IWDSC 2021

INTERNATIONAL WORKSHOP

ON DISTRIBUTED SMART CAMERAS

11TH OCTOBER 2021

Where Did I See It?

Object Instance Re-Identification with Attention



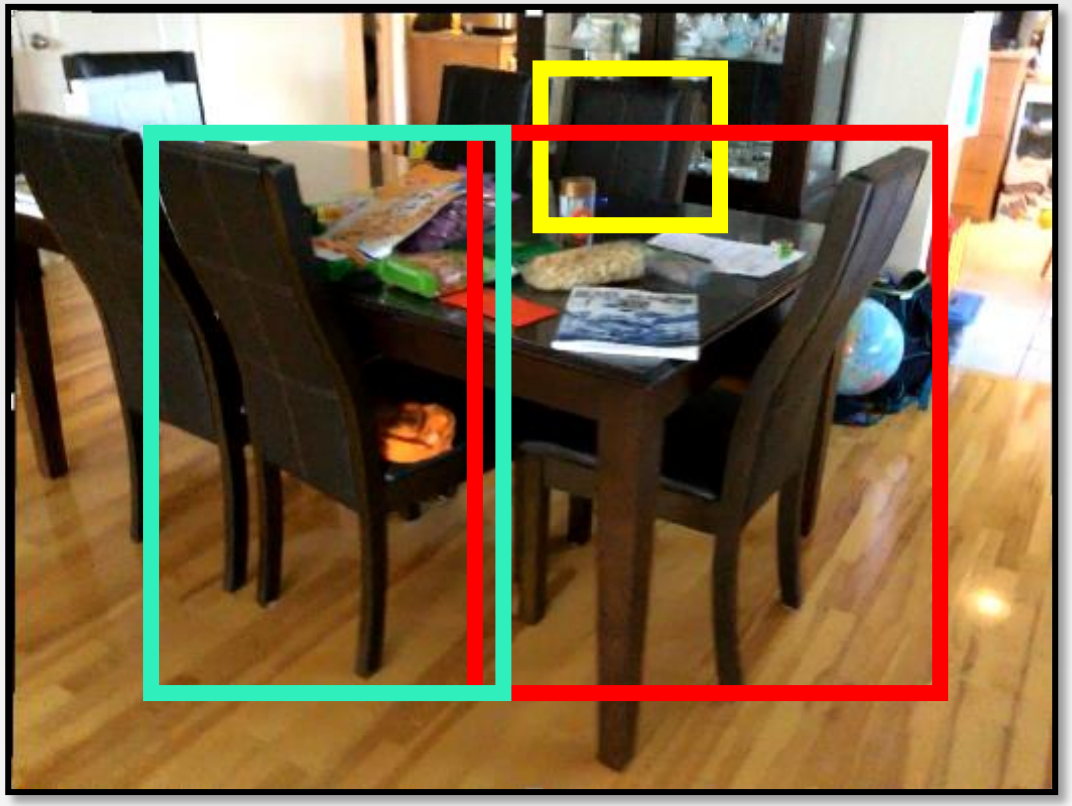
VAIBHAV BANSAL GIAN LUCA FORESTI NIKI MARTINEL

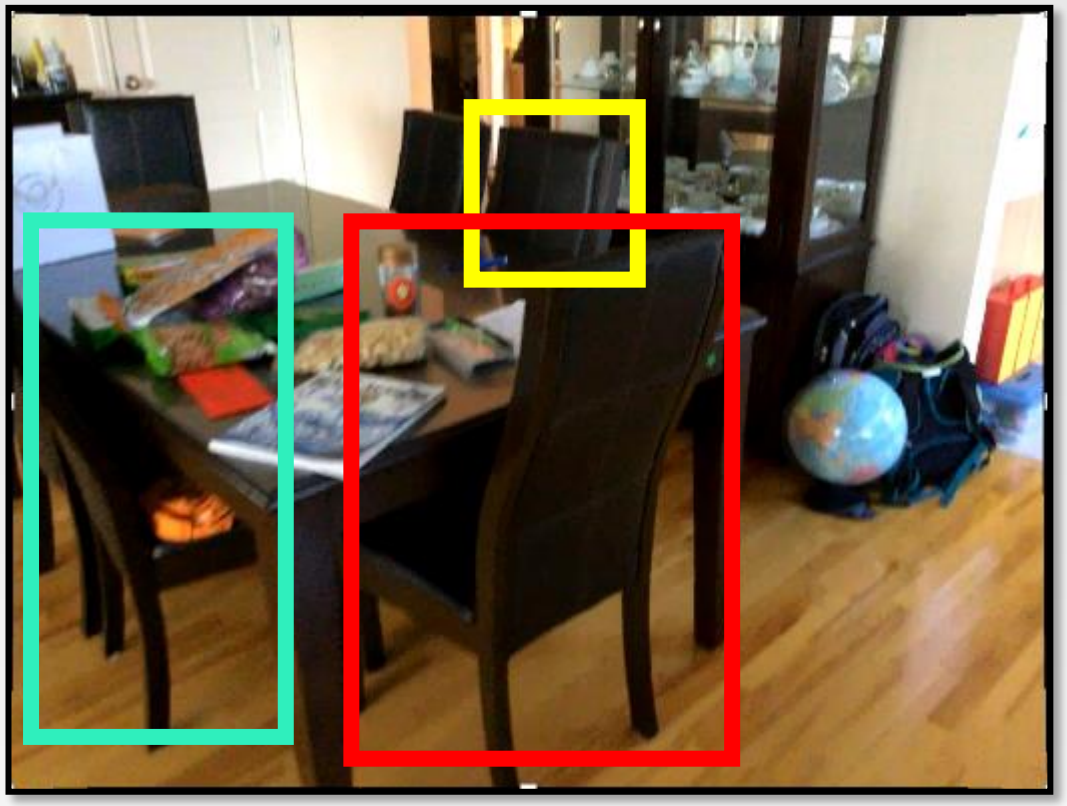
mlp UNIVERSITY OF UDINE

AViReS research laboratory

UNIVERSITÀ DEGLI STUDI DI UDINE

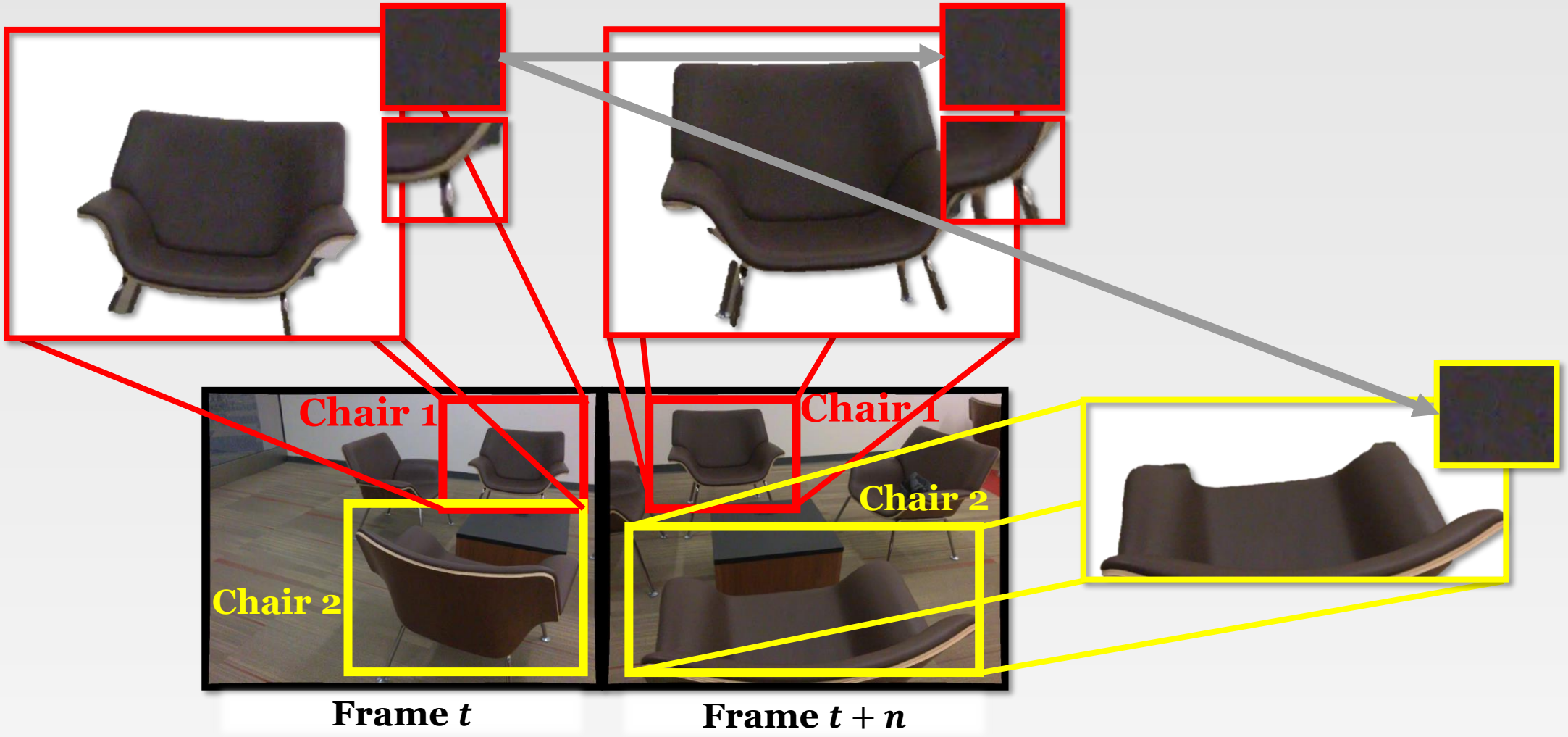
Problem





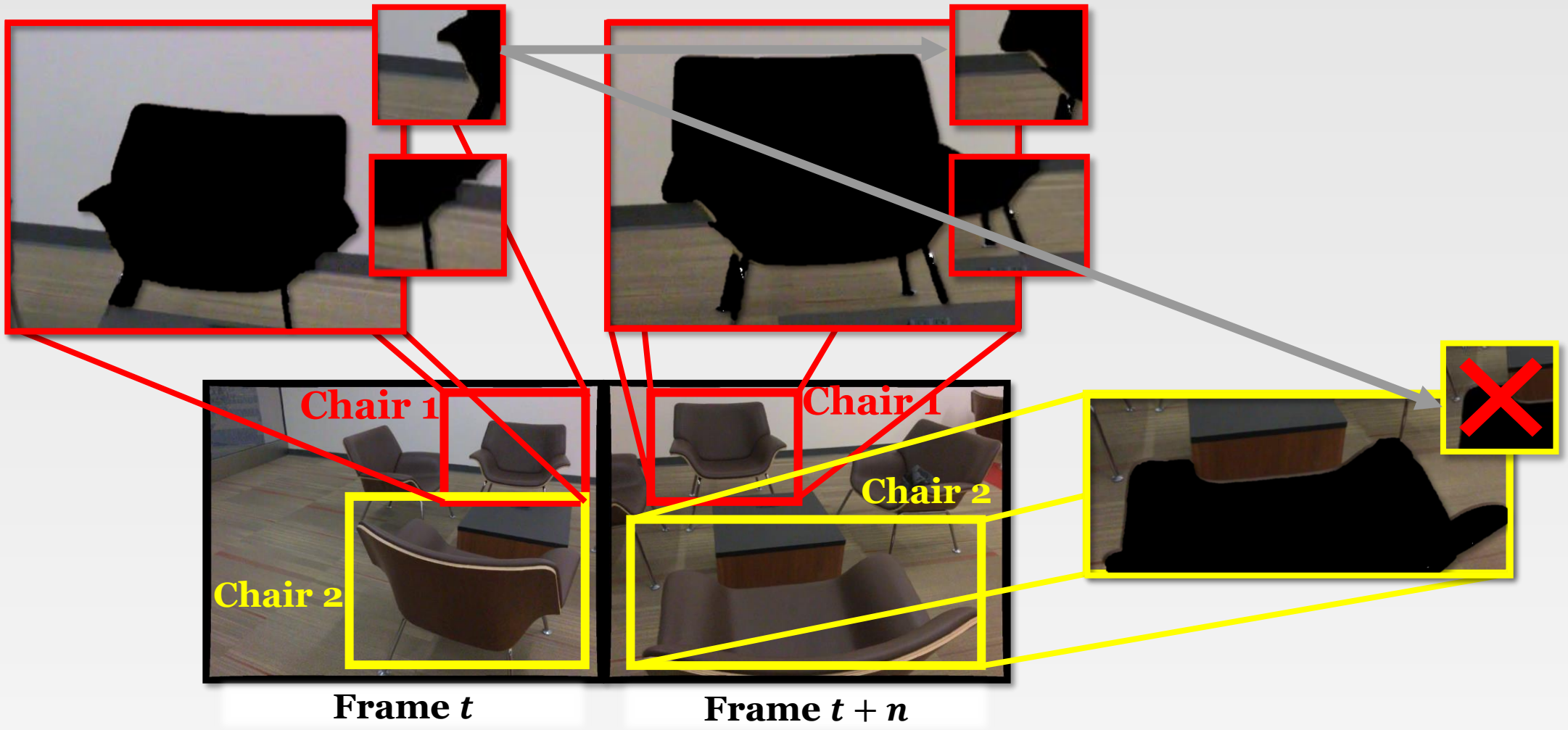
Dai, A., Chang, A.X., Savva, M., Halber, M., Funkhouser, T. and Nießner, M., 2017. Scannet: Richly-annotated 3d reconstructions of indoor scenes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 5828-5839).

Appearance-based: Using Foreground



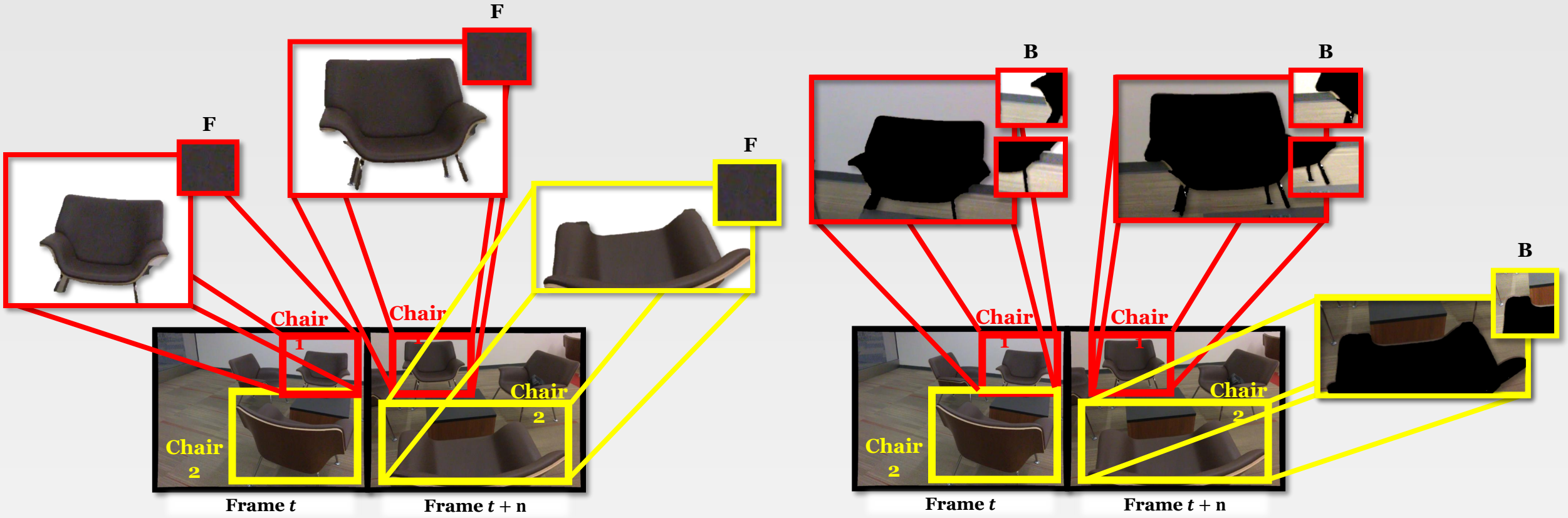
Dai, A., Chang, A.X., Savva, M., Halber, M., Funkhouser, T. and Nießner, M., 2017. Scannet: Richly-annotated 3d reconstructions of indoor scenes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 5828-5839).

Appearance-based: Using Background



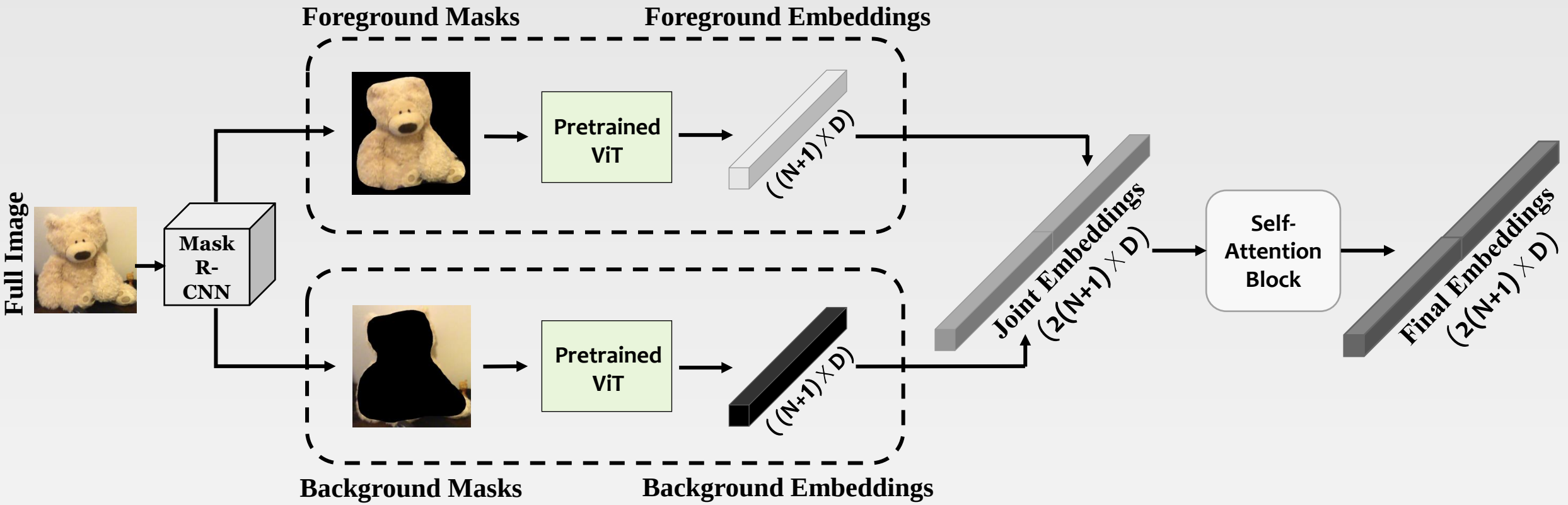
Dai, A., Chang, A.X., Savva, M., Halber, M., Funkhouser, T. and Nießner, M., 2017. Scannet: Richly-annotated 3d reconstructions of indoor scenes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 5828-5839).

Better Approach: Foreground-Background



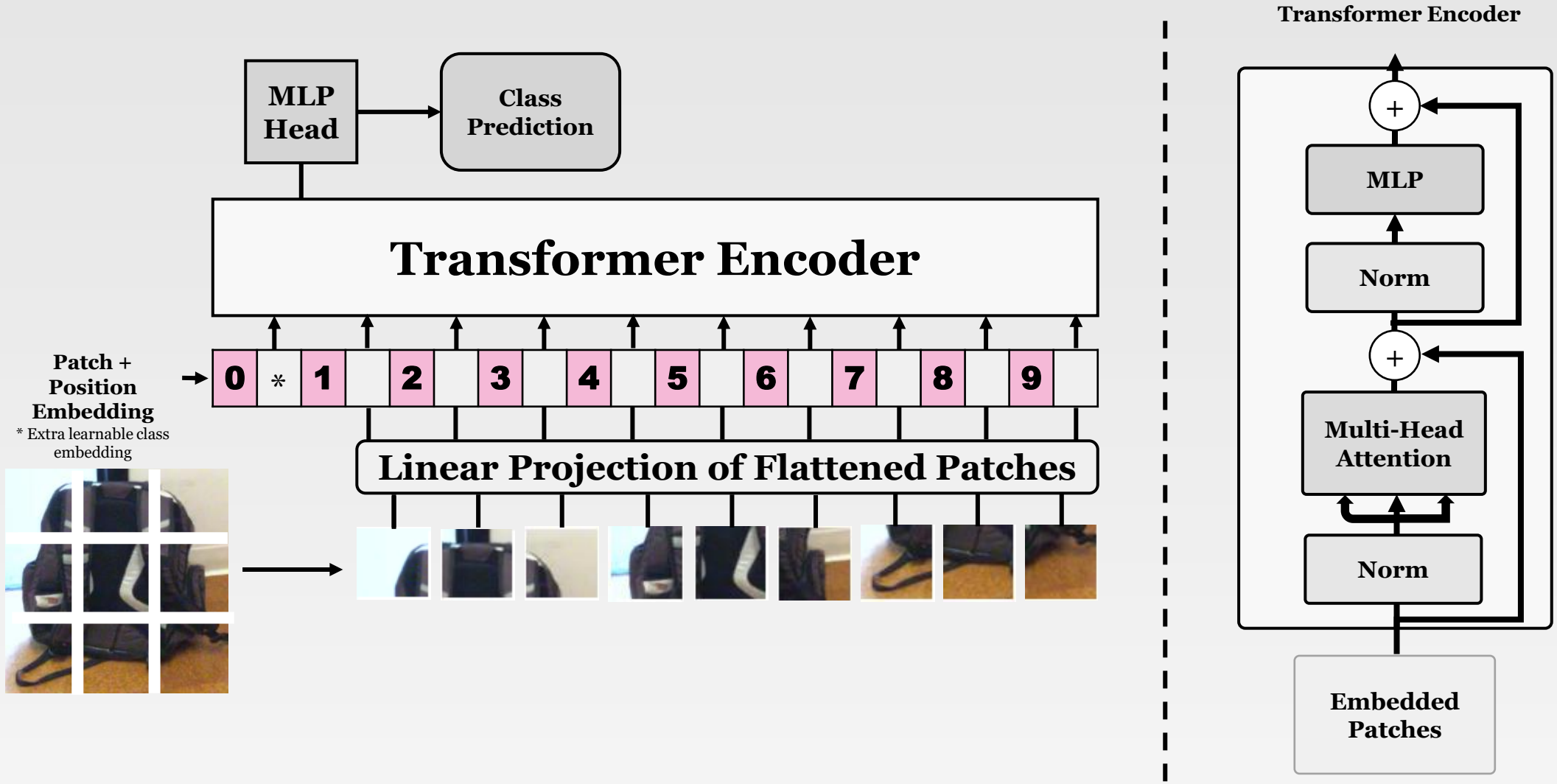
Dai, A., Chang, A.X., Savva, M., Halber, M., Funkhouser, T. and Nießner, M., 2017. Scannet: Richly-annotated 3d reconstructions of indoor scenes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 5828-5839).

Approach – Stage 1



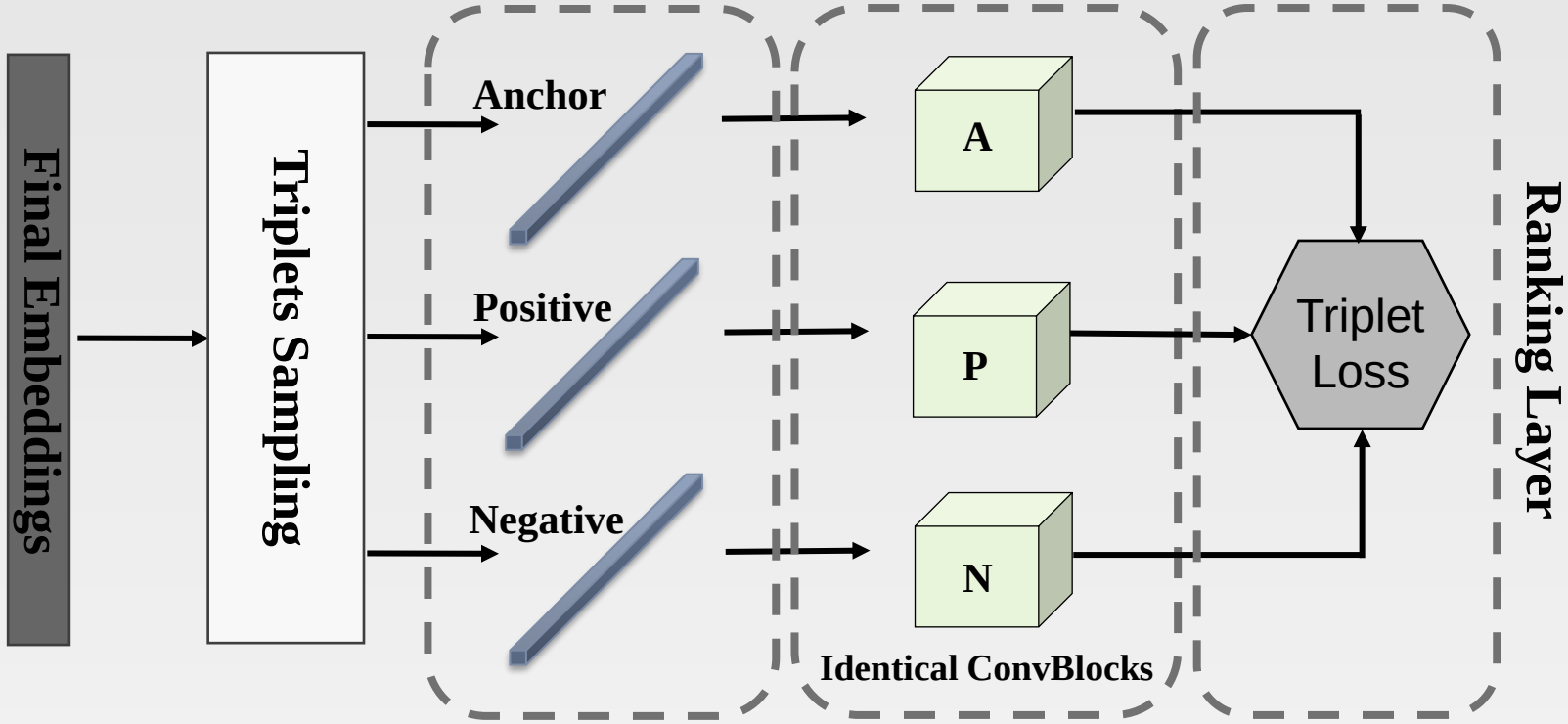
Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929, 2020.

Approach – **Vision Transformer (ViT)**

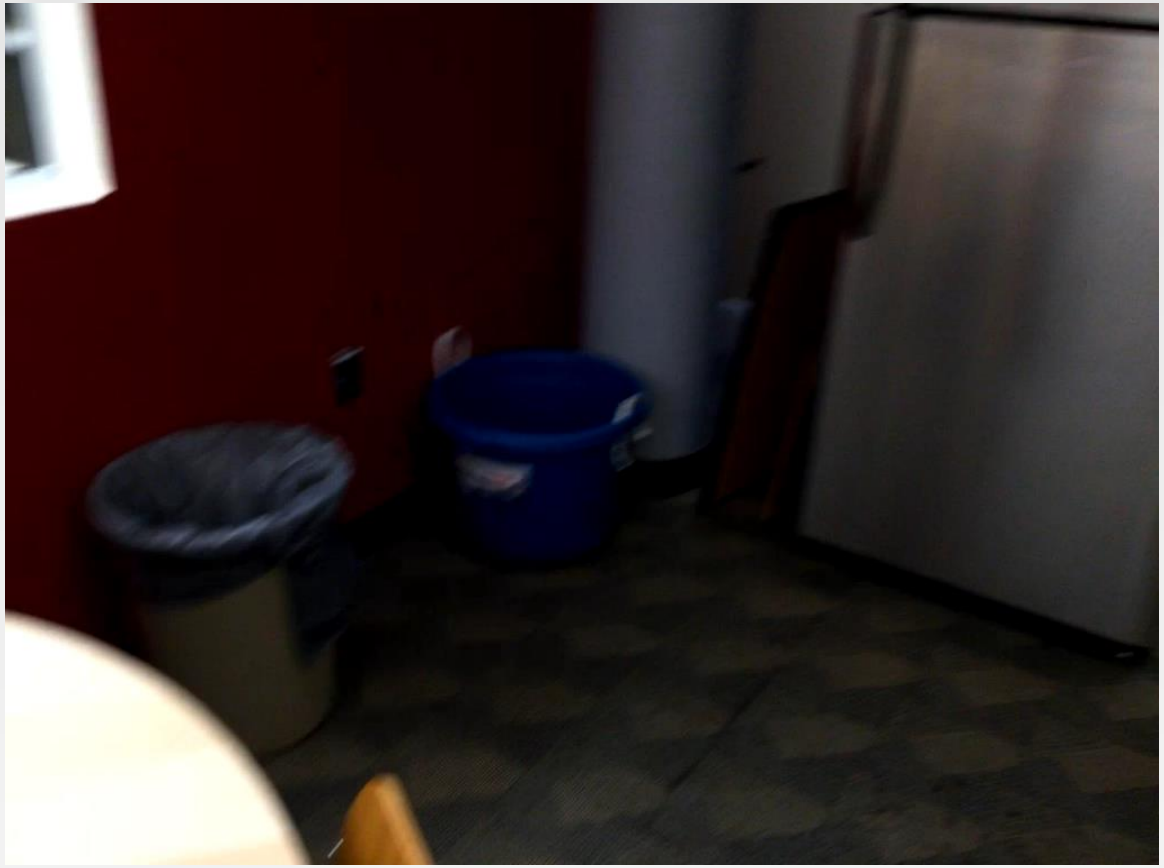


Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929, 2020.

Approach – Stage 2

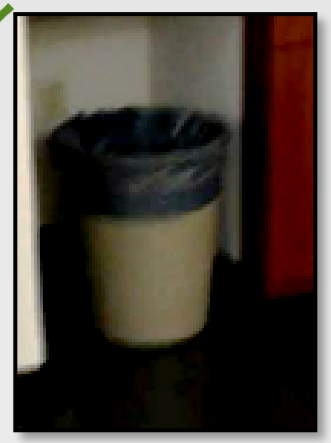


Triplets



Anchor or Probe

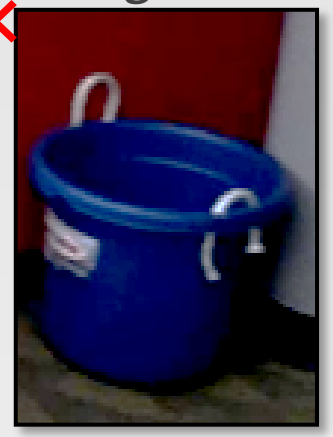
Positive



Different view

Different view

Negative



Different object

Different instance

Experimental setup

Type	Pre-trained ViT	Input images		Fine tuning	Feature concatenation
		Full	Mask		
no_train	✓	✓	✗	✗	✗
full	✓	✓	✗	✓	✗
concat	✓	✗	✓	✓	✓

Training data

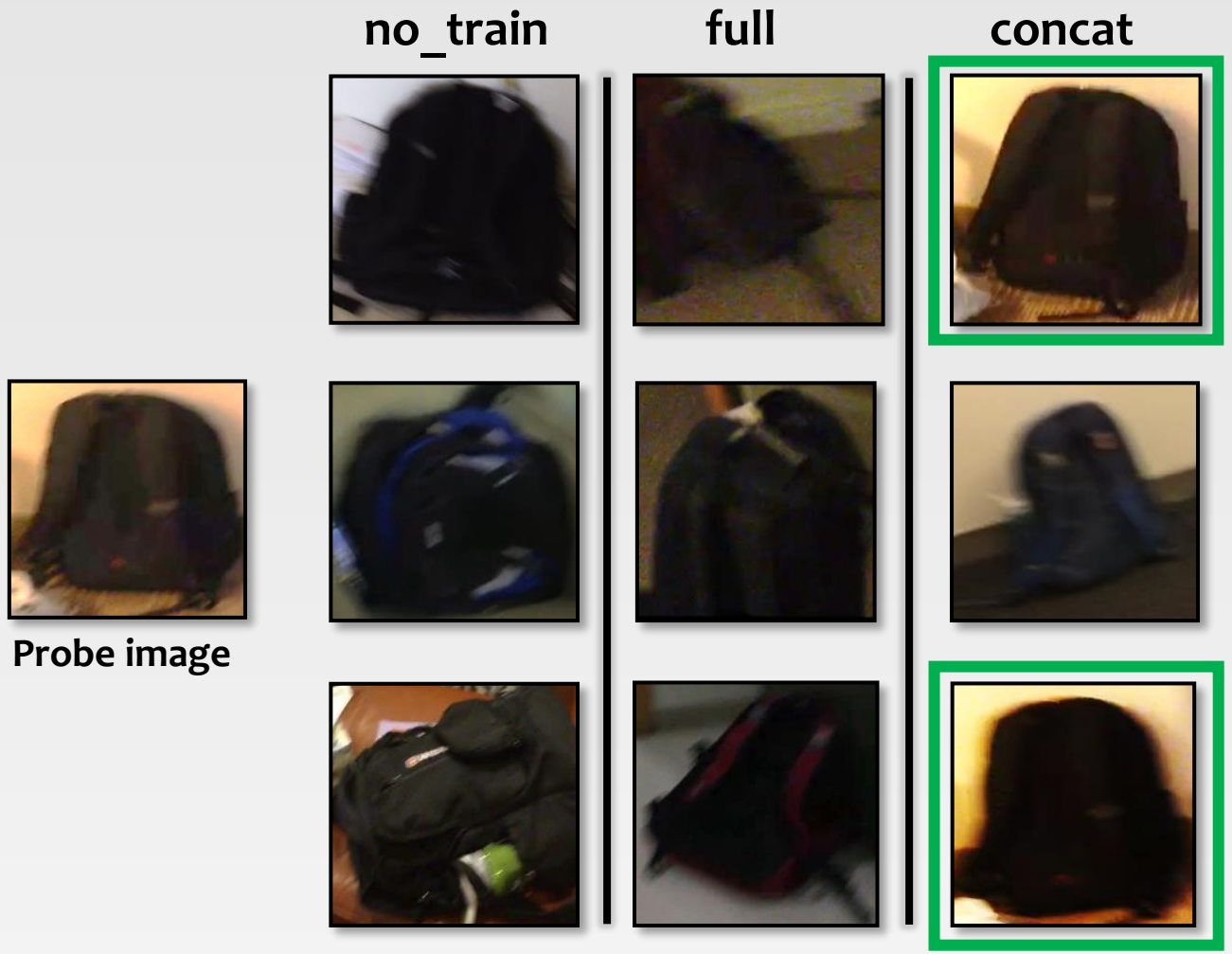
- ScanNet: Over 800 scans of rigid indoor scenes
- Total number of objects detected by Mask-RCNN: 646,156
- Number of object classes: 29
- Number of *Valid* detections : 58876 or 9.11% of the total
- Criteria for *Valid* detections: Bounding box overlap ratio > 60% and matching labels

Dai, A., Chang, A.X., Savva, M., Halber, M., Funkhouser, T. and Nießner, M., 2017. Scannet: Richly-annotated 3d reconstructions of indoor scenes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 5828-5839).

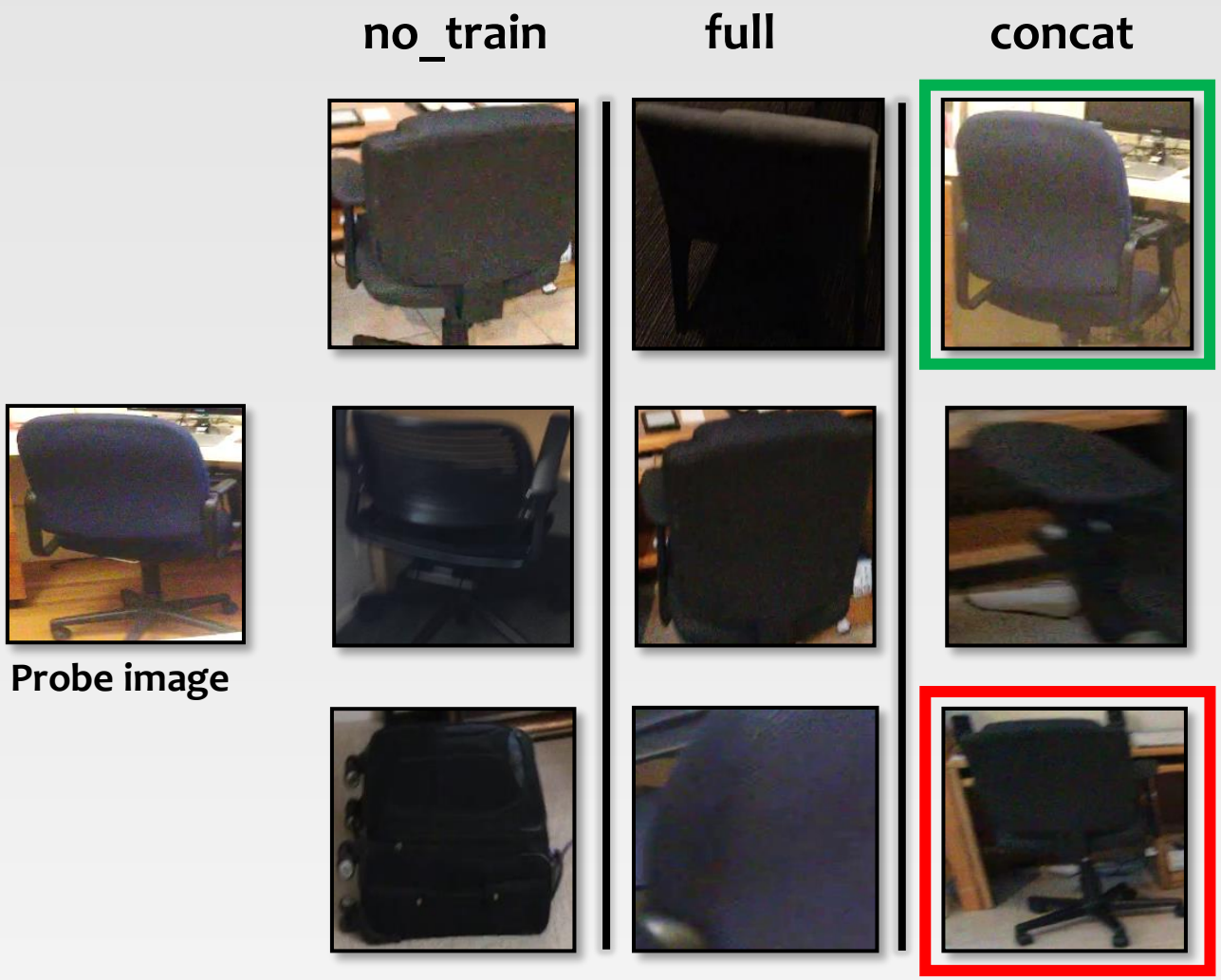
Results - Matching Accuracy

Type	Rank-1 (%)	Rank-5 (%)	Rank-20 (%)	Rank-50 (%)
no_train	68.7	77.06	81.78	92.71
full	75.79	88.0	91.22	96.25
concat	**83.89**	**94.61**	**99.42**	**100**

Visualization



Visualization



Results – Comparison with SOTA

Type	Rank-1 (%)	Rank-5 (%)	Rank-20 (%)
deepsort	49.60	-	-
re-OBJ	77.85	91.55	98.36
ours	**83.89**	**94.61**	**99.42**

- Wojke, N., Bewley, A. and Paulus, D., 2017, September. Simple online and realtime tracking with a deep association metric. In *2017 IEEE International Conference on Image Processing (ICIP)* (pp. 3645-3649). IEEE.
- Bansal, V., James, S. and Del Bue, A., 2019, September. re-OBJ: Jointly learning the foreground and background for object instance re-identification. In International Conference on Image Analysis and Processing (pp. 402-413). Springer, Cham.

Results – Comparison with Person ReID methods

Method	Rank-1 (%)	Rank-5 (%)	Rank-20 (%)
OSNet	69	85.7	91.3
DGNet	58.3	76	92.4
Ours	**83.89**	**94.61**	**99.42**

- Kaiyang Zhou, Yongxin Yang, Andrea Cavallaro, and Tao Xiang. Omni-scale feature learning for person reidentification. CoRR, abs/1905.00953, 2019.
- Zhedong Zheng, Xiaodong Yang, Zhiding Yu, Liang Zheng, Yi Yang, and Jan Kautz. Joint discriminative and generative learning for person re-identification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 2138–2147, 2019.

Thank You