Inverse Neural Rendering for Explainable Multi-Object Tracking

Julian Ost*, Tanushree Banerjee*, Mario Bijelic, Felix Heide

PRINCETON
COMPUTATIONAL IMAGING LAB

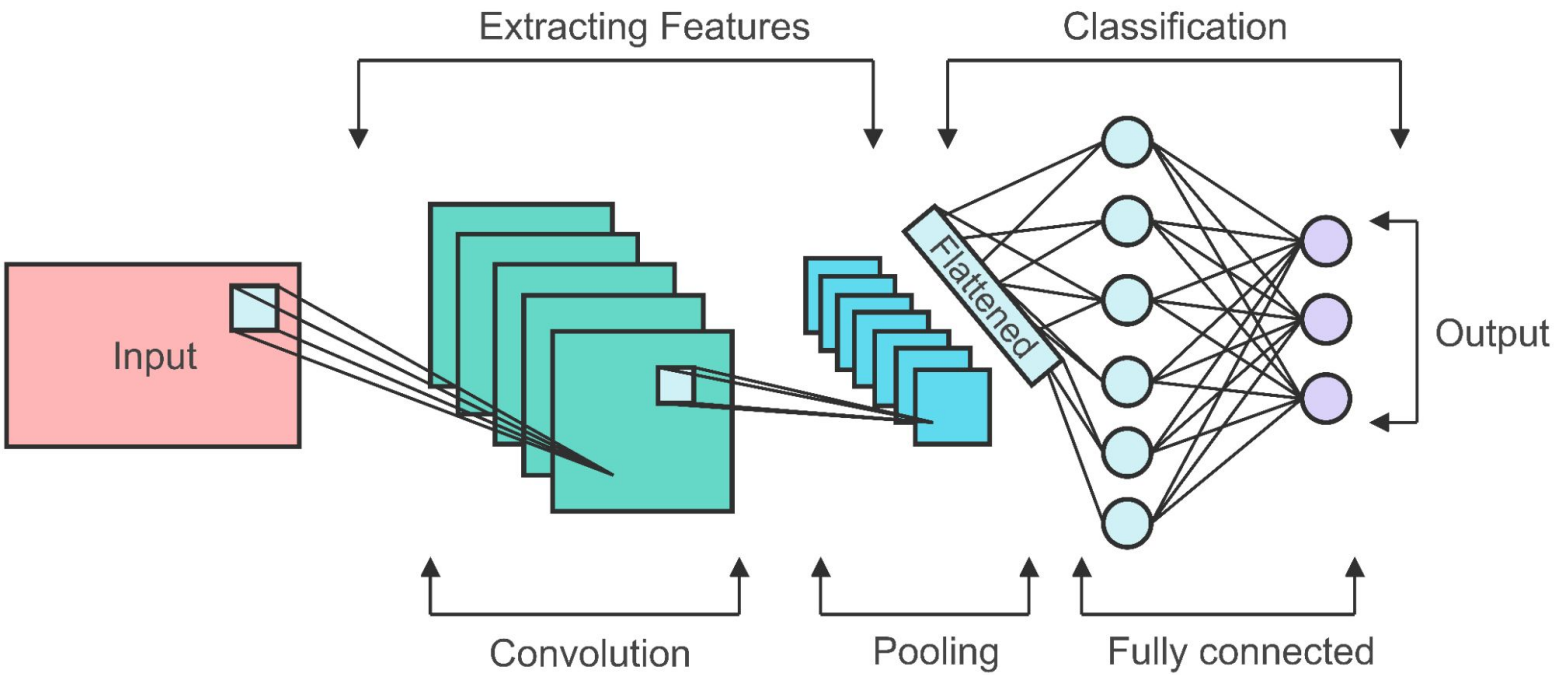
* denotes equal contribution

Agenda

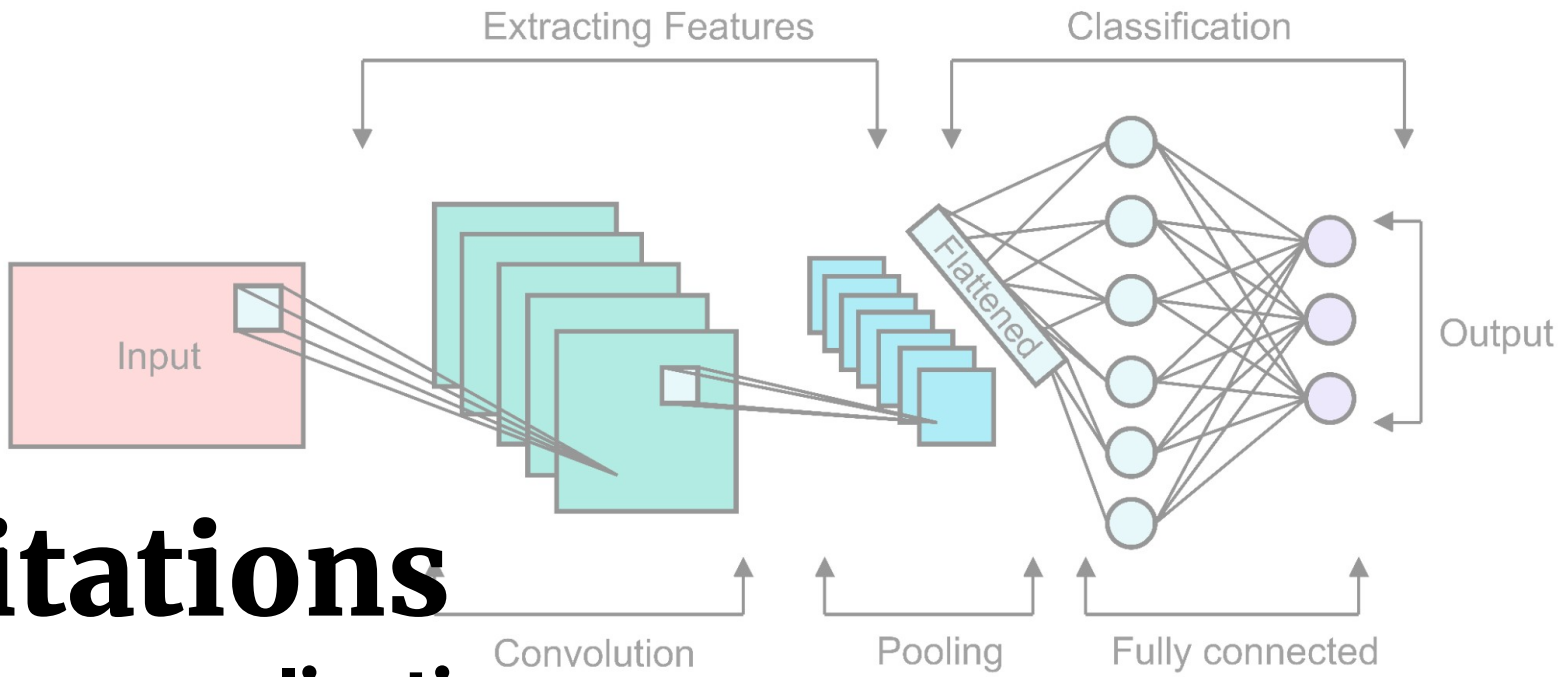
- ➔ Prior Work and Motivation
- ➔ Our Method
- ➔ Generalization
- ➔ Failure Cases and Interpretability
- ➔ 3D Interpretation
- ➔ Limitations
- ➔ Key Takeaway

Agenda

- → **Prior Work and Motivation**
- → Our Method
- → Generalization
- → Failure Cases and Interpretability
- → 3D Interpretation
- → Limitations
- → Key Takeaway



https://www.edrawmax.com/templates/1011715/



Limitations

1. **Poor generalization**
2. **Not interpretable**
3. **Hard to enforce 3D constraints**

https://www.edrawmax.com/templates/1011715/

Caesar, H., et al. (2020). nuScenes: A multimodal dataset for autonomous driving.

Monocular Tracking

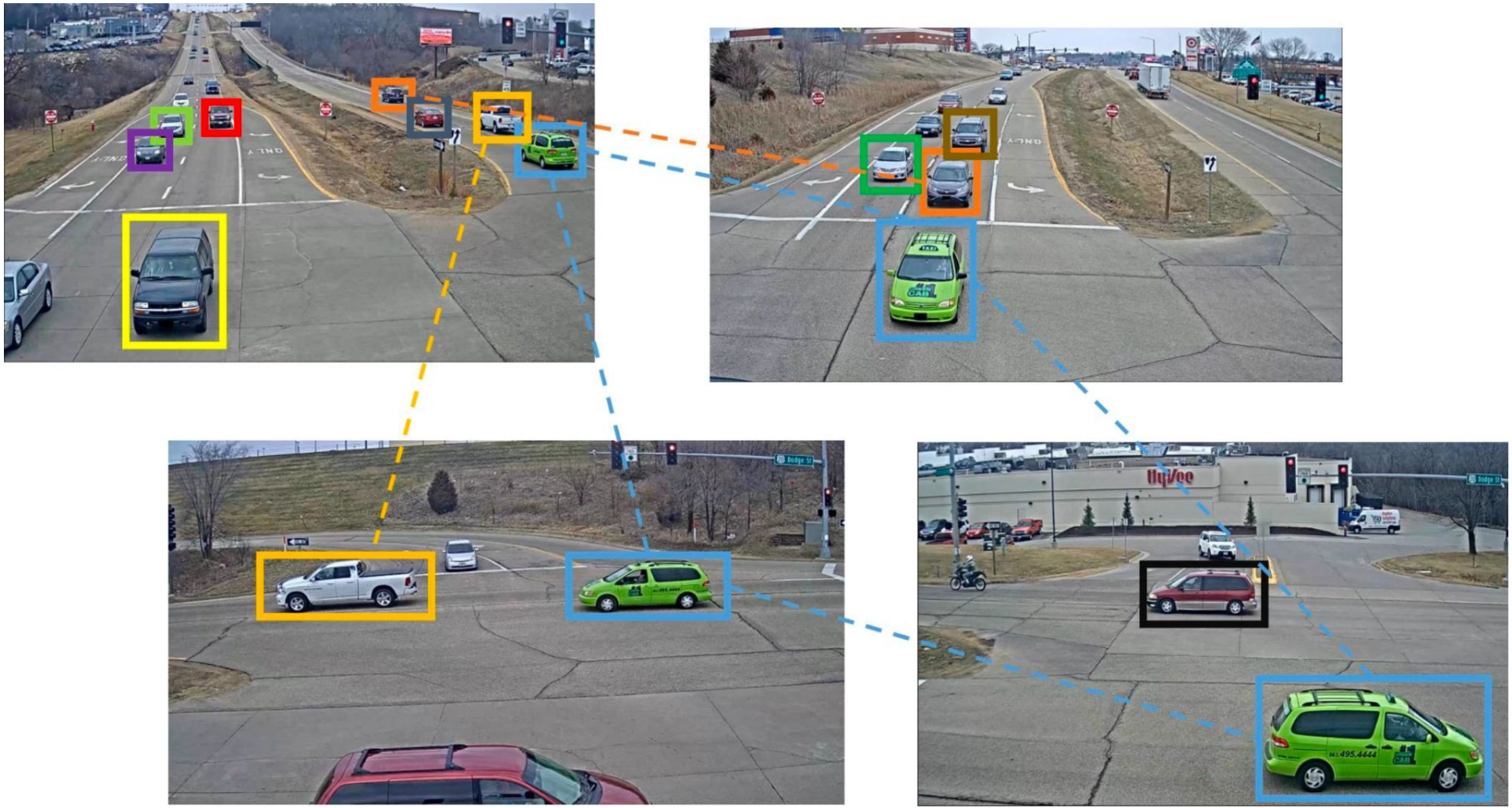
1. **Object Detection**
2. **Motion Prediction**
3. **Object Association**

Monocular Tracking

1. Object Detection

2. Motion Prediction

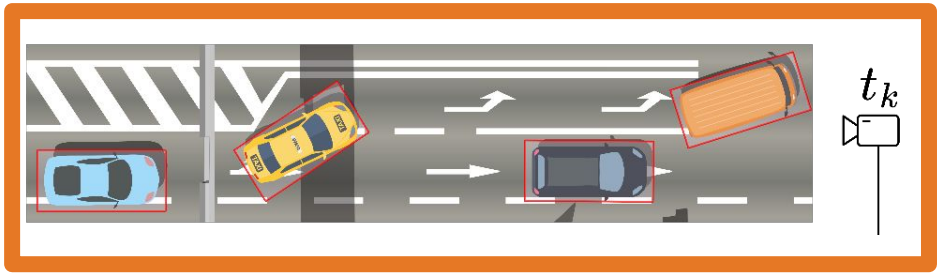
3. **Object Association**

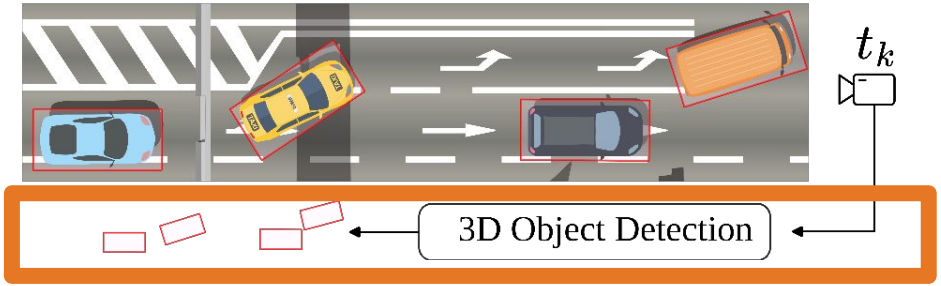


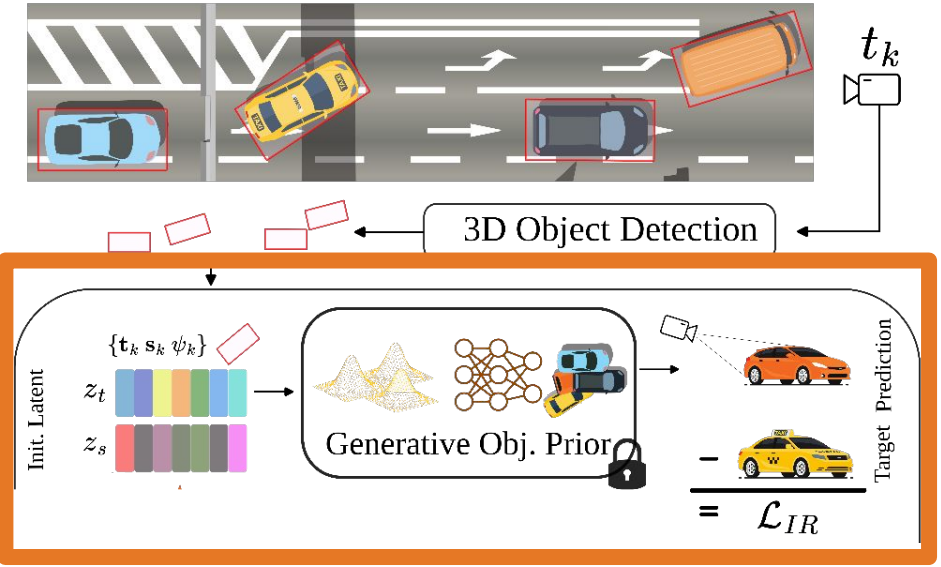
Li, Yun-Lun, et al. (2022). Multi-Camera Vehicle Tracking Based on Deep Tracklet Similarity Network.

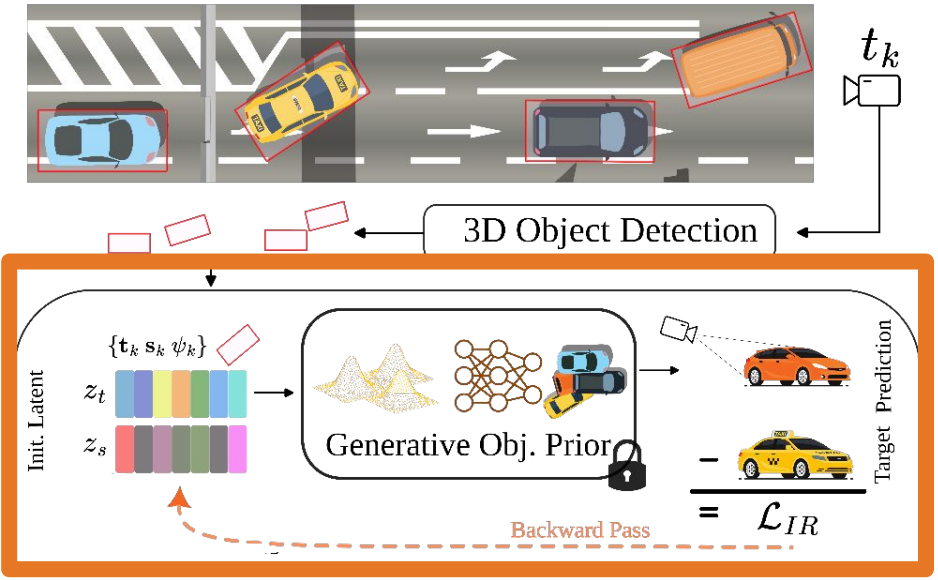
Agenda

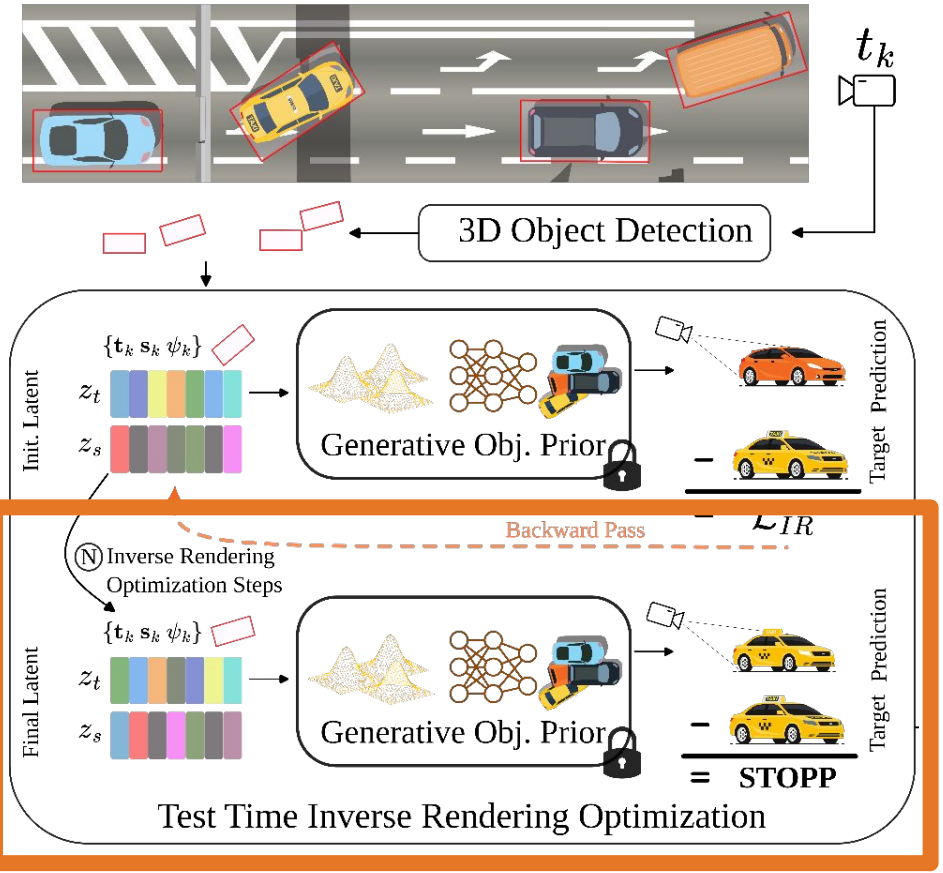
- → Prior Work and Motivation
- → **Our Method**
- → Generalization
- → Failure Cases and Interpretability
- → 3D Interpretation
- → Limitations
- → Key Takeaway

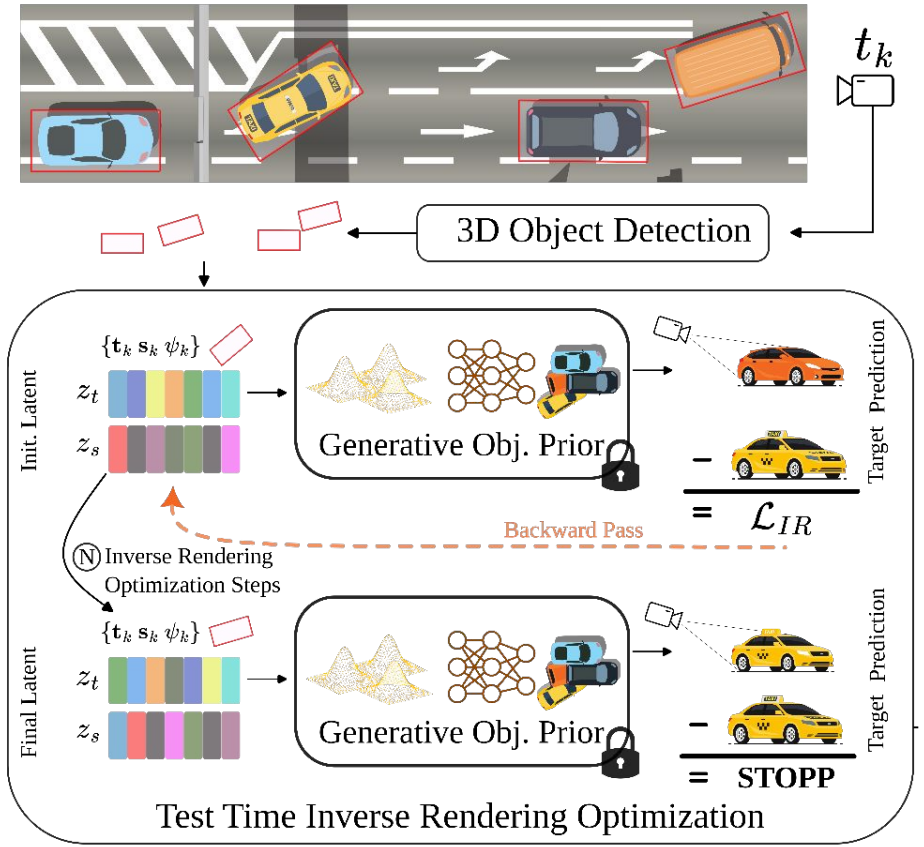

t_k

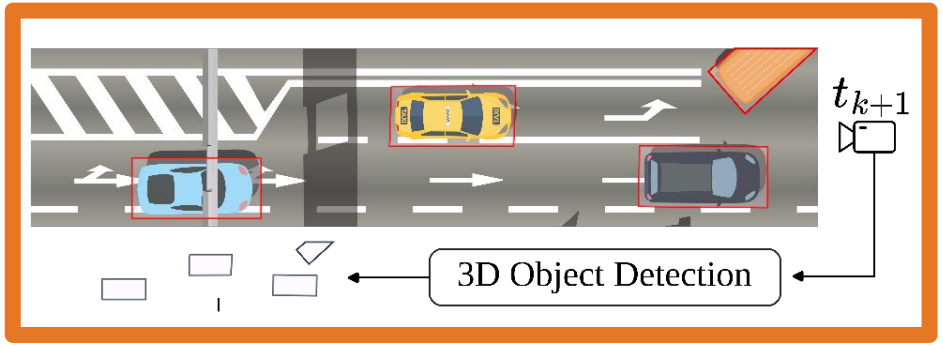

t_k
3D Object Detection

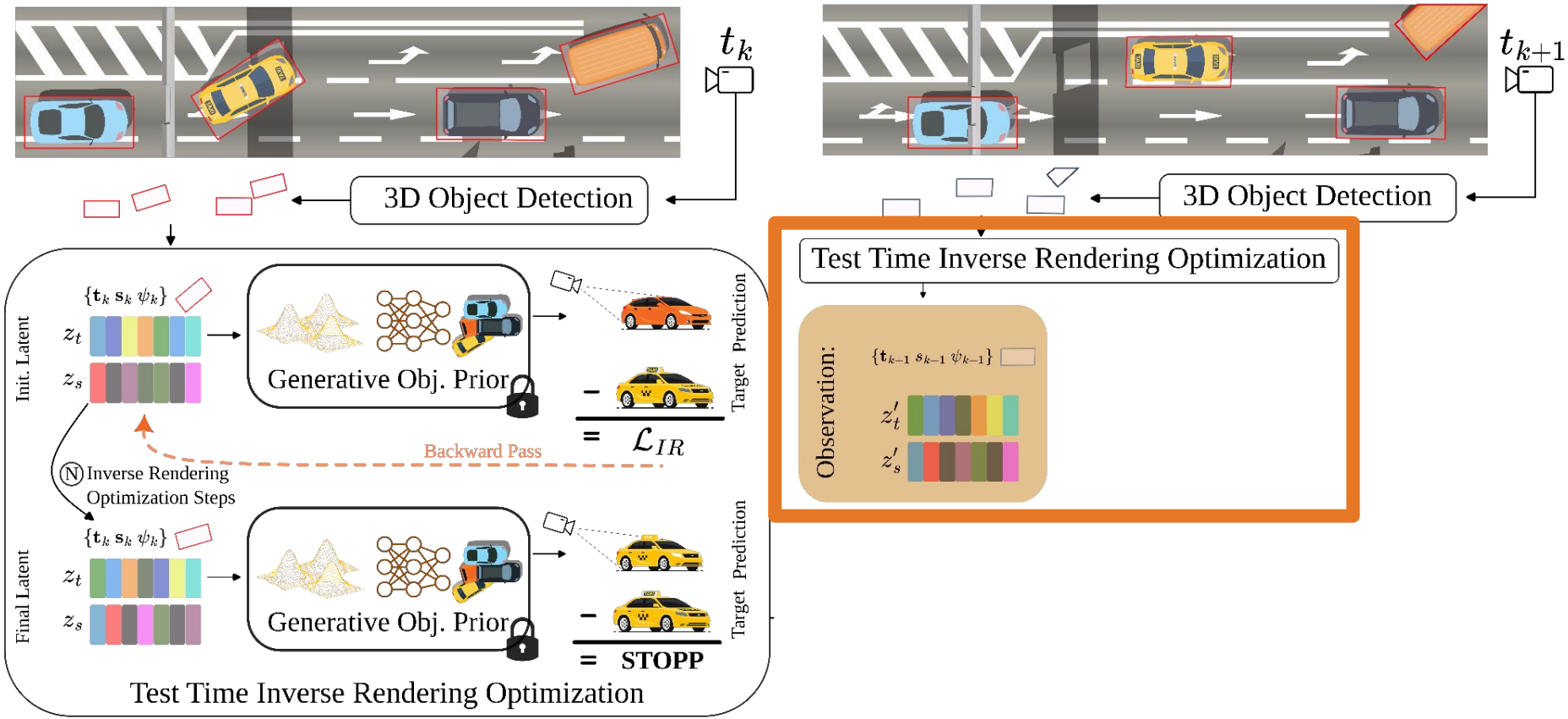

t_k
3D Object Detection
$\{\mathbf{t}_k\ \mathbf{s}_k\ \psi_k\}$
z_t
z_s
Init. Latent
Generative Obj. Prior
Target Prediction
$\mathcal{L}_{IR}$

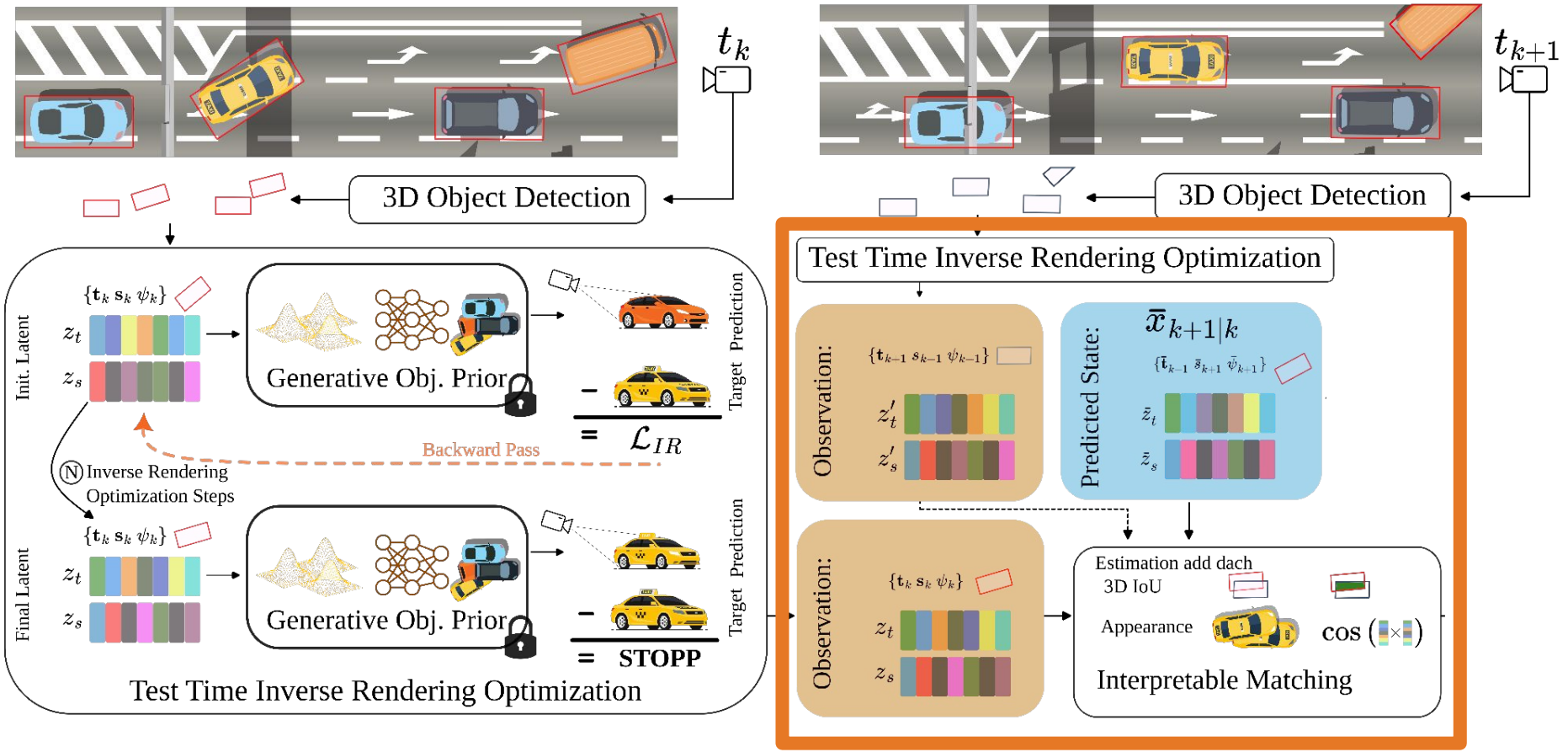

t_k
3D Object Detection
$\{\mathbf{t}_k\ \mathbf{s}_k\ \psi_k\}$
z_t
z_s
Init. Latent
Generative Obj. Prior
Prediction
Target
$\mathcal{L}_{IR}$
Backward Pass

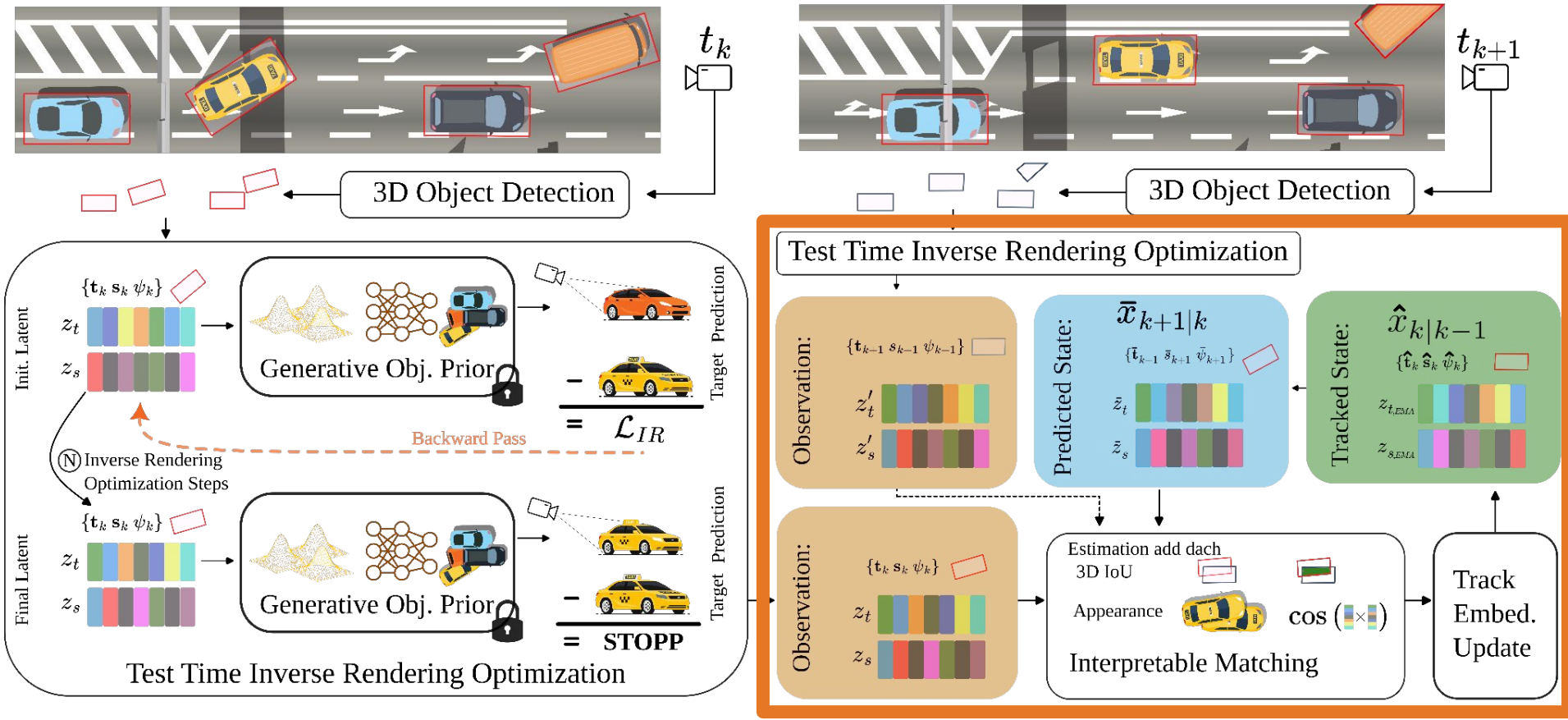

t_k
3D Object Detection
Init. Latent
$\{\mathbf{t}_k\ \mathbf{s}_k\ \psi_k\}$
z_t
z_s
Generative Obj. Prior
Target Prediction
$\mathcal{L}_{IR}$
Backward Pass
N Inverse Rendering Optimization Steps
Final Latent
$\{\mathbf{t}_k\ \mathbf{s}_k\ \psi_k\}$
z_t
z_s
Generative Obj. Prior
Target Prediction
= STOPP
Test Time Inverse Rendering Optimization


t_k
3D Object Detection
$\{\mathbf{t}_k\ \mathbf{s}_k\ \psi_k\}$
z_t
z_s
Init. Latent
Generative Obj. Prior
Prediction
Target
$\mathcal{L}_{IR}$
Backward Pass
N Inverse Rendering Optimization Steps
Final Latent
STOPP
Test Time Inverse Rendering Optimization

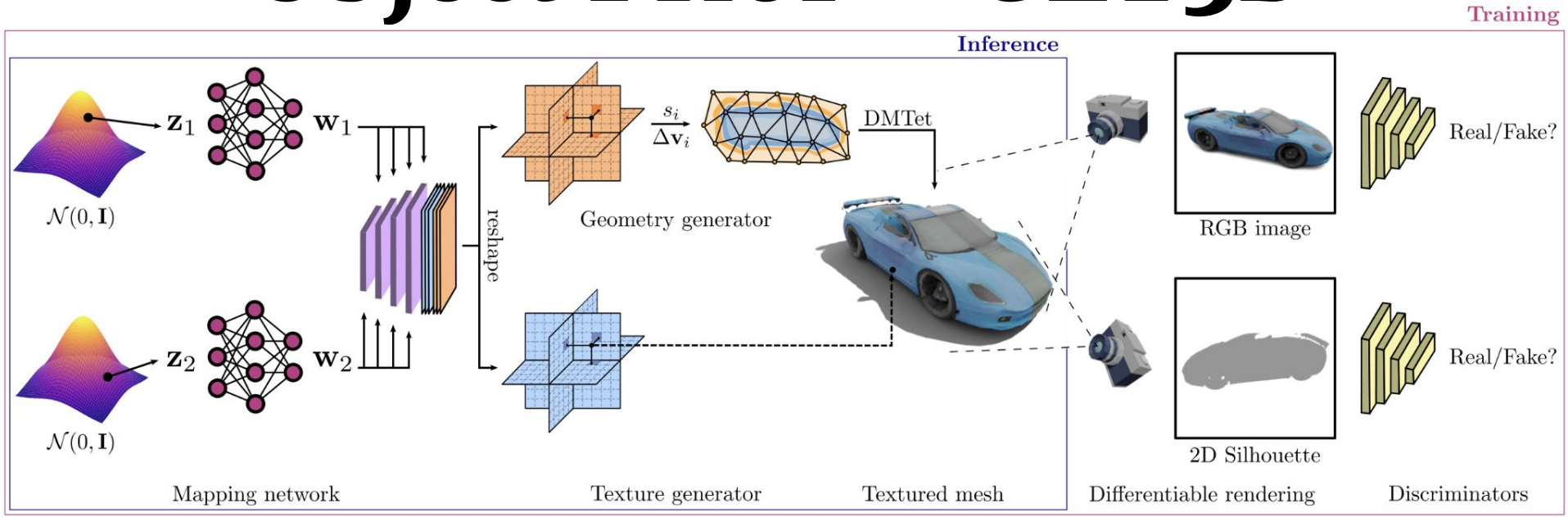

t_{k+1}
3D Object Detection


t_k
t_{k+1}
3D Object Detection
3D Object Detection
Test Time Inverse Rendering Optimization
Observation:
$\{\mathbf{t}_{k-1}\ s_{k-1}\ \psi_{k-1}\}$
z'_t
z'_s
Init. Latent
$\{\mathbf{t}_k\ \mathbf{s}_k\ \psi_k\}$
z_t
z_s
Generative Obj. Prior
Target Prediction
$\mathcal{L}_{IR}$
Backward Pass
N Inverse Rendering Optimization Steps
Final Latent
$\{\mathbf{t}_k\ \mathbf{s}_k\ \psi_k\}$
z_t
z_s
Generative Obj. Prior
Target Prediction
STOPP
Test Time Inverse Rendering Optimization


t_k
t_{k+1}
3D Object Detection
3D Object Detection
Test Time Inverse Rendering Optimization
Init. Latent
Final Latent
$\{\mathbf{t}_k\ \mathbf{s}_k\ \psi_k\}$
z_t
z_s
Generative Obj. Prior
Target Prediction
$\mathcal{L}_{IR}$
Backward Pass
N Inverse Rendering Optimization Steps
STOPP
Test Time Inverse Rendering Optimization
Observation:
$\{\mathbf{t}_{k-1}\ s_{k-1}\ \psi_{k-1}\}$
z'_t
z'_s
Predicted State:
$\bar{x}_{k+1|k}$
$\{\bar{\mathbf{t}}_{k-1}\ \bar{s}_{k+1}\ \bar{\psi}_{k+1}\}$
$\bar{z}_t$
$\bar{z}_s$
Estimation add dach
3D IoU
Appearance
cos
Interpretable Matching


t_k
t_{k+1}
3D Object Detection
3D Object Detection
Test Time Inverse Rendering Optimization
Init. Latent
Final Latent
$\{\mathbf{t}_k\ \mathbf{s}_k\ \psi_k\}$
z_t
z_s
Generative Obj. Prior
Target Prediction
$\mathcal{L}_{IR}$
Backward Pass
N Inverse Rendering Optimization Steps
STOPP
Test Time Inverse Rendering Optimization
Observation:
Predicted State:
Tracked State:
$\bar{x}_{k+1|k}$
$\hat{x}_{k|k-1}$
Estimation add dach
3D IoU
Appearance
cos
Interpretable Matching
Track Embed. Update

Object Prior – GET3D



Our method is <u>independent</u> of the object prior

G. Shen, et al. (2022). GET3D: A Generative Model of High Quality 3D Textured Shapes Learned from Images.

Real Dataset Unseen by Our Method – Generalizability

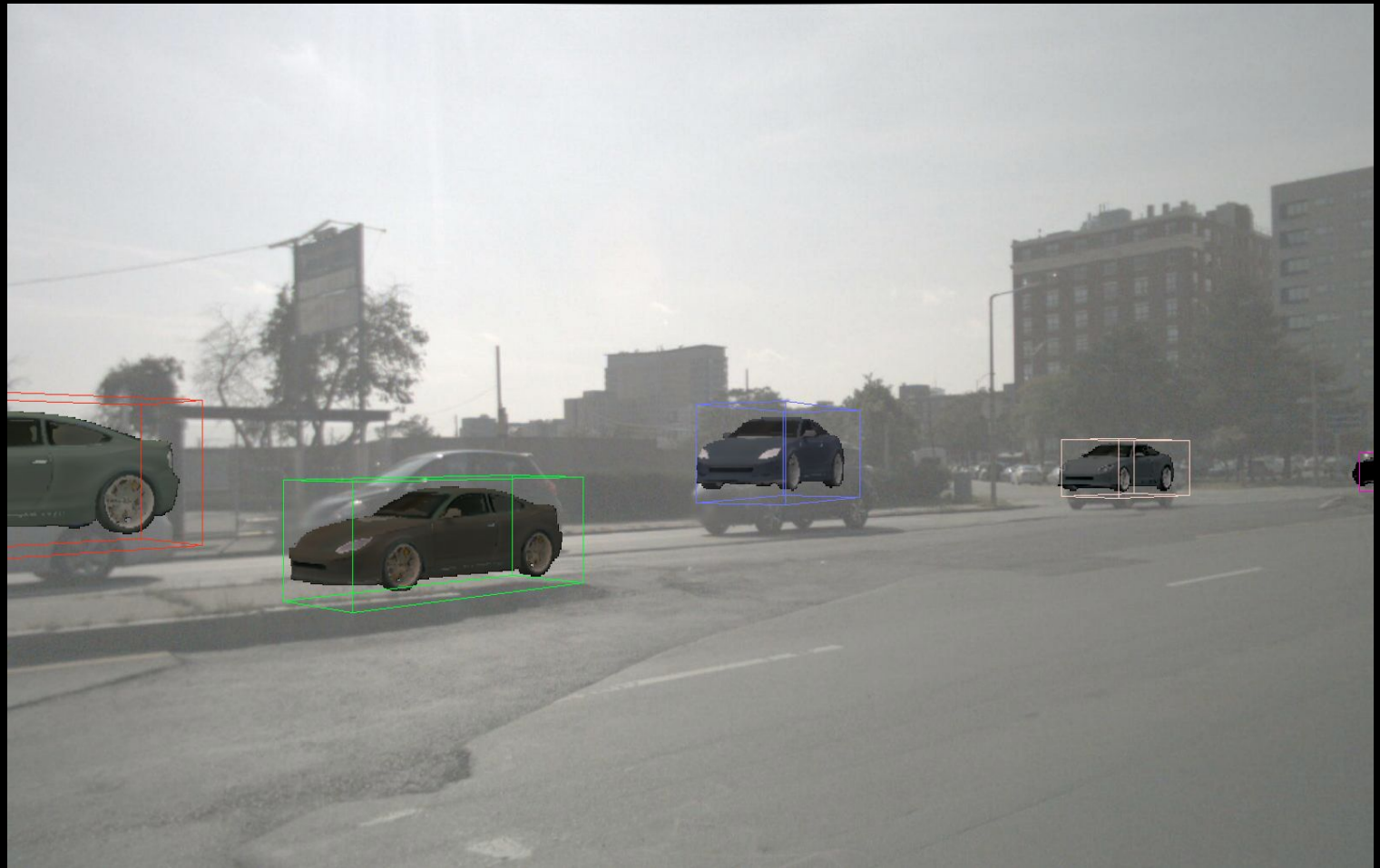
G. Shen, et al. (2022). GET3D: A Generative Model of High Quality 3D Textured Shapes Learned from Images.
Chang, A., et al. (2015). ShapeNet: An Information-Rich 3D Model Repository.

Optimization Process

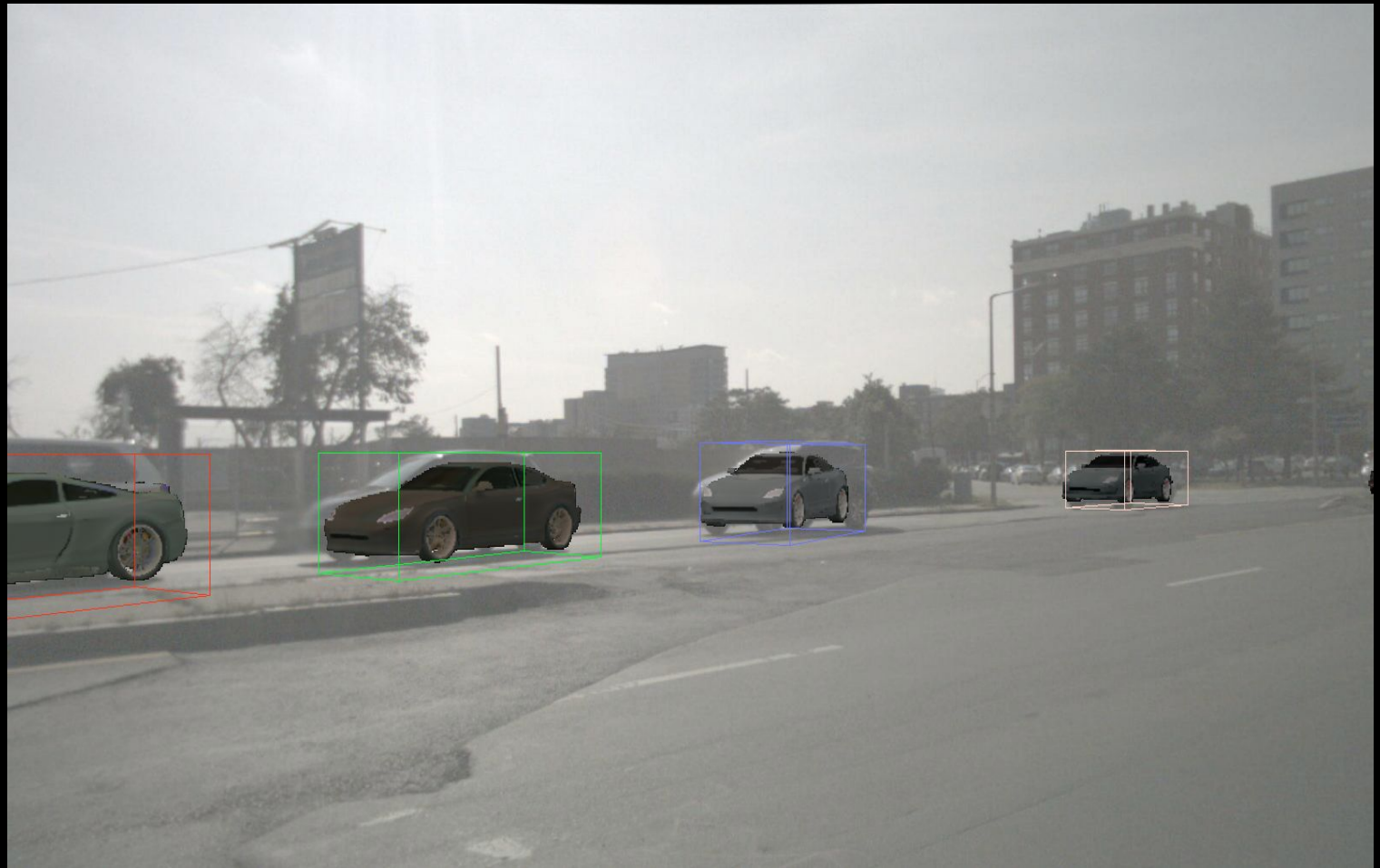
Input Frame

Initial Guess

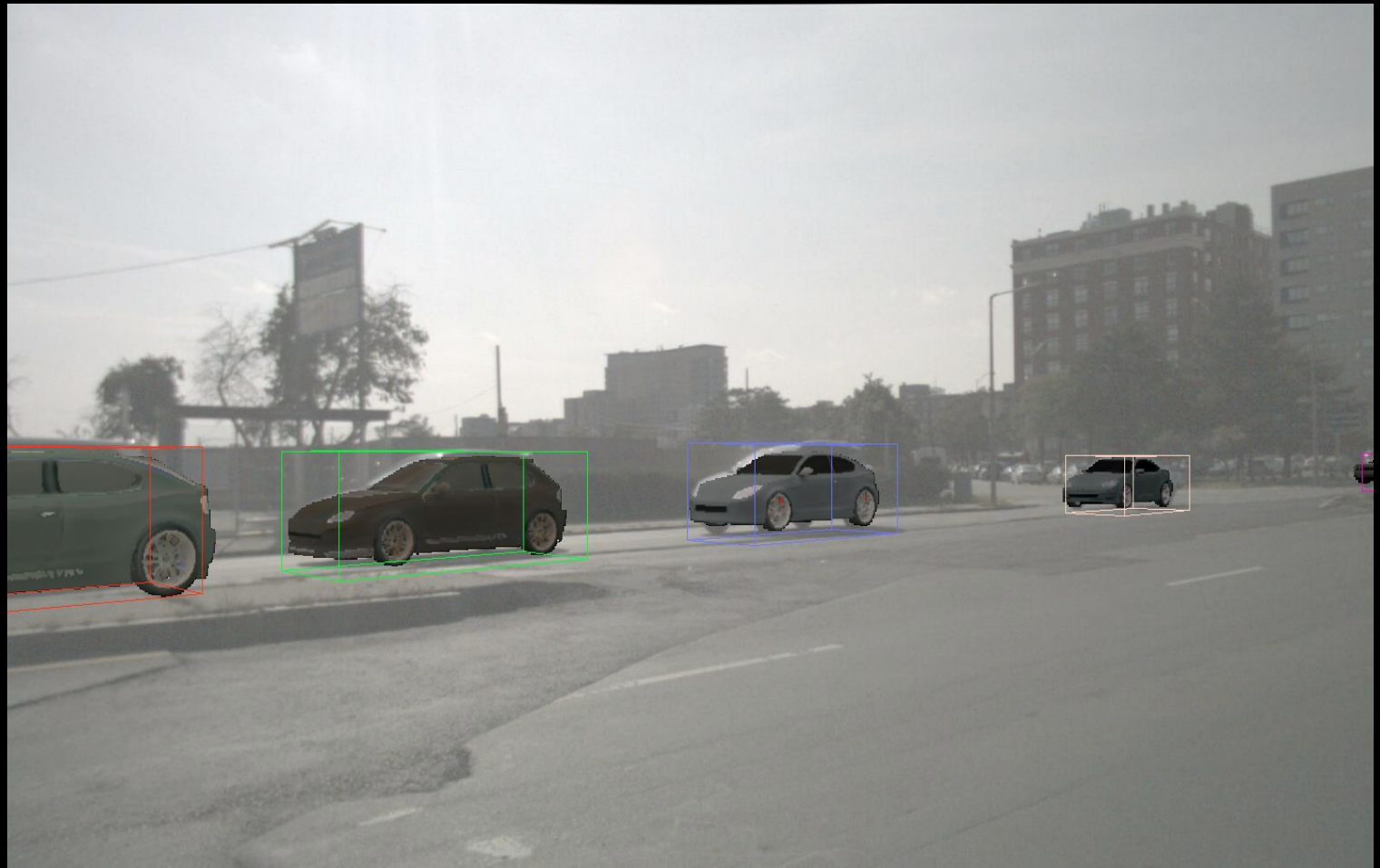
Texture Fit



Pose Fit



Shape Fit



Input

Our Schedule

Initial Guess

No Schedule

Loss Function

$$\mathcal{L}_{IR} = L_{RGB} + \lambda \mathcal{L}_{perceptual}$$

$$\mathcal{L}_{IR} = L_{RGB} + \lambda \mathcal{L}_{perceptual}$$

$$\left(I_c - \hat{I}_c \right) \circ \hat{M}_{I_c} \|_2$$

Ground Truth Image

Rendered Image

Rendered Object Masks

$$\mathcal{L}_{IR} = L_{RGB} + \lambda \mathcal{L}_{perceptual}$$

$$\text{LPIPS}_{patch}\left(I_c, \hat{I}_{c,p}\right)$$

Object-centered patches

Ground Truth Image

Rendered Image

$$\mathcal{L}_{IR} = L_{RGB} + \lambda \mathcal{L}_{perceptual}$$

$$\hat{\mathbf{z}}_{S,p}, \hat{\mathbf{z}}_{T,p}, \hat{s}_p \hat{\mathbf{t}}_p, \hat{\mathbf{R}}_p = \arg\min(\mathcal{L}_{IR})$$

Shape **Texture** **Scale** **Translation** **Rotation**

Agenda

- → Prior Work and Motivation
- → Our Method
- → **Generalization**
- → Failure Cases and Interpretability
- → 3D Interpretation
- → Limitations
- → Key Takeaway

Single-Shot Performance

Training Data Unseen	Method	AMOTA ↑	AMOTP (m) ↓	Recall ↑	MOTA ↑
×	PF-Track	0.622	0.916	0.719	0.558
×	QTrack	0.692	0.753	0.760	0.596
×	QD-3DT	0.425	1.258	0.563	0.358
✓	QD-3DT (trained on WOD)	0.000	1.893	0.226	0.000
× (CP)	CenterTrack	0.202	1.195	0.313	0.134
✓ (CP)	AB3DMOT	0.387	**1.158**	0.506	0.284
✓ (CP)	Inverse Neural Rendering (ours)	**0.413**	<u>1.189</u>	**0.536**	**0.321**

Bold denotes best
<u>Underlined</u> denotes second-best

Pang, Z., et al.: Standing between past and future: Spatio-temporal modeling for multi-camera 3d multi-object tracking. (2023).; Yang, J., et al.: Quality matters: Embracing quality clues for robust 3d multi-object tracking.; Hu, H.N., et al.: Monocular quasi-dense 3d object tracking. (2022); Zhou, X., et al.: Tracking objects as points. (2020); Weng, X., et al..: 3d multi-object tracking: A baseline and 677 new evaluation metrics. (2020)

Agenda

- → Prior Work and Motivation
- → Our Method
- → Generalization
- → **Failure Cases and Interpretability**
- → 3D Reasoning
- → Limitations
- → Key Takeaway

Rainy Scene

Reflection

Interpretability “for free”

Agenda

- → Prior Work and Motivation
- → Our Method
- → Generalization
- → Failure Cases and Interpretability
- → **3D Interpretation**
- → Key Takeaway
- → Current Project

Input Frame

INR Generation

INR BEV Layout

Agenda

- → Prior Work and Motivation
- → Our Method
- → Generalization
- → Failure Cases and Interpretability
- → 3D Interpretation
- → **Limitations**
- → Key Takeaway

Not real-time (yet)

 0.3 seconds / frame

Rendering pipeline not performance-optimized

Solution: Adaptive level-of-detail rendering?

Limited object prior

GET3D does not model shadows, reflectance, environmental conditions, etc.

Solution: Use object model that also models these?

Agenda

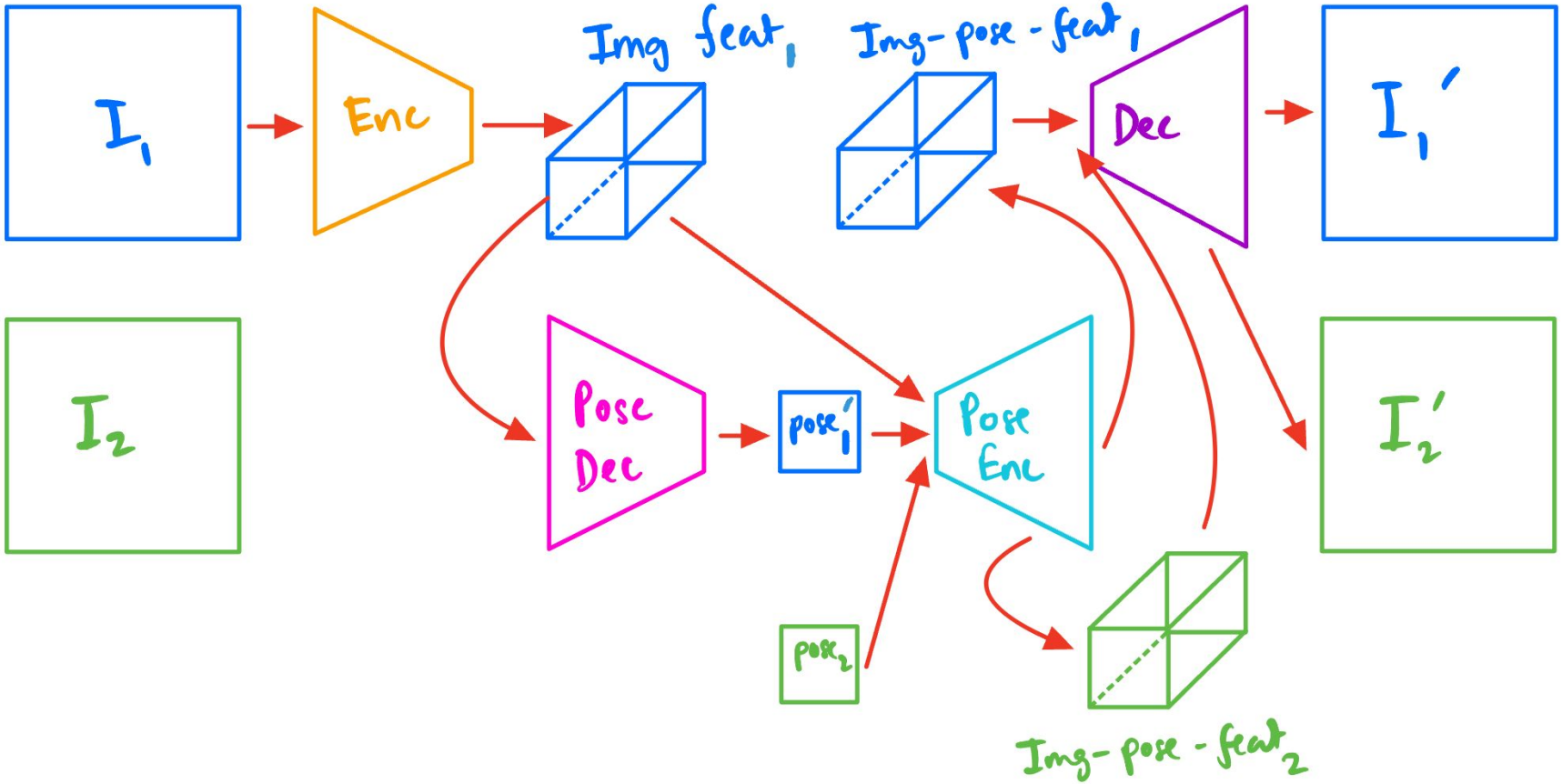
- → Prior Work and Motivation
- → Our Method
- → Generalization
- → Failure Cases and Interpretability
- → 3D Interpretation
- → Limitations
- → **Key Takeaway**

Leveraging 3D object priors can allow for interpretability

Leveraging 3D object priors can allow for generalization

Leveraging 3D object priors can allow for 3D interpretation

Object Detection Using VAEs



Thank you!

Inverse Neural Rendering for Explainable Multi-Object Tracking

Julian Ost*, Tanushree Banerjee*, Mario Bijelic, Felix Heide

PRINCETON
COMPUTATIONAL IMAGING LAB

* denotes equal contribution

References

Adelson, E. H., Bergen, J. R., & others. (1991). The plenoptic function and the elements of early vision (Vol. 2). Vision.
Agarwal, S., Furukawa, Y., Snavely, N., Simon, I., Curless, B., Seitz, S. M., & Szeliski, R. (2011). Building Rome in a Day. Commun. ACM, 54(10), 105–112. https://doi.org/10.1145/2001269.2001293
Aliev, K.-A., Sevastopolsky, A., Kolos, M., Ulyanov, D., & Lempitsky, V. (2020a). Neural Point-Based Graphics.
Aliev, K.-A., Sevastopolsky, A., Kolos, M., Ulyanov, D., & Lempitsky, V. (2020b). Neural point-based graphics. Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXII 16, 696–712.
Álvarez Aparicio, C., Guerrero-Higueras, Á. M., Rodríguez-Lera, F. J., Ginés Clavero, J., Martín Rico, F., & Matellán, V. (2019). People detection and tracking using LIDAR sensors. Robotics, 8(3), 75.
Ames, A. L., Nadeau, D. R., & Moreland, J. L. (1997). The VRML 2.0 sourcebook. John Wiley & Sons, Inc.
Arandjelović, R., & Zisserman, A. (2021). NeRF in detail: Learning to sample for view synthesis. ArXiv Preprint ArXiv:2106.05264.
Bastani, F., He, S., & Madden, S. (2021). Self-Supervised Multi-Object Tracking with Cross-Input Consistency. Advances in Neural Information Processing Systems, 34, 13695–13706.
Beker, D., Kato, H., Morariu, M. A., Ando, T., Matsuoka, T., Kehl, W., & Gaidon, A. (2020). Monocular differentiable rendering for self-supervised 3d object detection. European Conference on Computer Vision, 514–529.
Bergmann, P., Meinhardt, T., & Leal-Taixé, L. (2019, October). Tracking Without Bells and Whistles. The IEEE International Conference on Computer Vision (ICCV).
Bergmann, P., Meinhardt, T., & Leal-Taixe, L. (2019). Tracking without bells and whistles. Proceedings of the IEEE/CVF International Conference on Computer Vision, 941–951.
Bernardin, K., Elbs, A., & Stiefelhagen, R. (2006). Multiple object tracking performance metrics and evaluation in a smart room environment. Sixth IEEE International Workshop on Visual Surveillance, in Conjunction with ECCV, 90(91).
Bewley, A., Ge, Z., Ott, L., Ramos, F., & Upcroft, B. (2016). Simple online and realtime tracking. 2016 IEEE International Conference on Image Processing (ICIP), 3464–3468.
Bian, Z., Jabri, A., Efros, A. A., & Owens, A. (2022). Learning pixel trajectories with multiscale contrastive random walks. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 6508–6519.
Bianco, S., Ciocca, G., & Marelli, D. (2018). Evaluating the Performance of Structure from Motion Pipelines. Journal of Imaging, 4, 98. https://doi.org/10.3390/jimaging4080098
Bochinski, E., Eiselein, V., & Sikora, T. (2017). High-speed tracking-by-detection without using image information. 2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), 1–6.
Bolles, R. C., Baker, H. H., & Marimont, D. H. (1987). Epipolar-plane image analysis: An approach to determining structure from motion. International Journal of Computer Vision, 1(1), 7–55.
Boss, M., Braun, R., Jampani, V., Barron, J. T., Liu, C., & Lensch, H. (2021). NeRD: Neural reflectance decomposition from image collections. Proceedings of the IEEE/CVF International Conference on Computer Vision, 12684–12694.
Brasó, G., & Leal-Taixé, L. (2020). Learning a neural solver for multiple object tracking. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 6247–6257.
Breitenstein, M. D., Reichlin, F., Leibe, B., Koller-Meier, E., & Van Gool, L. (2009). Robust tracking-by-detection using a detector confidence particle filter. 2009 IEEE 12th International Conference on Computer Vision, 1515–1522.
Cabon, Y., Murray, N., & Humenberger, M. (2020). Virtual KITTI 2.
Caesar, H., Bankiti, V., Lang, A. H., Vora, S., Liong, V. E., Xu, Q., Krishnan, A., Pan, Y., Baldan, G., & Beijbom, O. (2020). nuscenes: A multimodal dataset for autonomous driving. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 11621–11631.
Cai, J., Xu, M., Li, W., Xiong, Y., Xia, W., Tu, Z., & Soatto, S. (2022). MeMOT: Multi-Object Tracking with Memory. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 8090–8100.
Cao, J., Weng, X., Khirodkar, R., Pang, J., & Kitani, K. (2022). Observation-Centric SORT: Rethinking SORT for Robust Multi-Object Tracking. ArXiv Preprint ArXiv:2203.14360.
Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., & Zagoruyko, S. (2020). End-to-end object detection with transformers. European Conference on Computer Vision, 213–229.
Chaabane, M., Zhang, P., Beveridge, J. R., & O'Hara, S. (2021). Deft: Detection embeddings for tracking. ArXiv Preprint ArXiv:2102.02267.
Chang, A. X., Funkhouser, T., Guibas, L., Hanrahan, P., Huang, Q., Li, Z., Savarese, S., Savva, M., Song, S., Su, H., Xiao, J., Yi, L., & Yu, F. (2015). ShapeNet: An Information-Rich 3D Model Repository (Techreport arXiv:1512.03012 [cs.GR]). Stanford University — Princeton University — Toyota Technological Institute at Chicago.
Chen, A., Wu, M., Zhang, Y., Li, N., Lu, J., Gao, S., & Yu, J. (2018). Deep Surface Light Fields. Proceedings of the ACM on Computer Graphics and Interactive Techniques, 1(1), 1–17. https://doi.org/10.1145/3203192
Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2014). Semantic image segmentation with deep convolutional nets and fully connected crfs. ArXiv Preprint ArXiv:1412.7062.
Chen, S. (2011). Kalman filter for robot vision: a survey. IEEE Transactions on Industrial Electronics, 59(11), 4409–4420.
Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020). A simple framework for contrastive learning of visual representations. International Conference on Machine Learning, 1597–1607.
Chen, Y., Yu, Z., Chen, Y., Lan, S., Anandkumar, A., Jia, J., & Alvarez, J. M. (2023). FocalFormer3D: Focusing on Hard Instance for 3D Object Detection. ICCV.
Chou, G., Chugunov, I., & Heide, F. (2022). GenSDF: Two-Stage Learning of Generalizable Signed Distance Functions. Proc. of Neural Information Processing Systems (NeurIPS).
Chu, P., & Ling, H. (2019). Famnet: Joint learning of feature, affinity and multi-dimensional assignment for online multiple object tracking. Proceedings of the IEEE/CVF International Conference on Computer Vision, 6172–6181.
Ciaparrone, G., Sánchez, F. L., Tabik, S., Troiano, L., Tagliaferri, R., & Herrera, F. (2020). Deep learning in video multi-object tracking: A survey. Neurocomputing, 381, 61–88.
Costeira, J., & Kanade, T. (1995). A multi-body factorization method for motion analysis. Proceedings of IEEE International Conference on Computer Vision, 1071–1076.
Cunningham, S., & Bailey, M. (2001). Lessons from scene graphs: using scene graphs to teach hierarchical modeling. Comput. Graph., 25, 703–711.
Dai, P., Zhang, Y., Li, Z., Liu, S., & Zeng, B. (2020, June). Neural Point Cloud Rendering via Multi-Plane Projection. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).
Deng, K., Liu, A., Zhu, J.-Y., & Ramanan, D. (2021). Depth-supervised NeRF: Fewer Views and Faster Training for Free. ArXiv Preprint ArXiv:2107.02791.
DeVries, T., Bautista, M. A., Srivastava, N., Taylor, G. W., & Susskind, J. M. (2021). Unconstrained Scene Generation with Locally Conditioned Radiance Fields. ArXiv Preprint ArXiv:2104.00670.
Dewan, A., Caselitz, T., Tipaldi, G. D., & Burgard, W. (2016). Motion-based detection and tracking in 3d lidar scans. 2016 IEEE International Conference on Robotics and Automation (ICRA), 4508–4513.
Dhamo, H., Farshad, Azade, Laina, I., Navab, N., Hager, Gregory D., Tombari, F., & Rupprecht, C. (2020). Semantic Image Manipulation Using Scene Graphs. CVPR.
Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., & others. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. ArXiv Preprint ArXiv:2010.11929.
Flynn, J., Broxton, M., Debevec, P., DuVall, M., Fyffe, G., Overbeck, R., Snavely, N., & Tucker, R. (2019). DeepView: View Synthesis With Learned Gradient Descent. Proceedings IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). https://doi.org/10.1109/cvpr.2019.00247
Fu, Z., Liu, Q., Fu, Z., & Wang, Y. (2021). Stmtrack: Template-free visual tracking with space-time memory networks. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 13774–13783.

References

Gao, J., Shen, T., Wang, Z., Chen, W., Yin, K., Li, D., Litany, O., Gojcic, Z., & Fidler, S. (2022). GET3D: A Generative Model of High Quality 3D Textured Shapes Learned from Images. Advances In Neural Information Processing Systems.
Garbin, S. J., Kowalski, M., Johnson, M., Shotton, J., & Valentin, J. (2021). Fastnerf: High-fidelity neural rendering at 200fps. ArXiv Preprint ArXiv:2103.10380.
Geiger, A., Lenz, P., & Urtasun, R. (2012a). Are we ready for Autonomous Driving? The KITTI Vision Benchmark Suite. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
Geiger, A., Lenz, P., & Urtasun, R. (2012b). Are we ready for Autonomous Driving? The KITTI Vision Benchmark Suite. Conference on Computer Vision and Pattern Recognition (CVPR).
Girshick, R. (2015). Fast r-cnn. Proceedings of the IEEE International Conference on Computer Vision, 1440–1448.
Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 580–587.
Gladkova, M., Korobov, N., Demmel, N., Ošep, A., Leal-Taixé, L., & Cremers, D. (2022). DirectTracker: 3D Multi-Object Tracking Using Direct Image Alignment and Photometric Bundle Adjustment. ArXiv Preprint ArXiv:2209.14965.
Goyal, A., Law, H., Liu, B., Newell, A., & Deng, J. (2021). Revisiting Point Cloud Shape Classification with a Simple and Effective Baseline. ArXiv Preprint ArXiv:2106.05304.
Groueix, T., Fisher, M., Kim, V. G., Russell, B. C., & Aubry, M. (2018). A papier-mâché approach to learning 3d surface generation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 216–224.
Guo, S., Wang, J., Wang, X., & Tao, D. (2021). Online multiple object tracking with cross-task synergy. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 8136–8145.
Guo, Y.-C., Kang, D., Bao, L., He, Y., & Zhang, S.-H. (2022). Nerfren: Neural radiance fields with reflections. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 18409–18418.
Hamdi, A., Giancola, S., & Ghanem, B. (2021). MVTN: Multi-View Transformation Network for 3D Shape Recognition. Proceedings of the IEEE/CVF International Conference on Computer Vision, 1–11.
Hao, Z., Mallya, A., Belongie, S., & Liu, M.-Y. (2021). GANcraft: Unsupervised 3D Neural Rendering of Minecraft Worlds. ArXiv Preprint ArXiv:2104.07659.
Hart, J. C. (1996). Sphere tracing: A geometric method for the antialiased ray tracing of implicit surfaces. The Visual Computer, 12(10), 527–545.
Hartley, R., & Zisserman, A. (2003). Multiple View Geometry in Computer Vision (2nd ed.). Cambridge University Press.
He, J., Huang, Z., Wang, N., & Zhang, Z. (2021). Learnable graph matching: Incorporating graph partitioning with deep feature learning for multiple object tracking. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 5299–5309.
He, K., Fan, H., Wu, Y., Xie, S., & Girshick, R. (2020). Momentum contrast for unsupervised visual representation learning. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 9729–9738.
He, K., Zhang, X., Ren, S., & Sun, J. (2015). Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 37(9), 1904–1916.
He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 770–778.
He, T., & Soatto, S. (2019). Mono3d++: Monocular 3d vehicle detection with two-scale 3d hypotheses and task priors. Proceedings of the AAAI Conference on Artificial Intelligence, 33(1), 8409–8416.
Hedman, P., Srinivasan, P. P., Mildenhall, B., Barron, J. T., & Debevec, P. (2021). Baking Neural Radiance Fields for Real-Time View Synthesis. ArXiv Preprint ArXiv:2103.14645.
Heung-Yeung Shum, Qifa Ke, & Zhengyou Zhang. (1999). Efficient bundle adjustment with virtual key frames: a hierarchical approach to multi-frame structure from motion. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 2, 538-543 Vol. 2.
Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. Advances in Neural Information Processing Systems, 33, 6840–6851.
Hu, H.-N., Cai, Q.-Z., Wang, D., Lin, J., Sun, M., Krahenbuhl, P., Darrell, T., & Yu, F. (2019). Joint monocular 3D vehicle detection and tracking. Proceedings of the IEEE/CVF International Conference on Computer Vision, 5390–5399.
Hu, H.-N., Yang, Y.-H., Fischer, T., Darrell, T., Yu, F., & Sun, M. (2022a). Monocular quasi-dense 3d object tracking. IEEE Transactions on Pattern Analysis and Machine Intelligence.
Hu, H.-N., Yang, Y.-H., Fischer, T., Darrell, T., Yu, F., & Sun, M. (2022b). Monocular Quasi-Dense 3D Object Tracking. IEEE Transactions on Pattern Analysis and Machine Intelligence.
Huang, K., & Hao, Q. (2021). Joint Multi-Object Detection and Tracking with Camera-LiDAR Fusion for Autonomous Driving. 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 6983–6989.
Hung, W.-C., Kretzschmar, H., Lin, T.-Y., Chai, Y., Yu, R., Yang, M.-H., & Anguelov, D. (2020). SoDA: Multi-object tracking with soft data association. ArXiv Preprint ArXiv:2008.07725.
Hyun, J., Kang, M., Wee, D., & Yeung, D.-Y. (2022). Detection Recovery in Online Multi-Object Tracking with Sparse Graph Tracker. ArXiv Preprint ArXiv:2205.00968.
Jabri, A., Owens, A., & Efros, A. (2020). Space-time correspondence as a contrastive random walk. Advances in Neural Information Processing Systems, 33, 19545–19560.
Jiang, Y., Ji, D., Han, Z., & Zwicker, M. (2020). Sdfdiff: Differentiable rendering of signed distance fields for 3d shape optimization. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 1251–1261.
Jiao, L., Zhang, F., Liu, F., Yang, S., Li, L., Feng, Z., & Qu, R. (2019). A survey of deep learning-based object detection. IEEE Access, 7, 128837–128868.
Johnson, J., Gupta, A., & Fei-Fei, L. (2018). Image Generation from Scene Graphs. CVPR.
Kalal, Z., Mikolajczyk, K., & Matas, J. (2011). Tracking-learning-detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 34(7), 1409–1422.
Kalman, R. E. (1960). A new approach to linear filtering and prediction problems.
Kang, M., Zhu, J.-Y., Zhang, R., Park, J., Shechtman, E., Paris, S., & Park, T. (2023). Scaling up gans for text-to-image synthesis. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 10124–10134.
Karras, T., Laine, S., & Aila, T. (2019). A style-based generator architecture for generative adversarial networks. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 4401–4410.
Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., & Aila, T. (2020). Analyzing and improving the image quality of stylegan. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 8110–8119.
Kellnhofer, P., Jebe, L. C., Jones, A., Spicer, R., Pulli, K., & Wetzstein, G. (2021). Neural lumigraph rendering. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 4287–4297.
Kim, A., Ošep, A., & Leal-Taixé, L. (2021). Eagermot: 3d multi-object tracking via sensor fusion. 2021 IEEE International Conference on Robotics and Automation (ICRA), 11315–11321.
Kingma, D. P., & Ba, J. (2015). Adam: A Method for Stochastic Optimization. CoRR, abs/1412.6980.
Koestler, L., Grittner, D., Moeller, M., Cremers, D., & Lähner, Z. (2022). Intrinsic neural fields: Learning functions on manifolds. ArXiv Preprint ArXiv:2203.07967, 2.
Kopanas, G., Philip, J., Leimkühler, T., & Drettakis, G. (2021). Point-Based Neural Rendering with Per-View Optimization. Computer Graphics Forum, 40(4), 29–43.
Ku, J., Mozifian, M., Lee, J., Harakeh, A., & Waslander, S. L. (2018). Joint 3d proposal generation and object detection from view aggregation. 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 1–8.
Ku, J., Pon, A. D., & Waslander, S. L. (2019). Monocular 3d object detection leveraging accurate proposals and shape reconstruction. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 11867–11876.
Kuhn, H. W. (1955). The Hungarian method for the assignment problem. Naval Research Logistics Quarterly, 2(1–2), 83–97.

References

Lai, Z., Lu, E., & Xie, W. (2020). Mast: A memory-augmented self-supervised tracker. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 6479–6488.
Lai, Z., & Xie, W. (2019). Self-supervised learning for video correspondence flow. ArXiv Preprint ArXiv:1905.00875.
Laine, S., Hellsten, J., Karras, T., Seol, Y., Lehtinen, J., & Aila, T. (2020). Modular primitives for high-performance differentiable rendering. ACM Transactions on Graphics (TOG), 39(6), 1–14.
Levoy, M., & Hanrahan, P. (1996). Light field rendering. Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques, 31–42.
Li, J., Gao, X., & Jiang, T. (2020). Graph networks for multiple object tracking. Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 719–728.
Li, S., Kong, Y., & Rezatofighi, H. (2022). Learning of Global Objective for Network Flow in Multi-Object Tracking. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 8855–8865.
Li, Y., Qi, H., Dai, J., Ji, X., & Wei, Y. (2017). Fully convolutional instance-aware semantic segmentation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2359–2367.
Liao, S., & Shao, L. (2021). TransMatcher: Deep Image Matching Through Transformers for Generalizable Person Re-identification. Advances in Neural Information Processing Systems, 34, 1992–2003.
Lin, C.-H., Ma, W.-C., Torralba, A., & Lucey, S. (2021). Barf: Bundle-adjusting neural radiance fields. Proceedings of the IEEE/CVF International Conference on Computer Vision, 5741–5751.
Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017). Feature pyramid networks for object detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2117–2125.
Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., & Zitnick, C. L. (2014). Microsoft coco: Common objects in context. European Conference on Computer Vision, 740–755.
Liu, C., Li, Z., Yuan, J., & Xu, Y. (2021). NeLF: Practical Novel View Synthesis with Neural Light Field. ArXiv Preprint ArXiv:2105.07112.
Liu, L., Gu, J., Lin, K. Z., Chua, T.-S., & Theobalt, C. (2020). Neural sparse voxel fields. ArXiv Preprint ArXiv:2007.11571.
Liu, Q., Chen, D., Chu, Q., Yuan, L., Liu, B., Zhang, L., & Yu, N. (2022). Online multi-object tracking with unsupervised re-identification learning and occlusion estimation. Neurocomputing, 483, 333–347.
Liu, S., Zhang, Y., Peng, S., Shi, B., Pollefeys, M., & Cui, Z. (2020). Dist: Rendering deep implicit signed distance function with differentiable sphere tracing. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019–2028.
Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., & Berg, A. C. (2016). Ssd: Single shot multibox detector. European Conference on Computer Vision, 21–37.
Liu, Ze, Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., & Guo, B. (2021). Swin transformer: Hierarchical vision transformer using shifted windows. Proceedings of the IEEE/CVF International Conference on Computer Vision, 10012–10022.
Liu, Zhijian, Tang, H., Amini, A., Yang, X., Mao, H., Rus, D., & Han, S. (2022). BEVFusion: Multi-Task Multi-Sensor Fusion with Unified Bird's-Eye View Representation. ArXiv.
Lombardi, S., Simon, T., Saragih, J., Schwartz, G., Lehrmann, A., & Sheikh, Y. (2019). Neural Volumes: Learning Dynamic Renderable Volumes from Images. ACM Trans. Graph., 38(4), 65:1-65:14. https://doi.org/10.1145/3306346.3323020
Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 3431–3440.
Lu, E., Cole, F., Dekel, T., Xie, W., Zisserman, A., Salesin, D., Freeman, W. T., & Rubinstein, M. (2020). Layered Neural Rendering for Retiming People in Video.
Luiten, J., Fischer, T., & Leibe, B. (2020). Track to reconstruct and reconstruct to track. IEEE Robotics and Automation Letters, 5(2), 1803–1810.
Ma, F., Shou, M. Z., Zhu, L., Fan, H., Xu, Y., Yang, Y., & Yan, Z. (2022). Unified transformer tracker for object tracking. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 8781–8790.
Majercik, A., Crassin, C., Shirley, P., & McGuire, M. (2018). A Ray-Box Intersection Algorithm and Efficient Dynamic Voxel Rendering. Journal of Computer Graphics Techniques (JCGT), 7(3), 66–81. http://jcgt.org/published/0007/03/04/
Mao, J., Shi, S., Wang, X., & Li, H. (2022). 3d object detection for autonomous driving: A review and new outlooks. ArXiv Preprint ArXiv:2206.09474.
Marinello, N., Proesmans, M., & Van Gool, L. (2022). TripletTrack: 3D Object Tracking Using Triplet Embeddings and LSTM. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 4500–4510.
Martin-Brualla, R., Radwan, N., Sajjadi, M. S. M., Barron, J. T., Dosovitskiy, A., & Duckworth, D. (2020). NeRF in the Wild: Neural Radiance Fields for Unconstrained Photo Collections. ArXiv.
Meinhardt, T., Kirillov, A., Leal-Taixe, L., & Feichtenhofer, C. (2022). Trackformer: Multi-object tracking with transformers. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 8844–8854.
Mescheder, L., Oechsle, M., Niemeyer, M., Nowozin, S., & Geiger, A. (2019, June). Occupancy Networks: Learning 3D Reconstruction in Function Space. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).
Milan, A., Leal-Taixé, L., Reid, I., Roth, S., & Schindler, K. (2016). MOT16: A benchmark for multi-object tracking. ArXiv Preprint ArXiv:1603.00831.
Mildenhall, B., Srinivasan, P. P., Ortiz-Cayon, R., Kalantari, N. K., Ramamoorthi, R., Ng, R., & Kar, A. (2019). Local Light Field Fusion: Practical View Synthesis with Prescriptive Sampling Guidelines. ACM Transactions on Graphics (TOG).
Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., & Ng, R. (2020). NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. In Proceedings of the IEEE European Conf. on Computer Vision (ECCV).
Nadeau, D. R. (2000). Volume Scene Graphs. 2000 IEEE Symposium on Volume Visualization (VV 2000), 49–56.
Neff, T., Stadlbauer, P., Parger, M., Kurz, A., Alla Chaitanya, C. R., Kaplanyan, A., & Steinberger, M. (2021). DONeRF: Towards real-time rendering of neural radiance fields using depth oracle networks. Computer Graphics Forum, 40.
Nguyen, H. T., & Smeulders, A. W. (2004). Fast occluded object tracking by a robust appearance filter. IEEE Transactions on Pattern Analysis and Machine Intelligence, 26(8), 1099–1104.
Nguyen, P., Quach, K. G., Duong, C. N., Le, N., Nguyen, X.-B., & Luu, K. (2022). Multi-Camera Multiple 3D Object Tracking on the Move for Autonomous Vehicles. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2569–2578.
Nichol, A. Q., & Dhariwal, P. (2021). Improved denoising diffusion probabilistic models. International Conference on Machine Learning, 8162–8171.
Niemeyer, M., & Geiger, A. (2021). GIRAFFE: Representing Scenes as Compositional Generative Neural Feature Fields. Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR).
Niemeyer, M., Mescheder, L., Oechsle, M., & Geiger, A. (2020). Differentiable Volumetric Rendering: Learning Implicit 3D Representations without 3D Supervision. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
Nimier-David, M., Dong, Z., Jakob, W., & Kaplanyan, A. (2021). Material and Lighting Reconstruction for Complex Indoor Scenes with Texture-space Differentiable Rendering. In A. Bousseau & M. McGuire (Eds.), Eurographics Symposium on Rendering - DL-only Track. The Eurographics Association. https://doi.org/10.2312/sr.20211292
Nimier-David, M., Vicini, D., Zeltner, T., & Jakob, W. (2019). Mitsuba 2: A Retargetable Forward and Inverse Renderer. Transactions on Graphics (Proceedings of SIGGRAPH Asia), 38(6). https://doi.org/10.1145/3355089.3356498
Osep, A., Mehner, W., Mathias, M., & Leibe, B. (2017). Combined image-and world-space tracking in traffic scenes. 2017 IEEE International Conference on Robotics and Automation (ICRA), 1988–1995.
Ost, J., Mannan, F., Thuerey, N., Knodt, J., & Heide, F. (2021). Neural scene graphs for dynamic scenes. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2856–2865.
Ozden, K. E., Schindler, K., & Van Gool, L. (2010). Multibody Structure-from-Motion in Practice. IEEE Transactions on Pattern Analysis and Machine Intelligence, 32(6), 1134–1141.
Pang, Z., Li, J., Tokmakov, P., Chen, D., Zagoruyko, S., & Wang, Y.-X. (2023). Standing Between Past and Future: Spatio-Temporal Modeling for Multi-Camera 3D Multi-Object Tracking. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.

References

Pang, Z., Li, Z., & Wang, N. (2021). Simpletrack: Understanding and rethinking 3d multi-object tracking. ArXiv Preprint ArXiv:2111.09621.
Park, J. J., Florence, P., Straub, J., Newcombe, R., & Lovegrove, S. (2019, June). DeepSDF: Learning Continuous Signed Distance Functions for Shape Representation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
Park, K., Sinha, U., Barron, J. T., Bouaziz, S., Goldman, D. B., Seitz, S. M., & Martin-Brualla, R. (2021). Nerfies: Deformable Neural Radiance Fields. Proceedings of the IEEE International Conference on Computer Vision.
Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., … Chintala, S. (2019). PyTorch: An Imperative Style, High-Performance Deep Learning Library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché Buc, E. Fox, & R. Garnett (Eds.), Advances in Neural Information Processing Systems 32 (pp. 8024–8035). Curran Associates, Inc. http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf
Pittaluga, F., Koppal, S. J., Bing Kang, S., & Sinha, S. N. (2019). Revealing scenes by inverting structure from motion reconstructions. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 145–154.
Possegger, H., Mauthner, T., Roth, P. M., & Bischof, H. (2014). Occlusion geodesics for online multi-object tracking. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1306–1313.
Qi, C. R., Liu, W., Wu, C., Su, H., & Guibas, L. J. (2018). Frustum pointnets for 3d object detection from rgb-d data. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 918–927.
Rakai, L., Song, H., Sun, S., Zhang, W., & Yang, Y. (2021). Data association in multiple object tracking: A survey of recent techniques. Expert Systems with Applications, 116300.
Ravi, N., Reizenstein, J., Novotny, D., Gordon, T., Lo, W.-Y., Johnson, J., & Gkioxari, G. (2020). Accelerating 3D Deep Learning with PyTorch3D. ArXiv:2007.08501.
Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 779–788.
Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. Advances in Neural Information Processing Systems, 28.
Riegler, G., & Koltun, V. (2020). Free view synthesis. European Conference on Computer Vision, 623–640.
Riegler, G., & Koltun, V. (2021). Stable view synthesis. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 12216–12225.
Rusinkiewicz, S., & Levoy, M. (2001). Efficient variants of the ICP algorithm. Proceedings Third International Conference on 3-D Digital Imaging and Modeling, 145–152.
Saleh, F., Aliakbarian, S., Rezatofighi, H., Salzmann, M., & Gould, S. (2021). Probabilistic tracklet scoring and inpainting for multiple object tracking. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 14329–14339.
Sauer, A., Schwarz, K., & Geiger, A. (2022). Stylegan-xl: Scaling stylegan to large diverse datasets. ACM SIGGRAPH 2022 Conference Proceedings, 1–10.
Scheidegger, S., Benjaminsson, J., Rosenberg, E., Krishnan, A., & Granström, K. (2018). Mono-camera 3d multi-object tracking using deep learning detections and pmbm filtering. 2018 IEEE Intelligent Vehicles Symposium (IV), 433–440.
Schönberger, J. L., & Frahm, J.-M. (2016). Structure-from-Motion Revisited. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
Schönberger, J. L., Zheng, E., Pollefeys, M., & Frahm, J.-M. (2016). Pixelwise View Selection for Unstructured Multi-View Stereo. Proceedings of the IEEE European Conf. on Computer Vision (ECCV).
Schwarz, K., Liao, Y., Niemeyer, M., & Geiger, A. (2020). GRAF: Generative Radiance Fields for 3D-Aware Image Synthesis. Advances in Neural Information Processing Systems (NeurIPS).
Sharma, S., Ansari, J. A., Murthy, J. K., & Krishna, K. M. (2018). Beyond pixels: Leveraging geometry and shape cues for online multi-object tracking. 2018 IEEE International Conference on Robotics and Automation (ICRA), 3508–3515.
Shen, B., Yan, X., Qi, C. R., Najibi, M., Deng, B., Guibas, L., Zhou, Y., & Anguelov, D. (2023). GINA-3D: Learning To Generate Implicit Neural Assets in the Wild. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 4913–4926.
Shen, T., Gao, J., Yin, K., Liu, M.-Y., & Fidler, S. (2021). Deep marching tetrahedra: a hybrid representation for high-resolution 3d shape synthesis. Advances in Neural Information Processing Systems, 34, 6087–6101.
Shi, S., Wang, X., & Li, H. (2019, June). PointRCNN: 3D Object Proposal Generation and Detection From Point Cloud. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
Simonyan, K., & Zisserman, A. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition.
Sitzmann, V., Rezchikov, S., Freeman, W. T., Tenenbaum, J. B., & Durand, F. (2021a). Light Field Networks: Neural Scene Representations with Single-Evaluation Rendering. ArXiv Preprint ArXiv:2106.02634.
Sitzmann, V., Rezchikov, S., Freeman, W. T., Tenenbaum, J. B., & Durand, F. (2021b). Light Field Networks: Neural Scene Representations with Single-Evaluation Rendering. Proc. NeurIPS.
Sitzmann, V., Thies, J., Heide, F., Nießner, M., Wetzstein, G., & Zollhöfer, M. (2019). DeepVoxels: Learning Persistent 3D Feature Embeddings. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).
Sitzmann, V., Zollhöfer, M., & Wetzstein, G. (2019). Scene Representation Networks: Continuous 3D-Structure-Aware Neural Scene Representations. Advances in Neural Information Processing Systems.
Smeulders, A. W., Chu, D. M., Cucchiara, R., Calderara, S., Dehghan, A., & Shah, M. (2013). Visual tracking: An experimental survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 36(7), 1442–1468.
Sowizral, H. (2000). Scene graphs in the new millennium. IEEE Computer Graphics and Applications, 20(1), 56–57.
Sowizral, H. A., Nadeau, D. R., Bailey, M. J., & Deering, M. F. (1998). Introduction to Programming with Java 3D. ACM SIGGRAPH 98 Course Notes.
Srinivasan, P. P., Tucker, R., Barron, J. T., Ramamoorthi, R., Ng, R., & Snavely, N. (2019). Pushing the Boundaries of View Extrapolation With Multiplane Images. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 175–184.
Srinivasan, Pratul P, Deng, B., Zhang, X., Tancik, M., Mildenhall, B., & Barron, J. T. (2021). Nerv: Neural reflectance and visibility fields for relighting and view synthesis. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 7495–7504.
Stadler, D., & Beyerer, J. (2021). Improving multiple pedestrian tracking by track management and occlusion handling. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 10958–10967.
Straßer, W. (1974). Schnelle kurven-und flächendarstellung auf grafischen sichtgeräten [Phdthesis].
Sun, Pei, Kretzschmar, H., Dotiwalla, X., Chouard, A., Patnaik, V., Tsui, P., Guo, J., Zhou, Y., Chai, Y., Caine, B., & others. (2020). Scalability in perception for autonomous driving: Waymo open dataset. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2446–2454.
Sun, Peize, Cao, J., Jiang, Y., Zhang, R., Xie, E., Yuan, Z., Wang, C., & Luo, P. (2020). Transtrack: Multiple object tracking with transformer. ArXiv Preprint ArXiv:2012.15460.
Tan, S., & Mayrovouniotis, M. L. (1995). Reducing data dimensionality through optimizing neural network inputs. AIChE Journal, 41(6), 1471–1480. https://doi.org/10.1002/aic.690410612
Tancik, M., Srinivasan, P. P., Mildenhall, B., Fridovich-Keil, S., Raghavan, N., Singhal, U., Ramamoorthi, R., Barron, J. T., & Ng, R. (2020). Fourier Features Let Networks Learn High Frequency Functions in Low Dimensional Domains.
Teed, Z., & Deng, J. (2020). Raft: Recurrent all-pairs field transforms for optical flow. European Conference on Computer Vision, 402–419.
Tewari, A., Fried, O., Thies, J., Sitzmann, V., Lombardi, S., Sunkavalli, K., Martin-Brualla, R., Simon, T., Saragih, J., Nießner, M., Pandey, R., Fanello, S., Wetzstein, G., Zhu, J.-Y., Theobalt, C., Agrawala, M., Shechtman, E., Goldman, D. B., & Zollhöfer, M. (2020). State of the Art on Neural Rendering. Computer Graphics Forum (EG STAR 2020).

References

Thies, J., Zollhöfer, M., & Nießner, M. (2019). Deferred neural rendering. ACM Transactions on Graphics, 38(4), 1–12. https://doi.org/10.1145/3306346.3323035
Tokmakov, P., Jabri, A., Li, J., & Gaidon, A. (2022). Object Permanence Emerges in a Random Walk along Memory. ArXiv Preprint ArXiv:2204.01784.
Tokmakov, P., Li, J., Burgard, W., & Gaidon, A. (2021). Learning to track with object permanence. Proceedings of the IEEE/CVF International Conference on Computer Vision, 10860–10869.
Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30.
Voigtlaender, P., Chai, Y., Schroff, F., Adam, H., Leibe, B., & Chen, L.-C. (2019). Feelvos: Fast end-to-end embedding learning for video object segmentation. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 9481–9490.
Vondrick, C., Shrivastava, A., Fathi, A., Guadarrama, S., & Murphy, K. (2018). Tracking emerges by colorizing videos. Proceedings of the European Conference on Computer Vision (ECCV), 391–408.
Wang, C., Xu, D., Zhu, Y., Martín-Martín, R., Lu, C., Fei-Fei, L., & Savarese, S. (2019). Densefusion: 6d object pose estimation by iterative dense fusion. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 3343–3352.
Wang, G., Song, M., & Hwang, J.-N. (2022). Recent Advances in Embedding Methods for Multi-Object Tracking: A Survey. ArXiv Preprint ArXiv:2205.10766.
Wang, J., Ancha, S., Chen, Y.-T., & Held, D. (2020). Uncertainty-aware self-supervised 3d data association. 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 8125–8132.
Wang, Q., Wang, Z., Genova, K., Srinivasan, P. P., Zhou, H., Barron, J. T., Martin-Brualla, R., Snavely, N., & Funkhouser, T. (2021). Ibrnet: Learning multi-view image-based rendering. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 4690–4699.
Wang, S., Liu, Y., Wang, T., Li, Y., & Zhang, X. (2023). Exploring Object-Centric Temporal Modeling for Efficient Multi-View 3D Object Detection. ArXiv Preprint ArXiv:2303.11926.
Wang, X., Jabri, A., & Efros, A. A. (2019). Learning correspondence from the cycle-consistency of time. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2566–2576.
Wang, Y., Kitani, K., & Weng, X. (2021). Joint object detection and multi-object tracking with graph neural networks. 2021 IEEE International Conference on Robotics and Automation (ICRA), 13708–13715.
Wang, Z., Simoncelli, E., & Bovik, A. (2003). Multiscale structural similarity for image quality assessment. Conference Record of the Asilomar Conference on Signals, Systems and Computers, 2, 1398-1402 Vol.2. https://doi.org/10.1109/ACSSC.2003.1292216
Wang, Zirui, Wu, S., Xie, W., Chen, M., & Prisacariu, V. A. (2021). NeRF–: Neural radiance fields without known camera parameters. ArXiv Preprint ArXiv:2102.07064.
Wanner, S., & Goldluecke, B. (2013). Variational light field analysis for disparity estimation and super-resolution. IEEE Transactions on Pattern Analysis and Machine Intelligence, 36(3), 606–619.
Wanner, S., & Goldluecke, B. (2012). Globally consistent depth labeling of 4D light fields. 2012 IEEE Conference on Computer Vision and Pattern Recognition, 41–48.
Wei, Y., Hu, H., Xie, Z., Zhang, Z., Cao, Y., Bao, J., Chen, D., & Guo, B. (2022). Contrastive Learning Rivals Masked Image Modeling in Fine-tuning via Feature Distillation. ArXiv Preprint ArXiv:2205.14141.
Weng, X., Wang, J., Held, D., & Kitani, K. (2020). 3d multi-object tracking: A baseline and new evaluation metrics. 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 10359–10366.
Weng, X., Wang, Y., Man, Y., & Kitani, K. M. (2020). Gnn3dmot: Graph neural network for 3d multi-object tracking with 2d-3d multi-feature learning. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 6499–6508.
Wernecke, J. (1993). The inventor mentor - programming object-oriented 3D graphics with Open Inventor, release 2.
Wilson, B., Qi, W., Agarwal, T., Lambert, J., Singh, J., Khandelwal, S., Pan, B., Kumar, R., Hartnett, A., Pontes, J. K., Ramanan, D., Carr, P., & Hays, J. (2021). Argoverse 2: Next Generation Datasets for Self-driving Perception and Forecasting. Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks (NeurIPS Datasets and Benchmarks 2021).
Wojke, N., & Bewley, A. (2018). Deep Cosine Metric Learning for Person Re-identification. 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), 748–756. https://doi.org/10.1109/WACV.2018.00087
Wojke, N., Bewley, A., & Paulus, D. (2017). Simple online and realtime tracking with a deep association metric. 2017 IEEE International Conference on Image Processing (ICIP), 3645–3649.
Wu, J., Cao, J., Song, L., Wang, Y., Yang, M., & Yuan, J. (2021). Track to detect and segment: An online multi-object tracker. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 12352–12361.
Wu, Y., Lim, J., & Yang, M.-H. (2013). Online object tracking: A benchmark. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2411–2418.
Xian, W., Huang, J.-B., Kopf, J., & Kim, C. (2021). Space-time Neural Irradiance Fields for Free-Viewpoint Video. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 9421–9431.
Xiang, F., Xu, Z., Hasan, M., Hold-Geoffroy, Y., Sunkavalli, K., & Su, H. (2021). Neutex: Neural texture mapping for volumetric neural rendering. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 7119–7128.
Xiang, Y., Choi, W., Lin, Y., & Savarese, S. (2015). Data-driven 3d voxel patterns for object category recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1903–1911.
Xiang, Y., Schmidt, T., Narayanan, V., & Fox, D. (2017). Posecnn: A convolutional neural network for 6d object pose estimation in cluttered scenes. ArXiv Preprint ArXiv:1711.00199.
Xiong, B., Fan, H., Grauman, K., & Feichtenhofer, C. (2021). Multiview pseudo-labeling for semi-supervised learning from video. Proceedings of the IEEE/CVF International Conference on Computer Vision, 7209–7219.
Xu, J., & Wang, X. (2021). Rethinking self-supervised correspondence learning: A video frame-level similarity perspective. Proceedings of the IEEE/CVF International Conference on Computer Vision, 10075–10085.
Xu, Q., Wang, W., Ceylan, D., Mech, R., & Neumann, U. (2019). DISN: Deep Implicit Surface Network for High-quality Single-view 3D Reconstruction.
Xuyang Bai, & Tai, C.-L. (2022). TransFusion: Robust LiDAR-Camera Fusion for 3D Object Detection with Transformers. CVPR.
Yan, X., Yang, J., Yumer, E., Guo, Y., & Lee, H. (2016). Perspective Transformer Nets: Learning Single-View 3D Object Reconstruction without 3D Supervision.
Yang, C., Lamdouar, H., Lu, E., Zisserman, A., & Xie, W. (2021). Self-supervised video object segmentation by motion grouping. Proceedings of the IEEE/CVF International Conference on Computer Vision, 7177–7188.
Yang, G., Manela, J., Happold, M., & Ramanan, D. (2019). Hierarchical deep stereo matching on high-resolution images. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 5515–5524.
Yang, J., Yu, E., Li, Z., Li, X., & Tao, W. (2022). Quality matters: Embracing quality clues for robust 3d multi-object tracking. ArXiv Preprint ArXiv:2208.10976.
Yang, T., & Chan, A. B. (2018). Learning dynamic memory networks for object tracking. Proceedings of the European Conference on Computer Vision (ECCV), 152–167.
Yang, Z., Chai, Y., Anguelov, D., Zhou, Y., Sun, P., Erhan, D., Rafferty, S., & Kretzschmar, H. (2020). SurfelGAN: Synthesizing realistic sensor data for autonomous driving. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 11118–11127.
Yariv, L., Kasten, Y., Moran, D., Galun, M., Atzmon, M., Ronen, B., & Lipman, Y. (2020). Multiview neural surface reconstruction by disentangling geometry and appearance. Advances in Neural Information Processing Systems, 33, 2492–2502.
Yen-Chen, L., Florence, P., Barron, J. T., Rodriguez, A., Isola, P., & Lin, T.-Y. (2021). iNeRF: Inverting neural radiance fields for pose estimation. 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 1323–1330.

References

Yilmaz, A., Javed, O., & Shah, M. (2006). Object tracking: A survey. Acm Computing Surveys (CSUR), 38(4), 13-es.
Yin, T., Zhou, X., & Krahenbuhl, P. (2021). Center-based 3d object detection and tracking. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 11784–11793.
Yoon, J. S., Kim, K., Gallo, O., Park, H. S., & Kautz, J. (2020). Novel View Synthesis of Dynamic Scenes with Globally Coherent Depths from a Monocular Camera. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
Yu, A., Ye, V., Tancik, M., & Kanazawa, A. (2021). pixelnerf: Neural radiance fields from one or few images. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 4578–4587.
Yuan, D., Chang, X., Huang, P.-Y., Liu, Q., & He, Z. (2020). Self-supervised deep correlation tracking. IEEE Transactions on Image Processing, 30, 976–985.
Yuan, W., Lv, Z., Schmidt, T., & Lovegrove, S. (2021). Star: Self-supervised tracking and reconstruction of rigid objects in motion with neural rendering. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 13144–13152.
Zaidi, S. S. A., Ansari, M. S., Aslam, A., Kanwal, N., Asghar, M., & Lee, B. (2022). A survey of modern deep learning based object detection models. Digital Signal Processing, 103514.
Zakharov, S., Kehl, W., Bhargava, A., & Gaidon, A. (2020). Autolabeling 3d objects with differentiable rendering of sdf shape priors. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 12224–12233.
Zbontar, J., Jing, L., Misra, I., LeCun, Y., & Deny, S. (2021). Barlow twins: Self-supervised learning via redundancy reduction. International Conference on Machine Learning, 12310–12320.
Zeng, F., Dong, B., Wang, T., Chen, C., Zhang, X., & Wei, Y. (n.d.). MOTR: End-to-End Multiple-Object Tracking with TRansformer. arXiv 2021. ArXiv Preprint ArXiv:2105.03247.
Zhang, K., Riegler, G., Snavely, N., & Koltun, V. (2020). Nerf++: Analyzing and improving neural radiance fields. ArXiv Preprint ArXiv:2010.07492.
Zhang, R., Isola, P., Efros, A. A., Shechtman, E., & Wang, O. (2018). The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. CVPR.
Zhang, W., Zhou, H., Sun, S., Wang, Z., Shi, J., & Loy, C. C. (2019). Robust multi-modality multi-object tracking. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2365–2374.
Zhang, X., Srinivasan, P. P., Deng, B., Debevec, P., Freeman, W. T., & Barron, J. T. (2021). NeRFactor: Neural Factorization of Shape and Reflectance Under an Unknown Illumination. ArXiv Preprint ArXiv:2106.01970.
Zhang, Y., Wang, C., Wang, X., Zeng, W., & Liu, W. (2022). Robust Multi-Object Tracking by Marginal Inference. ArXiv Preprint ArXiv:2208.03727.
Zhou, Q.-Y., Park, J., & Koltun, V. (2018). Open3D: A modern library for 3D data processing. ArXiv Preprint ArXiv:1801.09847.
Zhou, T., Tucker, R., Flynn, J., Fyffe, G., & Snavely, N. (2018). Stereo Magnification: Learning View Synthesis using Multiplane Images. SIGGRAPH.
Zhou, X., Koltun, V., & Krähenbühl, P. (2020). Tracking objects as points. European Conference on Computer Vision, 474–490.
Zhou, X., Wang, D., & Krähenbühl, P. (2019). Objects as Points. ArXiv Preprint ArXiv:1904.07850.
Zhou, X., Yin, T., Koltun, V., & Krähenbühl, P. (2022). Global tracking transformers. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 8771–8780.
Zhou, Y., & Tuzel, O. (2018). Voxelnet: End-to-end learning for point cloud based 3d object detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 4490–4499.
Zhu, X., Su, W., Lu, L., Li, B., Wang, X., & Dai, J. (2020). Deformable detr: Deformable transformers for end-to-end object detection. ArXiv Preprint ArXiv:2010.04159.