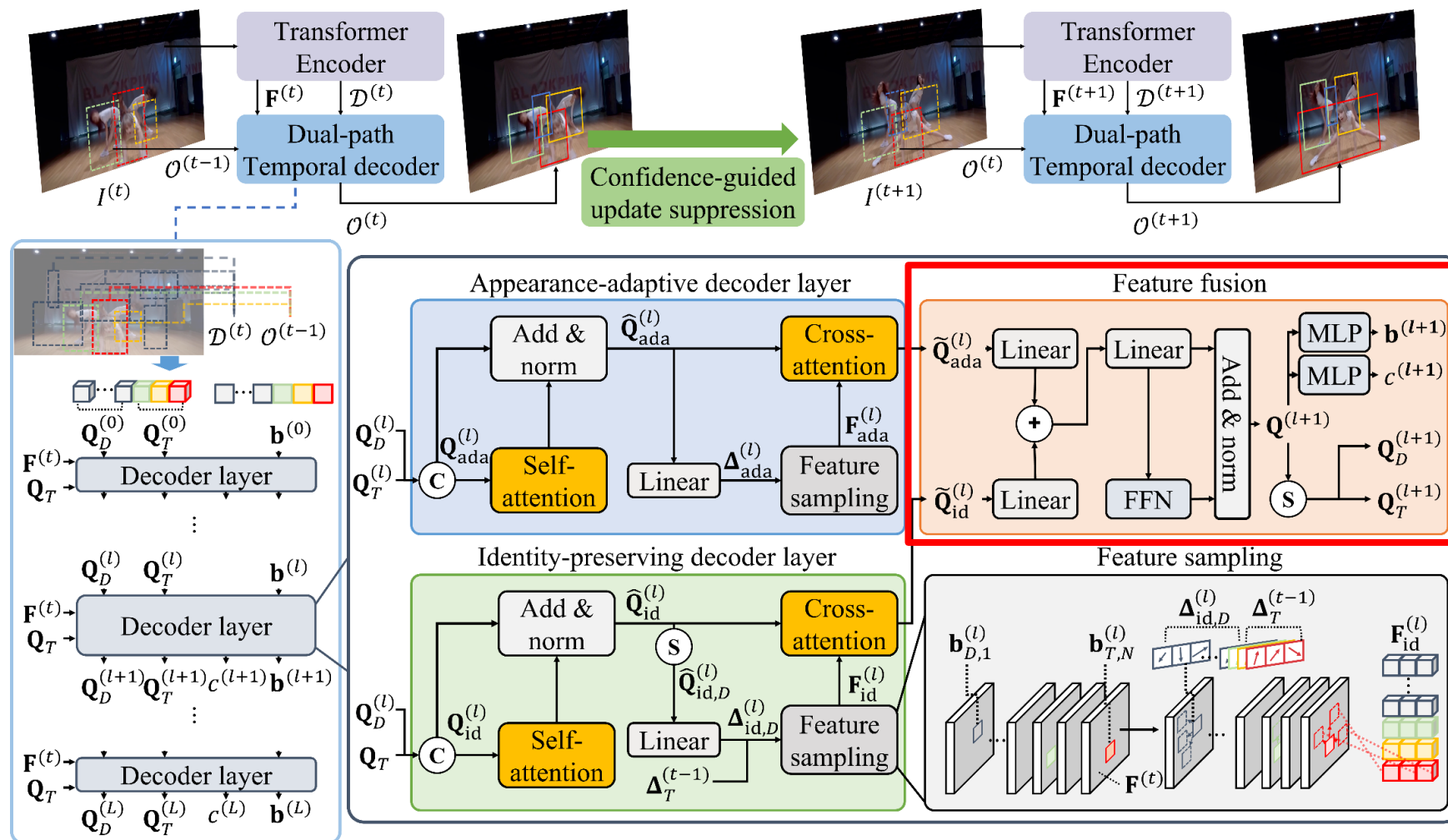
- Identity-preserving decoder layer (IDL)**





# Method

## • Feature fusion

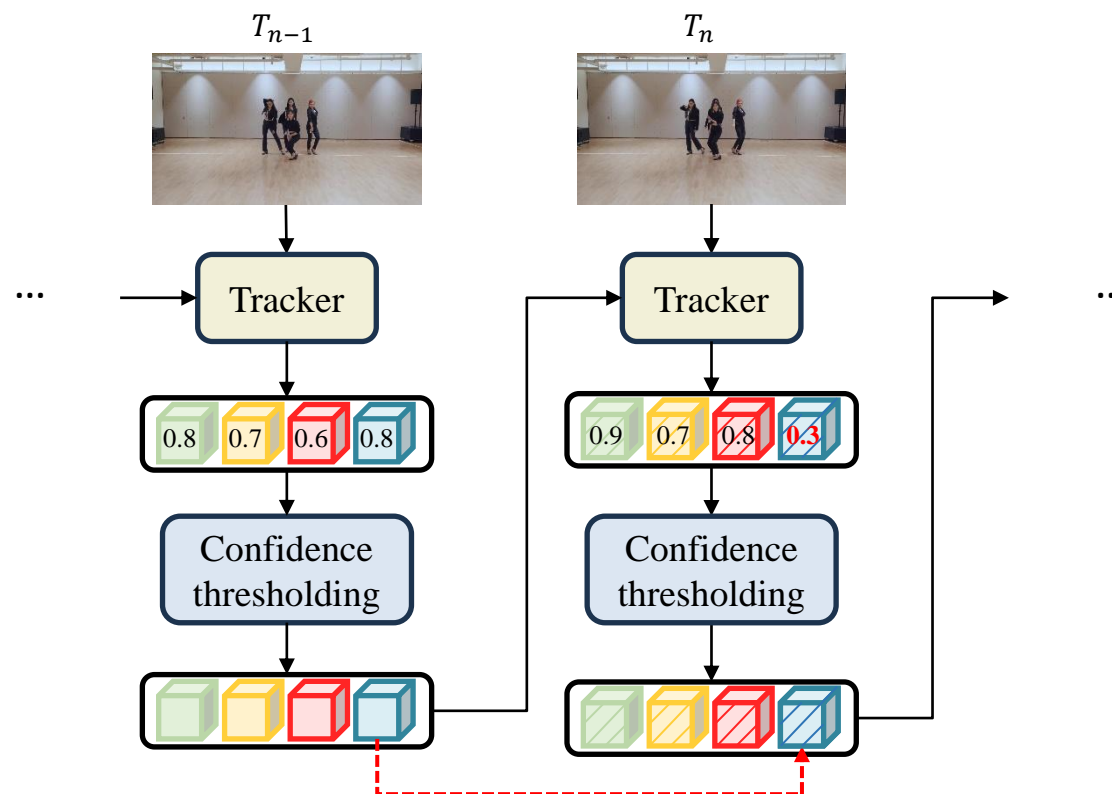




# Method

- **Confidence-guided update suppression**

- Applying for inference stage.
- update threshold = 0.4



# Experiments

- **Quantitative comparison on DanceTrack test set.**
  - SOTA performance on DanceTrack.

Methods	HOTA	DetA	AssA	MOTA	IDF1
<i>w/o valid data:</i>					
CenterTrack [40]	41.8	78.1	22.6	86.8	35.7
TransTrack [26]	45.5	75.9	27.5	88.4	45.2
ByteTrack [37]	47.7	71.0	32.1	89.6	53.9
QDTrack [22]	54.2	80.1	36.8	87.7	50.4
MOTR [34]	54.2	73.5	40.2	79.7	51.5
OC-SORT [6]	55.1	80.3	38.3	92.0	54.6
DiffMOT [18]	62.3	82.5	47.2	<b>92.8</b>	63.0
MeMOTR [12]	68.5	80.5	58.4	89.9	71.2
CO-MOT [32]	69.4	82.1	58.9	91.2	71.9
MOTRv2 [39]	69.9	83.0	59.0	91.9	71.7
MOTIP [11]	72.0	81.8	63.5	91.9	76.8
ColTrack [16]	72.6	-	62.3	92.1	74.0
Ours	<b>74.1</b>	<b>83.9</b>	<b>65.6</b>	92.5	<b>78.6</b>
<i>with valid data:</i>					
MOTRv2 [39]	73.4	83.7	64.4	92.1	76.0
ColTrack [16]	75.3	-	66.9	92.2	77.3
Ours	<b>76.2</b>	<b>85.0</b>	<b>68.3</b>	<b>92.5</b>	<b>79.9</b>

Table 1. Performance on DanceTrack test set

# Experiments

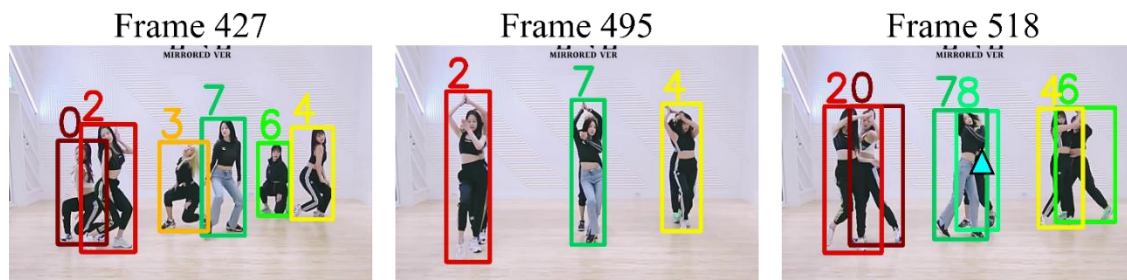
- **Quantitative comparison on SportsMOT test set.**
  - SOTA performance on SportsMOT.

Methods	HOTA	DetA	AssA	MOTA	IDF1
QDTrack [22]	60.4	77.5	47.2	90.1	62.3
CenterTrack [40]	62.3	82.1	48.0	90.8	60.0
ByteTrack [37]	62.8	77.1	51.2	94.1	69.8
TrackFormer [20]	63.3	66.0	61.1	74.1	72.4
BoT-SORT [1]	68.7	84.4	55.9	<b>94.5</b>	70.0
MeMOTR [12]	68.8	82.0	57.8	90.2	69.9
TransTrack [26]	68.9	82.7	57.5	92.6	71.5
ColTrack [16]	71.5	80.5	63.6	89.4	74.6
OC-SORT [6]	71.9	<b>86.4</b>	59.8	<b>94.5</b>	72.2
DiffMOT [18]	72.1	86.0	60.5	<b>94.5</b>	72.8
MOTIP [11]	72.6	83.5	63.2	92.4	77.1
Ours	<b>73.9</b>	82.2	<b>66.6</b>	91.5	<b>78.7</b>

Table 2. Performance on SportsMOT test set

# Experiments

- **Qualitative comparison on DanceTrack validation set.**
  - Failure examples are marked with blue triangles.



(a) MOTRv2



(b) ColTrack



(c) Ours

# Experiments

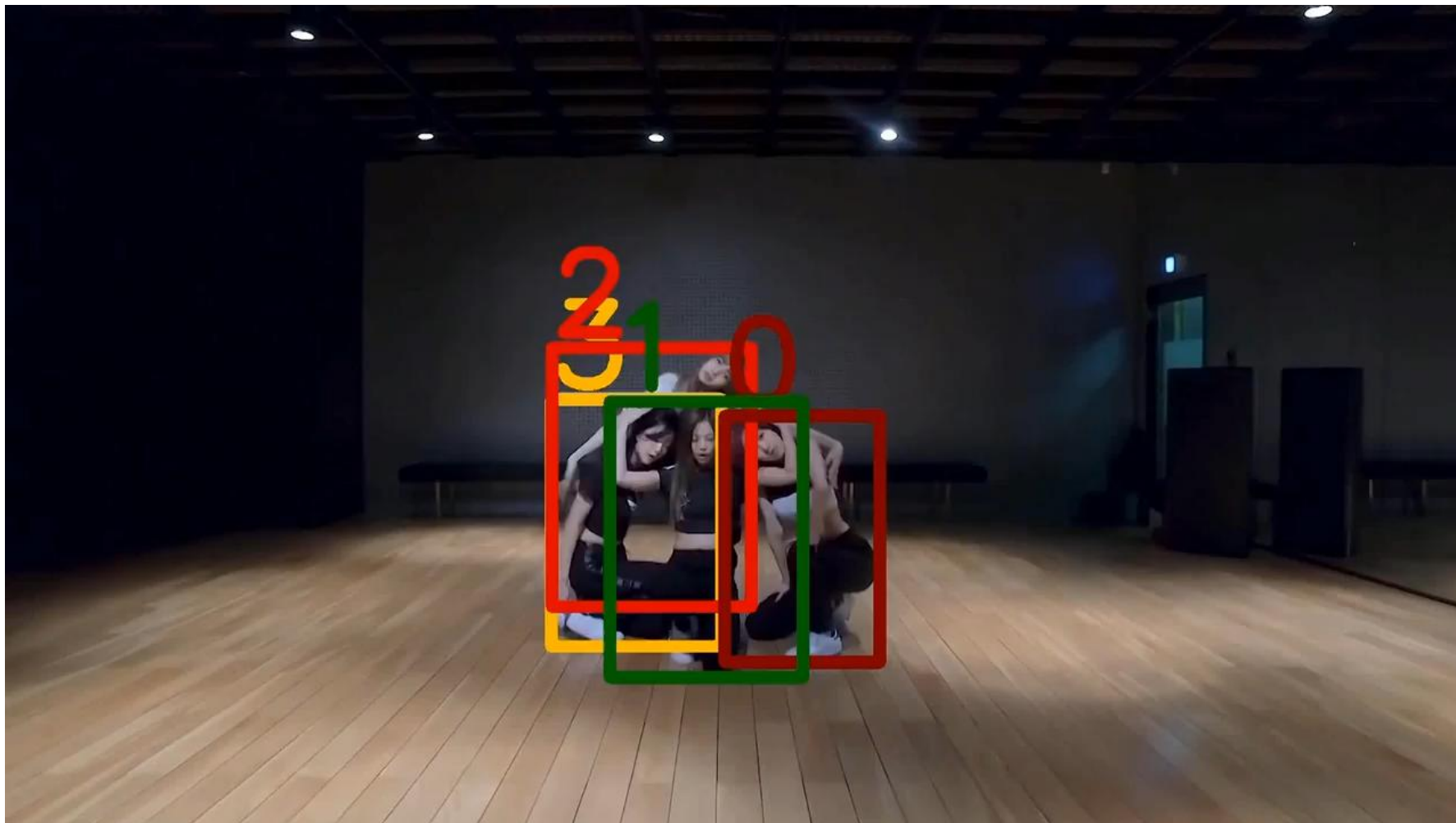
- **Ablation study**
  - Effect of IDL and historical offsets on tracking performance
    - Study conducted on DanceTrack validation set.

Method	HOTA	DetA	AssA	MOTA	IDF1
without IDL	66.7	76.1	56.3	87.0	69.5
IDL with varying offsets $\Delta_{\text{id},T}^{(l)}$	67.5	77.5	59.0	87.1	73.2
IDL with static historical offsets $\Delta_T^{(t-1)}$	<b>69.1</b>	<b>77.8</b>	<b>61.6</b>	<b>87.5</b>	<b>74.9</b>

Table 4. Ablation studies for the identity-preserving decoder layer (IDL)

# Demo

- DanceTrack



Thank you for your attention.

