Context-Aware Relative Object Queries to Unify Video Instance and Panoptic Segmentation

Anwesa Choudhuri, Girish Chowdhary, Alexander Schwing
University of Illinois at Urbana-Champaign

TUE-PM-215

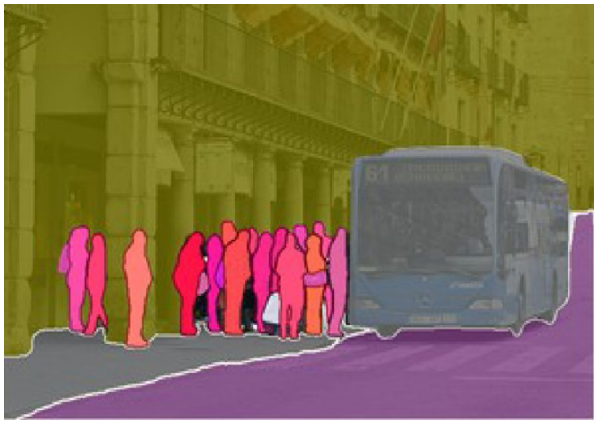
Overview

Transformers for Image Segmentation

- DETR [1]
- MaskFormer [2]
- Mask2Former [3]

uses object queries

[1]



[3]

[1] Carion et al., ECCV 2020
[2] Cheng et al., NeurIPS 2021
[3] Cheng et al., CVPR 2022

Transformers for Video Segmentation?

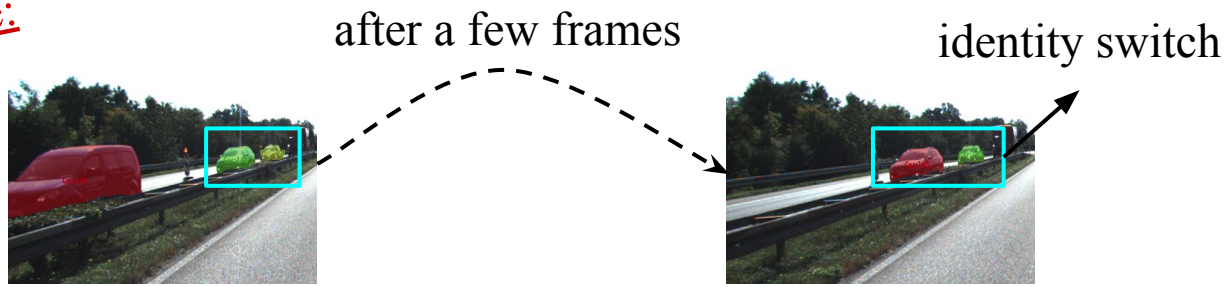
- MinVIS [1]
- IDOL [2]
- TrackFormer [3]
- VisTR [4]
- Mask2Former-VIS [5]
- TubeFormer-DeepLab [6]

[1]

[1] Huang et al., NeurIPS 2022
[2] Wu et al., ECCV 2022
[3] Meinhardt et al., CVPR 2022
[4] Wang et al., CVPR 2021
[5] Cheng et al., arXiv 2022
[6] Kim et. al., CVPR 2022

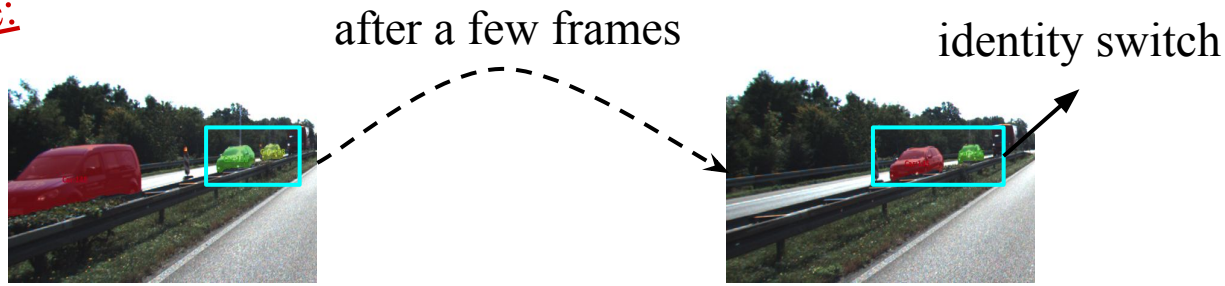
Transformers for Video Segmentation?

Issue:



Transformers for Video Segmentation?

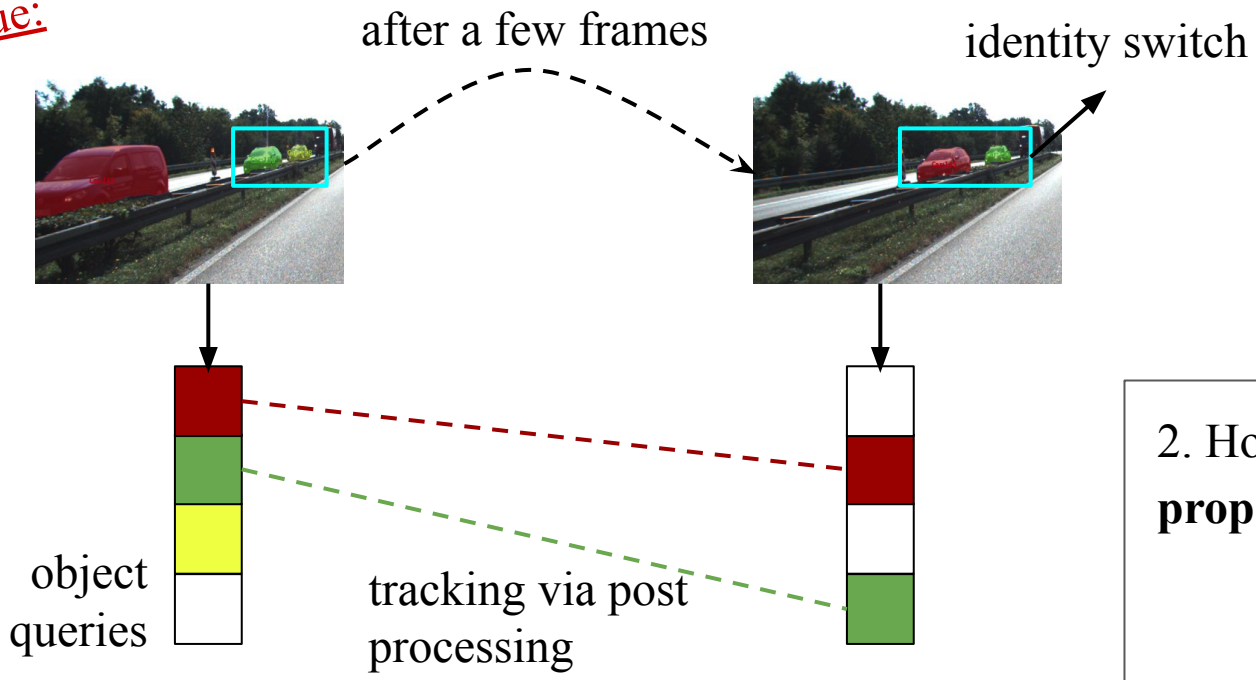
Issue:



1. How to model **appearance** and **position changes**?

Transformers for Video Segmentation?

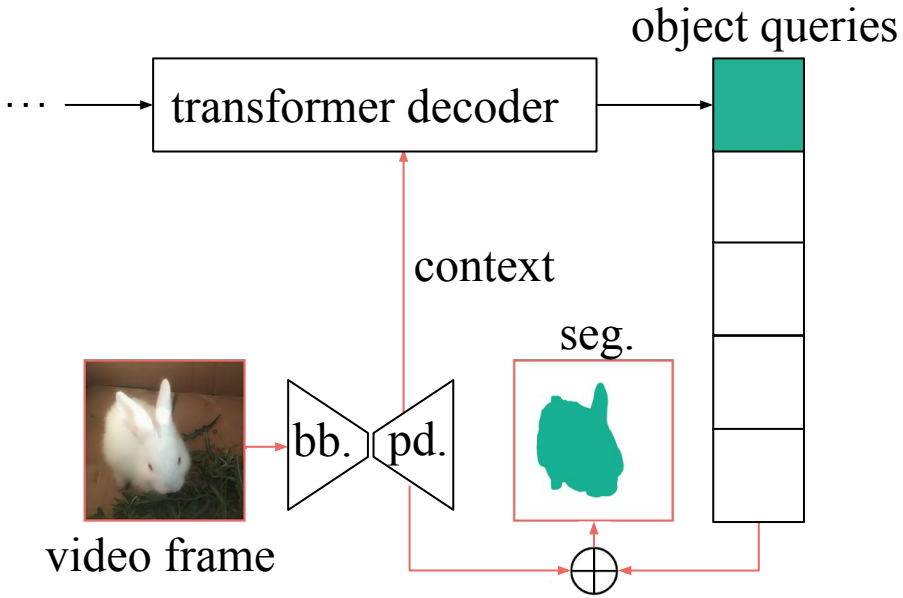
Issue:



2. How to **seamlessly propagate** object queries?

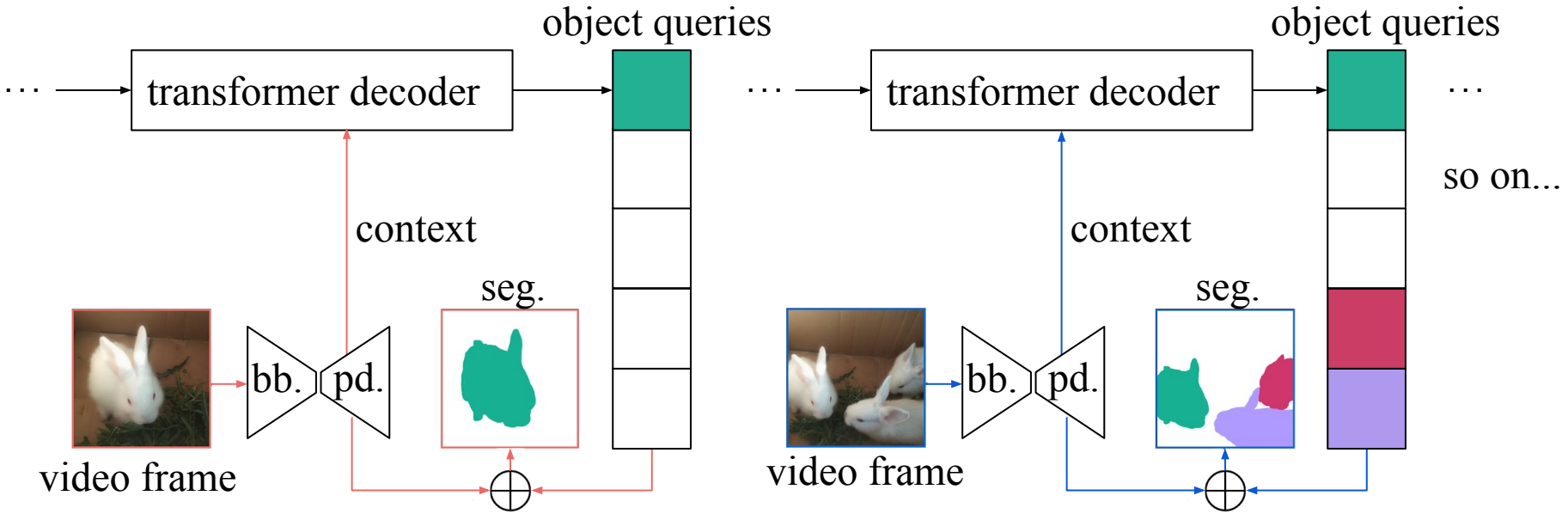
Context-Aware Relative Object Queries

Context-Aware Relative Object Queries



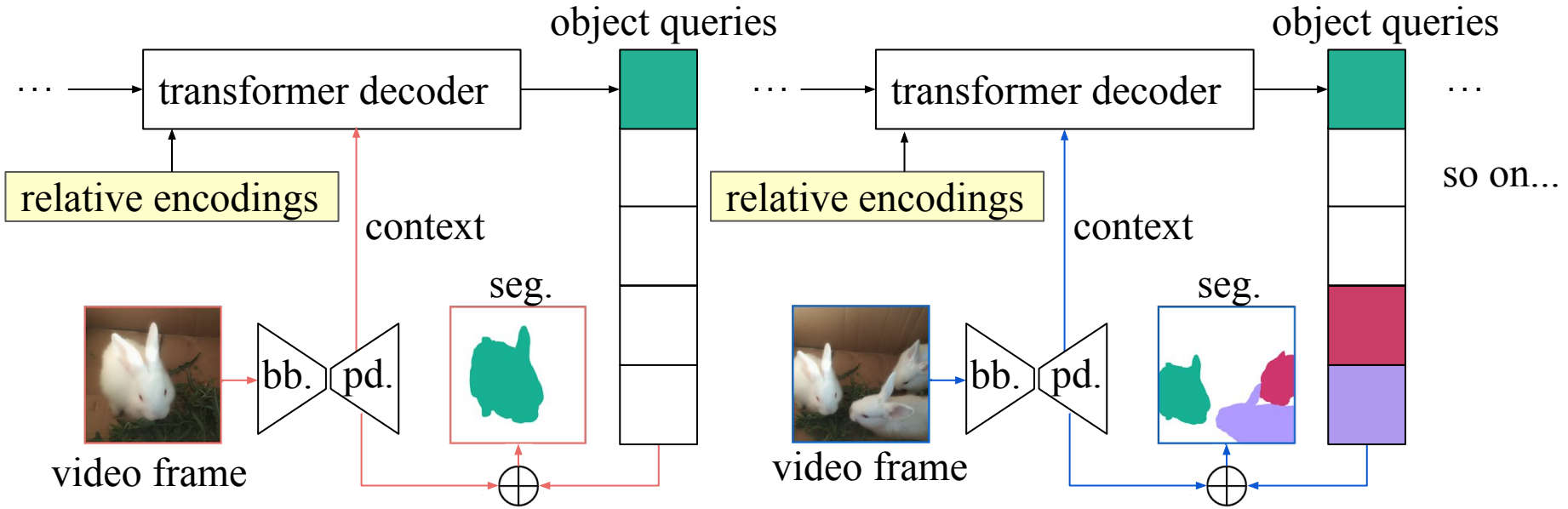
bb: backbone
pd: pixel decoder
seg: segmentations

Context-Aware Relative Object Queries



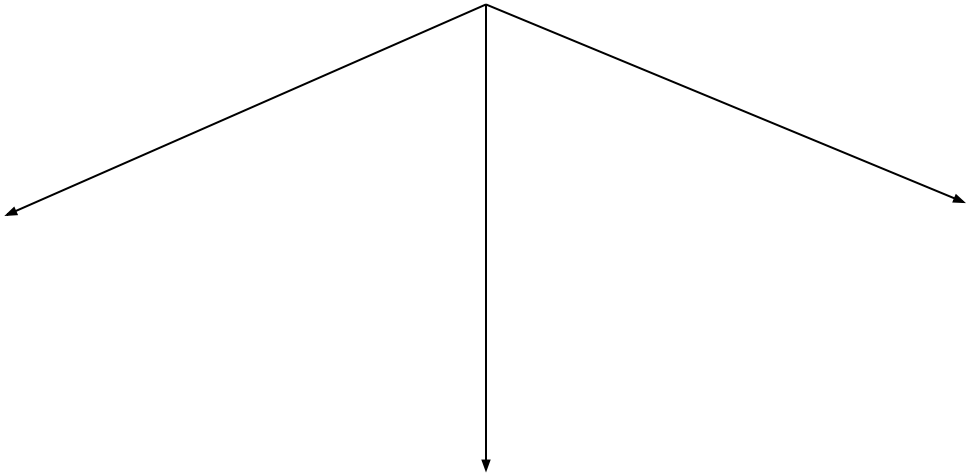
bb: backbone
pd: pixel decoder
seg: segmentations

Context-Aware Relative Object Queries



bb: backbone
pd: pixel decoder
seg: segmentations

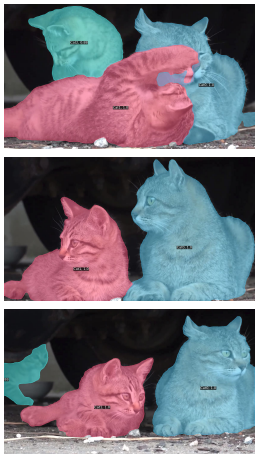
Context-Aware Relative Object Queries



Context-Aware Relative Object Queries

Video Instance Segmentation

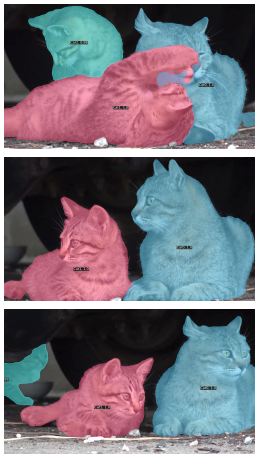
(Youtube-VIS 2019, Youtube-VIS 2021, OVIS)



Context-Aware Relative Object Queries

Video Instance Segmentation

(Youtube-VIS 2019, Youtube-VIS 2021, OVIS)



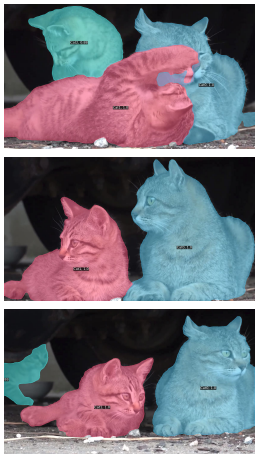
Multi-Object Tracking and Segmentation

(KITTI-MOTS, MOTS-2020)

Context-Aware Relative Object Queries

Video Instance Segmentation

(Youtube-VIS 2019, Youtube-VIS 2021, OVIS)

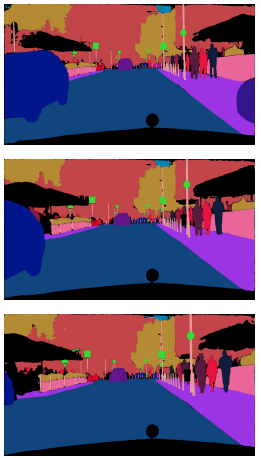


Multi-Object Tracking and Segmentation

(KITTI-MOTS, MOTS-2020)

Video Panoptic Segmentation

(Cityscapes-VPS)



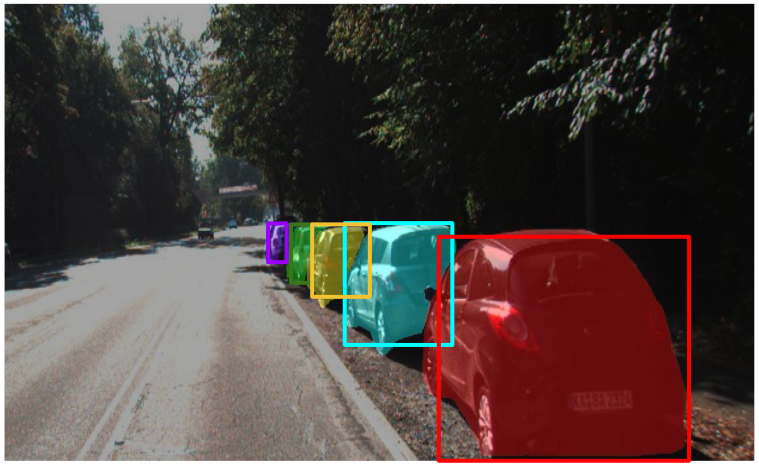
Video Segmentation

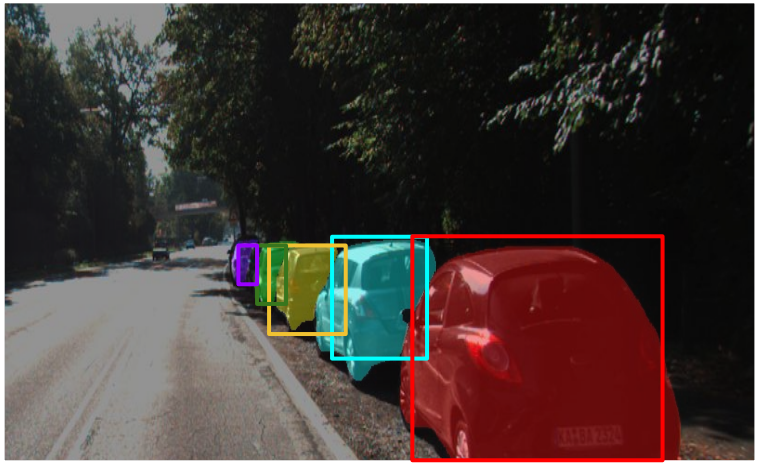
Video Segmentation

- Segmenting and associating objects of interest in a video

Video Segmentation

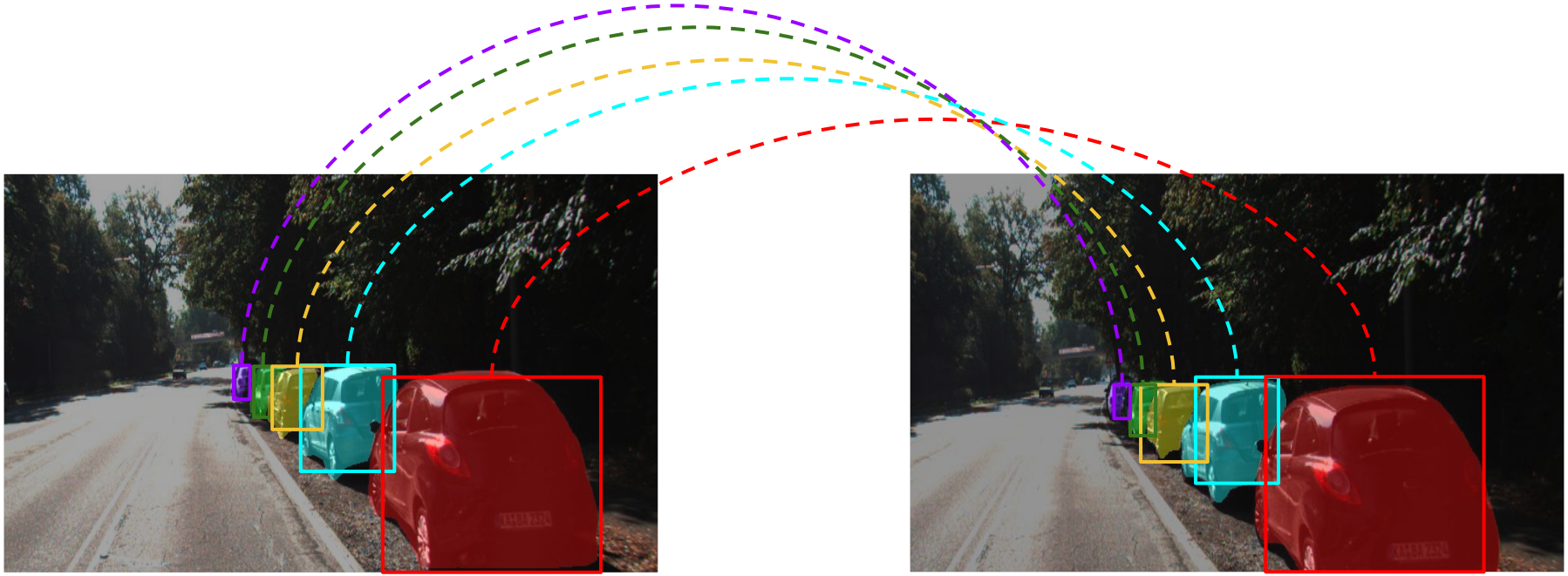
- Segmenting and associating objects of interest in a video





Video Segmentation

- Segmenting and associating objects of interest in a video



Video Segmentation: Sub-tasks

- Segmenting and associating objects of interest in a video

Video Segmentation: Sub-tasks

- Segmenting and associating objects of interest in a video
  - Video Instance Segmentation (VIS)

(e.g., 40 categories for Youtube-VIS
25 categories for OVIS)

Example from OVIS dataset

Video Segmentation: Sub-tasks

- Segmenting and associating objects of interest in a video
  - Video Instance Segmentation (VIS)
  - Multi-Object Tracking and Segmentation (MOTS)

(cars, pedestrians)

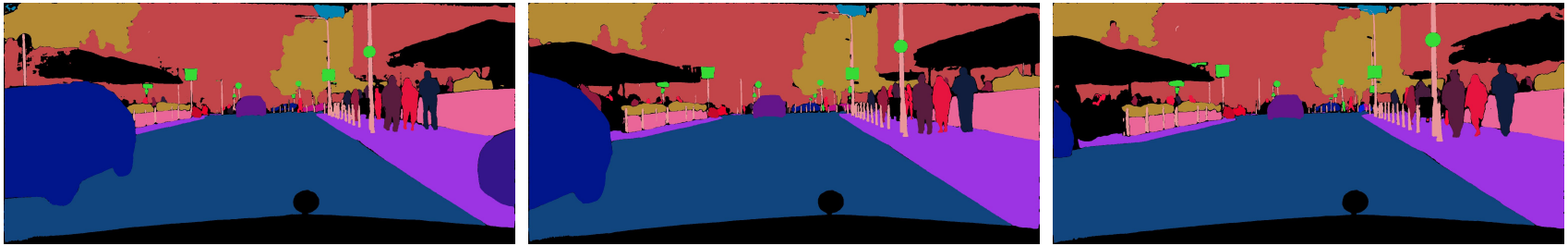
Example from KITTI-MOTS dataset

Video Segmentation: Sub-tasks

- Segmenting and associating objects of interest in a video
  - Video Instance Segmentation (VIS)
  - Multi-Object Tracking and Segmentation (MOTS)
  - Video Panoptic Segmentation (VPS)

(e.g., 18 categories in Cityscapes-VPS,
also includes sky, road, etc)

Example from Cityscapes-VPS dataset



Video Segmentation: Sub-tasks

- Segmenting and associating objects of interest in a video
  - Video Instance Segmentation (VIS)
  - Multi-Object Tracking and Segmentation (MOTS)
  - Video Panoptic Segmentation (VPS)

} known categories

Prior Work

Transformers for Video Segmentation?

- MinVIS [1]
- IDOL [2]
- TrackFormer [3]
- VisTR [4]
- Mask2Former-VIS [5]
- TubeFormer-DeepLab [6]

[1]

[1] Huang et al., NeurIPS 2022
[2] Wu et al., ECCV 2022
[3] Meinhardt et al., CVPR 2022
[4] Wang et al., CVPR 2021
[5] Cheng et al., arXiv 2022
[6] Kim et. al., CVPR 2022

Transformers for Video Segmentation?

Local object queries

- MinVIS [1]
- IDOL [2]
- TrackFormer [3]

- VisTR [4]
- Mask2Former-VIS [5]
- TubeFormer-DeepLab [6]

[1]

[1] Huang et al., NeurIPS 2022
[2] Wu et al., ECCV 2022
[3] Meinhardt et al., CVPR 2022
[4] Wang et al., CVPR 2021
[5] Cheng et al., arXiv 2022
[6] Kim et. al., CVPR 2022

Transformers for Video Segmentation?

Local object queries

- MinVIS [1]
- IDOL [2]
- TrackFormer [3]

Global object queries

- VisTR [4]
- Mask2Former-VIS [5]
- TubeFormer-DeepLab [6]

[1]

[1] Huang et al., NeurIPS 2022
[2] Wu et al., ECCV 2022
[3] Meinhardt et al., CVPR 2022
[4] Wang et al., CVPR 2021
[5] Cheng et al., arXiv 2022
[6] Kim et. al., CVPR 2022

Transformers for Video Segmentation?

Local object queries

- MinVIS [1]
- IDOL [2]
- TrackFormer [3]

[1]

[1] Huang et al., NeurIPS 2022
[2] Wu et al., ECCV 2022
[3] Meinhardt et al., CVPR 2022
[4] Wang et al., CVPR 2021
[5] Cheng et al., arXiv 2022
[6] Kim et. al., CVPR 2022

Transformers for Video Segmentation?

Local object queries

- MinVIS [1]
- IDOL [2]

} 2-steps: segmentation followed by association

- TrackFormer [3]

[1]

[1] Huang et al., NeurIPS 2022
[2] Wu et al., ECCV 2022
[3] Meinhardt et al., CVPR 2022

Transformers for Video Segmentation?

Local object queries

- MinVIS [1]
- IDOL [2]
- TrackFormer [3]

} 2 types of object queries

[1] Huang et al., NeurIPS 2022
[2] Wu et al., ECCV 2022
[3] Meinhardt et al., CVPR 2022

Transformers for Video Segmentation?

Global object queries

- VisTR [1]
- Mask2Former-VIS [2]
- TubeFormer-DeepLab [3]

[1] Wang et al., CVPR 2021
[2] Cheng et al., arXiv 2022
[3] Kim et. al., CVPR 2022

Transformers for Video Segmentation?

Global object queries

Offline processing

- VisTR [1]
- Mask2Former-VIS [2]
- TubeFormer-DeepLab [3]

[1] Wang et al., CVPR 2021
[2] Cheng et al., arXiv 2022
[3] Kim et. al., CVPR 2022

Local vs. Global Object Queries

Local object queries

- Local temporal field of view

Global object queries

- Large temporal field of view

Local vs. Global Object Queries

Local object queries

- Local temporal field of view

Global object queries

- Large temporal field of view

However:

Local vs. Global Object Queries

Local object queries

↳ Local temporal field of view

Global object queries

↳ Large temporal field of view

However:

Method	Type	YTVIS 2021 (AP)	YTVIS 2019 (AP)	OVIS (AP)
MinVIS	Image-level queries	**44.2**	**47.4**	**25.0**
Mask2Former-VIS	Video-level queries	40.6	46.4	17.3*

* : 30 frames at a time

Open Questions!

1. Why do global object queries fail to accurately represent objects spatiotemporally?

Open Questions!

1. Why do global object queries fail to accurately represent objects spatiotemporally?

Object queries are often too reliant on the static spatial positions!

How to address this?

Open Questions!

1. Why do global object queries fail to accurately represent objects spatiotemporally?

2. How to extend object queries to temporal domain while processing frames sequentially?

Context-Aware Relative Object Queries to Unify Video Instance and Panoptic Segmentation

Context-Aware Relative Object Queries

- Continuous refinement and propagation of object queries frame-by-frame

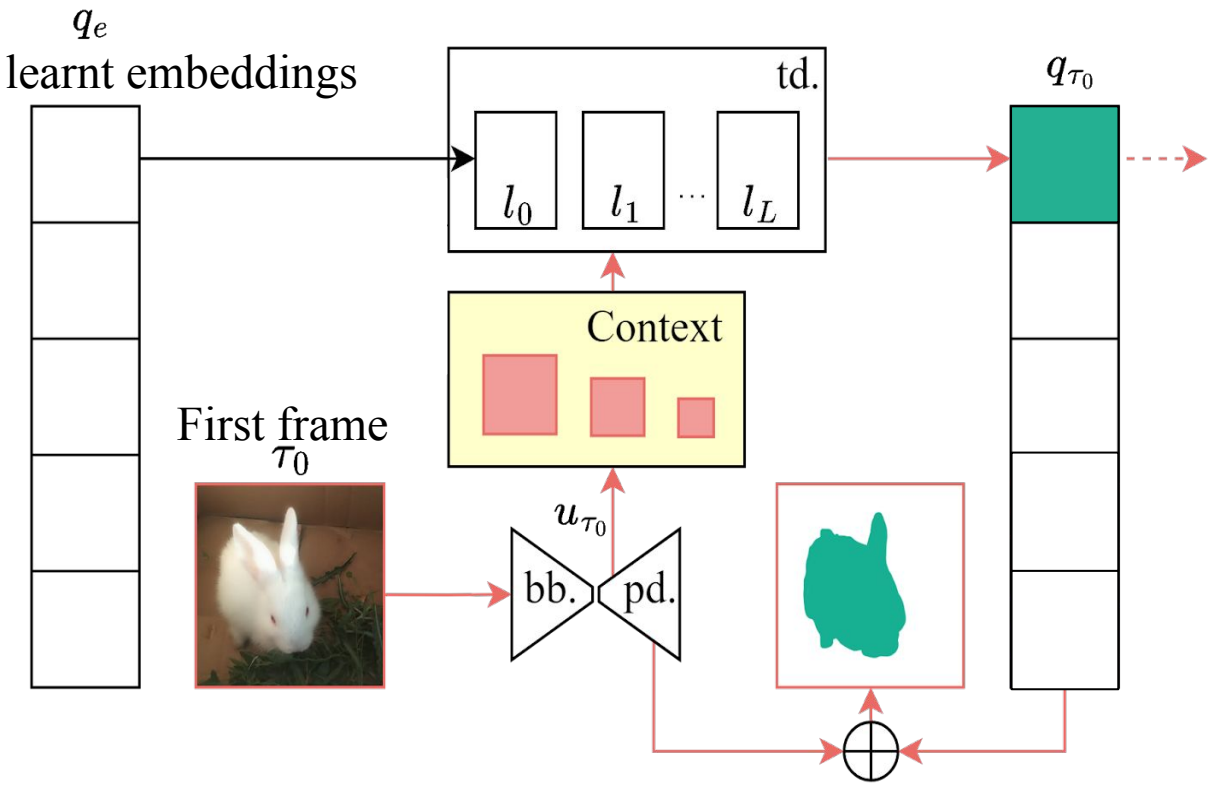
Context-Aware Relative Object Queries

- Continuous refinement and propagation of object queries frame-by-frame
- Context-Aware Queries: Spatio-temporal context even for frame-by-frame processing

Context-Aware Relative Object Queries

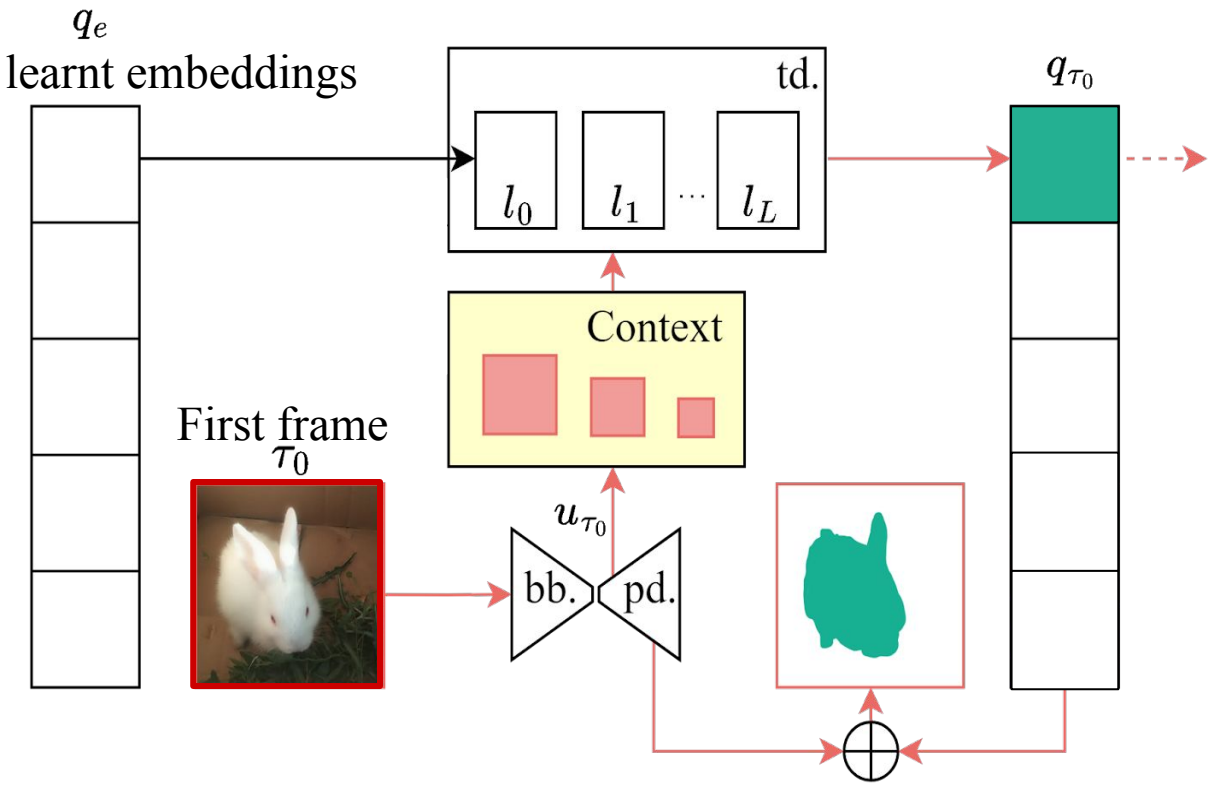
- Continuous refinement and propagation of object queries frame-by-frame
- Context-Aware Queries: Spatio-temporal context even for frame-by-frame processing
- Relative Object Queries: Use of relative positional encodings

Context-Aware Relative Object Queries



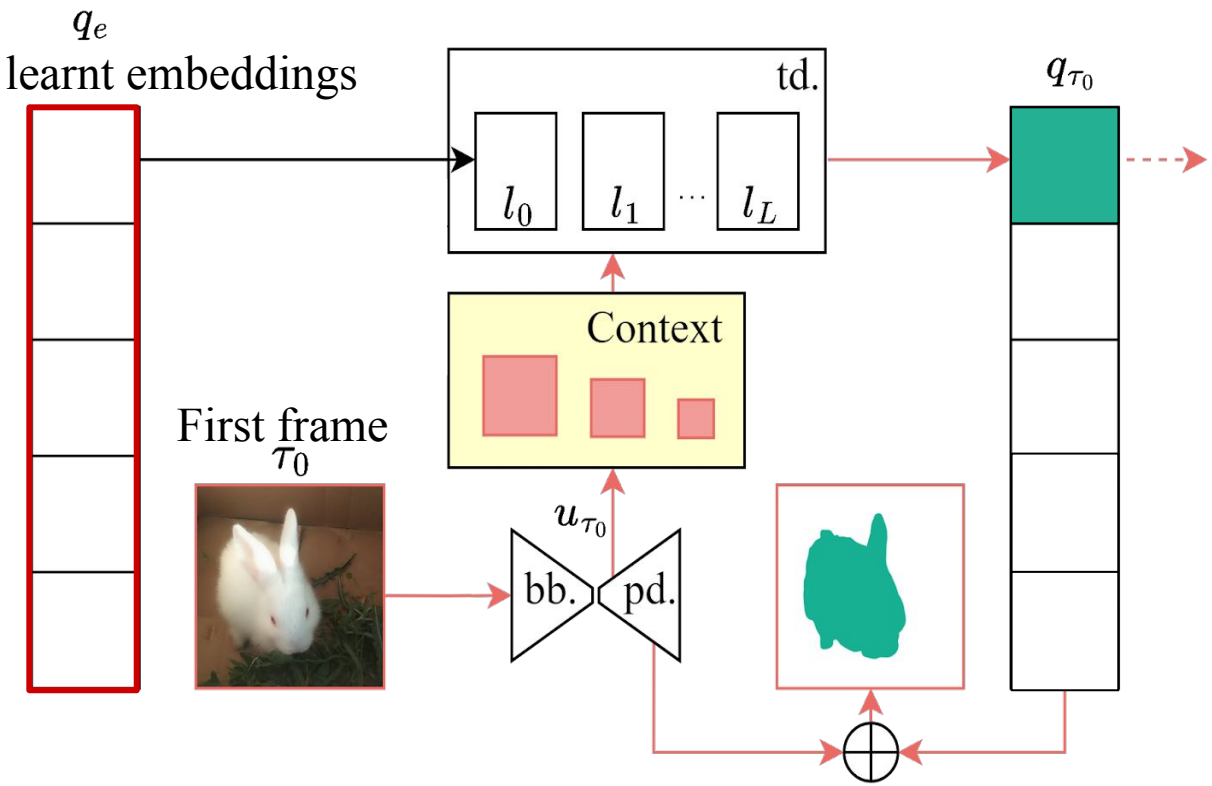
bb: backbone
pd: pixel decoder
td: transformer decoder

Context-Aware Relative Object Queries



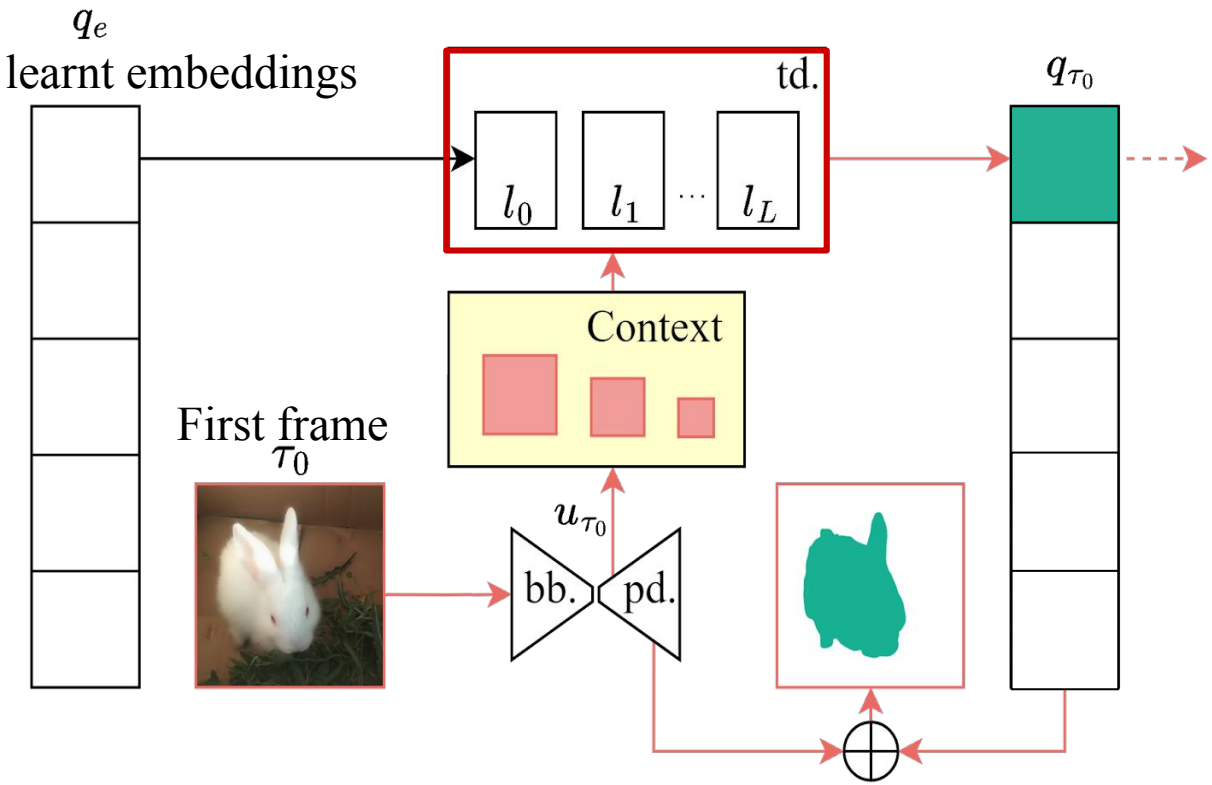
bb: backbone
pd: pixel decoder
td: transformer decoder

Context-Aware Relative Object Queries



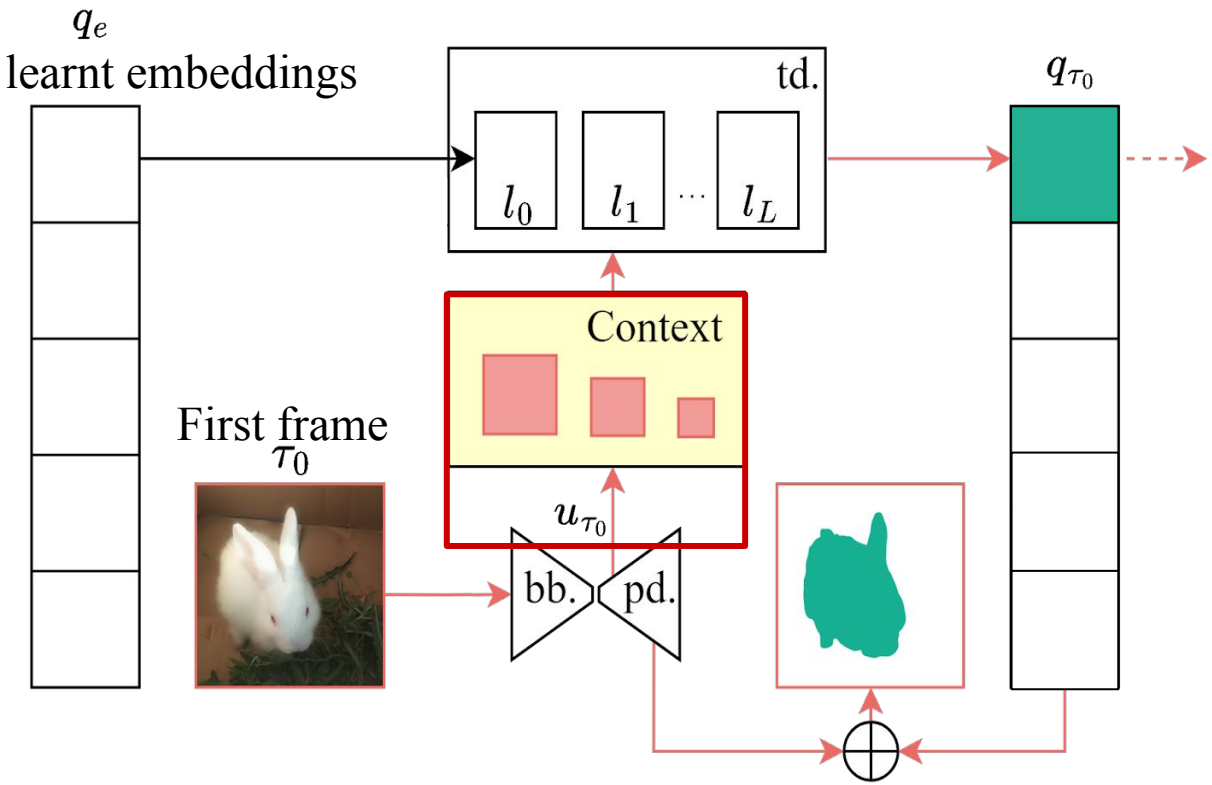
bb: backbone
pd: pixel decoder
td: transformer decoder

Context-Aware Relative Object Queries



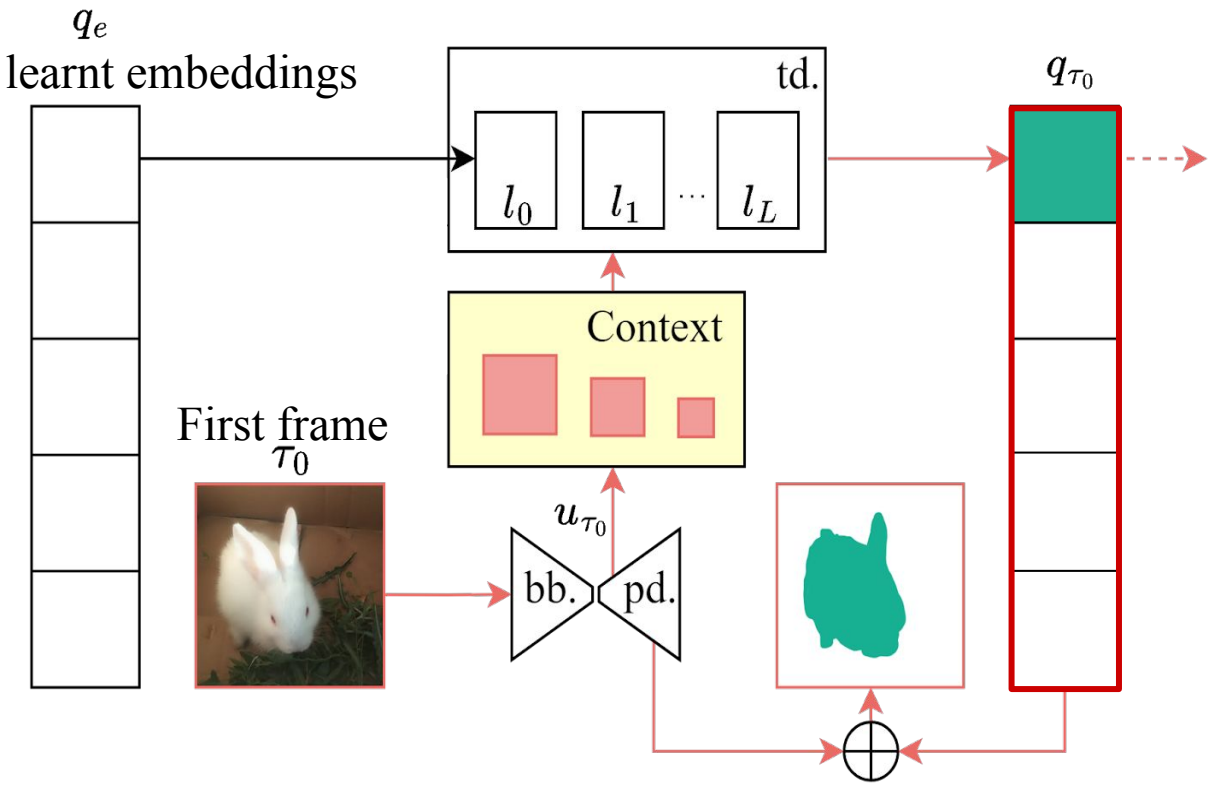
bb: backbone
pd: pixel decoder
td: transformer decoder

Context-Aware Relative Object Queries



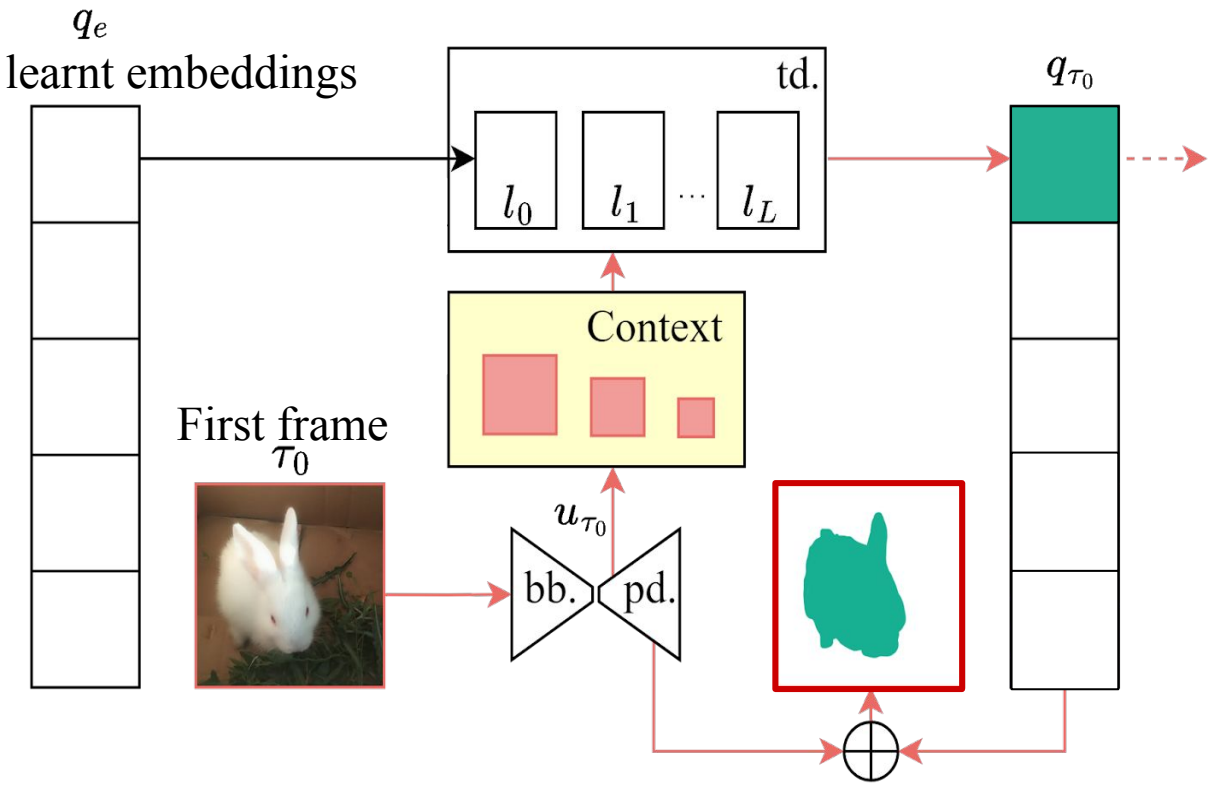
bb: backbone
pd: pixel decoder
td: transformer decoder

Context-Aware Relative Object Queries



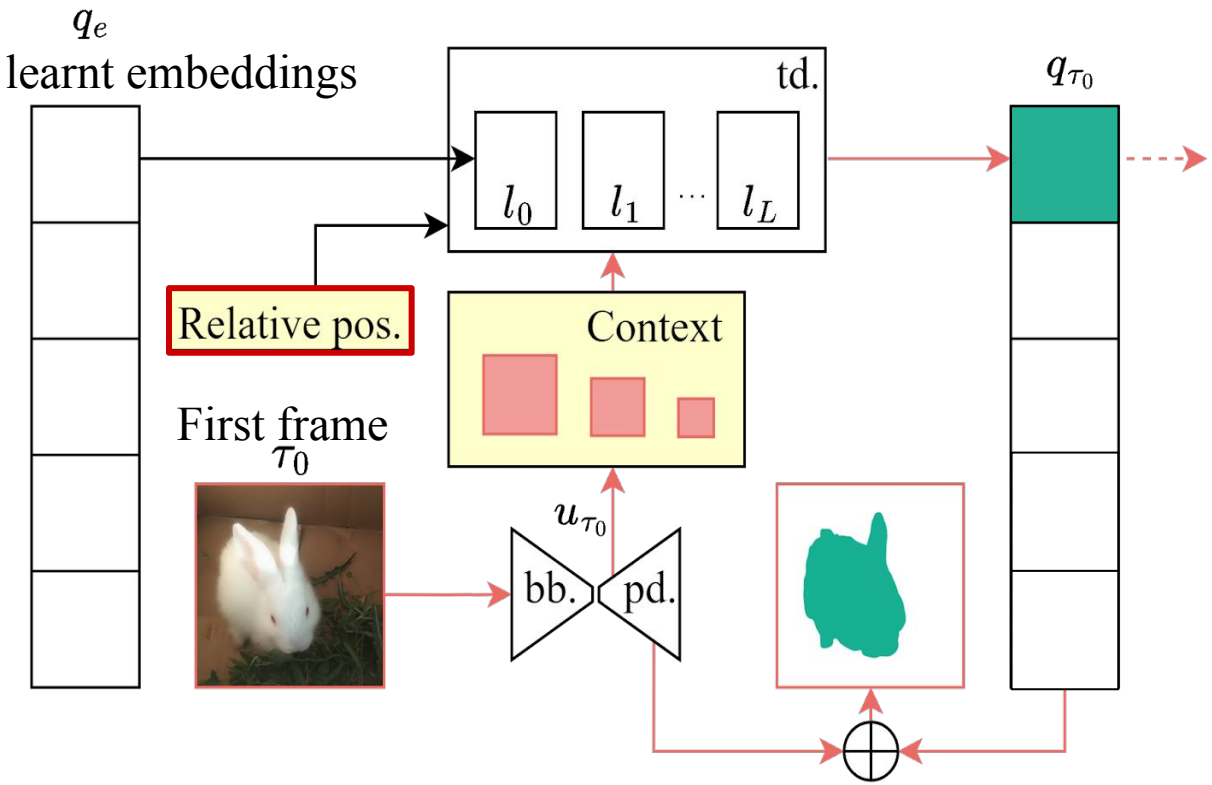
bb: backbone
pd: pixel decoder
td: transformer decoder

Context-Aware Relative Object Queries



bb: backbone
pd: pixel decoder
td: transformer decoder

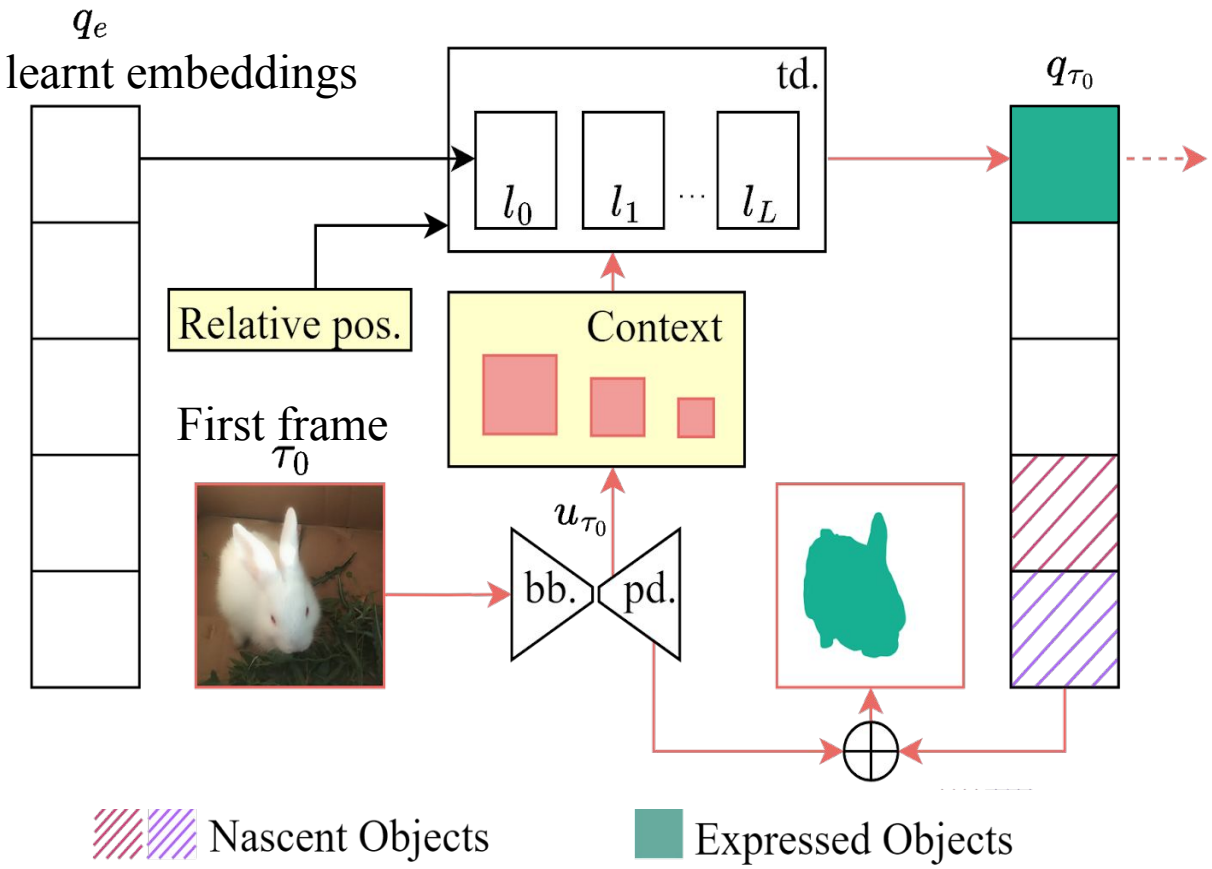
Context-Aware Relative Object Queries



bb: backbone
pd: pixel decoder
td: transformer decoder

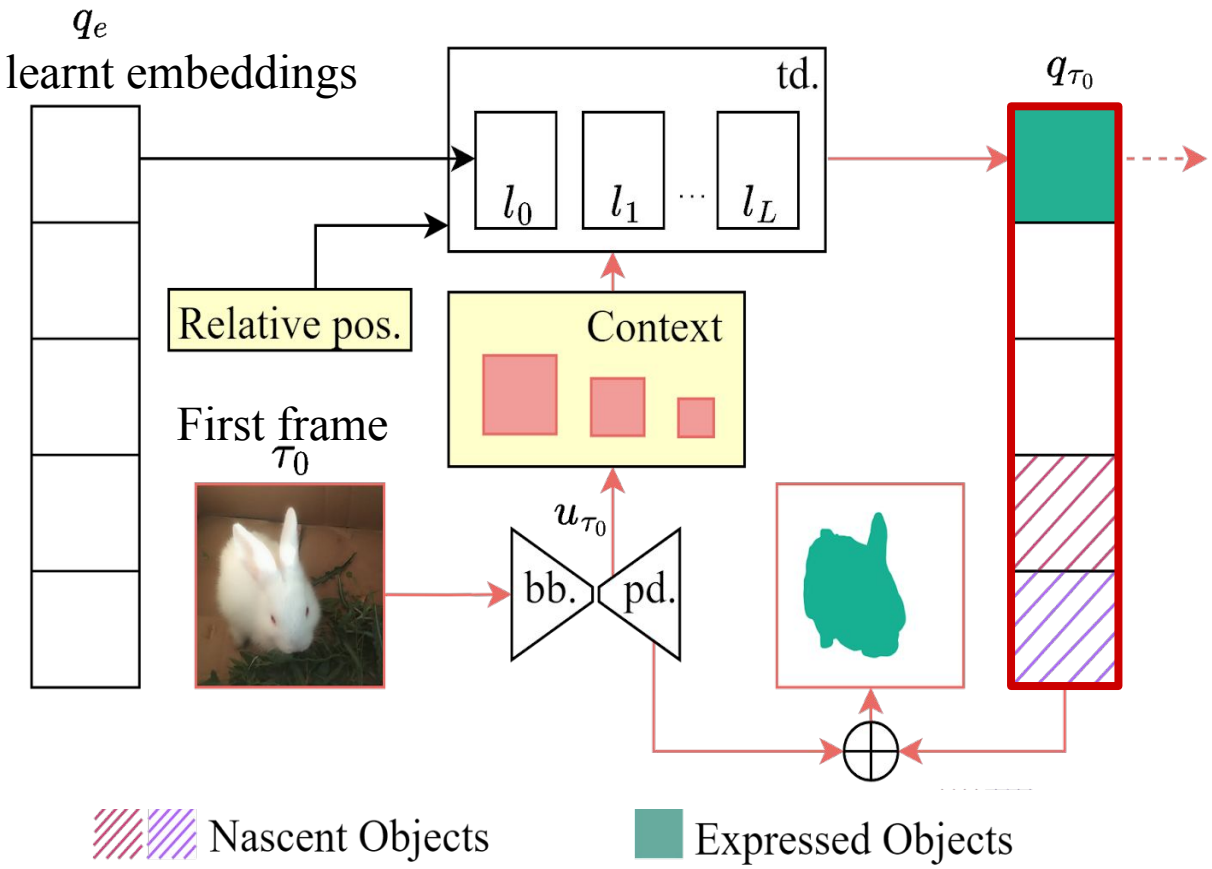
a) **Propagation** of Context-Aware Relative Object Queries

Propagation of Context-Aware Relative Object Queries



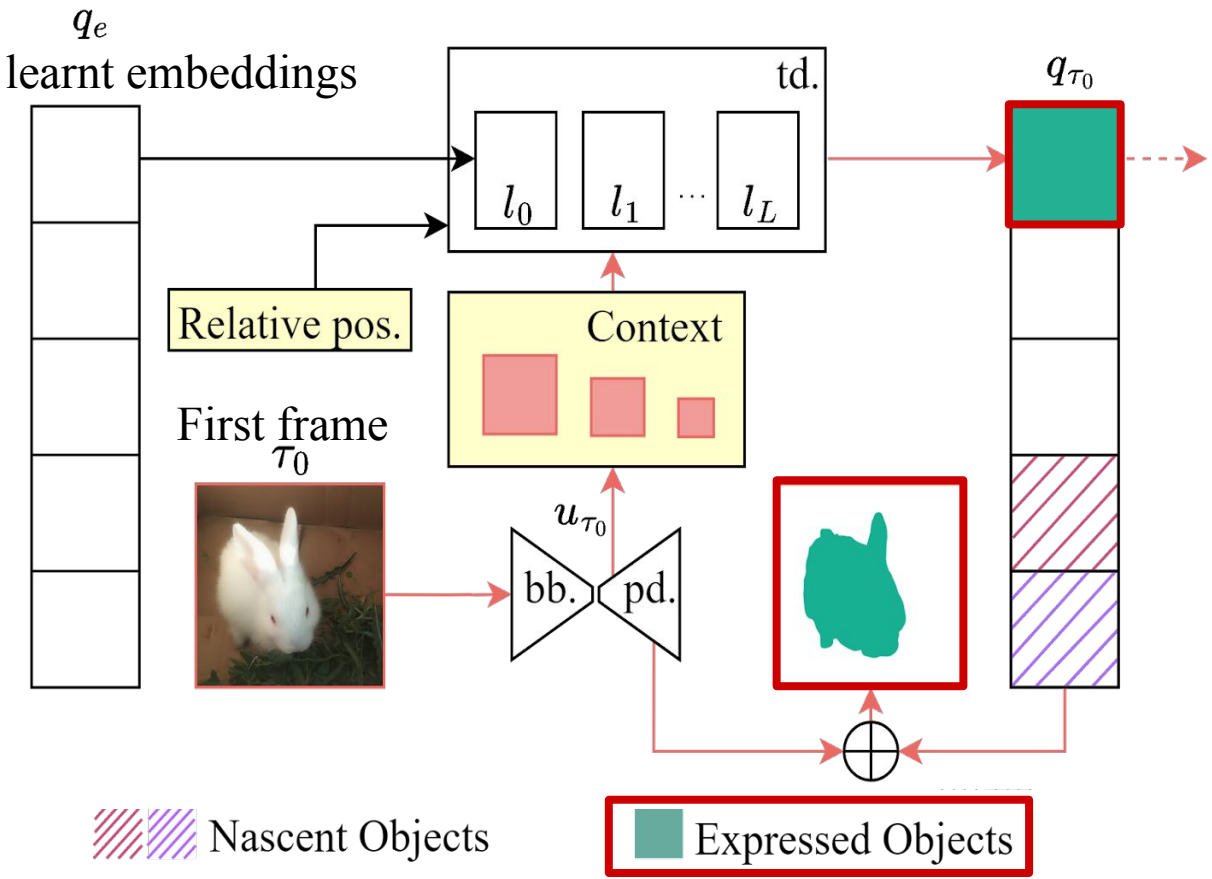
bb: backbone
pd: pixel decoder
td: transformer decoder

Propagation of Context-Aware Relative Object Queries



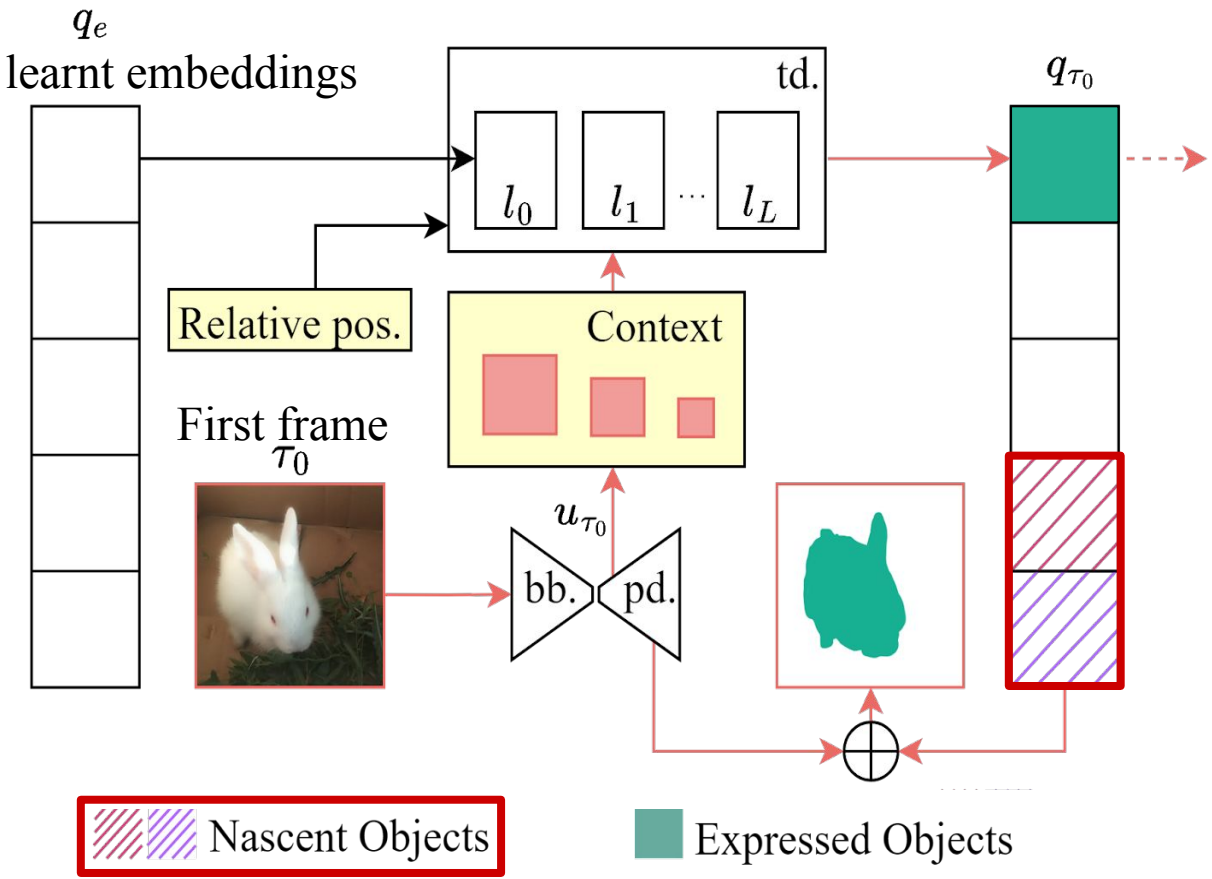
bb: backbone
pd: pixel decoder
td: transformer decoder

Propagation of Context-Aware Relative Object Queries



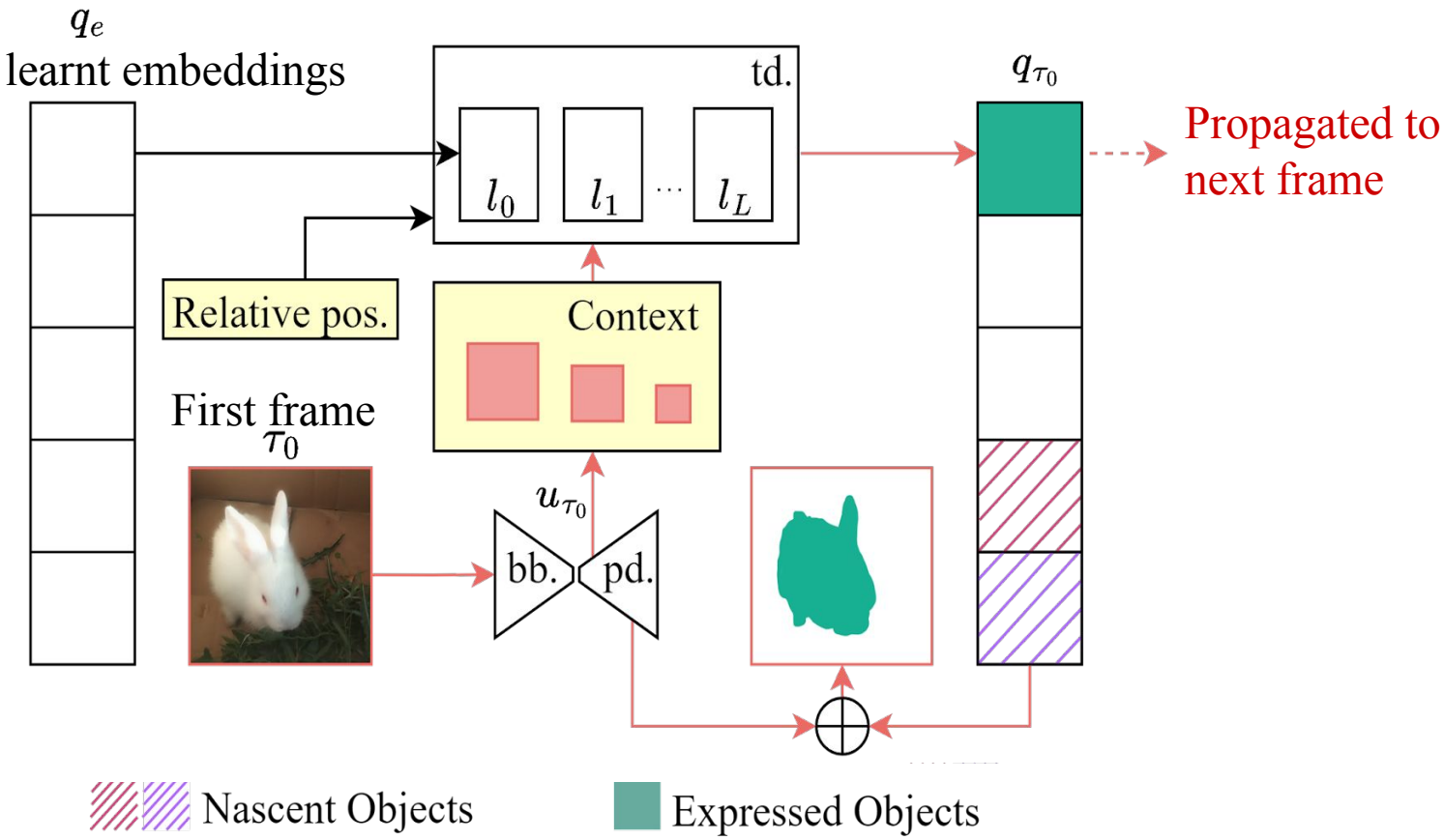
bb: backbone
pd: pixel decoder
td: transformer decoder

Propagation of Context-Aware Relative Object Queries



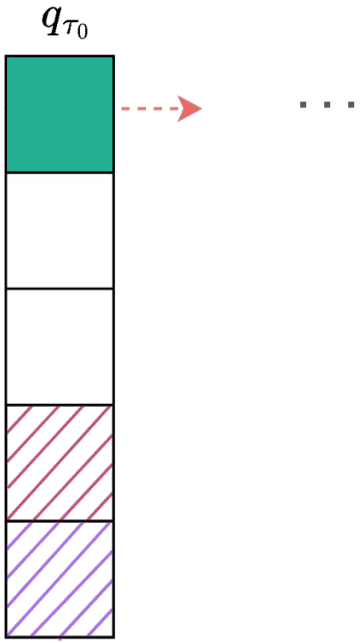
bb: backbone
pd: pixel decoder
td: transformer decoder

Propagation of Context-Aware Relative Object Queries

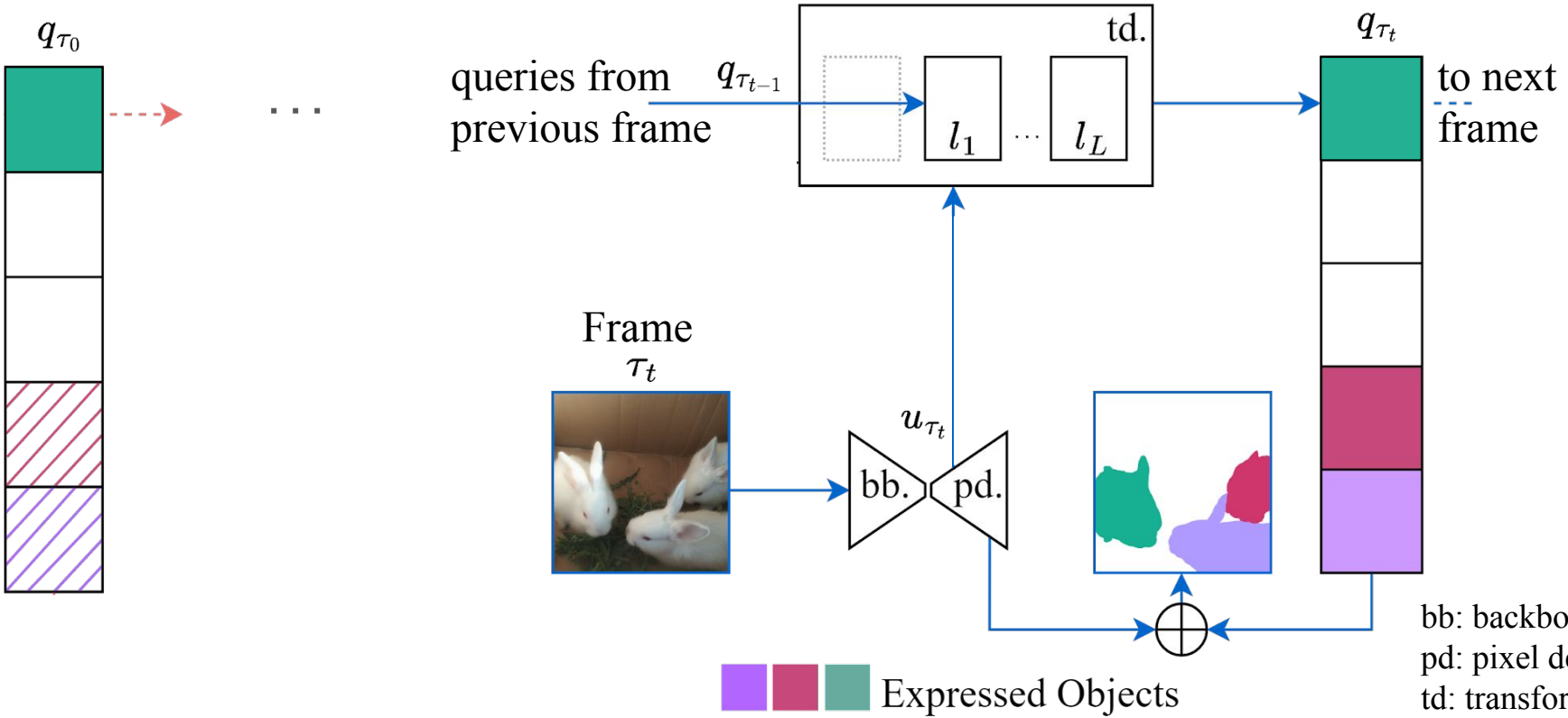


bb: backbone
pd: pixel decoder
td: transformer decoder

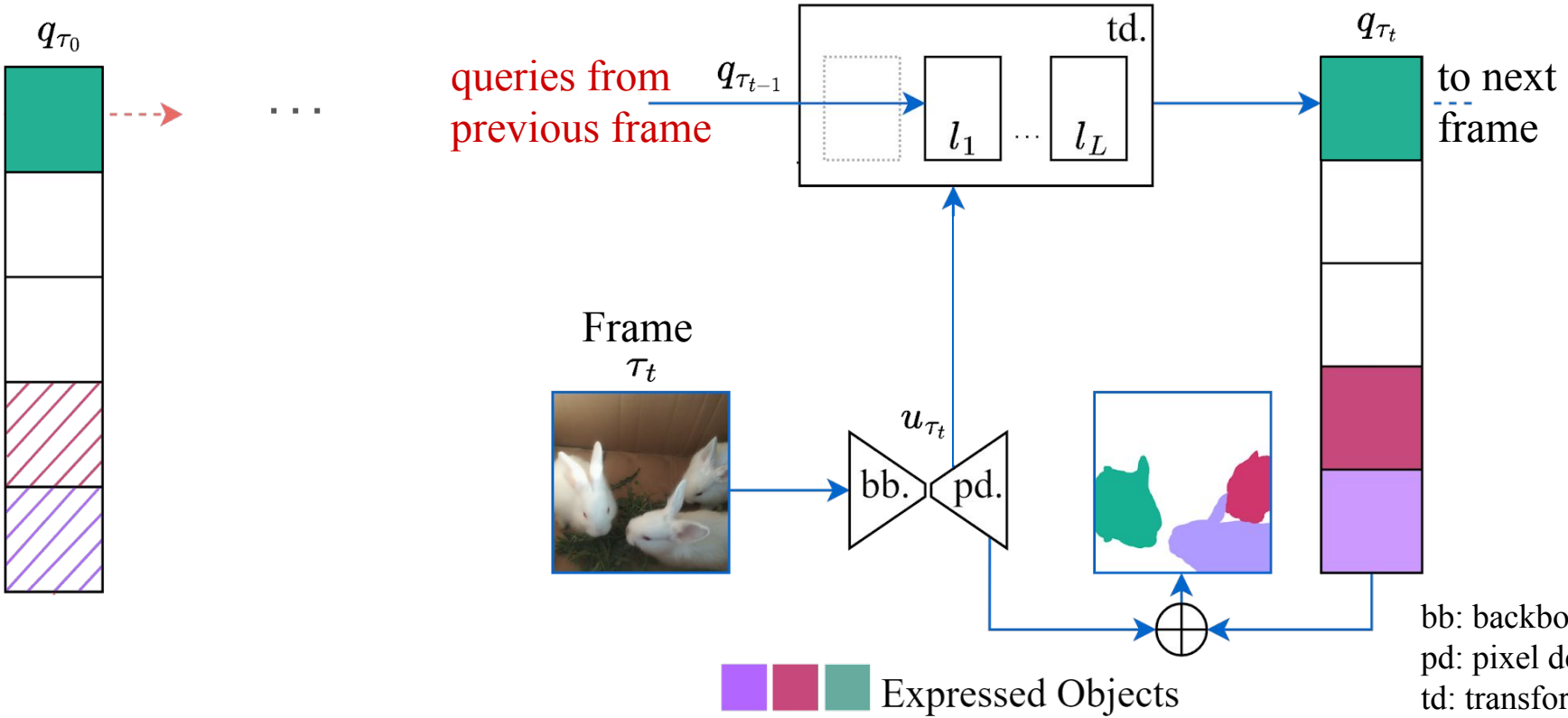
Propagation of Context-Aware Relative Object Queries



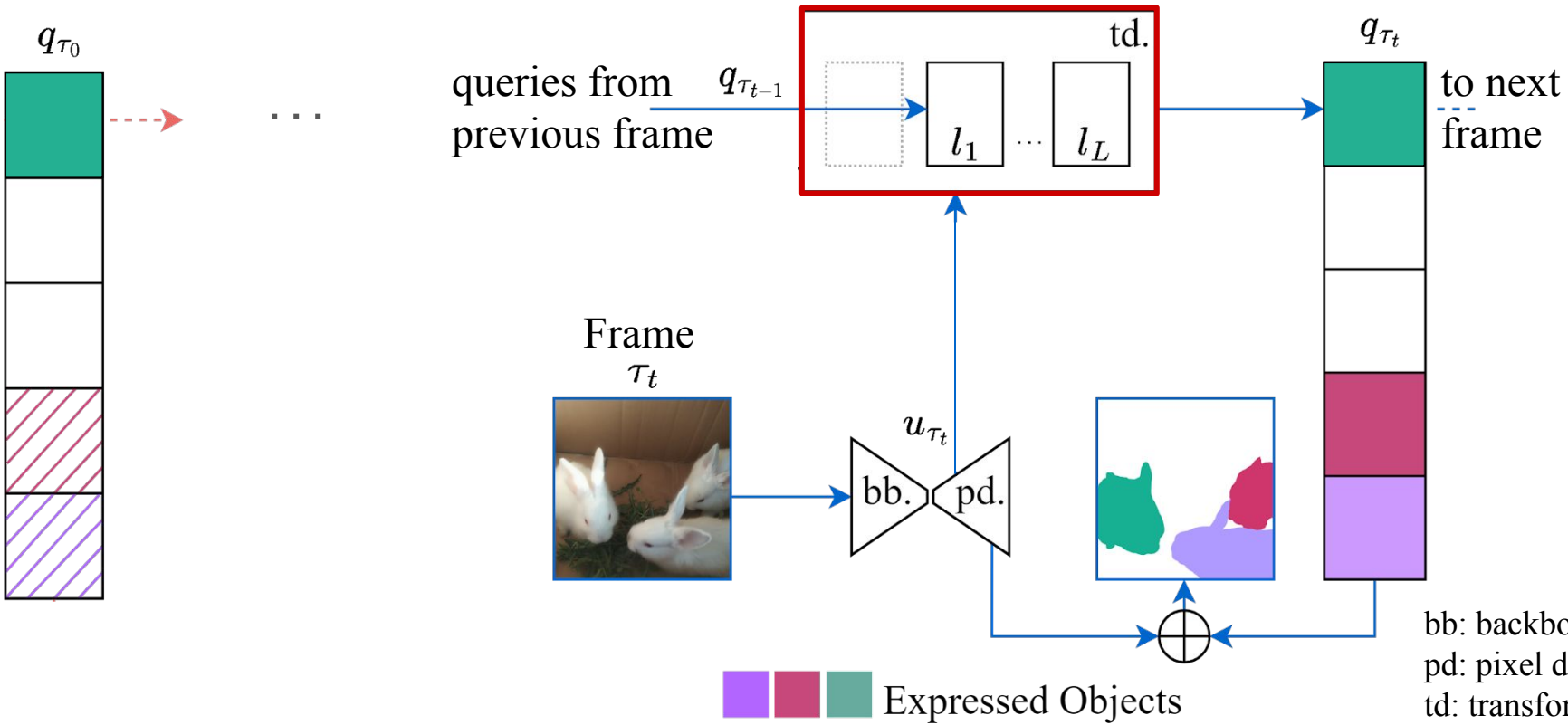
Propagation of Context-Aware Relative Object Queries



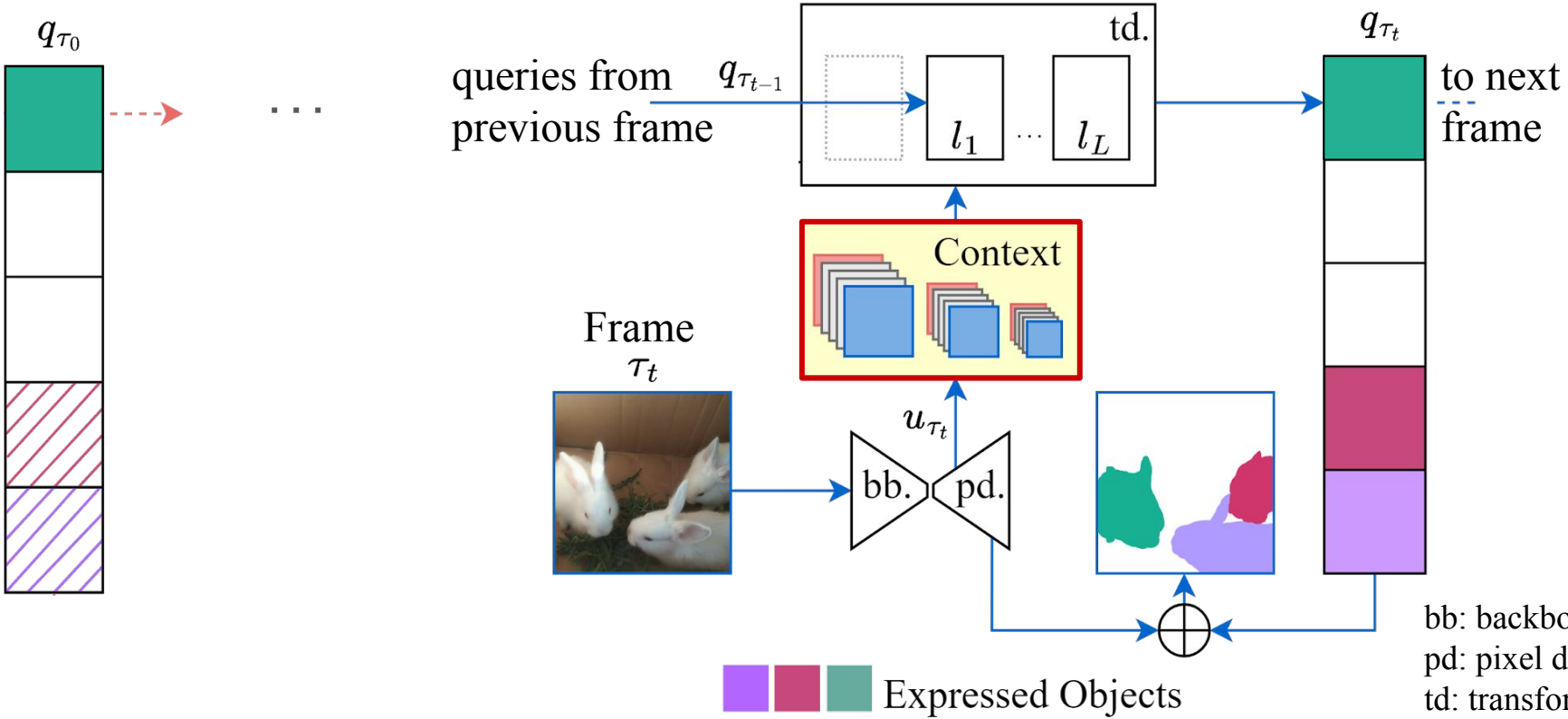
Propagation of Context-Aware Relative Object Queries



Propagation of Context-Aware Relative Object Queries

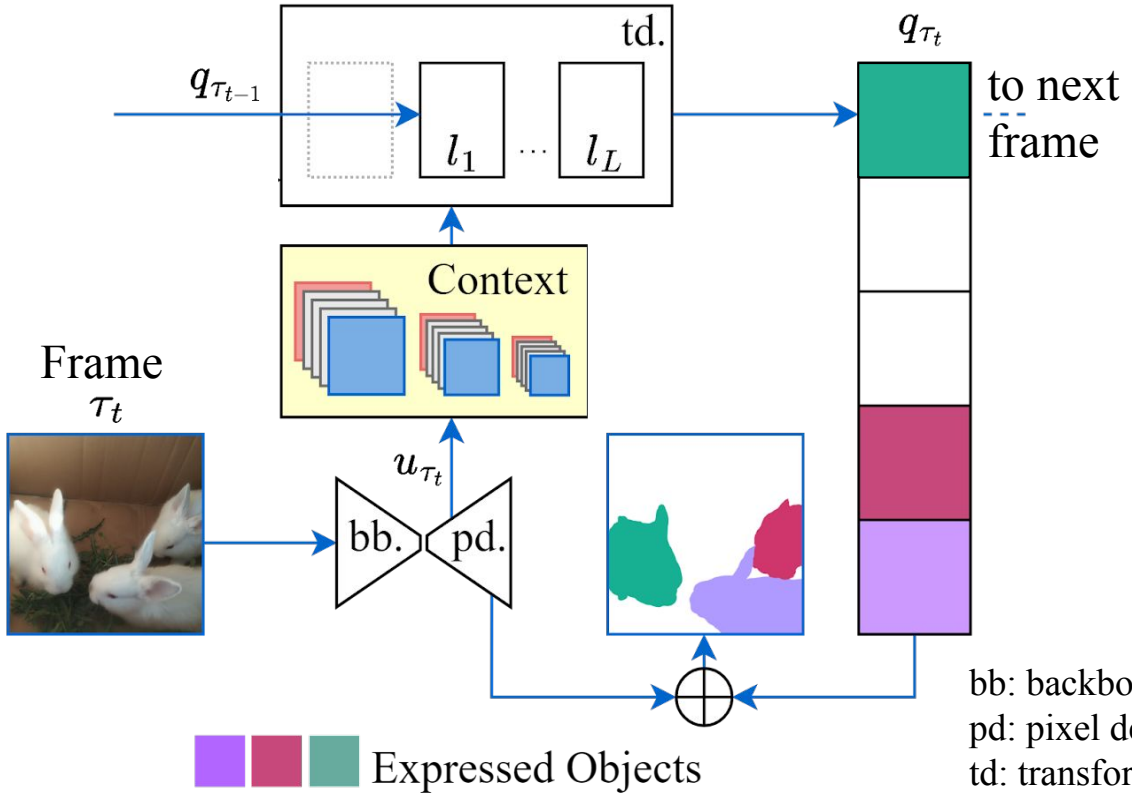


Propagation of Context-Aware Relative Object Queries



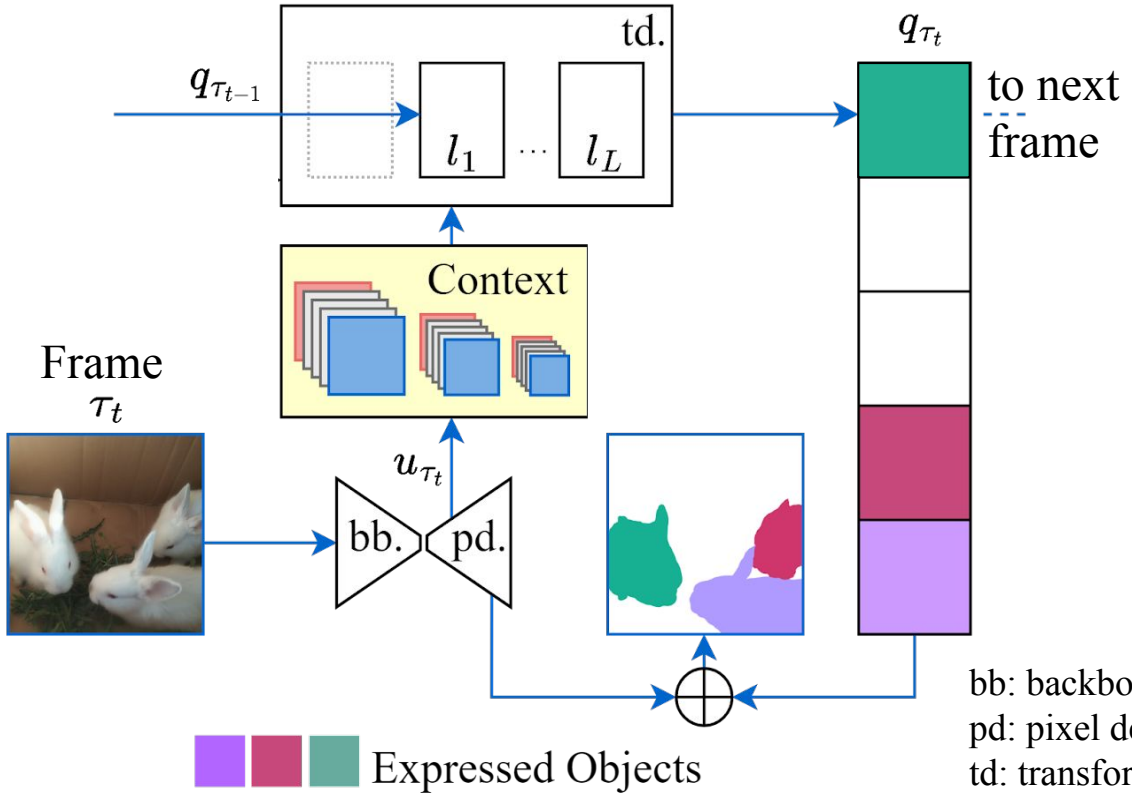
Propagation of Context-Aware Relative Object Queries

- Model gradual appearance change



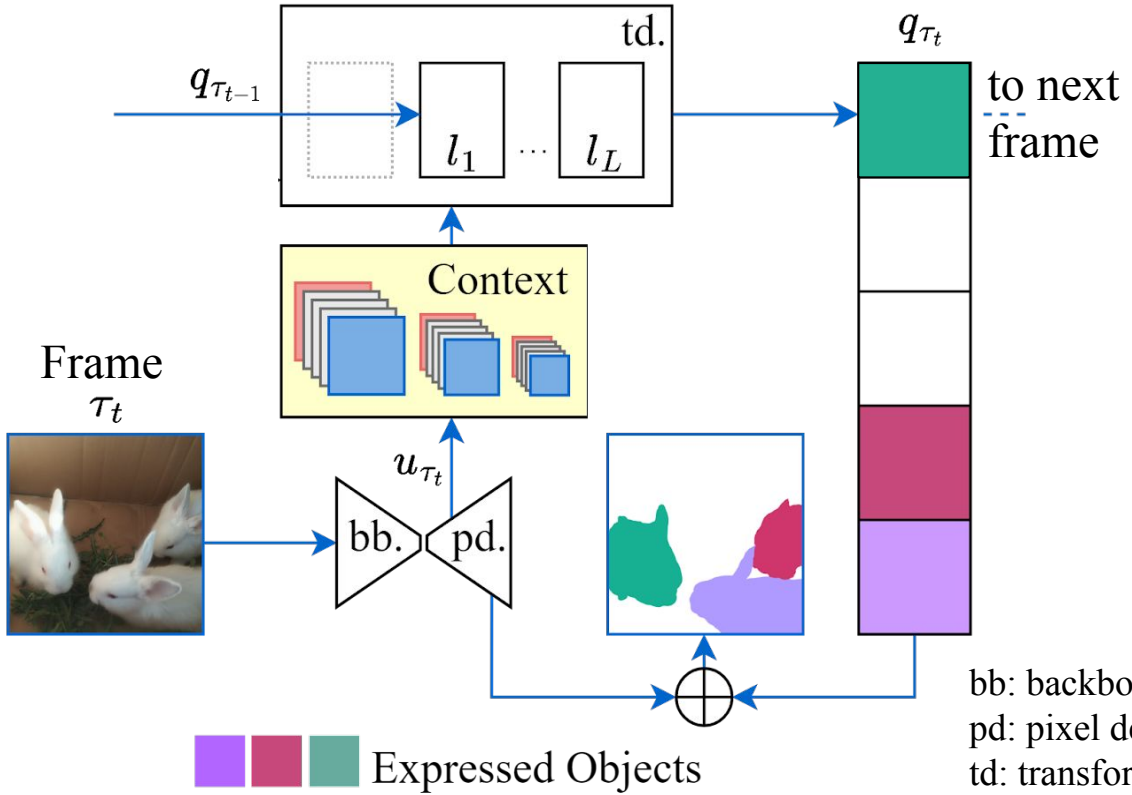
Propagation of Context-Aware Relative Object Queries

- Model gradual appearance change
- Carry long-term temporal information
- Handle occlusions, absence, viewpoint changes.



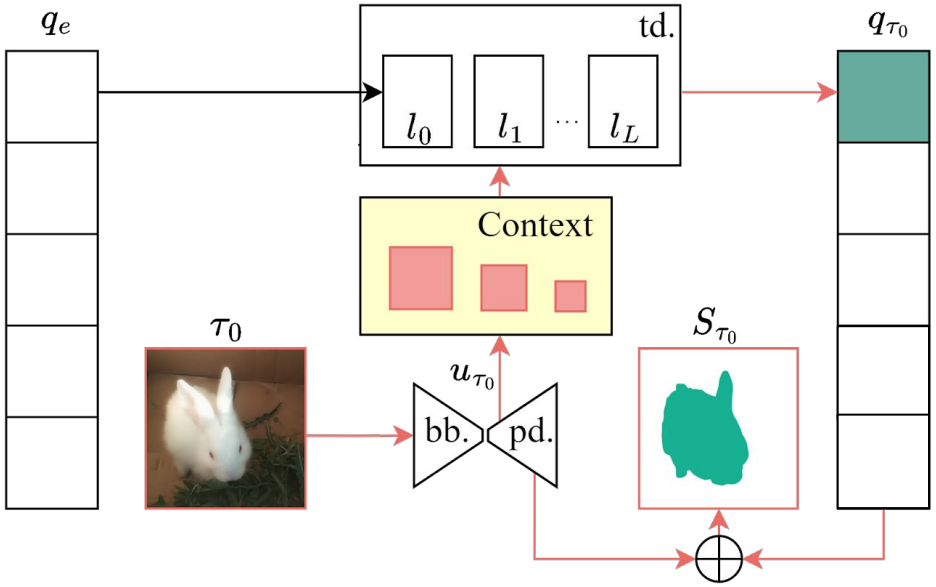
Propagation of Context-Aware Relative Object Queries

- Model gradual appearance change
- Carry long-term temporal information
- Handle occlusions, absence, viewpoint changes.

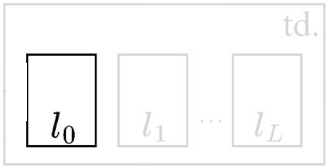


b) **Relative** Object Queries

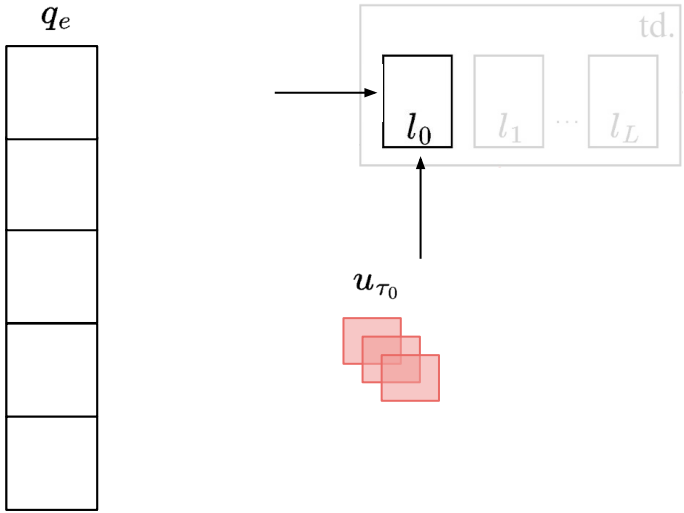
Prior Work: Absolute Positional Encoding



Prior Work: Absolute Positional Encoding

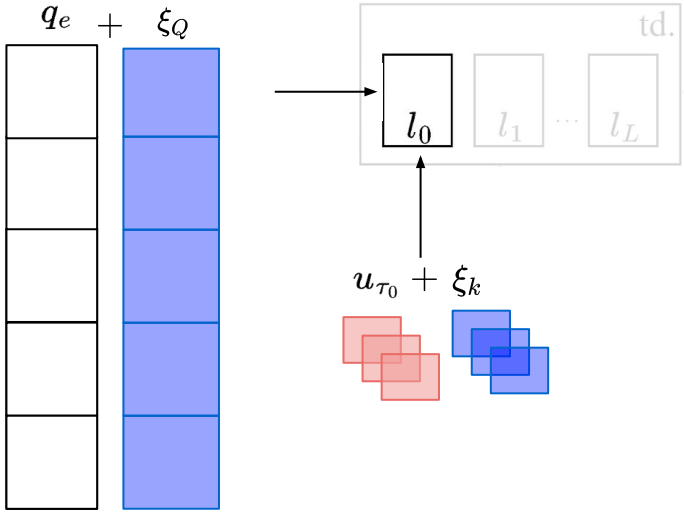


Prior Work: Absolute Positional Encoding



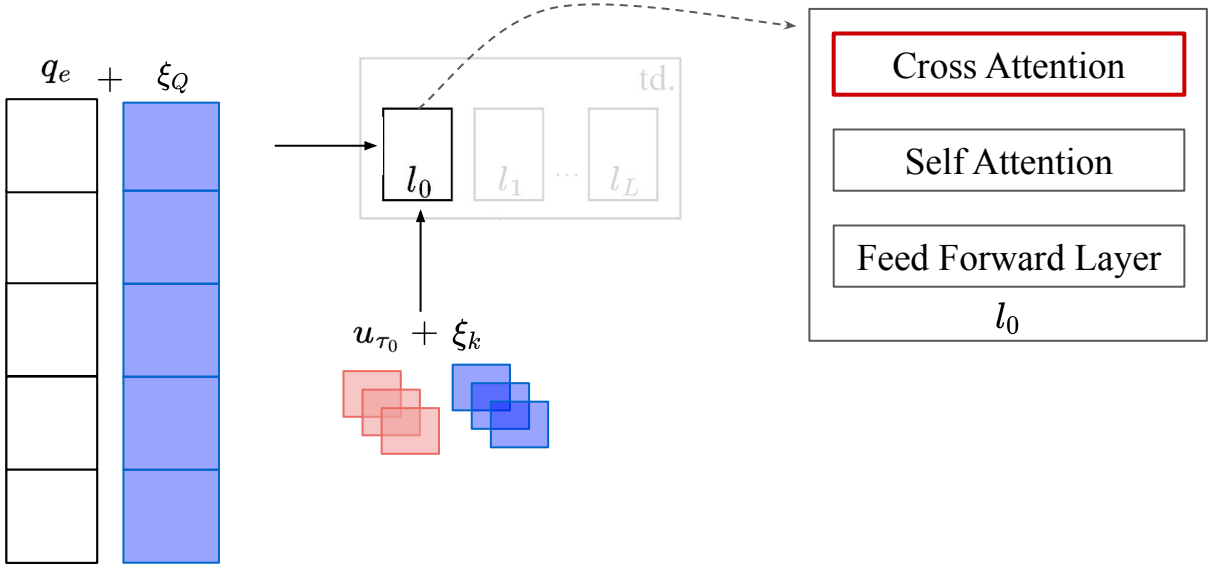
u_{τ_0}: intermediate features
q_e : learnt embeddings

Prior Work: Absolute Positional Encoding



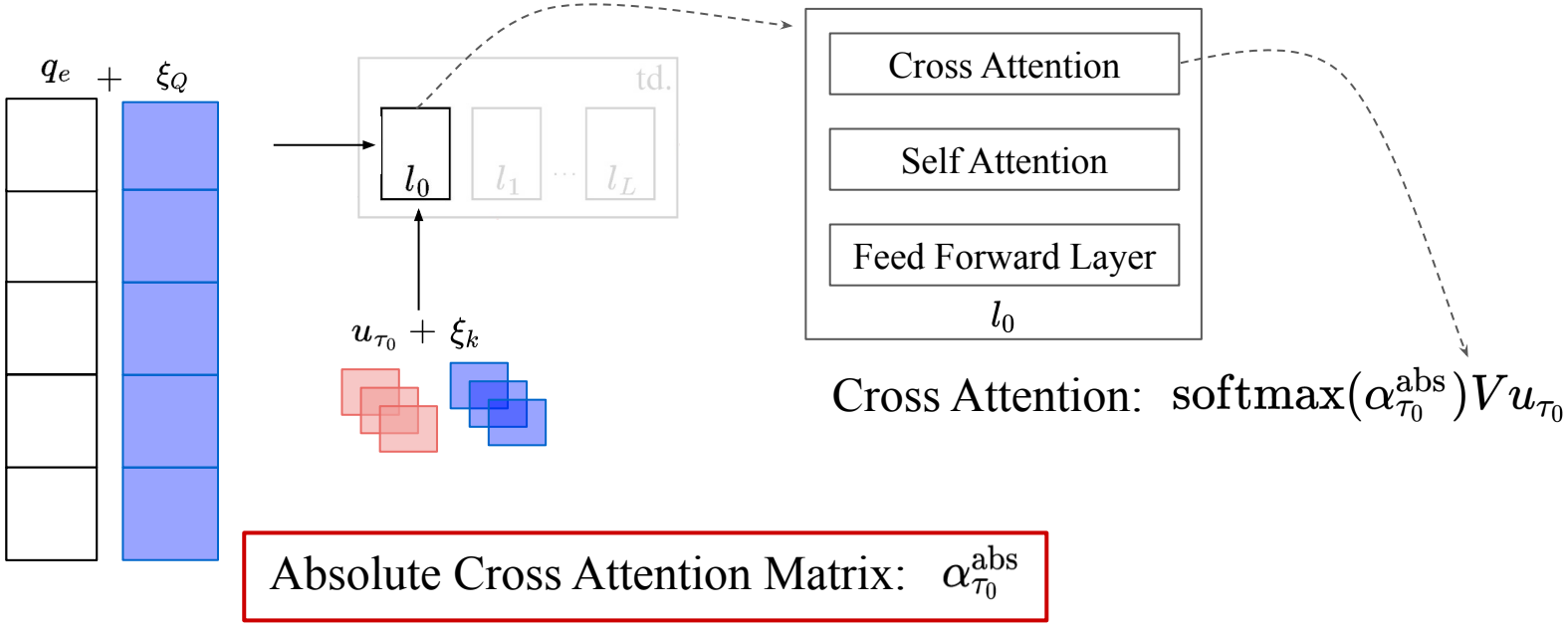
u_{τ_0}: intermediate features
q_e : learnt embeddings

Prior Work: Absolute Positional Encoding



u_{τ_0}: intermediate features
q_e : learnt embeddings

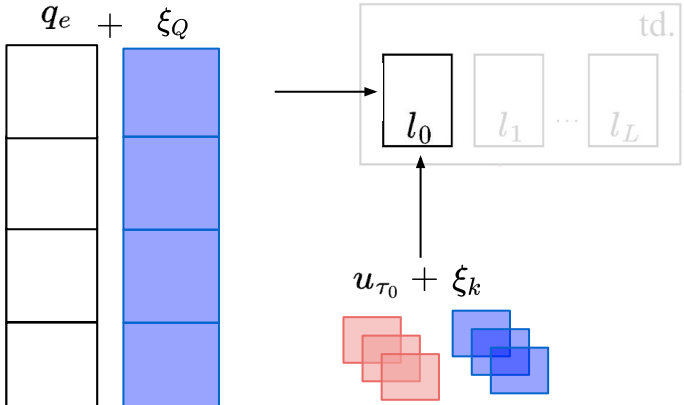
Prior Work: Absolute Positional Encoding



Absolute Cross Attention Matrix: $\alpha^{\mathrm{abs}}_{\tau_0}$

u_{τ_0}: intermediate features
q_e : learnt embeddings
V: learnt weights

Prior Work: Absolute Positional Encoding



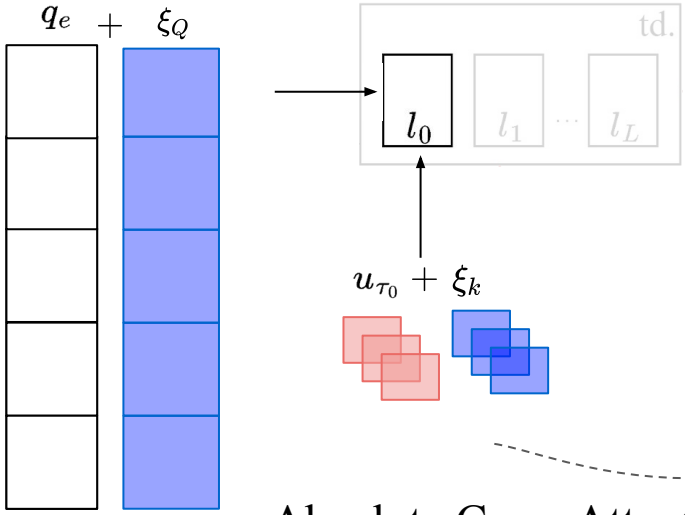
Cross Attention: $\text{softmax}(\alpha^{\text{abs}}_{\tau_0})Vu_{\tau_0}$

Absolute Cross Attention Matrix: $\alpha^{\text{abs}}_{\tau_0} = Q(\quad)[K(\quad)]^\top$

u_{τ_0}: intermediate features
q_e : learnt embeddings

ξ_Q, ξ_k : absolute position encodings
V, Q, K : learnt weights

Prior Work: Absolute Positional Encoding



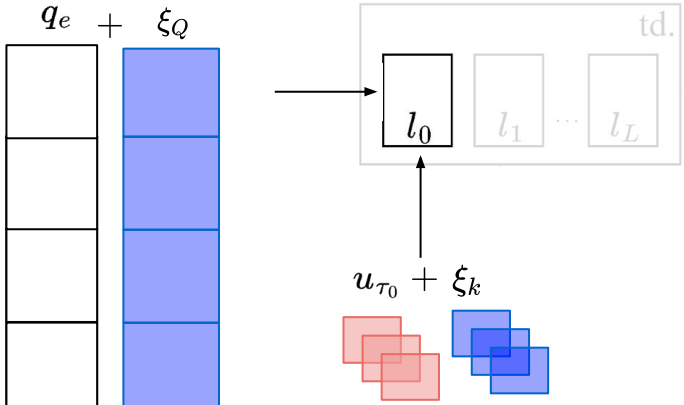
Cross Attention: $\mathrm{softmax}(\alpha^{\mathrm{abs}}_{\tau_0})Vu_{\tau_0}$

Absolute Cross Attention Matrix: $\alpha^{\mathrm{abs}}_{\tau_0} = Q($ $)[K($ $)]^\top$

u_{τ_0}: intermediate features
q_e : learnt embeddings

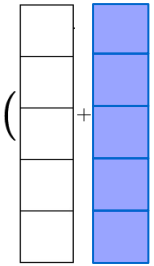
ξ_Q, ξ_k : absolute position encodings
V, Q, K : learnt weights

Prior Work: Absolute Positional Encoding



Cross Attention: $\mathrm{softmax}(\alpha^{\mathrm{abs}}_{\tau_0})Vu_{\tau_0}$

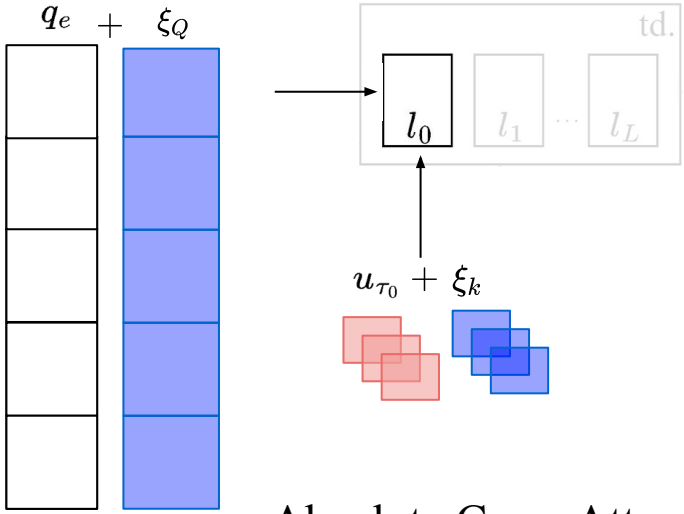
Absolute Cross Attention Matrix: $\alpha^{\mathrm{abs}}_{\tau_0} = Q(\quad)[K(\quad)]^\top$



u_{τ_0}: intermediate features
q_e : learnt embeddings

ξ_Q, ξ_k : absolute position encodings
V, Q, K : learnt weights

Prior Work: Absolute Positional Encoding



$q_e \in \mathbb{R}^{N \times C}$
$u_{\tau_0} \in \mathbb{R}^{H \times W \times C}$
$\alpha^{\text{abs}}_{\tau_0} \in \mathbb{R}^{N \times H \times W}$

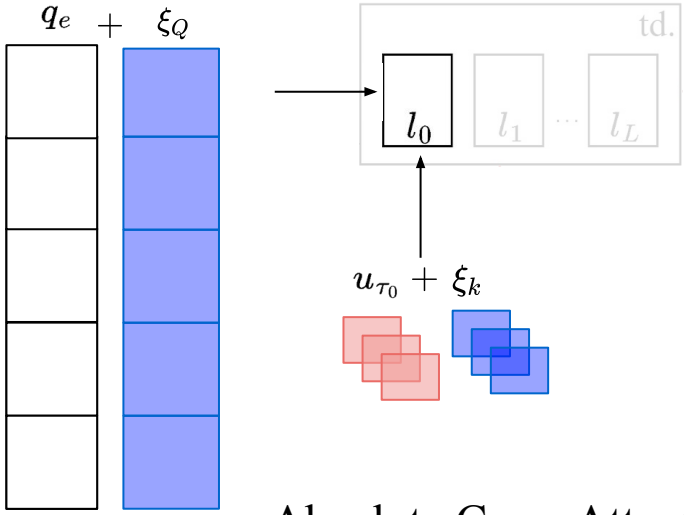
Cross Attention: $\text{softmax}(\alpha^{\text{abs}}_{\tau_0}) V u_{\tau_0}$

Absolute Cross Attention Matrix: $\alpha^{\text{abs}}_{\tau_0} = Q(q_e + \xi_Q)[K(u_{\tau_0} + \xi_K)]^\top$

u_{τ_0}: intermediate features
q_e : learnt embeddings

ξ_Q, ξ_k : absolute position encodings
V, Q, K : learnt weights

Prior Work: Absolute Positional Encoding



$q_e \in \mathbb{R}^{N \times C}$
$u_{\tau_0} \in \mathbb{R}^{H \times W \times C}$
$\alpha^{\text{abs}}_{\tau_0} \in \mathbb{R}^{N \times H \times W}$

Cross Attention: $\text{softmax}(\alpha^{\text{abs}}_{\tau_0}) V u_{\tau_0}$

Absolute Cross Attention Matrix: $\alpha^{\text{abs}}_{\tau_0} = Q(q_e + \xi_Q)[K(u_{\tau_0} + \xi_K)]^\top$

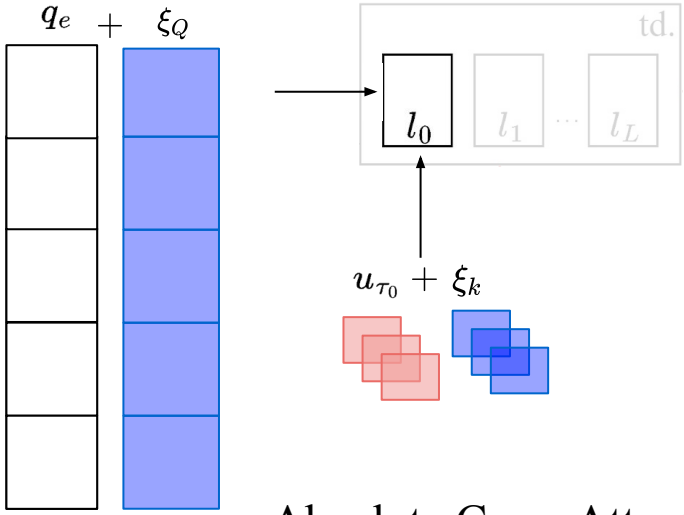
$$= Q q_e u_{\tau_0}^\top K^\top + Q q_e \xi_K^\top K^\top + Q \xi_Q u_{\tau_0}^\top K^\top + Q \xi_Q \xi_K^\top K^\top$$

u_{τ_0}: intermediate features
q_e : learnt embeddings

ξ_Q, ξ_k : absolute position encodings
V, Q, K : learnt weights

: content
: position

Prior Work: Absolute Positional Encoding



$$q_e \in \mathbb{R}^{N \times C}$$
$$u_{\tau_0} \in \mathbb{R}^{H \times W \times C}$$
$$\alpha^{\text{abs}}_{\tau_0} \in \mathbb{R}^{N \times H \times W}$$

Cross Attention: $\text{softmax}(\alpha^{\text{abs}}_{\tau_0}) V u_{\tau_0}$

Absolute Cross Attention Matrix:

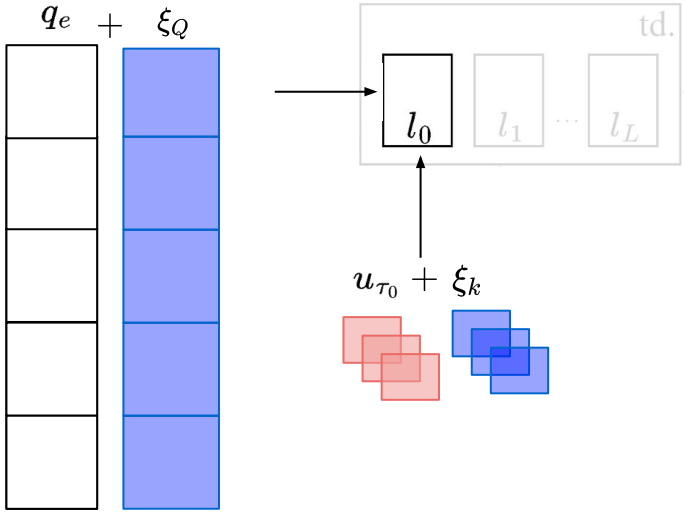
$$\alpha^{\text{abs}}_{\tau_0} = Q(q_e + \xi_Q)[K(u_{\tau_0} + \xi_K)]^\top$$
$$= Q q_e u_{\tau_0}^\top K^\top + Q q_e \xi_K^\top K^\top + Q \xi_Q u_{\tau_0}^\top K^\top + Q \xi_Q \xi_K^\top K^\top$$

u_{τ_0}: intermediate features
q_e : learnt embeddings

ξ_Q, ξ_k : absolute position encodings
V, Q, K : learnt weights

: content
: position

Prior Work: Absolute Positional Encoding



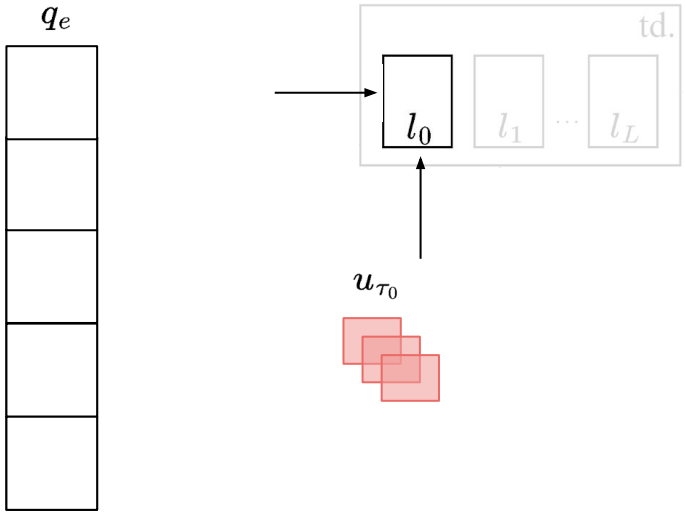
$q_e \in \mathbb{R}^{N \times C}$
$u_{\tau_0} \in \mathbb{R}^{H \times W \times C}$
$\alpha^{\text{abs}}_{\tau_0} \in \mathbb{R}^{N \times H \times W}$

Cross Attention: $\text{softmax}(\alpha^{\text{abs}}_{\tau_0})Vu_{\tau_0}$

u_{τ_0}: intermediate features
q_e : learnt embeddings

ξ_Q, ξ_k : absolute position encodings
V, Q, K : learnt weights

Prior Work: Absolute Positional Encoding



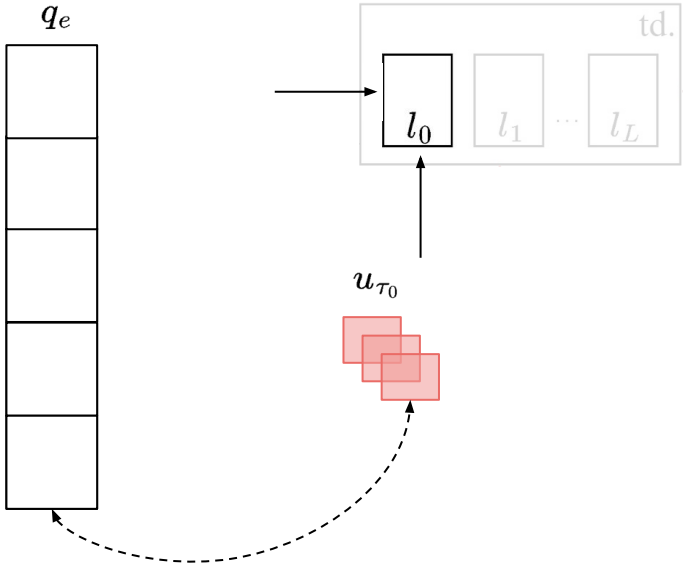
$$q_e \in \mathbb{R}^{N \times C}$$
$$u_{\tau_0} \in \mathbb{R}^{H \times W \times C}$$
$$\alpha^{\text{abs}}_{\tau_0} \in \mathbb{R}^{N \times H \times W}$$

Cross Attention: $\text{softmax}(\alpha^{\text{abs}}_{\tau_0})Vu_{\tau_0}$

u_{τ_0}: intermediate features
q_e : learnt embeddings

ξ_Q, ξ_k : absolute position encodings
V, Q, K : learnt weights

Prior Work: Absolute Positional Encoding



$$q_e \in \mathbb{R}^{N \times C}$$
$$u_{\tau_0} \in \mathbb{R}^{H \times W \times C}$$
$$\alpha^{\text{abs}}_{\tau_0} \in \mathbb{R}^{N \times H \times W}$$

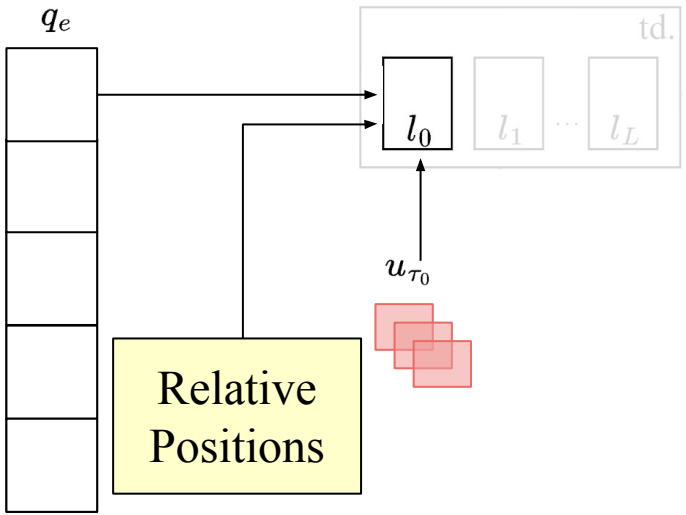
Cross Attention: $\text{softmax}(\alpha^{\text{abs}}_{\tau_0}) V u_{\tau_0}$

Relative positions!

u_{τ_0}: intermediate features
q_e : learnt embeddings

ξ_Q, ξ_k : absolute position encodings
V, Q, K : learnt weights

Our Approach: Relative Positional Encoding

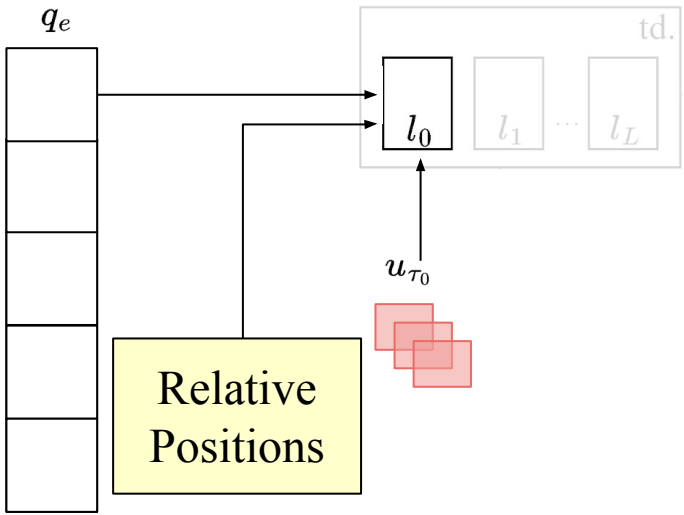


Cross Attention: $\mathrm{softmax}(\alpha_{\tau_0}^{\mathrm{rel}})Vu_{\tau_0}$

Relative Cross Attention Matrix:

$\alpha_{\tau_0}^{\mathrm{rel}}$

Our Approach: Relative Positional Encoding



Cross Attention: $\mathrm{softmax}(\alpha^{\mathrm{rel}}_{\tau_0})Vu_{\tau_0}$

Relative Cross Attention Matrix:

$$\alpha^{\mathrm{rel}}_{\tau_0} = Qq_e{u_{\tau_0}}^{\top}K^{\top} + Qq_e \odot {\xi^{\mathrm{rel}}}^{\top}K^{\top}$$

Similar to Dai et al., ACL 2019

$q_e \in \mathbb{R}^{N\times C}$
$u_{\tau_0} \in \mathbb{R}^{H\times W\times C}$
$\alpha^{\mathrm{rel}}_{\tau_0} \in \mathbb{R}^{N\times H\times W}$
$\xi^{\mathrm{rel}} \in \mathbb{R}^{N\times H\times W\times C}$

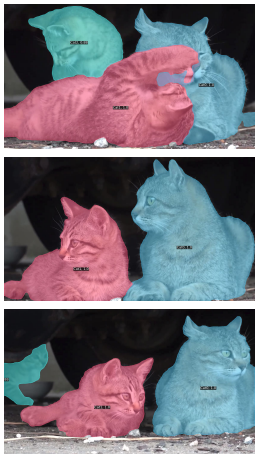
u_{τ_0}: intermediate features
q_e : learnt embeddings
: content
: relative position
$\odot$: dot product along C,
pointwise multiplication along N

Results

Context-Aware Relative Object Queries

Video Instance Segmentation

(Youtube-VIS 2019, Youtube-VIS 2021, OVIS)

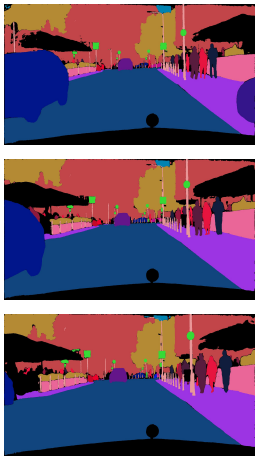


Multi-Object Tracking and Segmentation

(KITTI-MOTS, MOTS-2020)

Video Panoptic Segmentation

(Cityscapes-VPS)



Video Instance Segmentation

Method	Type	Mode	OVIS (AP)
MinVIS [1]	Image-level queries	Onl.	25.0
Mask2Former-VIS [2]	Video-level queries	Off.	17.3*
SeqFormer [3]	Video-level queries	Off.	15.1*
Ours	Video-level queries	Onl.	**25.8**

* : 30 frames at a time

[1] Huang et al., NeurIPS 2022
[2] Cheng et al., arXiv 2022
[3] Wu et al., ECCV 2022

Multi-Object Tracking and Segmentation

	KITTI-MOTS		MOTS 2020	
Method	HOTA (car, ped.)	AssA (car, ped.)	sMOTSA	MOTSA
AS-MOTS [1]	68.7, 58.8	62.2, 57.6	**63.2**	72.9
PointTrack [2]	61.6, 54.4	48.8, 48.0	58.1	70.6
TrackFormer [3]	-	-	58.7	-
Ours	**74.5, 62.2**	**64.0, 58.4**	61.2	**73.2**

[1] Choudhuri et al., ICCV 2021
[2] Xu et al., ECCV 2020
[3] Meinhardt et al., CVPR 2022

Video Panoptic Segmentation

Cityscapes-VPS				
Method	Depth	VPQ	VPQ_{th}	VPQ_{st}
Vip-DeepLab [1]	✓	63.1	49.5	73.0
VPS-Net [2]		57.5	44.8	66.7
Ours		**63.0**	**48.0**	**72.8**

[1] Qiao et al., CVPR 2021
[2] Kim et al., CVPR 2020

Importance of Relative Positional Encoding

KITTI-MOTS						
	Car			Pedestrian		
Method	HOTA	DetA	AssA	HOTA	DetA	AssA
Ours (Rel. pos.)	**83.2**	**84.5**	**85.0**	**64.1**	**64.4**	**63.7**
Ours (Abs. pos.)	70.0	78.6	62.7	52.0	58.0	46.4

Importance of Relative Positional Encoding

KITTI-MOTS						
	Car			Pedestrian		
Method	HOTA	DetA	AssA	HOTA	DetA	AssA
Ours (Rel. pos.)	**83.2**	**84.5**	**85.0**	**64.1**	**64.4**	**63.7**
Ours (Abs. pos.)	70.0	78.6	62.7	52.0	58.0	46.4

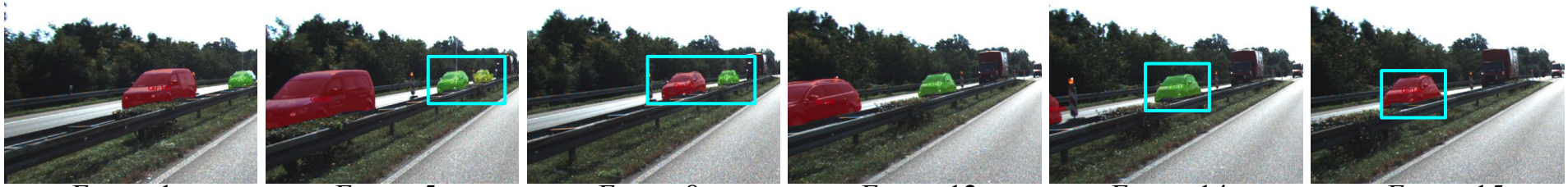
Importance of Relative Positional Encoding

KITTI-MOTS						
	Car			Pedestrian		
Method	HOTA	DetA	AssA	HOTA	DetA	AssA
Ours (Rel. pos.)	**83.2**	**84.5**	**85.0**	**64.1**	**64.4**	**63.7**
Ours (Abs. pos.)	70.0	78.6	62.7	52.0	58.0	46.4

Importance of Relative Positional Encoding

KITTI-MOTS						
	Car			Pedestrian		
Method	HOTA	DetA	AssA	HOTA	DetA	AssA
Ours (Rel. pos.)	**83.2**	**84.5**	**85.0**	**64.1**	**64.4**	**63.7**
Ours (Abs. pos.)	70.0	78.6	62.7	52.0	58.0	46.4

Importance of Relative Positional Encoding

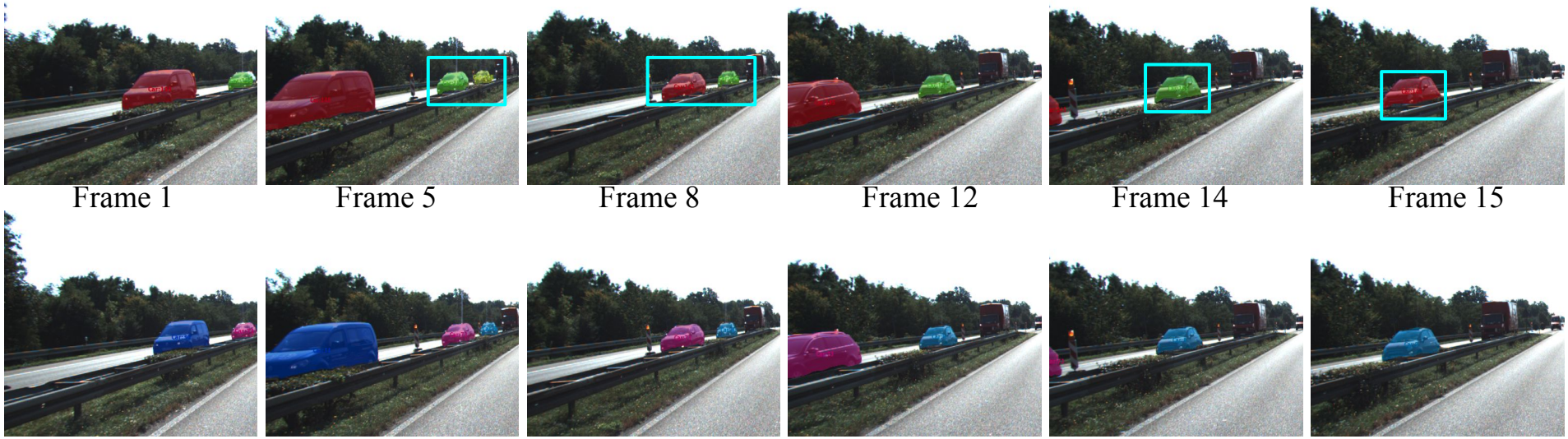


Frame 1 Frame 5 Frame 8 Frame 12 Frame 14 Frame 15

Top Row: Mask2Former-VIS [1]

[1] Cheng et al., arXiv 2022

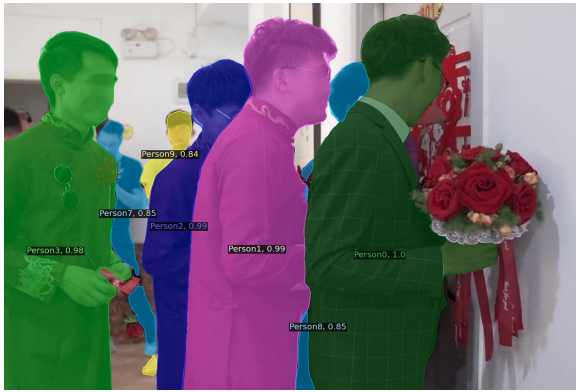
Importance of Relative Positional Encoding



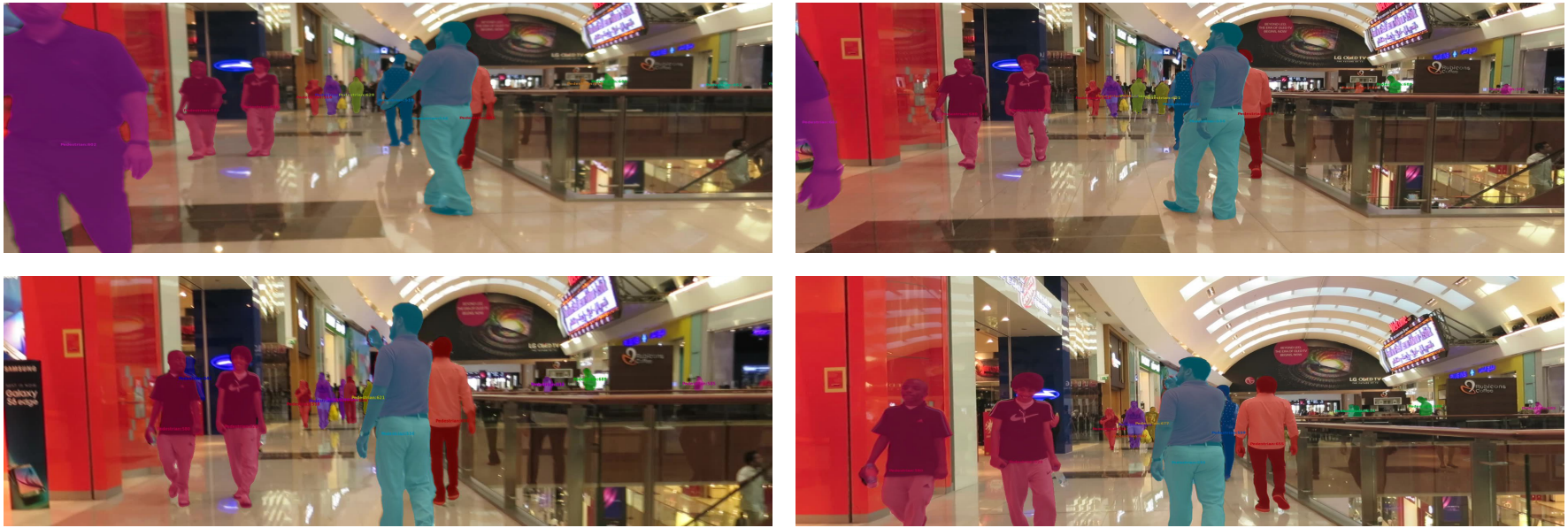
Top Row: Mask2Former-VIS [1]
Bottom Row: Our Approach

[1] Cheng et al., arXiv 2022

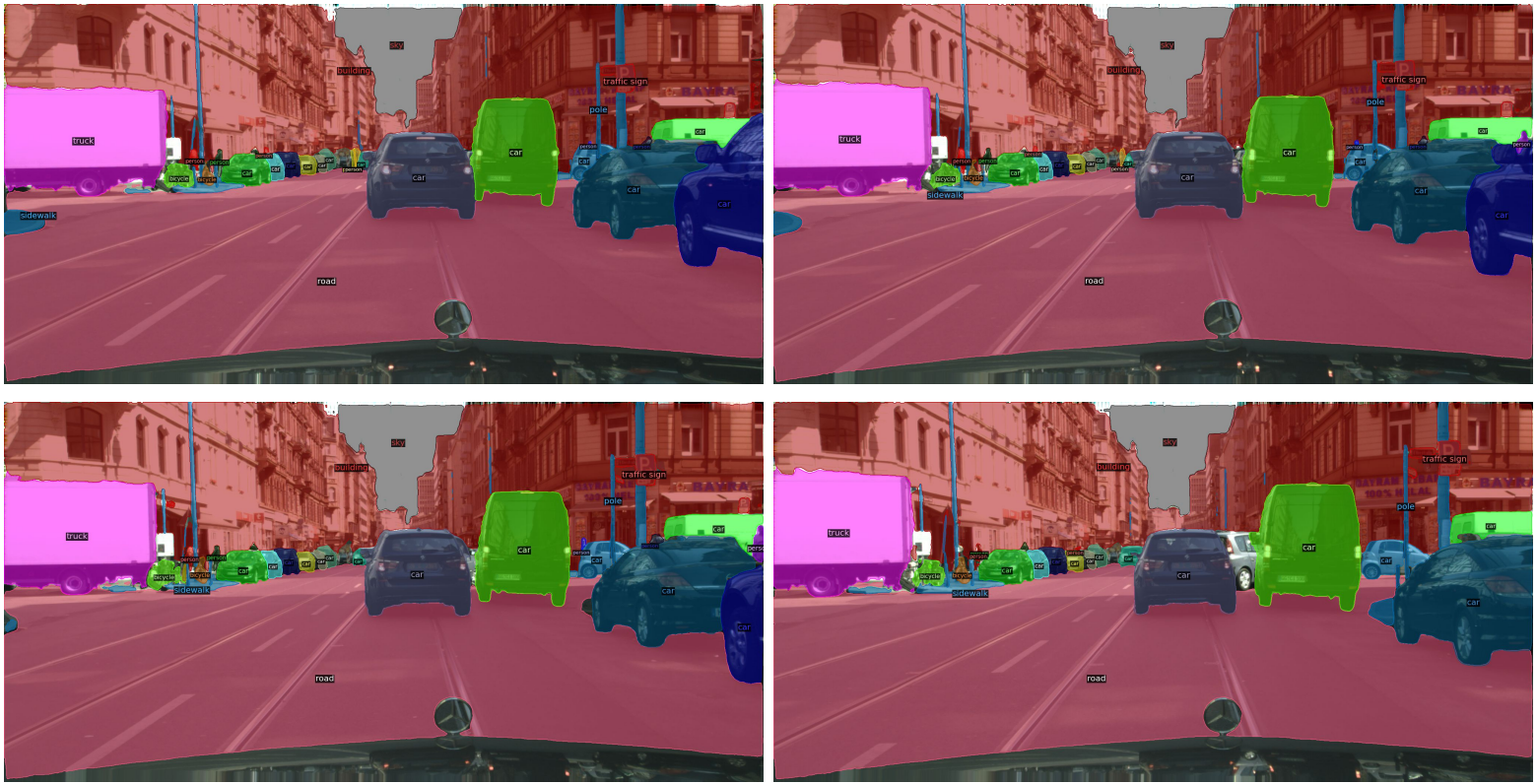
Qualitative Examples

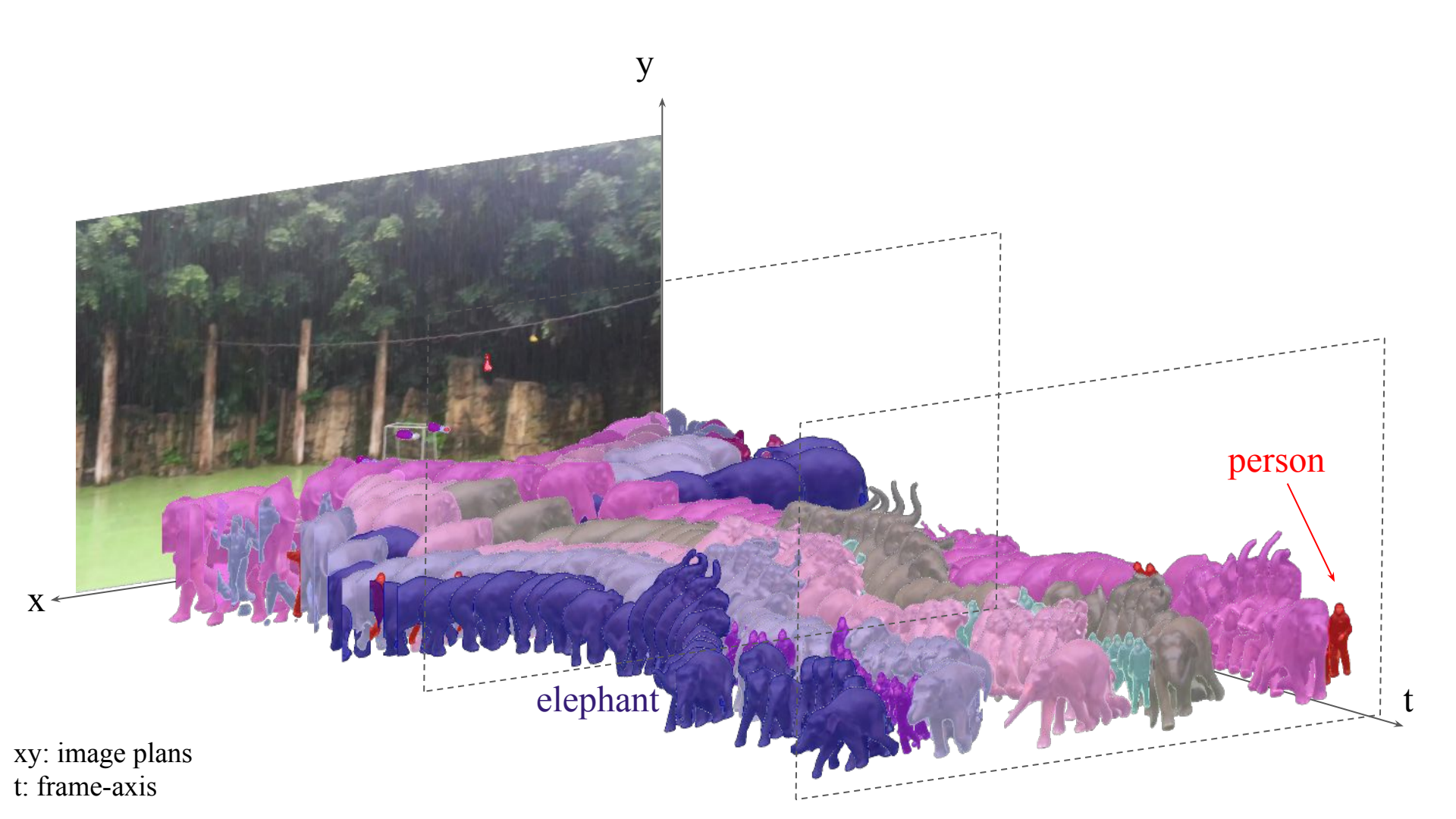


Qualitative Examples

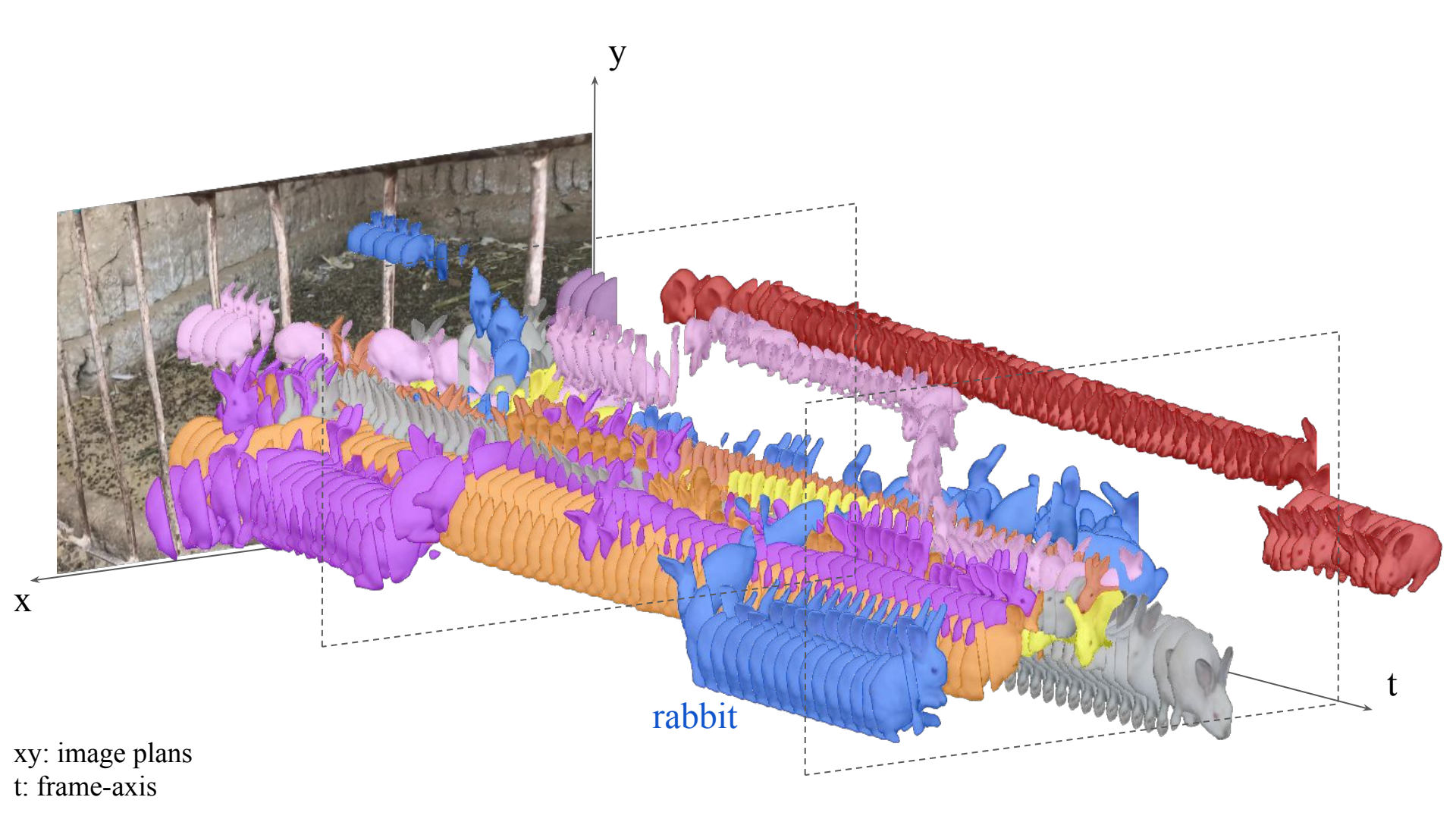


Qualitative Examples





xy: image plans
t: frame-axis



xy: image plans
t: frame-axis

To Summarize

Context-aware relative object queries:

To Summarize

Context-aware relative object queries:

- Continuous refinement and propagation of object queries for **seamless tracking.**

To Summarize

Context-aware relative object queries:

- Continuous refinement and propagation of object queries for **seamless tracking.**
- **Relative positional encoding** to better capture position changes of objects.

To Summarize

Context-aware relative object queries:

- Continuous refinement and propagation of object queries for **seamless tracking.**
- **Relative positional encoding** to better capture position changes of objects.
- Demonstrate the **effectiveness and generalizability** by evaluating on 3 challenging tasks.

Thank You!

Please visit our poster!

Poster ID: TUE-PM-215

Website: https://anwesachoudhuri.github.io/ContextAwareRelativeObjectQueries/