# Papers

## Essential

1. [Cognitive Mapping and Planning for Visual Navigation](https://arxiv.org/pdf/1702.03920.pdf) \*
2. [Camera-based vehicle velocity estimation from monocular video](https://arxiv.org/pdf/1802.07094.pdf) \*
3. [Vision meets Robotics: The KITTI Dataset](http://www.cvlibs.net/publications/Geiger2013IJRR.pdf) \*
4. [Microsoft COCO: Common Objects in Context](https://arxiv.org/pdf/1405.0312.pdf) \*
5. [COCO-Stuff: Thing and Stuff Classes in Context](https://arxiv.org/pdf/1612.03716.pdf) \*
6. [Machine Learning Methods for Solving Assignment Problems in Multi-Target Tracking](https://arxiv.org/pdf/1802.06897.pdf) \*
7. [Size to Depth: A New Perspective for Single Image Estimation](https://arxiv.org/pdf/1801.04461.pdf) \*
8. [RTSEG: Real-time Semantic Segmentation Comparative Study](https://arxiv.org/pdf/1803.02758.pdf) \*
9. [MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications](https://arxiv.org/pdf/1704.04861.pdf)
10. [SSD: Single Shot MultiBox Detector](https://arxiv.org/pdf/1512.02325.pdf)
11. [Translating Videos to Natural Language Using Deep Recurrent Neural Networks](https://arxiv.org/pdf/1412.4729.pdf)
12. [VizWiz Grand Challenge: Answering Visual Questions from Blind People](https://arxiv.org/pdf/1802.08218.pdf)
13. [How would surround vehicles move? A Unified Framework for Maneuver Classification and Motion Prediction](https://arxiv.org/pdf/1801.06523.pdf)
14. [Fusion of stereo and still monocular depth estimates in a self-supervised learning context](https://arxiv.org/pdf/1803.07512.pdf)
15. [Explicit Reasoning over End-to-End Neural Architectures for Visual Question Answering](https://arxiv.org/pdf/1803.08896.pdf)
16. [Particle-based pedestrian path prediction using LSTM-MDL models](https://arxiv.org/pdf/1804.05546.pdf)

## Computer Vision

1. [A General Pipeline for 3D Detection of Vehicles](https://arxiv.org/pdf/1803.00387.pdf)
2. [Deep-6DPose: Recovering 6D Object Pose from a Single RGB Image](https://arxiv.org/pdf/1802.10367.pdf)
3. [Mono-Camera 3D Multi-Object Tracking Using Deep Learning Detections and PMBM Filtering](https://arxiv.org/pdf/1802.09975.pdf)
4. [Learning Image Conditioned Label Space for Multilabel Classification](https://arxiv.org/pdf/1802.07460.pdf)
5. [Real-Time Dense Stereo Matching with ELAS on FPGA Accelerated Embedded Devices](https://arxiv.org/pdf/1802.07210.pdf)
6. [Structured Label Inference for Visual Understanding](https://arxiv.org/pdf/1802.06459.pdf)
7. [Tiny SSD: A Tiny Single-shot Detection Deep Convolutional Neural Network for Real-time Embedded Object Detection](https://arxiv.org/pdf/1802.06488.pdf)
8. [Structured Label Inference for Visual Understanding](https://arxiv.org/pdf/1802.06459.pdf)
9. [Learning to Count Objects in Natural Images for Visual Question Answering](https://arxiv.org/pdf/1802.05766.pdf)
10. [Unsupervised Learning of Depth and Ego-Motion from Monocular Video Using 3D Geometric Constraints](https://arxiv.org/pdf/1802.05522.pdf)
11. [Joint 3D Reconstruction of a Static Scene and Moving Objects](https://arxiv.org/pdf/1802.04738.pdf)
12. [Answerer in Questioner’s Mind for Goal-Oriented Visual Dialogue](https://arxiv.org/pdf/1802.03881.pdf)
13. [TSViz: Demystification of Deep Learning Models for Time-Series Analysis](https://arxiv.org/pdf/1802.02952.pdf)
14. [Tracking Multiple Moving Objects Using Unscented Kalman Filtering Techniques](https://arxiv.org/ftp/arxiv/papers/1802/1802.01235.pdf)
15. [Explaining First Impressions: Modeling, Recognizing, and Explaining Apparent Personality from Videos](https://arxiv.org/pdf/1802.00745.pdf)
16. [Dual Recurrent Attention Units for Visual Question Answering](https://arxiv.org/pdf/1802.00209.pdf)
17. [Parallel Tracking and Verifying](https://arxiv.org/pdf/1801.10496.pdf)
18. [Object-based reasoning in VQA](https://arxiv.org/pdf/1801.09718.pdf)
19. [Object Detection in Videos by Short and Long Range Object Linking](https://arxiv.org/pdf/1801.09823.pdf)
20. [Open3D: A Modern Library for 3D Data Processing](https://arxiv.org/pdf/1801.09847.pdf)
21. [Image Captioning at Will: A Versatile Scheme for Effectively Injecting Sentiments into Image Descriptions](https://arxiv.org/pdf/1801.10121.pdf)
22. [Improving Multiple Object Tracking with Optical Flow and Edge Preprocessing](https://arxiv.org/pdf/1801.09646.pdf)
23. [Structured Triplet Learning with POS-tag Guided Attention for Visual Question Answering](https://arxiv.org/pdf/1801.07853.pdf)
24. [The challenge of simultaneous object detection and pose estimation: a comparative study](https://arxiv.org/pdf/1801.08110.pdf)
25. [What Makes Good Synthetic Training Data for Learning Disparity and Optical Flow Estimation?](https://arxiv.org/pdf/1801.06397.pdf)
26. [Monocular Depth Estimation using Multi-Scale Continuous CRFs as Sequential Deep Networks](https://arxiv.org/pdf/1803.00891.pdf)
27. [Single View Stereo Matching](https://arxiv.org/pdf/1803.02612.pdf)
28. [Intentions of Vulnerable Road Users – Detection and Forecasting by Means of Machine Learning](https://arxiv.org/pdf/1803.03577.pdf)
29. [Indoor Scene Understanding in 2.5/3D: A Survey](https://arxiv.org/pdf/1803.03352.pdf)
30. [Unsupervised Learning of Monocular Depth Estimation and Visual Odometry with Deep Feature Reconstruction](https://arxiv.org/pdf/1803.03893.pdf)
31. [Transparency by Design: Closing the Gap Between Performance and Interpretability in Visual Reasoning](https://arxiv.org/pdf/1803.05268.pdf)
32. [Vision-Aided Absolute Trajectory Estimation Using an Unsupervised Deep Network with Online Error Correction](https://arxiv.org/pdf/1803.05850.pdf)
33. [Complex-YOLO: An Euler-Region-Proposal for Real-time 3D Object Detection on Point Clouds](https://arxiv.org/pdf/1803.06199.pdf)
34. [Monocular Fisheye Camera Depth Estimation Using Semi-supervised Sparse Velodyne Data](https://arxiv.org/pdf/1803.06192.pdf)
35. [Object Captioning and Retrieval with Natural Language](https://arxiv.org/pdf/1803.06152.pdf)
36. [Live Target Detection with Deep Learning Neural Network and Unmanned Aerial Vehicle on Android Mobile Device](https://arxiv.org/ftp/arxiv/papers/1803/1803.07015.pdf)
37. [VQA-E: Explaining, Elaborating, and Enhancing Your Answers for Visual Questions](https://arxiv.org/pdf/1803.07464.pdf)
38. [Text Detection and Recognition in images: A survey](https://arxiv.org/ftp/arxiv/papers/1803/1803.07278.pdf)
39. [Monocular Depth Estimation by Learning from Heterogeneous Datasets](https://arxiv.org/pdf/1803.08018.pdf)
40. [End-to-End Video Captioning with Multitask Reinforcement Learning](https://arxiv.org/pdf/1803.07950.pdf)
41. [Fast Semantic Segmentation on Video Using Motion Vector-Based Feature Interpolation](https://arxiv.org/pdf/1803.07742.pdf)
42. [Context Encoding for Semantic Segmentation](https://arxiv.org/pdf/1803.08904.pdf)
43. [Learning Driving Models with a Surround-View Camera System and a Route Planner](https://arxiv.org/pdf/1803.10158.pdf)
44. [Learning Depth from Single Images with Deep Neural Network Embedding Focal Length](https://arxiv.org/pdf/1803.10039.pdf)
45. [Iterative Visual Reasoning Beyond Convolutions](https://arxiv.org/pdf/1803.11189.pdf)
46. [Two can play this Game: Visual Dialog with Discriminative Question Generation](https://arxiv.org/pdf/1803.11186.pdf)
47. [Reconstruction Network for Video Captioning](https://arxiv.org/pdf/1803.11438.pdf)
48. [MegaDepth: Learning Single-View Depth Prediction from Internet Photos](https://arxiv.org/pdf/1804.00607.pdf)
49. [Transferable Pedestrian Motion Prediction Models at Intersections](https://arxiv.org/pdf/1804.00495.pdf)
50. [Question Type Guided Attention in Visual Question Answering](https://arxiv.org/pdf/1804.02088.pdf)
51. [Monocular Semantic Occupancy Grid Mapping with Convolutional Variational Auto-Encoders](https://arxiv.org/pdf/1804.02176.pdf)
52. [Blazingly Fast Video Object Segmentation with Pixel-Wise Metric Learning](https://arxiv.org/pdf/1804.03131.pdf)
53. [Viewpoint-aware Video Summarization](https://arxiv.org/pdf/1804.02843.pdf)
54. [YOLOv3: An Incremental Improvement](https://arxiv.org/pdf/1804.02767.pdf)
55. [Scaling Egocentric Vision: The EPIC-KITCHENS Dataset](https://arxiv.org/pdf/1804.02748.pdf)
56. [Audio-Visual Scene Analysis with Self-Supervised Multisensory Features](https://arxiv.org/pdf/1804.03641.pdf)
57. [Classification of Point Cloud Scenes with Multiscale Voxel Deep Network](https://arxiv.org/pdf/1804.03583.pdf)
58. [EVALUATION OF THE VISUAL ODOMETRY METHODS FOR SEMI-DENSE REAL-TIME](https://arxiv.org/ftp/arxiv/papers/1804/1804.03558.pdf)
59. [Two Stream 3D Semantic Scene Completion](https://arxiv.org/pdf/1804.03550.pdf)
60. [Deep Learning For Computer Vision Tasks: A review](https://arxiv.org/ftp/arxiv/papers/1804/1804.03928.pdf)
61. [Pix3D: Dataset and Methods for Single-Image 3D Shape Modeling](https://arxiv.org/pdf/1804.04610.pdf)
62. [Personalized Classifier for Food Image Recognition](https://arxiv.org/pdf/1804.04600.pdf)
63. [Towards High Performance Video Object Detection for Mobiles](https://arxiv.org/pdf/1804.05830.pdf)
64. [IterGANs: Iterative GANs to Learn and Control 3D Object Transformation](https://arxiv.org/pdf/1804.05651.pdf)
65. [Im2Struct: Recovering 3D Shape Structure from a Single RGB Image](https://arxiv.org/pdf/1804.05469.pdf)
66. [Comparative study of motion detection methods for video surveillance systems](https://arxiv.org/ftp/arxiv/papers/1804/1804.05459.pdf)
67. [FDMO: Feature Assisted Direct Monocular Odometry](https://arxiv.org/pdf/1804.05422.pdf)

## Computation and Language

1. [DP-GAN: Diversity-Promoting Generative Adversarial Network for Generating Informative and Diversified Text](https://arxiv.org/pdf/1802.01345.pdf)
2. [Texygen: A Benchmarking Platform for Text Generation Models](https://arxiv.org/pdf/1802.01886.pdf)
3. [An End-to-End Goal-Oriented Dialog System with a Generative Natural Language Response Generation](https://arxiv.org/pdf/1803.02279.pdf)
4. [Translating Questions into Answers using DBPedia n-triples](https://arxiv.org/pdf/1803.02914.pdf)
5. [Syntax-Aware Language Modeling with Recurrent Neural Networks](https://arxiv.org/pdf/1803.03665.pdf)
6. [Attention on Attention: Architectures for Visual Question Answering (VQA)](https://arxiv.org/pdf/1803.07724.pdf)
7. [Scene Graph Parsing as Dependency Parsing](https://arxiv.org/pdf/1803.09189.pdf)
8. [Jointly Discovering Visual Objects and Spoken Words from Raw Sensory Input](https://arxiv.org/pdf/1804.01452.pdf)
9. [Speech Commands: A Dataset for Limited-Vocabulary Speech Recognition](https://arxiv.org/pdf/1804.03209.pdf)
10. [Pragmatically Informative Image Captioning with Character-Level Inference](https://arxiv.org/pdf/1804.05417.pdf)

## Robotics

1. [Action Anticipation: Reading the Intentions of Humans and Robots](https://arxiv.org/pdf/1802.02788.pdf)
2. [Real Time Collision Detection and Identification for Robotic Manipulators](https://arxiv.org/pdf/1802.00546.pdf)
3. [The Earth ain’t Flat: Monocular Reconstruction of Vehicles on Steep and Graded Roads from a Moving Camera](https://arxiv.org/pdf/1803.02057.pdf)
4. [Explain Yourself: A Natural Language Interface for Scrutable Autonomous Robots](https://arxiv.org/pdf/1803.02088.pdf)
5. [DeepMoTIon: Learning to Navigate Like Humans](https://arxiv.org/pdf/1803.03719.pdf)
6. [A Survey of Deep Learning Techniques for Mobile Robot Applications](https://arxiv.org/pdf/1803.07608.pdf)
7. [Mapping Walls of Indoor Environment Using Moving RGB-D Sensor](https://arxiv.org/pdf/1803.10687.pdf)
8. [Design of an Autonomous Racecar: Perception, State Estimation and System Integration](https://arxiv.org/pdf/1804.03252.pdf)
9. [Geometric Consistency for Self-Supervised End-to-End Visual Odometry](https://arxiv.org/pdf/1804.03789.pdf)
10. [Tightly-coupled Monocular Visual-odometric SLAM using Wheels and a MEMS Gyroscope](https://arxiv.org/pdf/1804.04854.pdf)
11. [Safe Motion Planning in Unknown Environments: Optimality Benchmarks and Tractable Policies](https://arxiv.org/pdf/1804.05804.pdf)
12. [LoST? Appearance-Invariant Place Recognition for Opposite Viewpoints using Visual Semantics](https://arxiv.org/pdf/1804.05526.pdf)
13. [Efficient Computation of Collision Probabilities for Safe Motion Planning∗](https://arxiv.org/pdf/1804.05384.pdf)

## Other