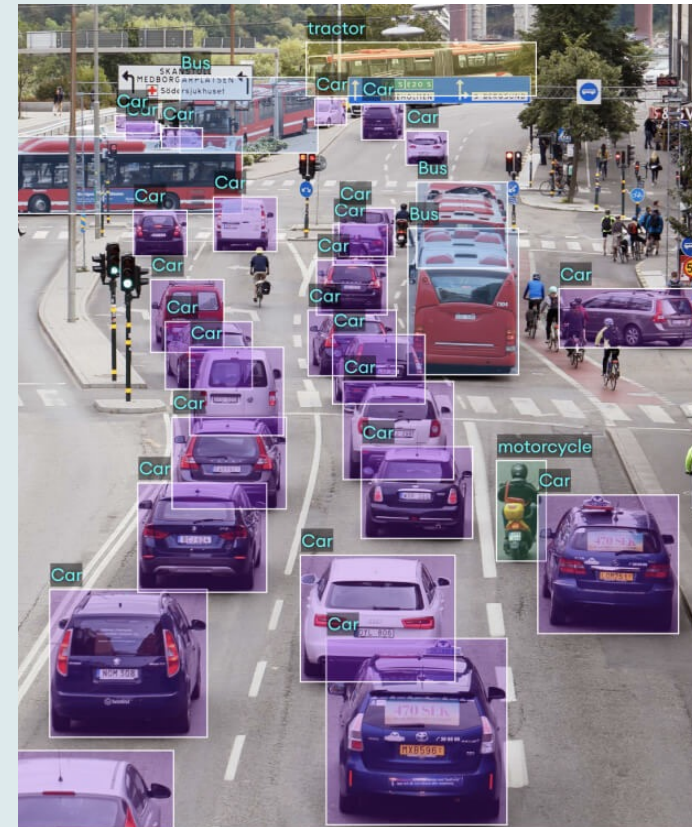


# ادغام بدون نظارت دورین و لیدار برای پیشنهاد ناحیه سریع

ارائه دهنده: بردیا اردکانیان  
استاد راهنما: دکتر مهدی جوانمردی

مهر ۱۴۰۳



دانشکده مهندسی کامپیوتر



دانشگاه صنعتی امیرکبیر  
( پلی تکنیک تهران )

## اهداف ارائه

### ادغام داده‌ها

ادغام دوربین و لیدار  
تأثیر ادغام بر تشخیص  
چالش‌ها و مزایا

### سنسورهای بینایی در خودرو خودران

دوربین  
لیدار  
محدودیت و مزایا

### پیشنهاد ناحیه در تشخیص شیء

تعریف  
تکنیک‌های رایج  
اهمیت پیشنهاد ناحیه

### تشخیص شیء در خودرو خودران

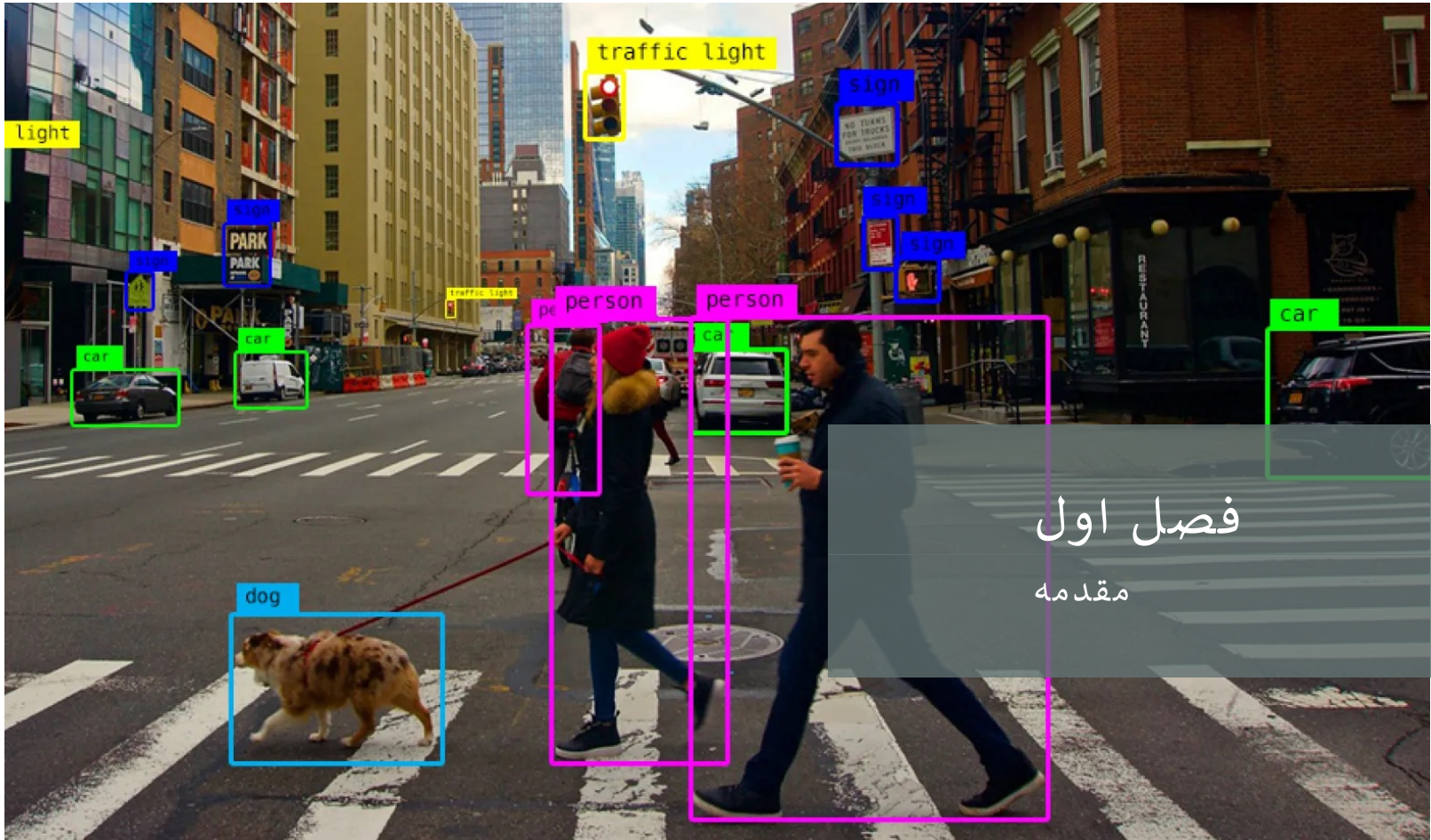
یک مرحله‌ای  
- مدل YOLO  
دو مرحله‌ای  
- مدل RCNN  
- مدل Fast-RCNN  
- مدل Faster-RCNN

### خودرو خودران

تعریف  
اهمیت  
اجزای اصلی

# سیر مطالب





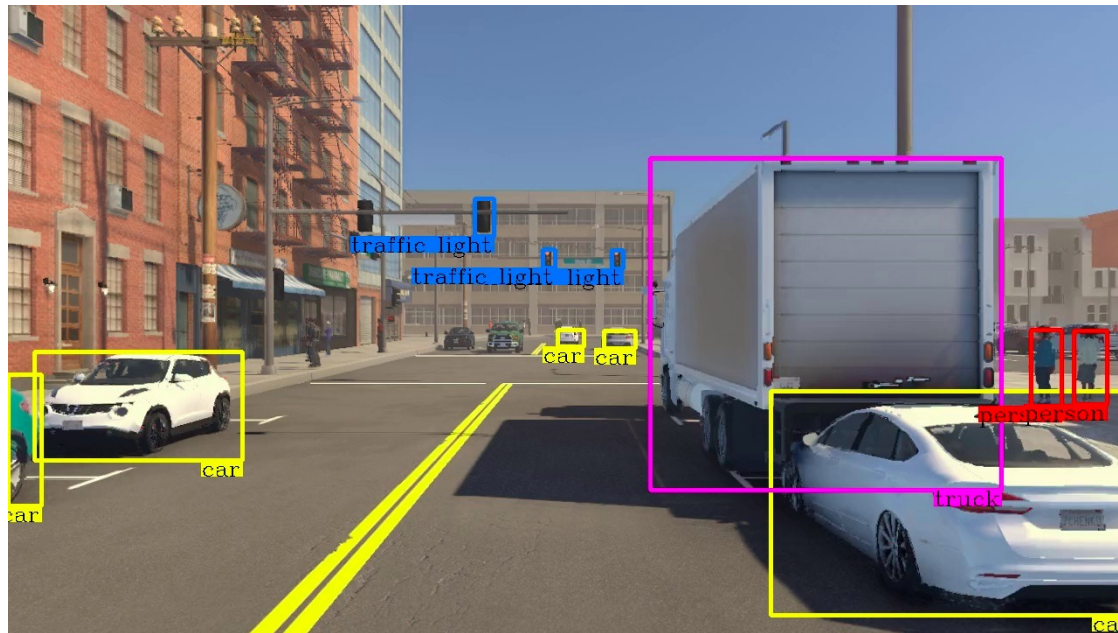
# فصل اول

مقدمه

## مقدمه

### تشخیص شی

تشخیص شیء فرآیندی است که در آن موقعیت و نوع اشیاء در تصاویر یا ویدئوها شناسایی و مشخص می‌شود.



## مقدمه

### انواع تشخیص شیء

#### تشخیص شیء تک مرحله ای

فرآیندی که به صورت مستقیم اشیاء را در یک مرحله تشخیص و مکانیابی می کند.

❖ مدل YOLO

❖ سریع و کارآمد

❖ مناسب برای کاربردهای بلادرنگ

❖ دقت کمتری نسبت به روش های دومرحله ای

❖ مشکلات در تشخیص اشیاء کوچک و نزدیک

#### تشخیص شیء دومرحله ای

فرآیندی که ابتدا نواحی پیشنهادی را شناسایی کرده و سپس اشیاء را در این نواحی تشخیص می دهد.

❖ مدل RCNN، مدل Fast-RCNN، مدل Faster-RCNN

❖ استفاده از شبکه پیشنهاد ناحیه

❖ دقت بالاتر در تشخیص اشیاء، به ویژه در اشیاء کوچک و پیچیده

❖ زمان بر بودن و پیچیدگی بالاتر معماری

❖ نیاز بیشتر به داده های آموزشی



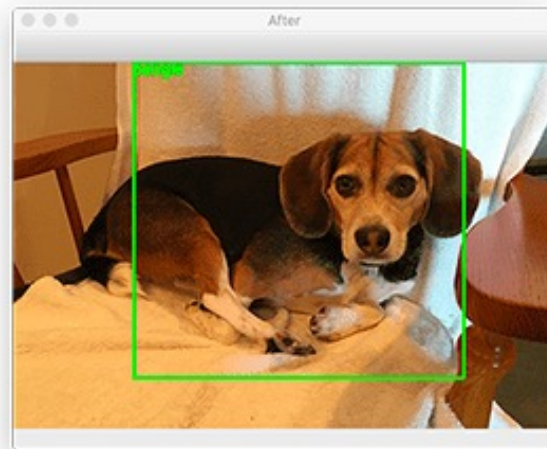
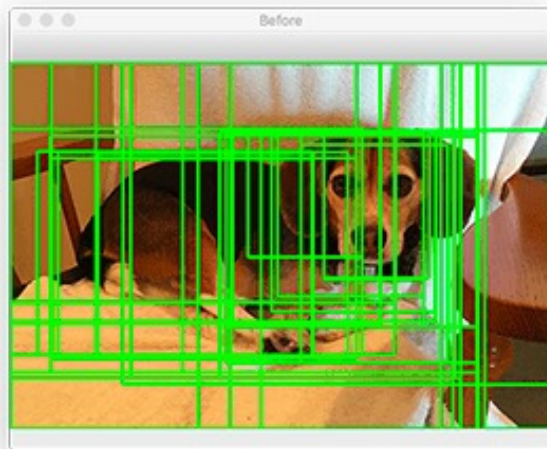
## مقدمه

### اهمیت و تعریف پیشنهاد ناحیه

فرآیندی برای شناسایی و تعیین نواحی محتمل در تصویر که ممکن است اشیاء باشند.

مزایای استفاده از پیشنهاد ناحیه:

- ❖ افزایش دقت
- ❖ بهبود کارایی مدل‌های تشخیص شیء



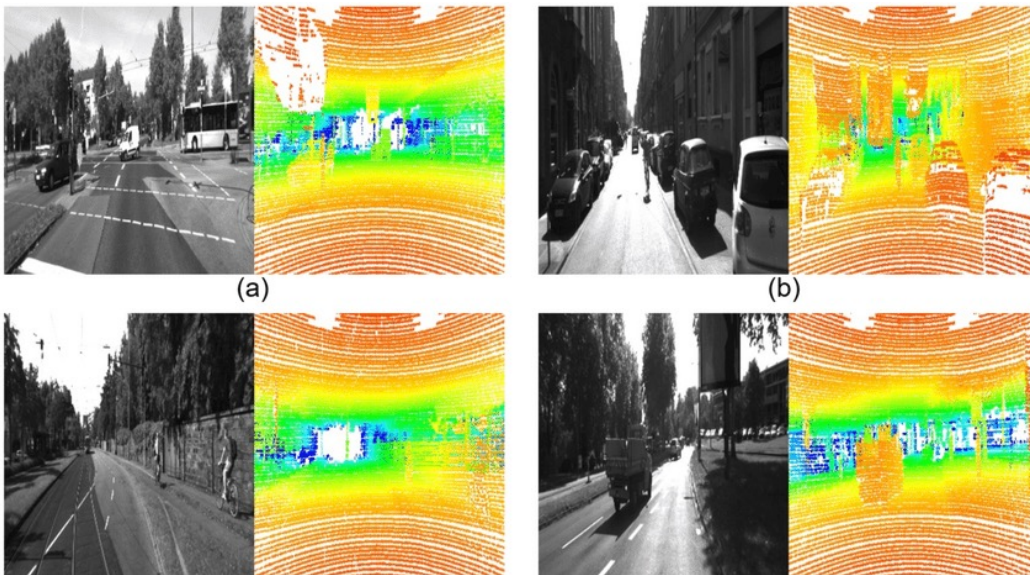
## حسگرهای کلیدی در خودروهای خودران

### دوربین:

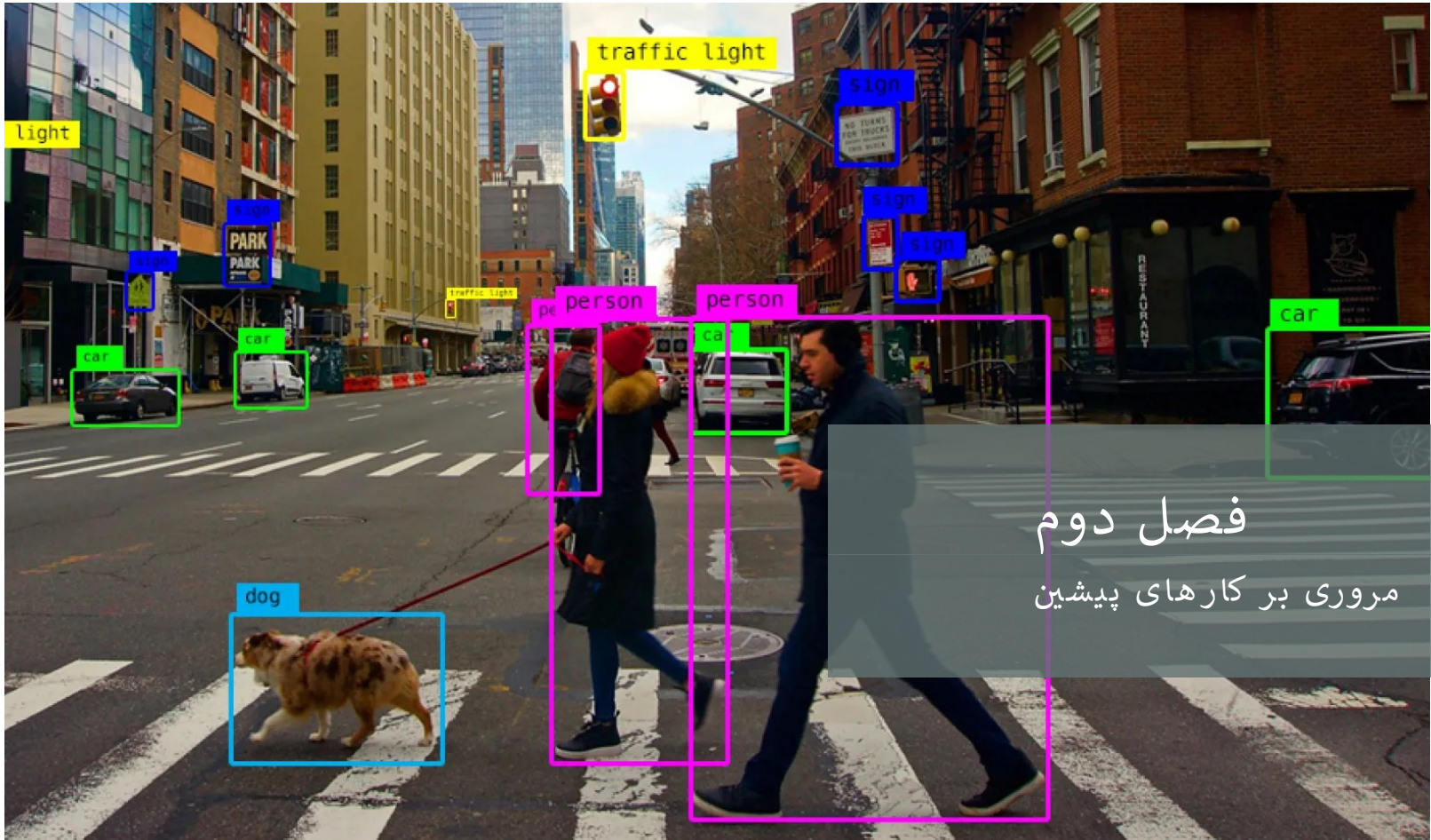
- ❖ فراهم آوری اطلاعات بصری دقیق از محیط
- ❖ هزینه پایین، وضوح بالا، قابلیت تشخیص ویژگی‌های بصری
- ❖ حساس به شرایط نوری، مشکلات در تشخیص عمق

### لیدار:

- ❖ اندازه‌گیری فاصله و ایجاد نقشه‌های سه بعدی از محیط
- ❖ دقت بالا در اندازه‌گیری فاصله، عملکرد در تاریکی
- ❖ هزینه بالا، حساسیت به شرایط آب و هوایی



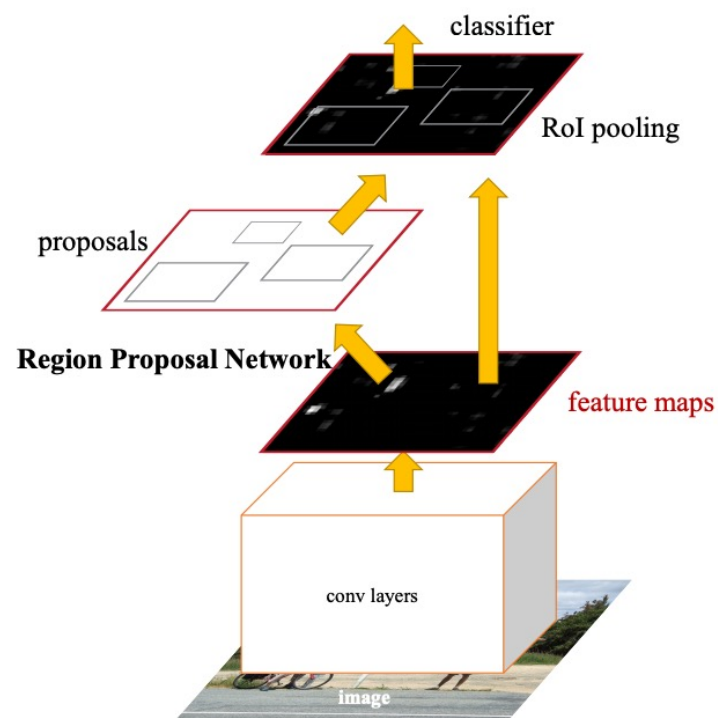




## فصل دوم

مروری بر کارهای پیشین

## تشخیص شیء دو مرحله‌ای

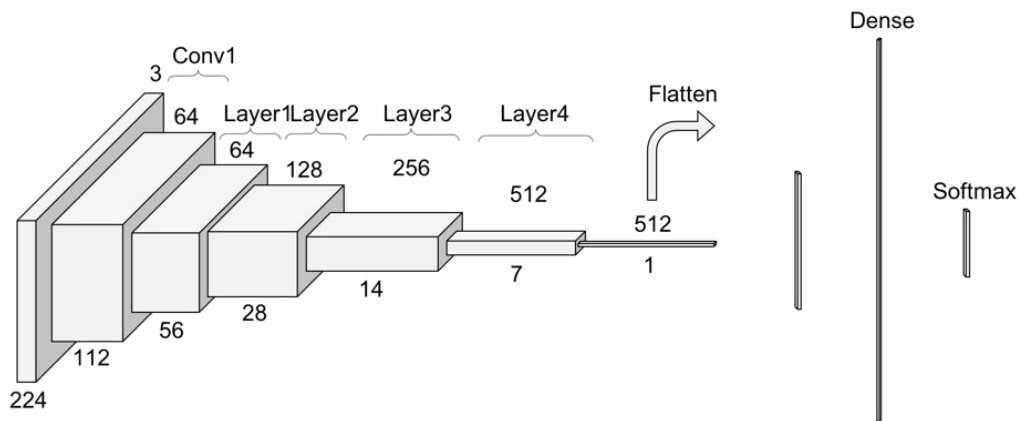


### معماری مدل Faster-RCNN

- ❖ دو مرحله اصلی: پیشنهاد ناحیه و تشخیص/طبقه‌بندی
- ❖ استفاده از ResNet به عنوان ستون فقرات برای استخراج ویژگی‌ها
- ❖ شبکه پیشنهاد ناحیه برای شناسایی نواحی محتمل اشیاء
- ❖ دقت بالاتر در تشخیص اشیاء نسبت به روش تک مرحله‌ای

# استخراج ویژگی تصویر

## استخراج ویژگی با استفاده از مدل ResNet



معماری عمیق و قوی برای استخراج ویژگی های تصویر

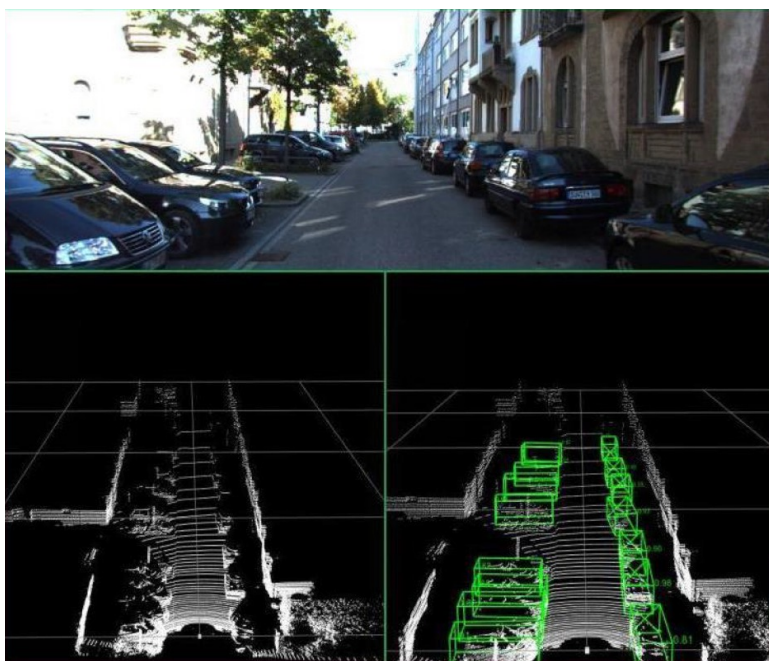
استفاده از اتصالات میانبر برای جلوگیری از مشکل ناپدید شدن گرادیان

به عنوان ستون فقرات در مدل های تشخیص شی مانند Faster-RCNN

استخراج ویژگی های غنی و چند سطحی برای تشخیص دقیق تر اشیاء

بهبود عملکرد در ادغام داده های دوربین و لیدار

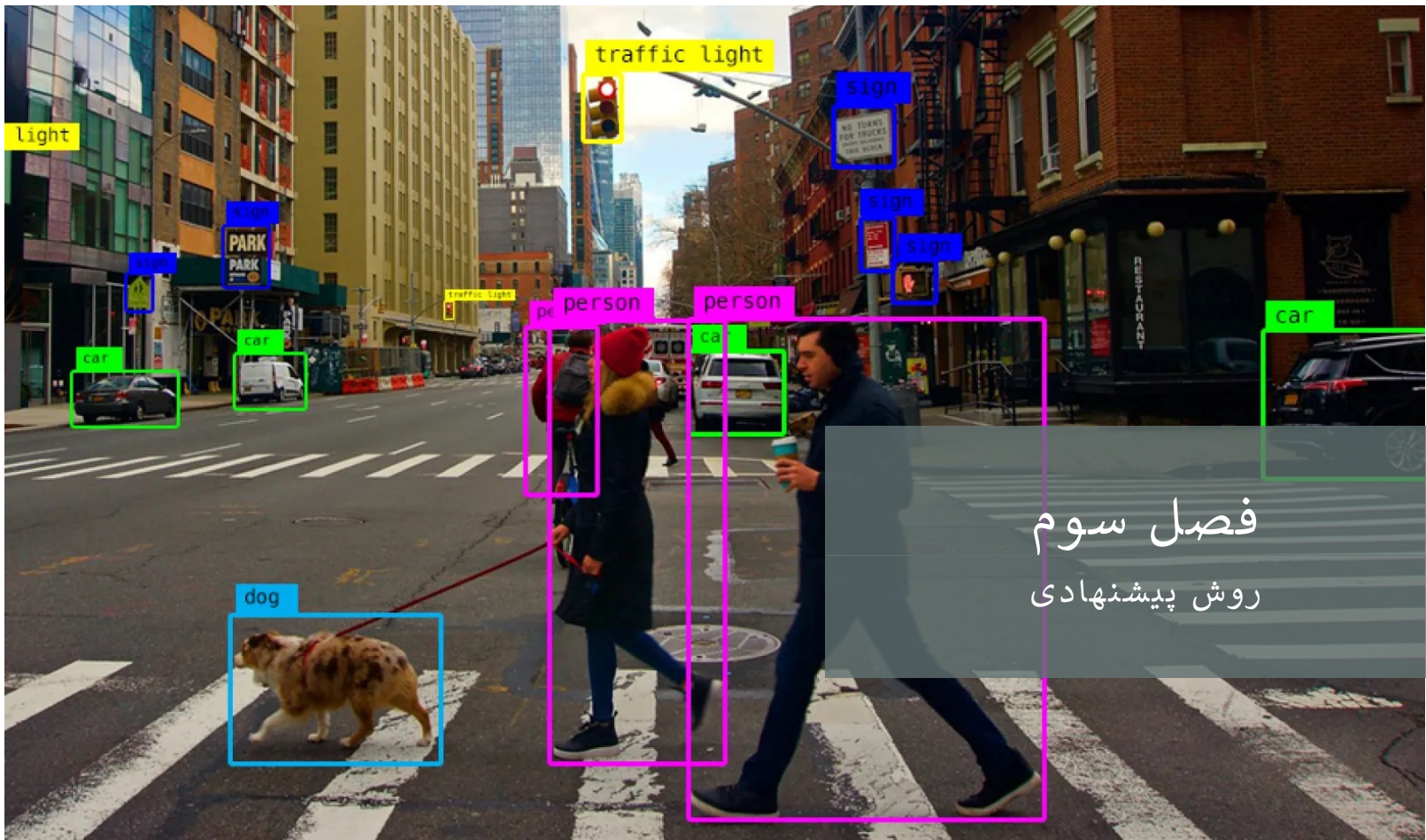
## ادغام داده‌های دوربین و لیدار



### ادغام داده‌های دوربین و لیدار با Deep Fusion

- ❖ ادغام داده‌های دوربین و لیدار برای تشخیص شیء دقیق‌تر
- ❖ ترکیب ویژگی‌های بصری دوربین با داده‌های عمق لیدار
- ❖ استفاده از شبکه عصبی برای هم تراز و ادغام داده‌ها
- ❖ افزایش دقت و قابلیت اطمینان در شرایط نوری و آب و هوایی متفاوت
- ❖ جبران نقاط ضعف هر حسگر با مزایای حسگرهای دیگر
- ❖ بهبود تشخیص اشیاء در سامانه‌های خودران



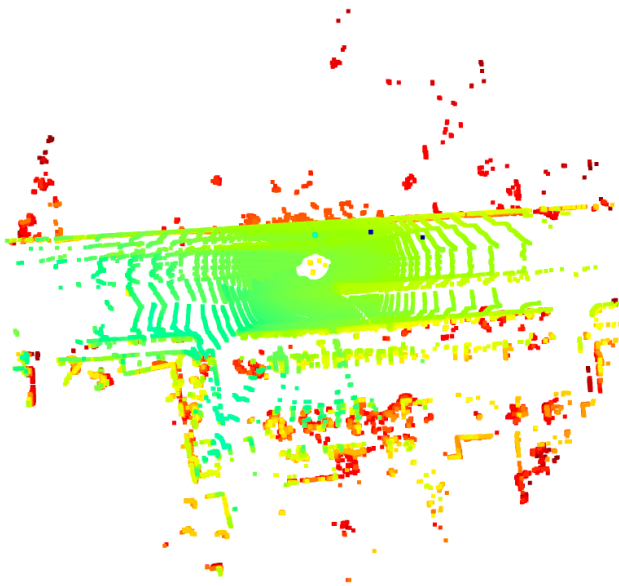


فصل سوم  
روش پیشنهادی





## تبدیل داده‌های خام لیدار به ابر نقاط



داده‌های خام لیدار:

- ❖ شامل مختصات سه‌بعدی نقاط و شدت بازتاب
- ❖ اندازه‌گیری فاصله با استفاده از پرتوهای لیزری

فرآیند تبدیل به ابر نقاط:

- ❖ حذف نویز و نقاط نامربوط
- ❖ تبدیل مختصات لیدار به سیستم مختصات خودرو
- ❖ ایجاد نمای بالا برای تحلیل بهتر محیط

## هم ترازى داده ها با كاليراسيون دوربين

2D Projected Points Overlaid on Image



❖ فیلتر کردن ابر نقاط موجود در تصویر

❖ استفاده از كاليراسيون دوربين برای هم ترازى دقيق

❖ تطبيق مختصات سه بعدى لیدار با دوربين دوبعدى دوربين

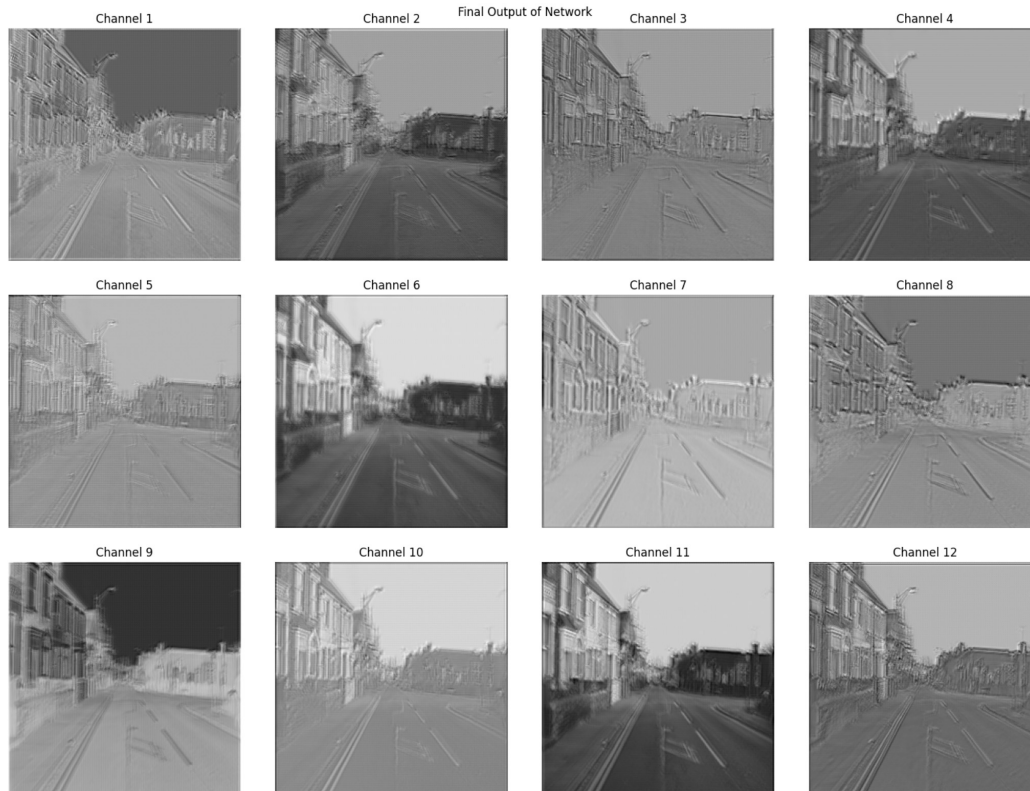
Original Image



2D Projected Points



# استخراج ویژگی های تصویر



- ❖ استفاده از مدل های پایه مانند ResNet و DeepLabv3
- ❖ استخراج ویژگی های چند سطحی و پیچیده از تصاویر دوربین
- ❖ نمونه افزایی ویژگی ها برای هم تراز
- ❖ هماهنگی با ابعاد ابر نقاط لیدار
- ❖ آماده سازی برای ادغام داده ها

## ادغام ویژگی‌های تصویر با ابر نقاط و خوشه بندی

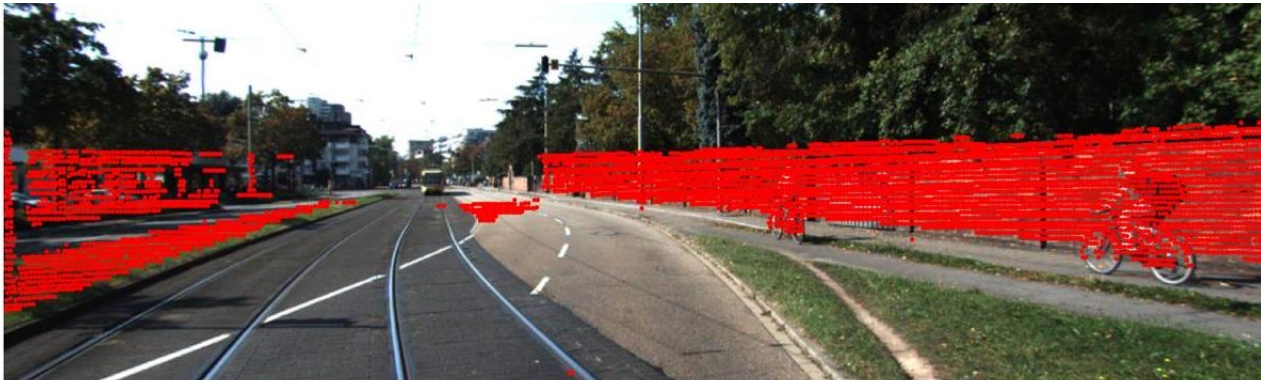


- ❖ ادغام ویژگی‌های استخراج شده از تصویر با ابر نقاط
- ❖ اعمال الگوریتم خوشه‌بندی DBSCAN بر روی داده‌های ادغام شده
- ❖ مشاهده نتایج اولیه خوشه‌بندی
- ❖ شناسایی مشکلات دقت و پراکندگی خوشه‌ها
- ❖ نیاز به بهبود عملکرد خوشه‌بندی
- ❖ عموم ابر نقاط مربوط به زمین می‌باشند
- ❖ نیاز به حذف نقاط زمین از ابر نقاط





## حذف نقاط زمین

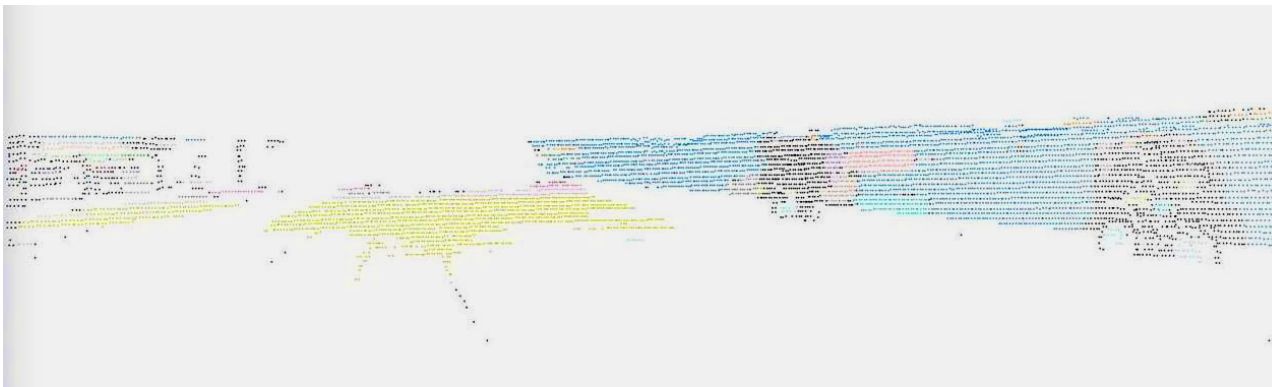


❖ شناسایی و حذف نقاط زمین از ابر نقاط

❖ استفاده از الگوریتم RANSAC برای تخمین صفحه زمین

❖ کاهش نویز و افزایش دقت خوشه بندی

❖ تمرکز بر نقاط مربوط به اشیاء واقعی

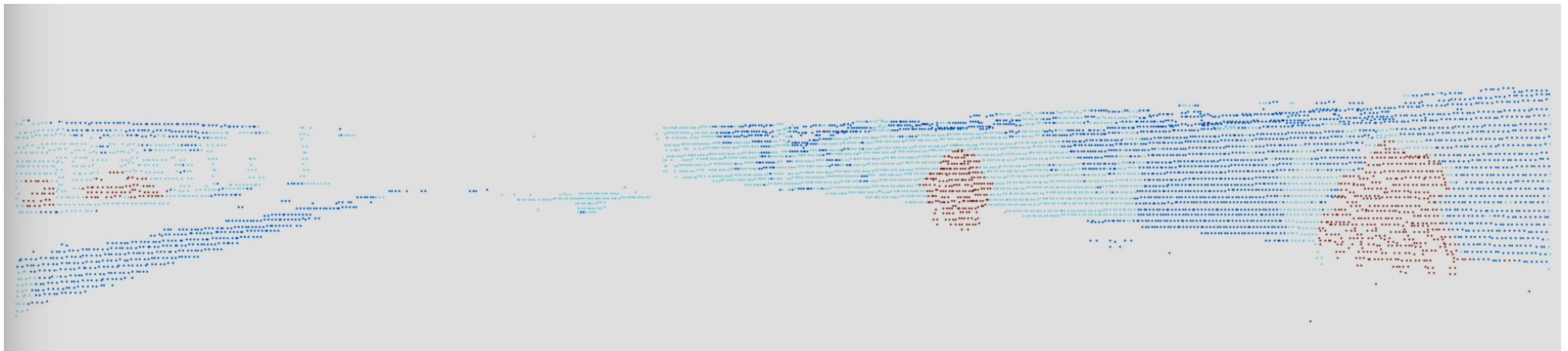






## تنظیم وزن بین ابر نقاط و ویژگی های تصویر

- ❖ استفاده از روش جستجوی شبکه‌ای برای بهینه‌سازی وزن‌ها
- ❖ توازن بین تاثیر ابر نقاط و ویژگی‌ها در خوشه بندی
- ❖ بهبود دقت خوشه‌بندی با وزن‌های بهینه





## خوشه‌بندی با الگوریتم‌های دیگر

- ❖ استفاده از الگوریتم K-Means برای خوشه‌بندی
- ❖ استفاده از الگوریتم Spatial برای خوشه‌بندی
- ❖ عملکرد بهتر الگوریتم K-Means در خوشه‌بندی در مقایسه با DBSCAN
- ❖ عملکرد بهتر الگوریتم Spatial در خوشه‌بندی در مقایسه با دو الگوریتم  
پیشین
- ❖ کندتر بودن الگوریتم Spatial برای خوشه‌بندی
- ❖ انتخاب کردن الگوریتم K-Means به عنوان بهینه‌ترین روش خوشه‌بندی

# خوشبه بندی با الگوریتم های دیگر



K-Means

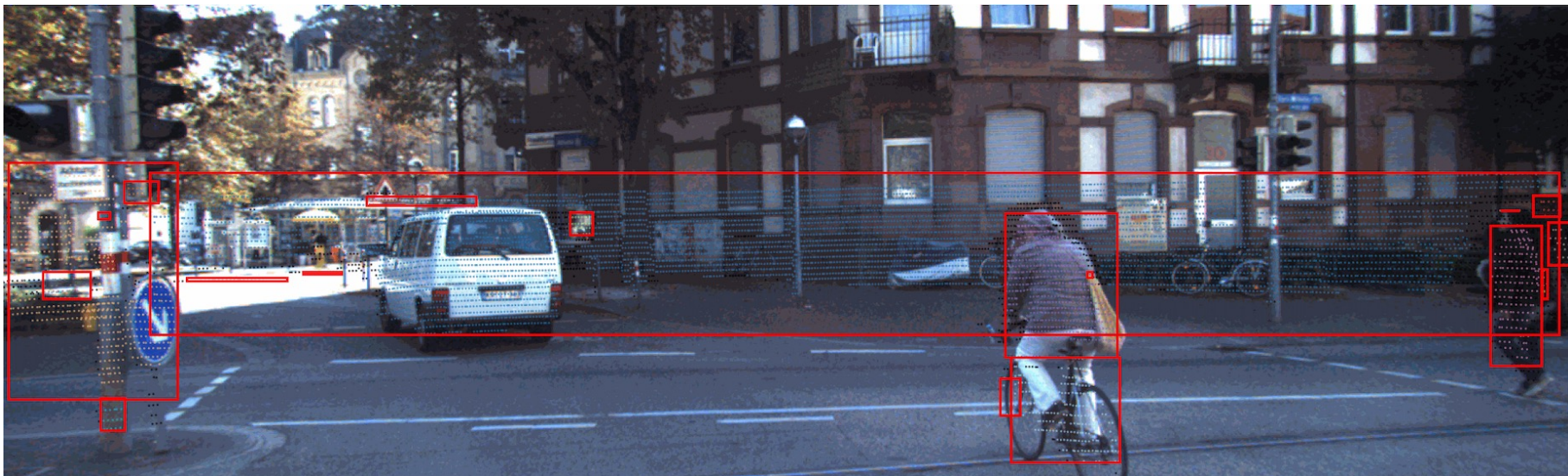


Spatial Clustering

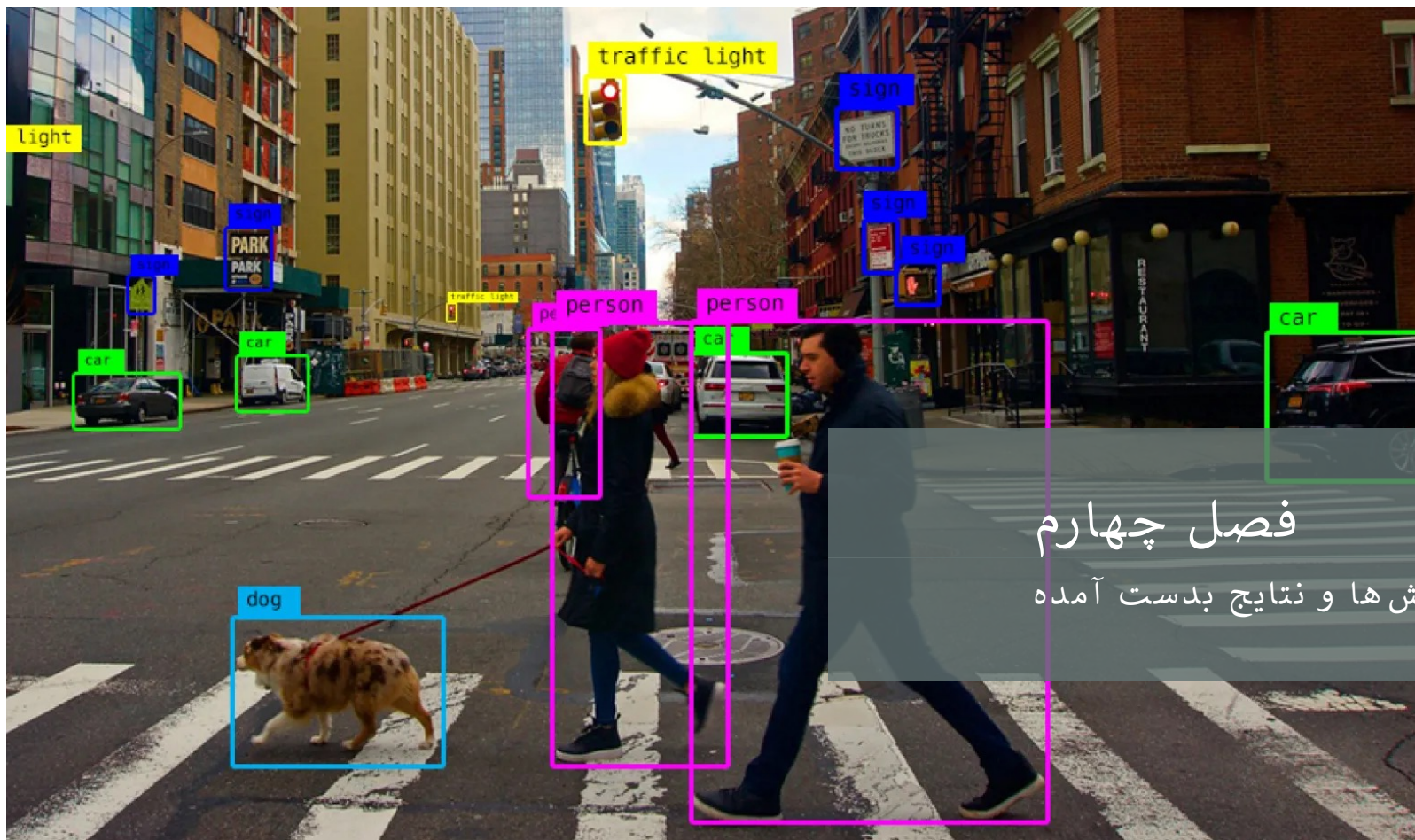


## تبدیل خوشه بندی به پیشنهاد ناحیه

- ❖ تعیین جعبه‌های محدودکننده برای هر خوشه
- ❖ تبدیل جعبه‌های محدودکننده به پیشنهاد ناحیه



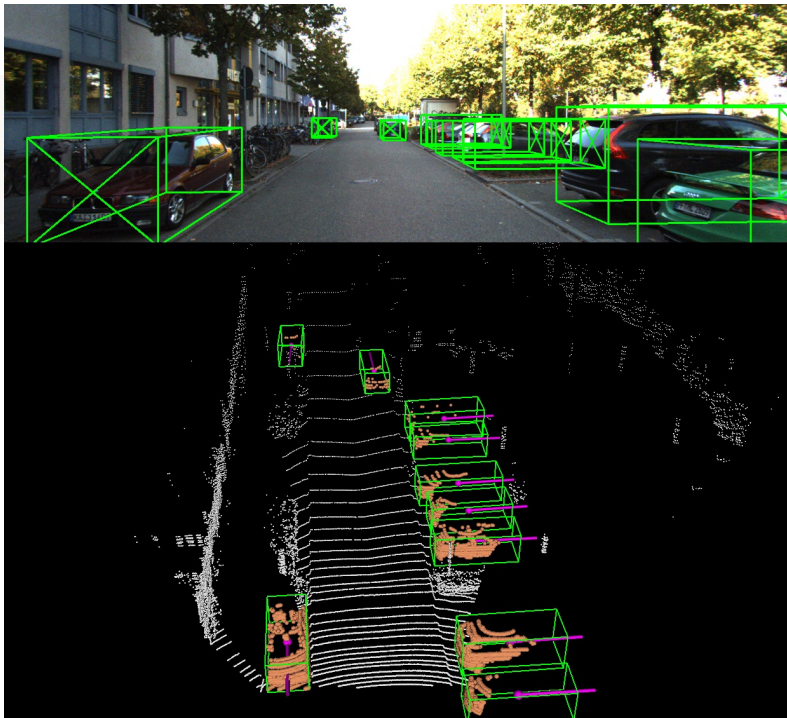




فصل چهارم  
آزمایش‌ها و نتایج بدست آمده

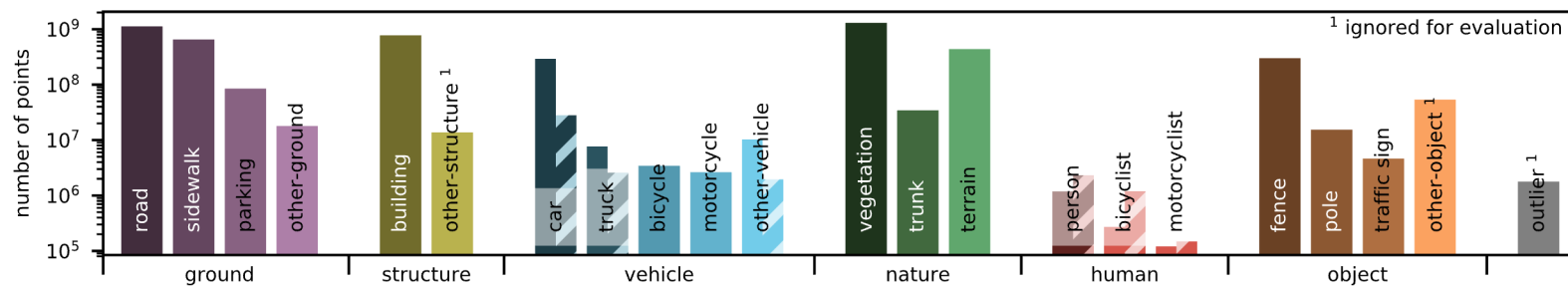


# معرفی دادگان



## مجموعه داده KITTI

- ❖ یکی از معتبرترین منابع کامپیوتر خودران
- ❖ شامل ۲۸ کلاس
- ❖ استفاده از تمامی ۵۰۰۰ تصویر آموزش برای ارزیابی روش پیشنهادی
- ❖ ۵۰۰۰ تصویر برچسب گذاری شده برای آموزش
- ❖ ۲۴۱۸ تصویر آزمایشی بدون برچسب



# معیارهای ارزیابی

## معیارهای ارزیابی پیشنهاد ناحیه

تعداد مثبت واقعی، منفی واقعی، و منفی کاذب

دقت: نسبت مثبت‌های واقعی به مجموع مثبت‌های واقعی و کاذب

بازخوانی: نسبت مثبت‌های واقعی به مجموع مثبت‌های واقعی و

منفی‌های کاذب

IoU: میزان هم‌پوشانی بین جعبه‌های پیشنهادی و واقعی

زمان اجرا

## معیارهای ارزیابی خوشه‌بندی

NMI: ارزیابی کیفیت خوشه‌بندی با استفاده از مدل SAM به عنوان

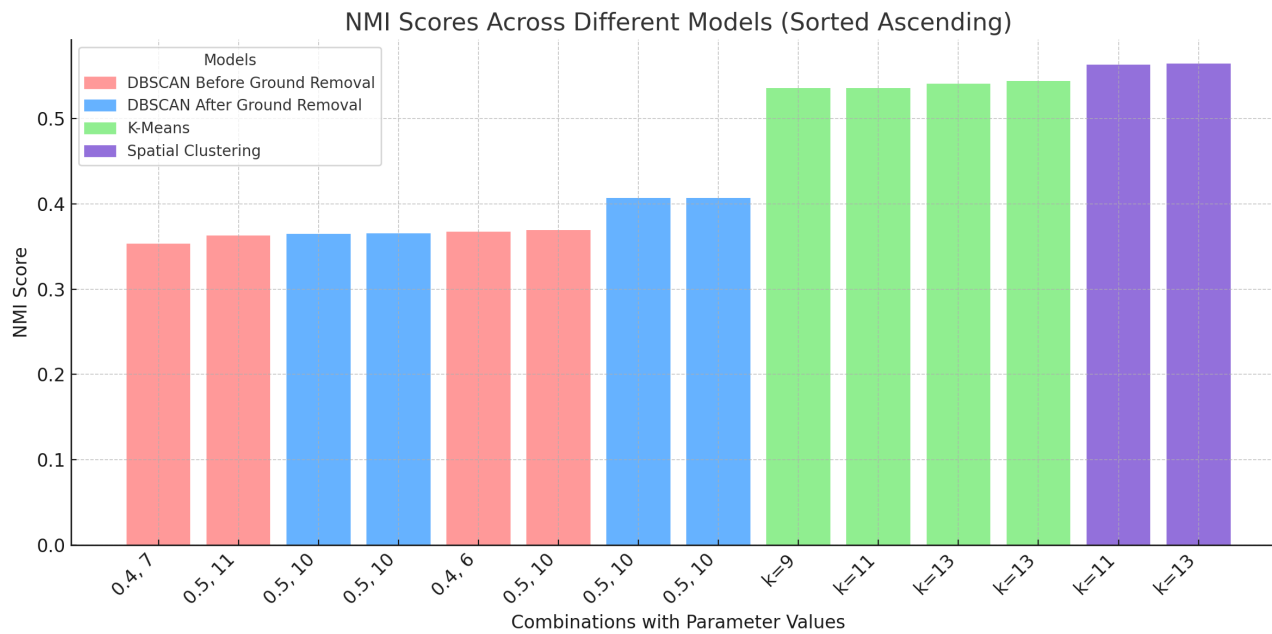
داده واقعی

استفاده از مدل SAM برای بدست آوردن تقسیم‌بندی معنایی

مقایسه خوشه‌بندی انجام شده با تقسیم‌بندی معنایی مدل SAM

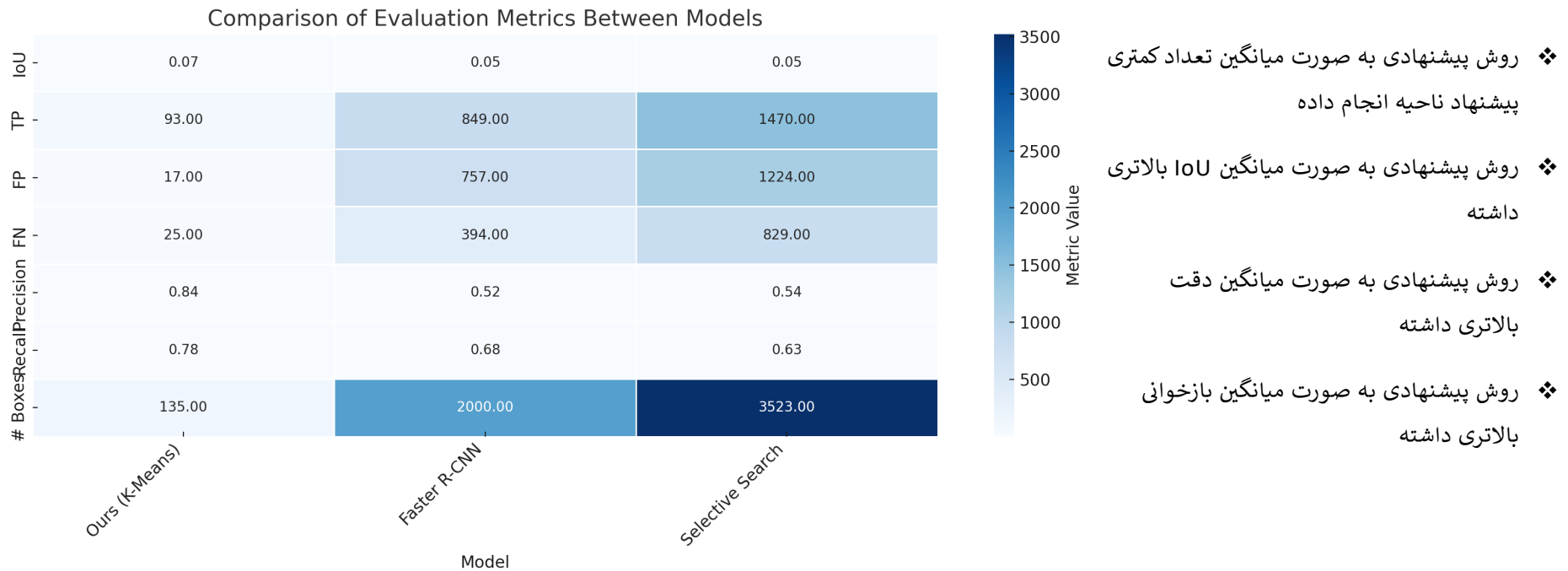


## ارزیابی خوشه بندی با جستجوی شبکه ای



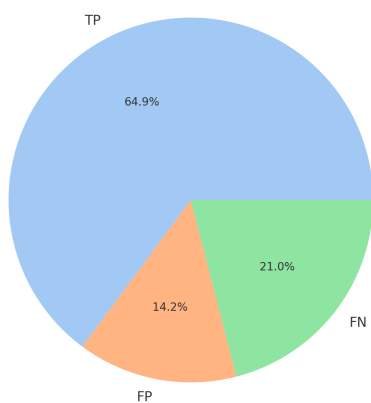
- ❖ مدل بهترین عملکرد را نشان داد Spatial Clustering
- ❖ مدل Spatial Clustering پنج برابر از K-Means کند تر بود
- ❖ مدل بهترین تعادل را بین عملکرد و زمان اجرا نشان داد K-Means
- ❖ مدل DBSCAN ضعیف ترین عملکرد را بین دیگر مدل ها نشان داد

## مقایسه روش پیشنهادی با مدل های پایه - میانگین

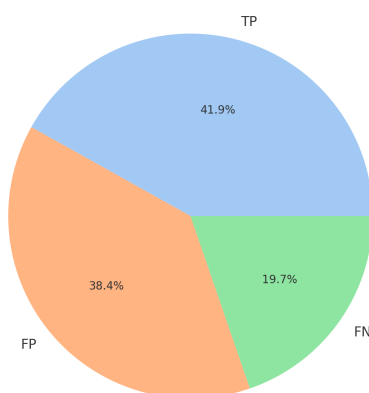


## مقایسه روش پیشنهادی با مدل های پایه - تجمعی

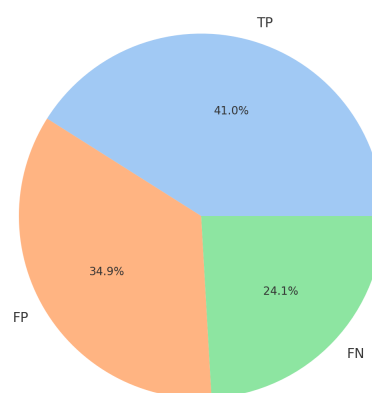
Ours (K-Means) Proposals Breakdown



Faster R-CNN Proposals Breakdown



Selective Search Proposals Breakdown



Model	Total Proposals	TP	FP	FN	FP + FN
Ours (K-Means)	685137	10058	2197	3253	5450
Faster R-CNN	10000000	10071	9227	4740	13967
Selective Search	17612358	9332	7932	5479	13411

❖ به طور کلی ۱۴۸۱۱ شیء در دادگان وجود دارد

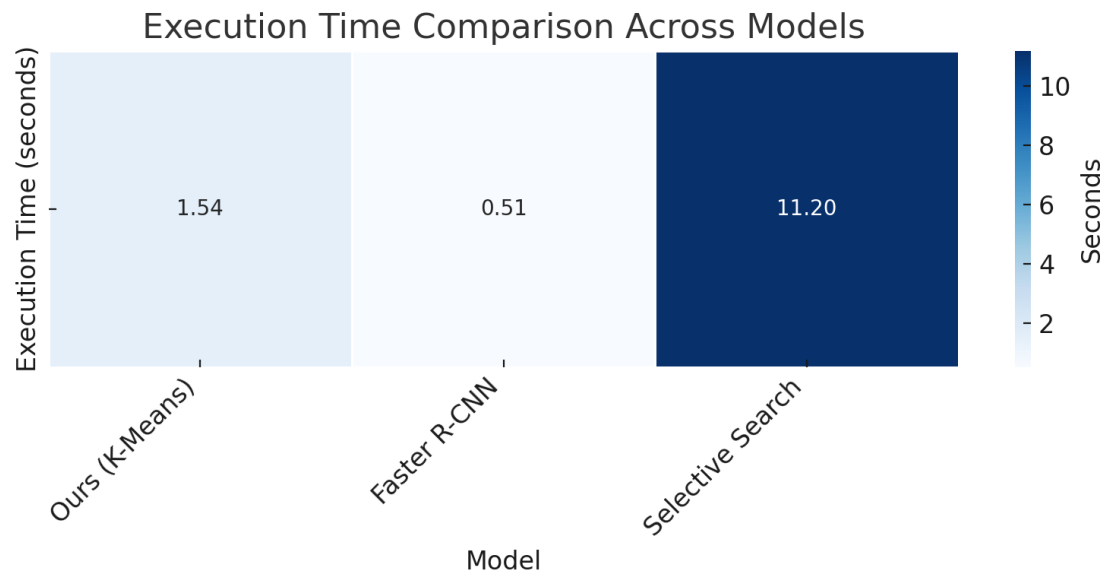
❖ روش پیشنهادی نسبت مثبت واقعی مشابه ای به مدل های پایه دارد.

❖ روش پیشنهادی نسبت منفی کاذب و منفی واقعی کمتری به مدل های پایه دارد

❖ مدل پیشنهادی تعداد پیشنهاد ناحیه کمتری در مقایسه با مدل های پایه دارد



## مقایسه روش پیشنهادی با مدل های پایه – زمان اجرا



- ❖ مدل پیشنهادی زمان اجرا کند تری نسبت به مدل Faster-RCNN دارد
- ❖ مدل پیشنهادی زمان اجرا بسیار سریع تری نسبت به مدل Selective Search دارد

## جمع بندی

- ❖ تعریف مسئله پیشنهاد ناحیه و چالش‌های آن
- ❖ مروری بر مسائل مرتبط و بررسی نقاط قوت و ضعف هرکدام
- ❖ ارائه مدل‌های پیشنهادی جهت ارزیابی کارایی ادغام بدون نظارت در پیشنهاد ناحیه
- ❖ مقایسه مدل‌های ارائه شده با مدل‌های پایه



آموزش مدل تشخیص شیء با روش پیشنهادی  
❖ نظیر آموزش مدل Faster-RCNN با پیشنهاد ناحیه پیشنهاد  
شده

استفاده از تکمیل عمق برای بهبود کیفیت پیشنهادات ناحیه  
❖ نظیر مدل Depth Completion و Sparse2Dense

## پیشنهادها

- 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. (2016). Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6), 1137–1149.
- 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.
- Meschtscherjakov, A., Tscheligi, M., Pfleging, B., Borojeni, S. S., Ju, W., Palanque, P., ... & Kun, A. L. (2018). Interacting with autonomous vehicles: Learning from other domains. *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems (CHI EA '18)*, 1–8.
- Pomerleau, D. (1989). Alvin: An autonomous land vehicle in a neural network. *Proceedings of Neural Information Processing Systems*, 305–313.
- Thrun, S., Montemerlo, M., Dahlkamp, H., Stavens, D., Aron, A., Diebel, J., ... & Mahoney, P. (2006). Stanley: The robot that won the DARPA grand challenge. *Journal of Field Robotics*, 23(9), 661–692.
- Levinson, J., Askeland, J., Becker, J., Dolson, J., Held, D., Kammel, S., ... & Thrun, S. (2011). Towards fully autonomous driving: Systems and algorithms. *2011 IEEE Intelligent Vehicles Symposium*, 163–168.
- Liu, L., Ouyang, W., Wang, X., Fieguth, P., Chen, J., Liu, X., & Pietikäinen, M. (2020). Deep learning for generic object detection: A survey. *International Journal of Computer Vision*, 128(2), 261–318.

- Viola, P., & Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1, 1–511–1–518.
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6), 84–90.
- Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016). SSD: Single shot multibox detector. In B. Leibe, J. Matas, N. Sebe, & M. Welling (Eds.), *Computer Vision – ECCV 2016* (pp. 21–37). Cham, Switzerland: Springer International Publishing.
- Uijlings, J. R. R., van de Sande, K. E. A., Gevers, T., & Smeulders, A. W. M. (2013). Selective search for object recognition. *International Journal of Computer Vision*, 104(2), 154–171.
- Zitnick, C. L., & Dollár, P. (2014). Edge boxes: Locating object proposals from edges. In D. Fleet, T. Pajdla, B. Schiele, & T. Tuytelaars (Eds.), *Computer Vision – ECCV 2014* (pp. 391–405). Cham, Switzerland: Springer International Publishing.
- Girshick, R. (2015). Fast R-CNN. *Proceedings of the IEEE International Conference on Computer Vision*.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. In I. Guyon, U. von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems* (Vol. 30).
- Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., & Zagoruyko, S. (2020). End-to-end object detection with transformers. In A. Vedaldi, H. Bischof, T. Brox, & J.-M. Frahm (Eds.), *Computer Vision – ECCV 2020* (pp. 213–229). Cham, Switzerland: Springer International Publishing.



Liu, S., Li, F., Zhang, H., Yang, X., Qi, X., Su, H., & Zhang, L. (2022). DAB-DETR: Dynamic anchor boxes are better queries for DETR. *International Conference on Learning Representations*.

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278–2324.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 770–778).

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *Clinical Orthopaedics and Related Research*, abs/1409.1556.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In N. Navab, J. Hornegger, W. M. Wells, & A. F. Frangi (Eds.), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* (pp. 234–241). Cham, Switzerland: Springer International Publishing.

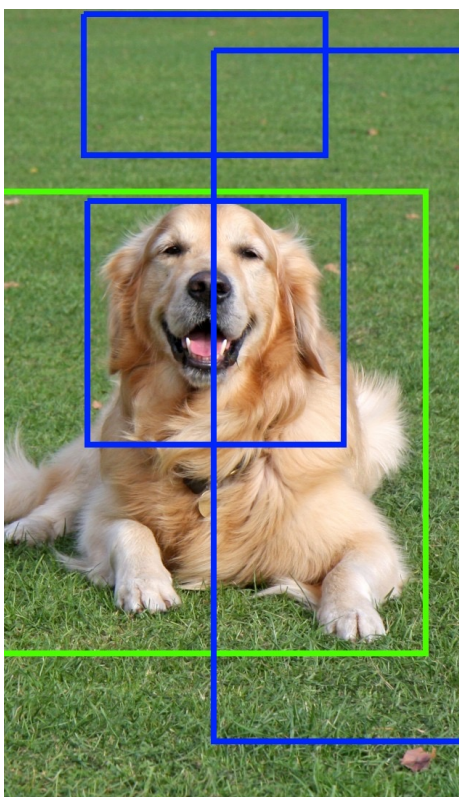
Chen, L.-C., Papandreou, G., Schroff, F., & Adam, H. (2017). Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587*.

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. In V. Ferrari, M. Hebert, C. Sminchisescu, & Y. W. M. Weiss (Eds.), *Computer Vision – ECCV 2018* (pp. 833–851). Cham, Switzerland: Springer International Publishing.

Tian, H., Chen, Y., Dai, J., Zhang, Z., & Zhu, X. (2021). Unsupervised object detection with lidar clues. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5958–5968.

- Zhang, L., Yang, A. J. J., Xiong, Y., Casas, S., Yang, B., Ren, M., & Urtasun, R. (2023). Towards unsupervised object detection from lidar point clouds. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 9317–9328). Vancouver, Canada: IEEE.
- Bai, X., Hu, Z., Zhu, X., Huang, Q., Chen, Y., Fu, H., & Tai, C.-L. (2022). Transfusion: Robust lidar-camera fusion for 3D object detection with transformers. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 1080–1089).
- Li, Y., Yu, A. W., Meng, T., Caine, B., Ngiam, J., Peng, D., ... & Tan, M. (2022). Deepfusion: Lidar-camera deep fusion for multi-modal 3D object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 17161–17170). New Orleans, USA: IEEE.
- Liu, L., He, J., Ren, K., Xiao, Z., & Hou, Y. (2022). A lidar-camera fusion 3D object detection algorithm. *Information*, 13(4).
- Geiger, A., Lenz, P., Stiller, C., & Urtasun, R. (2013). Vision meets robotics: The KITTI dataset. *International Journal of Robotics Research*.
- Fischler, M. A., & Bolles, R. C. (1981). Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6), 381–395.
- Bergstra, J., Bardenet, R., Bengio, Y., & Kégl, B. (2011). Algorithms for hyper-parameter optimization. In J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira, & K. Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems* (Vol. 24).
- Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., ... & Girshick, R. B. (2023). Segment anything. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 3992–4003.

- Strehl, A., & Ghosh, J. (2002). Cluster ensembles – a knowledge reuse framework for combining multiple partitions. *Journal of Machine Learning Research*, 3, 583–617.
- Ma, F., Cavalheiro, G. V., & Karaman, S. (2019). Self-supervised sparse-to-dense: Self-supervised depth completion from lidar and monocular camera. *Proceedings of the International Conference on Robotics and Automation*, 3288–3295.
- Uhrig, J., Schneider, N., Schneider, L., Franke, U., Brox, T., & Geiger, A. (2017). Sparsity invariant CNNs. In *Proceedings of the International Conference on 3D Vision* (pp. 11–20).
- Comaniciu, D., & Meer, P. (2002). Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5), 603–619.
- Frey, B. J., & Dueck, D. (2007). Clustering by passing messages between data points. *Science*, 315(5814), 972–976.
- Lin, T.-Y., Maire, M., Belongie, S. J., Hays, J., Perona, P., Ramanan, D., Dollár, P., & Zitnick, C. L. (2014). Microsoft COCO: Common objects in context. In *Proceedings of the European Conference on Computer Vision*.



با تشکر از توجه شما

Email: [bardia.ardakanian@gmail.com](mailto:bardia.ardakanian@gmail.com)

