IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Vision Guided Robotic Systems using Deep Learning

[1]Salabha Joy Jacob

[1]Assistant Professor

[1]Electronics and Computer Science Department,

[1]Shah and Anchor Kutchhi Engineering College, Mumbai, India

Abstract: A robotic vision system is a technology that enables a robot to "see." These systems enable the machine to be able to identify, navigate, inspect or handle parts or tasks. To interact with the real world, robots require various sensory inputs from their surroundings, and the use of vision is rapidly increasing nowadays, as vision is unquestionably a rich source of information for a robotic system. In recent years, robotic manipulators have made significant progress towards achieving human-like abilities. There is still a large gap between human and robot dexterity.. The development of deep learning methods for vision applications has become a hot research topic. Given that deep learning has already attracted the attention of the robot vision community, the main purpose of this survey is to address the use of deep learning in robot vision.

Index Terms – Robotic Vision System, Deep Learning

I. INTRODUCTION

A robotic vision system consists of one or more cameras connected to a computer. The computer contains a processing software program that helps the robot interpret what it sees. Then, the robot follows the program's instructions to complete the specified task. Additional elements, such as lighting, image sensors, communications devices or other components, can be incorporated to add to the machine's overall capabilities.There are different modules to build a vision-guided system namely perception, localization, path planning, and control. As far as robot planning is concerned it is difficult for a robot to react to sudden environmental changes or avoid any kind of obstacles whereas humans can easily achieve these tasks. In robot planning, a sequence of actions is planned from start to goal point by using planning algorithms simultaneously avoiding obstacles. However, designing an efficient navigation strategy is the most important issue in creating intelligent robots. Therefore, the robot planning problem using vision sensors is one of the most interesting and wide research areas where the ultimate goal is to achieve a safe and optimal route for robot navigation. In general, vision-based robotic systems can be applied to several industrial applications like spray painting, pick and place, assembly task in the optical industry, automotive, robotic welding like pipe and spot welding , payload identification and much more where efficient planning and control are of prime importance. The modules involved in vision-based autonomous robotic systems are illustrated in Figure 1.

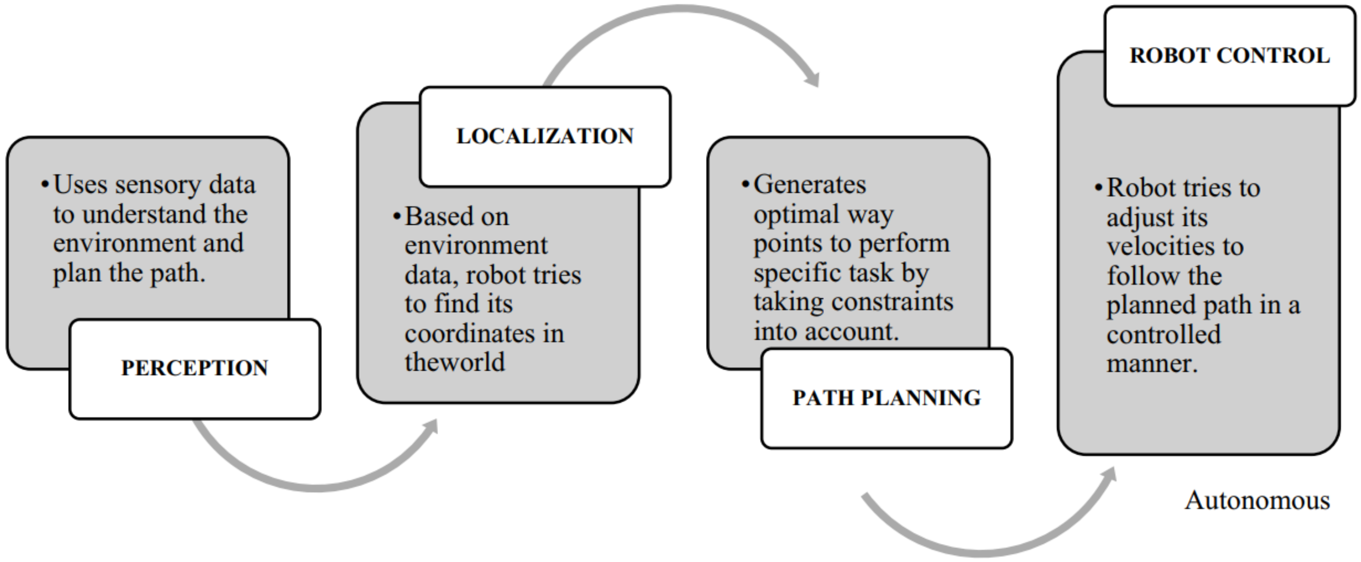

Figure1: Process flow diagram of vision-based autonomous robots.

II. VISION SENSORS USED FOR ROBOT PLANNING

Table 2.1: List of Vision Sensors used for Robot Planning

Sr.No	Sensor Type	Resolution	Frame Rate(fps)	Advantages	Disadvantages
1	2D Sensing	1920x1080	30	Less expensive, easy to integrate and highly reliable.	Cannot operate fully autonomously and poor working under low light conditions
2	RGBD Camera	RGB = 640x480 Depth = 1280x1024	15–30	Adaptable to complex varying work conditions, highly flexible, increases productivity.	More expensive, need proper system like GPUs for point cloud processing
3	Stereo	3840x1080, 2560x720, 4416x1242, 1280 x 720	15–100	Simulate human binocular vision, works better for long distance and moving objects, works better in outdoor applications.	Needs powerful processor, CPU intensive due to two cameras.
4	CCD	795x596, 4096x4112	20	Good Pixel to pixel reproducibility, Has high quality ADC	Expensive, Slow to readout, Needs more circuitry
5	CMOS	2592x1944	14–53	Less expensive, Faster read speed	12-bit ADC reduces the image quality, Linearity and sensitivity variations are high.
6	Ultrasonic Camera	1280x800	12.5	Longer and wider range, not affected by light conditions dust, colors, transparency, or the	Poor in defining edges, affected by temperature, humidity, and pressure

				reflective properties or surface texture of the object.	
7	Infrared Camera	160x120, 1080p,720p	9	Good in capturing edges, Cost effective	Sensitive to IR light and sunlight

III. Deep learning methods for robot vision

The use of the deep learning paradigm has facilitated addressing several computer vision problems in a more successful way than with traditional approaches. In fact, in several computer vision benchmarks, such as the ones addressing image classification, object detection and recognition, semantic segmentation, and action recognition, just to name a few, most of the competitive methods are now based on the use of deep learning techniques

The main motivation of this survey is to address the use of deep learning in robot vision. Table 3.1shows a summary of relevant work in deep learning for computer vision.

Table 3.1 Summary of Relevant Work in Deep Learning for Computer Vision

Paper	Contribution	Model/Algorithm
Fukushima, 1980 [4]	The Neocognitron network is proposed.	Neocognitron
LeCun et al., 1998 [19]	LeNet-5, the first widely known convolutional neural network, is proposed. Previous versions of this networkwere proposed in 1989-1990. [17][18].	LeNet-5
Hochreiter & Schmidhuber, 1997 [68]	LSTM recurrent networks are introduced.	LSTM (Long Short-Term Memory)
Nair & Hinton, 2010 [5]	The paper introduced the ReLU activation functions.	ReLU (Rectified linear unit)
Glorot, Bordes, & Bengio, 2010 [7]	The paper demonstrated that the training of a network ismuch faster when ReLU activation functions are used.	ReLU (Rectified linear unit)
Bottou, 2010 [67]	The Stochastic Gradient Descent is proposed as a learning algorithm for large-scale networks.	Stochastic Gradient Descent
Hinton et al. 2012 [62]	Dropout, a simple and effective method to preventoverfitting, is proposed.	Dropout
Krizhevsky et al., 2012 [20]	AlexNet is proposed and, thanks to its outstanding performance in the ILSVRC 2012 benchmark [69], the boom of deep learning in the computer vision communitystarted. AlexNet has 8 layers.	AlexNet
Simonyan & Zisserman, 2014 [21]	The VGG network is proposed. It has 16/18 layers.	VGG
Girshick et al., 2014 [27]	The Regions with CNN features network is proposed.	R-CNN
Szegedy et al., 2015 [22]	The GoogleNet network is proposed. It has 22 layers.	GoogleNet

He et al., 2015 [23][24]	The ResNet network is proposed. It is based on the use of residuals and is able to use up to 1001 layers.	ResNet
Badrinarayanan et al., 2015 [31]	SegNet, a fully convolutional network for imagesegmentation applications, is proposed.	SegNet
Van de Sande et al., 2016 [235]	The DenseNet network is proposed. It includes skip connections between every layer and its previous layers.	DenseNet
Hu et al [221]	The Squeeze-and-excitation network is proposed. It introduces squeeze-and-excitation blocks used for representing contextual information.	SENet

IV. DESIGNING DEEP-LEARNING BASED VISION APPLICATIONS

For designers of a deep-learning based vision system, it is important to know which learning frameworks and databases are available to be used in the design process, which DNN models are best suited for their application, whether an already trained DNN can be adapted to the new application, and how to approach the learning process. All these issues are addressed in this section. These guidelines can be used in the design process of any kind of vision system, including robot vision systems.

Given the complexity in the design and training of DNNs, special learning frameworks/tools are used for these tasks. There is a wide variety of frameworks available for training and using deep neural networks, as well as several public databases suited to various applications and used for training. Table 4.1 presents some of these frameworks, showing the DNN models included in each case.

Table 4.1 Tools for Designing and Developing deep learning based Vision Applications.

Tool	Description	Included DNN models
Caffe [60]	Deep learning framework, written in C++ and able to use CUDA/CuDNN for multi-GPU. Support for Python and MATLAB.	Caffe Model Zoo [3] includes pre-trained reference models of CaffeNet, AlexNet, R-CNN, GoogLeNet, NiN, VGG, Places-CNN, FCN, ParseNet, SegNet, among others.
Torch [130]	Scientific computing framework, written in C and able to use CUDA/CuDNN for multi- GPU. Support for LuaJIT.	Torch Model Zoo includes pre-trained models of OverFeat, DeepCompare, and models loaded from Caffe into Torch.
Theano [131][132]	Python library for large-scale computing, able to use CUDA/CuDNN. It has started experimental multi-GPU support.	Auto-encoders, RBMs (through Pylearn2), CNN, LSTM (through Theano Lights), models from Caffe Zoo (through Lasagne)
TensorFlow [137]	Framework for computation using data flow graphs, written in C++ with Python APIs. Able to use CUDA/CuDNN for multi- GPU/multi-machine.	It has examples of small CNNs and RNN with LSTM in its tutorial. Can load models from Caffe Zoo.
MXNet [138]	Deep learning framework. It is portable, allows multi-GPU/multi-machine use, and has bindings for Python, R, C++ and Julia.	Includes three pre-trained models: Inception-BN Network,Inception-V3 Network, and Full ImageNet Network.
Deeplearning4j [139]	Deep learning library written in Java and Scala. Has multi-GPU/multi-machine support.	Examples of RBM, DBM, LSTM. Its Model Zoo includes AlexNet, LeNet, VGGNetA, VGGNetD.

Chainer [140]	Deep Learning Framework written in Python. Implements CuPy for multi-GPU support.	AlexNet, GoogLeNet, NiN, MLP. Can import some pre-trained models from the Caffe Zoo.
CNTK [141]	Computational Network Toolkit. Allows efficient multi-GPU/multi-machine use.	It has no pre-trained models.
OverFeat [142]	Feature extractor and classifier based on CNNs. No support for GPUs.	OverFeat Network (pre-trained on ImageNet)
SINGA [143][144]	Distributed deep learning platform written in C++, Has support for multi-GPU/multi-machine.	It has no pre-trained models.
ConvNetJS [145]	Javascript library for training deep learning models entirely in a web browser. No support for GPUs.	Browser demos of CNNs
Cuda-convnet2 [146]	A fast C++/CUDA implementation of convolutional neural networks. Has support for multi-GPU.	It has no pre-trained models.
MatConvNet [147]	MATLAB toolbox implementing CNNs. Able to use CUDA/CuDNN on multi-GPU.	Pre-trained models of VGG-Face, FCNs, ResNet, GoogLeNet, VGG-VD [21], VGG-S,M,F [74] CaffeNet, AlexNet.
Neon [148]	Python based Deep Learning framework. Able to use CUDA for multi-GPU.	Pre-trained models of VGG, Reinforcement learning, ResNet, Image Captioning, Sentiment analysis, and more.
Veles [149]	Distributed platform, able to use CUDA/CuDNN on multi-GPU/multi-machine. Written in Python.	AlexNet, FCNs, CNNs, auto-encoders.

The availability of training data is also a crucial issue. There are several datasets available for popular computer vision tasks such as image classification, object detection and recognition, semantic segmentation, and action recognition (see Table 4.2).

Table 4. 2Training Databases used in some selected Computer Vision Applications.

Application	**Selected databases**
Image classification	MNIST (70,000 images, handwritten digits) [171], CIFAR-10 (60,000 tiny images) [172], CIFAR-100 (60,000 tiny images) [172], ImageNet [163] (14M+ images), SUN (scene classification, 130,519 images) [162]
Object detection and recognition	Pascal VOC 2012 (11,540 train/val images) [154], KITTI (7,481 train images, car/pedestrian/cyclist) [157], MS COCO (300,000+ images) [158], LabelMe (2,920 train images) [159]
Semantic segmentation	Cityscapes (5,000 fine + 20,000 coarse images) [160], Pascal VOC 2012 (2,913 images) [154], NYU2 (RGBD, 1,449 images) [155]
Action recognition	MPII (25,000+ images) [156], Pascal VOC 2012 (5,888 images) [154]

V. Selection of the DNN Model

The selection of the network model to be used must consider a trade-off between classification performance, computational requirements, and processing speed.

When selecting a DNN model, there are two main options: using a pre-existing model, or using a novel one. But, the use of a fully novel DNN model is not recommended because of the difficulties in predicting its behavior, unless the purpose of the developer is to propose a new optimized neural model. When using a pre-existing model there are three options: (i) using a pre- trained neural model for solving the task directly, (ii) fine-tuning the parameters of a pre-trained model for adapting it to the new task, or (iii) training the model from scratch. In cases (ii) and (iii), the dimensionality of the network output can be different from the dimensionality required by the task. In that case, the last fully-connected layers of the network in charge of the classification can be replaced by new ones, or by statistical classifiers such as SVMs or random forests.

Table 5.1 shows popular DNNs, the number of layers and parameters used in each case, the main applications in which the DNNs have been used, and the sizes of the datasets that have been used for training.

Table 5.1 Popular CNN-based Architectures used in Computer-vision Applications. The dataset size does not consider data augmentation.

Name	Layers	Parameters	Application	Dataset size used for training (# images)
Le-Net5	2 conv, 2 fc, 1 Gaussian	60 K	Image classification	70 K
AlexNet [20]	5 conv, 3 fc	60 M	Image classification	1.2 M
VGG-Fast/Medium/Slow	5 conv, 3 fc	77M / 102M / 102M	Image classification	1.2 M
VGG16	13 conv, 3 fc	138 M	Image classification	1.2 M
VGG19	16 conv, 3 fc	144 M	Image classification	1.2 M
GoogleNet	22 layers	7 M	Image classification	1.2 M
ResNet-50 [23]	49 res-conv + 1 fc	0.75 M[4]	Image classification (also used in other applications)	1.2 M
DenseNet [235]	1 conv + 1 pooling + N	0.8M – 27.2M	Image classification	1.2 M
	dense blocks + 1 fc			
SENet [221]	1 conv + 1 pooling + N squeeze-and-excitation blocks + 1 fc	103MB – 137MB	Image Classification	1.2 M
Faster R-CNN (ZF) [29]	5 conv shared, 2 conv RPN, 1 fc reg, 1 fc cls	54 M	Object detection and recognition	200 K
Faster R-CNN (VGG16) [29]	13 conv shared, 2 conv RPN, 1 fc reg, 1 fc cls	138 M	Object detection and recognition	200 K
Faster R-CNN (ResNet-50) [23]	49 res-conv shared, 2 conv RPN, 1 fc reg, 1 fc cls	0.75 M4	Object detection and recognition	200 K
SegNet [31]	13 conv encoder, 13 conv decoder	29.45 M	Semantic segmentation	200 K

Adelaide_context [166]	VGG-16 based, unary and pairwise nets	Not available	Semantic segmentation	200 K
DeepLabv3 [208]	ResNet based, atrous convolution	Not available	Semantic segmentation	200 K
Pyramid Scene Parsing network [209]	Pyramid pooling modules	188 MB	Semantic segmentation	200K
Stacked hourglass networks [167]	42 layers, multiscale skipping connections	166 MB	Action Recognition	24 K
R*CNN [168]	5 conv shared, 2 fc for main object, 2 fc for secondary object	500 MB	Action Recognition	24 K

VI. Deep Neural Networks in Robot Vision

Deep learning has already attracted the attention of the robot vision community, and during the last couple of years studies addressing the use of deep learning in robots have been published in robotics conferences. Table 6.1 shows some of the recently published papers on robot vision applications based on DNN.

Table 6. Selected Papers on Robot Vision Applications based on DNN.

Paper	DNN Model and Distinctive Methods	Application
Pasquale et al., 2015 [38]	Use of CaffeNet in humanoid robots for recognizing objects.	Object Detection and Categorization
Bogun et al., 2015 [39]	DNN with LSTM for recognizing objects in videos.	Object Detection and Categorization
Hosang at al., 2015 [40]	AlexNet-based R-CNN for pedestrian detection with SquaresChnFtrs as person proposal.	Object Detection and Categorization (Pedestrian detection)
Tome et al., 2016 [42]	CNN (Alexnet v/s GoogleNet) for pedestrian detection with LCDF as person proposal.	Object Detection and Categorization (Pedestrian detection)
Lenz at al., 2013 [45]	CNN trained on hand-labeled data, two-stage with two hidden layers each.	Object Grasping and Manipulation
Redmon et al., 2015 [46]	CNN based on AlexNet trained on hand-labeled data,rectangle regression.	Object Grasping and Manipulation
Sung et al., 2016 [47]	Transfer manipulation strategy in embedding space by using a deep neural network.	Object Grasping and Manipulation
Levine at al., 2016 [48]	Learn hand-eye coordination independently of camera calibration or robot pose, visual/motor DNN.	Object Grasping and Manipulation
Zhou 2014 et al., [50]	CNNs for place recognition based on CaffeNet.	Scene Representation and Classification (Place recognition)
Gomez-Ojeda et al., 2015 [49]	CNN for appearance-invariant place recognition, based on CaffeNet.	Scene Representation and Classification (Place recognition)

Hou et al., 2015 [51]	CNN for loop closing, based on CaffeNet.	Scene Representation and Classification (Place recognition)
Sundehauf et al., 2015 [52]	Place categorization and semantic mapping, based on CaffeNet.	Scene Representation and Classification (Scene categorization)
Ye et al., 2016 [53]	R-CNN for functional scene understanding, based on Selective Search and VGG.	Scene Representation and Classification (Scene categorization)
Cadena et al., 2016 [97]	Multi-modal auto-encoders for semantic segmentation. The inputs are RGB-D, LIDAR and stereo data. It uses inverse depth parametrization.	Scene Representation and Classification (Semantic segmentation, Scene Depth Estimation)
Li et al., 2016 [98]	FCN for vehicle detection. Input is a point map generatedusing a LIDAR.	Object Detection and Categorization
Alcantarilla et al., 2016 [99]	Deconvolutional Networks for Street-View ChangeDetection.	Scene Representation and Classification (Street-View Change Detection)
Sunderhauf et al., 2015 [100]	R-CNN used for creating region landmarks for describing an image. AlexNet (up to conv3) as feature extractor.	Scene Representation and Classification (Place Recognition)
Albani et al., 2016 [101]	CNN as a validating step for humanoid robot detection.	Object Detection and Categorization (humanoid soccer robot detection)
Speck et al., 2016 [102]	CNNs for ball localization in robotics soccer.	Object Detection and Categorization (humanoid soccer ball detection)
Finn et al., 2016 [103]	Deep Spatial Auto-encoder for learning state representation. Used for reinforcement learning.	Object Grasping and Manipulation
Gao et al., 2016 [104]	Visual CNN and Haptic CNN combined for haptic classification.	Spatiotemporal Vision (Object Understanding)
Husain et al., 2016 [105]	Temporal concatenation of the output of pre-trained VGG-16 into a 3D convolutional layer.	Spatiotemporal Vision (Action Recognition)
Oliveira et al., 2016 [106]	FCN-based architecture for human body part segmentation.	Object Detection and Categorization
Zaki et al. 2016 [107]	Convolutional Hypercube Pyramid based on VGG-f as feature extractor for RGBD.	Object Detection and Categorization
Mendes et al., 2016 [108]	Network-in-Network converted into FCN for road segmentation.	Scene Representation and Categorization (Semantic Segmentation)
Pinto et al., 2016 [109]	AlexNet based architecture to predict grasp location and angle from image patches, based on self-supervision.	Object Grasping and Manipulation

Jain et al., 2016 [110]	Fusion RNN based on LSTM units fed by visual features.	Spatiotemporal Vision (Human Action Prediction).
Schlosser et al., 2016 [111]	R-CNN for pedestrian detection by using RGB-D from camera and LIDAR. The depth represented using HHA.RGB-D deformable parts model as object proposals.	Object Detection and Recognition (Pedestrian detection)
Costante et al., 2016 [112]	CNNs for learning feature representation and frame to frame motion estimation from optical flow.	Spatiotemporal Vision (Visual Odometry)
Liao et al., 2016 [113]	AlexNet-based scene classifier with a semantic segmentation branch.	Scene Representation and Classification (Place Classification / Semantic Segmentation)
Mahler et al., 2016 [114]	Multi-View Convolutional Neural Networks with Pre-trained AlexNet as CNNs for computing shape descriptors for 3D objects. Used for selecting object grasps.	Object Grasping and Manipulation
Guo et al., 2016 [115]	AlexNet-based CNN for detecting object and grasp by regression on an image patch.	Object Grasping and Manipulation
Yin et al., 2016 [116]	Deep Autoencoder for nonlinear time alignment of human skeleton representations.	Spatiotemporal vision (Human Action Recognition)
Giusti et al., 2016 [117]	CNN that receives a trail image as input and classifies it asthe kind of motion needed for remaining on the trail.	Scene Representation and Classification (Trail Direction Classification)
Held et al., 2016 [118]	CaffeNet, pre-trained on ImageNet, fine-tuned for viewpoint invariance.	Object Detection and Categorization (Single-view Object Recognition)
Yang et al., 2016 [119]	FCN and DSN based Network with AlexNet as basis. Usedwith CRF for estimating 3D scene layout from monocular camera.	Scene Representation and Classification (Semantic Segmentation, Scene Depth Estimation)
Uršic et al., 2016 [120]	R-CNN used for generating histograms of part-based models. Places-CNN (CaffeNet trained on Places 205) as region feature extractor.	Scene Representation and Classification (Place Classification)
Bunel et al., 2016 [121]	CNN architecture of 4 convolutional layers with PReLU as activation function and 2 FC layers. Used for detecting pedestrians at far distance.	Object Detection and Categorization (Pedestrian Detection)
Murali et al., 2016 [122]	VGG architecture as feature extractor for unsupervised segmentation of image sequences.	Spatiotemporal Vision (Segmentation of trajectories in robot-assisted surgery).
Kendall et al., 2016 [123]	Bayesian PoseNet (Modified GoogLeNet) with poseregression, use dropout for estimating uncertainty.	Scene Representation and Classification (Camera re-localization)
Husain et al., 2016 [124]	CNN using layers from OverFeat Network with multiple pooling sizes. RGB-D inputs. Use of HHA and distance-from-wall for depth.	Scene Representation and Classification (Semantic Segmentation)

Hoffman et al., 2016 [125]	R-CNN using RGB-D data, based on AlexNet and HHA. Proposals are based on RGB-D Selective Search.	Object Detection and Categorization
Sunderhauf et al., 2016 [126]	Places-CNN as image classifier for building 2D grid semantic maps. LIDAR used for SLAM. Bayesian filtering over class labels.	Scene Representation and Classification (Place Classification)
Saxena et al., 2017 [189]	CNN used for image-based visual servoing. Inputs are monocular images from current and desired poses. Outputsare velocity commands for reaching the desired pose.	Object Grasping and Manipulation (Visual servoing)
Lei et al., [191]	Robot exploration by using a CNN, trained first by supervised learning, later by using deep reinforcement learning. Tested on simulated and real experiments.	Scene Representation and Classification (Scene Exploration)
Zhu et al., [192]	Robot navigation by learning a scene-dependent Siamese network, which receives images from two places as input, and generates motion commands for travelling between them.	Scene Representation and Classification (Visual Navigation)
Mirowski et al., [193]	Robot navigation in complex maze-like environments from raw monocular images and inertial information. Use of deep reinforcement learning on a network composed of a CNN followed by two LSTM layers. Multi-task loss considering reward prediction and depth prediction improves learning.	Scene Representation and Classification (Visual Navigation)

VII. Conclusions

This survey provides a valuable guidance for the developers of robot vision systems, since it promotes understanding of the basic concepts behind the application of deep-learning in vision applications, explains the tools and frameworks used in the development process of vision systems, and shows current tendencies in the use of DNN models in robot vision.

References

1. Y. Bengio, P. Simard, P. Frasconi. Learning long-term dependencies with gradient descent is difficult. IEEE Transactions on Neural Networks (Volume:5 , Issue: 2), pages 157 - 166, Mar 1994.
2. Corinna Cortes , Vladimir Vapnik. Support-Vector Networks. Machine Learning, vol 20, number 3, pages 273-297 (1995)
3. D. H. Hubel and T. N. Wiesel, Receptive fields and functional architecture of monkey striate cortex, The Journal of physiology, 1968.
4. K. Fukushima. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biological cybernetics, 36(4):193–202, 1980.
5. V. Nair, G.E. Hinton, Rectified Linear Units Improve Restricted Boltzmann Machines, Proc. of the ICML 2010.
6. X. Glorot, Y. Bengio. Understanding the difficulty of training deep feedforward neural networks. In Proceedings of the International Conference on Artificial Intelligence and Statistics (AISTATS'10). Society for Artificial Intelligence and Statistics. 2010.
7. X. Glorot, A. Bordes & Y. Bengio. Deep sparse rectifier neural networks. In Proc. 14th International Conference on Artificial Intelligence and Statistics 315–323 (2011).
8. Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. IEEE International Conference on Computer Vision (ICCV), 2015.
9. Yanming Guo, Yu Liu, Ard Oerlemans, Songyang Lao, Song Wu, Michael S. Lew. Deep learning for visual understanding: A review. Neurocomputing, 187 (2016), pp. 27-48.
10. Soren Goyal, Paul Benjamin. Object Recognition Using Deep Neural Networks: A Survey. http://arxiv.org/abs/1412.3684. Dec 2014.

11. Jiuxiang Gu, Zhenhua Wang, Jason Kuen, Lianyang Ma, Amir Shahroudy, Bing Shuai, Ting Liu, Xingxing Wang, Gang Wang. Recent Advances in Convolutional Neural Networks. http://arxiv.org/abs/1512.07108. Dec 2015.
12. Yann LeCun, Yoshua Bengio, Geoffrey Hinton. Deep learning. Nature 521, 436–444 (28 May 2015), doi:10.1038/nature14539
13. Li Deng. A Tutorial Survey of Architectures, Algorithms, and Applications for Deep Learning. APSIPA Transactions on Signal and Information Processing.
14. J. Schmidhuber. Deep Learning in Neural Networks: An Overview. Neural Networks, Volume 61, January 2015, Pages 85-117 (DOI: 10.1016/j.neunet.2014.09.003), published online in 2014.
15. Seyed-Mahdi Khaligh-Razavi. What you need to know about the state-of-the-art computational models of object-vision: A tour through the models. ArXiv:1407.2776
16. Suraj Srinivas, Ravi Kiran Sarvadevabhatla, Konda Reddy Mopuri, Nikita Prabhu, Srinivas S. S. Kruthiventi and R. Venkatesh Babu. A Taxonomy of Deep Convolutional Neural Nets for Computer Vision. Front. Robot. AI, 11 January 2016 | http://dx.doi.org/10.3389/frobt.2015.00036
17. Y. LeCun, B. Boser, J. Denker, D. Henderson, R. Howard, W. Hubbard, and L. Jackel. Backpropagation applied to handwritten zip coderecognition. Neural Comp., 1989.
18. Y. Le Cun, B. Boser, J. S. Denker, R. E. Howard, W. Habbard, L. D. Jackel, D. Henderson. Handwritten digit recognition with a back-propagation network. In Proc. Advances in Neural Information Processing Systems 396–404 (1990).
19. Y. Lecun, L. Bottou, Y. Bengio, P. Haffner. Gradient-based learning applied to document recognition. Proc. of the IEEE (Volume:86 , Issue: 11). Pages 2278 - 2324. Nov 1998.
20. Alex Krizhevsky and Sutskever, Ilya and Geoffrey E. Hinton. ImageNet Classification with Deep Convolutional Neural Networks.Advances in Neural Information Processing Systems 25. Pages 1097—1105. 2012.
21. Karen Simonyan and Andrew Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition. ILSVRC-2014. http://arxiv.org/pdf/1409.1556.
22. Christian Szegedy and Wei Liu and Yangqing Jia and Pierre Sermanet and Scott Reed and Dragomir Anguelov and Dumitru Erhan and Vincent Vanhoucke and Andrew Rabinovich. Going Deeper with Convolutions. CVPR 2015.
23. Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun. Deep Residual Learning for Image Recognition. Microsoft Research, 2015. ArXiv:1512.03385.
24. Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun. Identity Mappings in Deep Residual Networks. arXiv:1603.05027
25. https://github.com/KaimingHe/deep-residual-networks
26. Koen E.A.Van de Sande, Jasper R.R. Uijlings, Theo Gevers, Arnold W.M. Smeulders. Segmentation As Selective Search for Object Recognition. Proceedings of the 2011 International Conference on Computer Vision, pages 1879--1886, 2011.
27. R. Girshick, J. Donahue, T. Darrell, J. Malik. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014.
28. R. Girshick. Fast R-CNN. Proceedings of the 2015 IEEE International Conference on Computer Vision, 1440-1448
29. Shaoqing Ren, Kaiming He, Ross Girshick, Jian Sun. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. arXiv:1506.01497v3, 2015.
30. Jon Long, Evan Shelhamer, Trevor Darrell. Fully Convolutional Networks for Semantic Segmentation. CVPR 2015 (best paper honorable mention)
31. Vijay Badrinarayanan, Alex Kendall, Roberto Cipolla. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation, 2015. arXiv:1511.00561.
32. Alex Kendall, Vijay Badrinarayanan, Roberto Cipolla. Bayesian SegNet: Model Uncertainty in Deep Convolutional Encoder-Decoder Architectures for Scene Understanding. arXiv:1511.02680.
33. Max Schwarz, Hannes Schulz, and Sven Behnke. RGB-D Object Recognition and Pose Estimation based on Pre-trained Convolutional Neural Network Features. ICRA 2015.
34. Saurabh Gupta, Ross Girshick, Pablo Arbeláez, Jitendra Malik. Learning Rich Features from RGB-D Images for Object Detection and Segmentation. ArXiv:1407.5736 (ECCV 2014)
35. Asako Kanezaki. RotationNet: Learning Object Classification Using Unsupervised Viewpoint Estimation. arXiv:1603.06208v1 [cs.CV] 20 Mar 2016
36. Sepp Hochreiter, Jürgen Schmidhuber. Long Short-Term Memory. Neural Computation, Volume 9 Issue 8, November 15, 1997. Pages 1735-1780.
37. A. Gers F., J. Schmidhuber, F. Cummins. Learning to Forget: Continual Prediction with LSTM. Istituto Dalle Molle Di Studi Sull Intelligenza Artificiale. 1999.
38. Giulia Pasquale, Carlo Ciliberto, Francesca Odone, Lorenzo Rosasco, Lorenzo Natale. Real-world Object Recognition with Off-the-shelf Deep Conv Nets: How Many Objects can iCub Learn? arXiv:1504.03154v2
39. Ivan Bogun, Anelia Angelova, Navdeep Jaitly. Object Recognition from Short Videos for Robotic Perception. arXiv:1509.01602v1 [cs.CV] 4 Sep 2015
40. J. Hosang, M. Omran, R. Benenson, B. Schiele. Taking a Deeper Look at Pedestrians. CVPR 2015.
41. Rodrigo Benenson, Markus Mathias, Tinne Tuytelaars, Luc Van Gool. Seeking the Strongest Rigid Detector. Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on.

42. D. Tomè, F. Monti, L. Baroffio, L. Bondi, M. Tagliasacchi, S. Tubaro. Deep Convolutional neural networks for pedestrian detection. arXiv:1510.03608v5 [cs.CV] 7 Mar 2016
43. Woonhyun Nam, Piotr Dollár, and Joon Hee Han. Local Decorrelation For Improved Pedestrian Detection. NIPS 2014, Montreal, Quebec.
44. P. Dollar, R. Appel, S. Belongie, P. Perona. Fast feature pyramids for object detection. IEEE Trans. Pattern Anal. Mach. Intell., vol 36, number 8. August 2014.
45. Ian Lenz, Honglak Lee, Ashutosh Saxena. Deep Learning for Detecting Robotic Grasps. ArXiv:1301.3592.
46. Joseph Redmon, Anelia Angelova. Real-Time Grasp Detection Using Convolutional Neural Networks. arXiv:1412.3128v2 [cs.RO] 28 Feb 2015
47. Jaeyong Sung, Seok Hyun Jin, Ian Lenz, and Ashutosh Saxena. Robobarista: Learning to Manipulate Novel Objects via Deep Multimodal Embedding. arXiv:1601.02705v1 [cs.RO] 12 Jan 2016
48. Sergey Levine, Peter Pastor, Alex Krizhevsky, Deirdre Quillen. Learning Hand-Eye Coordination for Robotic Grasping with Deep Learningand Large-Scale Data Collection. rXiv:1603.02199v2 [cs.LG] 24 Mar 2016
49. Ruben Gomez-Ojeda, Manuel Lopez-Antequera, Nicolai Petkov, Javier Gonzales-Jimenez. Training a Convolutional Neural Network for Appearance-Invariant Place Recognition. arXiv:1505.07428v1 [cs.CV] 27 May 2015
50. Bolei Zhou, Agata Lapedriza, Jianxiong Xiao, Antonio Torralba, and Aude Oliva. Learning Deep Features for Scene Recognition using Places Database. Advances in Neural Information Processing Systems (NIPS) 27, 2014.
51. Yi Hou, Hong Zhang, Shilin Zhou. Convolutional Neural Network-Based Image Representation for Visual Loop Closure Detection. arXiv:1504.05241v1 [cs.RO] 20 Apr 2015
52. Niko Sunderhauf, Feras Dayoub, Sean McMahon, Ben Talbot, Ruth Schulz, Peter Corke, Gordon Wyeth, Ben Upcroft, and Michael Milford. Place Categorization and Semantic Mapping on a Mobile Robot. arXiv:1507.02428v1 [cs.RO] 9 Jul 2015
53. Chengxi Ye, Yezhou Yang, Cornelia Fermüller, Yiannis Aloimonos. What Can I Do Around Here? Deep Functional Scene Understanding for Cognitive Robots. ArXiv:1602.00032 , Feb 2016
54. P. Loncomilla, J. Ruiz-del-Solar and L. Martínez, Object Recognition using Local Invariant Features for Robotic Applications: A Survey, Pattern Recognition, Vol. 60, Dec. 2016, pp. 499-514.
55. Workshop Geometry Meets Deep Learning, ECCV 2016. Available (July 2016) in https://sites.google.com/site/deepgeometry/
56. Workshop Deep Learning for Autonomous Robots, RSS 2016. Available (July 2016) inhttp://www.umiacs.umd.edu/~yzyang/deeprobotics.html
57. Workshop Are the Skeptics Right? Limits and Potentials of Deep Learning in Robotics, RSS 2016. Available (July 2016) in http://juxi.net/workshop/deep-learning-rss-2016/
58. Workshop The Future of Real-Time SLAM: Sensors, Processors, Representations, and Algorithms, https://wp.doc.ic.ac.uk/thefutureofslam/
59. Keynote talk "Deep Grasping: Can large datasets and reinforcement learning bridge the dexterity gap?", Ken Goldberg at ICRA 2016
60. Caffe deep learning framework. Available (July 2016) in http://caffe.berkeleyvision.org/
61. M. Egmont-Petersen, D. de Ridder, H. Handels, Image processing with neural networks—a review, Pattern Recognition, Vol. 35, Issue 10, Oct. 2002, pp. 2279–2301.
62. G.E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, Improving neural networks by preventing co-adaptation of feature detectors, arXiv preprint, arXiv: 1207.0580, 2012.
63. N. Srivastava, G.E. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, Dropout: A Simple Way to Prevent Neural Networks from Overfitting, Journal of Machine Learning Research 15 (2014) 1929-1958.
64. L. Wan, M. Zeiler, S. Zhang, Y. LeCun, R. Fergus. Regularization of neural networks using DropConnect, Proc. of the ICML, 2013.
65. K. Hornik, Approximation capabilities of multilayer feedfoward networks, Neural Networks, Vol. 4, pp. 251-257, 1991.
66. B.T. Polyak, Some methods of speeding up the convergence of iteration methods, USSR Computational Mathematics and Mathematical Physics 4 (5), pp. 1-17, 1964.
67. L. Bottou, Large-Scale Machine Learning with Stochastic Gradient Descent, Proc. COM PSTAT 2010.
68. S. Hochreiter, and J. Schmidhuber. Long short-term memory. Neural Comput. 9, pp. 1735–1780, 1997.
69. Large Scale Visual Recognition Challenge (ILSVRC) Official website. Available (July 2016) in: http://www.image-net.org/challenges/LSVRC/
70. P. Werbos. Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences. PhD thesis, Harvard Univ. (1974).
71. G.E. Hinton, S. Osindero & Y.-W. Teh. A fast learning algorithm for deep belief nets. Neural Comp. 18, 1527–1554 (2006).
72. Y. Bengi, P. Lamblin, D. Popovici & H. Larochelle. Greedy layer-wise training of deep networks. In Proc. Advances in Neural Information Processing Systems 19, 153–160 (2006).
73. M. Ranzato, C. Poultney, S. Chopra & Y. LeCun. Efficient learning of sparse representations with an energy-based model. In Proc. Advances in Neural Information Processing Systems 19 1137–1144 (2006).
74. K Chatfield, K Simonyan, A Vedaldi, A Zisserman. Return of the devil in the details: Delving deep into convolutional nets, in Proc. of the British Machine Vision Conference 2014.

75. R. Verschae, J. Ruiz-del-Solar, M. Correa (2008). A Unified Learning Framework for object Detection and Classification using Nested Cascades of Boosted Classifiers. Machine Vision and Applications, Vol. 19, No. 2, pp. 85-103, 2008.
76. J. Yosinski, J. Clune, Y. Bengio, and H. Lipson. How transferable are features in deep neural networks? In Advances in Neural Information Processing Systems 27 (NIPS ’14), NIPS Foundation, 2014.
77. Olivier Delalleau and Yoshua Bengio. Shallow vs. Deep Sum-Product Networks. Advances in Neural Information Processing Systems 24. Pages 666-674, 2012
78. M. Bianchini and F. Scarselli. On the Complexity of Neural Network Classifiers: A Comparison Between Shallow and Deep Architectures. In IEEE Transactions on Neural Networks and Learning Systems, vol. 25, no. 8, pp. 1553-1565, Aug. 2014.
79. Y.N. Dauphin, R. Pascanu, C. Gulcehre, K. Cho, S. Ganguli, Y. Bengio. Identifying and attacking the saddle point problem in high- dimensional non-convex optimization. Advances in neural information processing systems 2014, 2933-2941.
80. Ian Goodfellow, Yoshua Bengio and Aaron Courville. Deep Learning. Book in preparation for MIT Press. 2016. http://www.deeplearningbook.org.
81. G.E. Hinton, & R.R. Salakhutdinov. Reducing the dimensionality of data with neural networks. Science, 313(5786), 504-507, year 2006.
82. P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio & P.A. Manzagol. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. Journal of Machine Learning Research, 11(Dec), 3371-3408, year 2010.
83. F. Yu, V. Koltun. Multi-Scale Context Aggregation by Dilated Convolutions. ICLR 2016.
84. G. Ghiasi, C. Fowlkes. Laplacian Reconstruction and Refinement for Semantic Segmentation. arXiv:1605.02264.
85. L. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A.L. Yuille. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. arXiv:1606.00915
86. G. Lin, C. Shen, A. Van Dan Hengel, I. Reid. Efficient piecewise training of deep structured models for semantic segmentation. IEEE Conf.Computer Vision and Pattern Recognition (CVPR) 2016.
87. M. Jaderberg, A. Vedaldi, A. Zisserman. Speeding up Convolutional Neural Networks with Low Rank Expansions. arXiv:1405.3866[cs.CV]
88. B. Liu, M. Wang, H. Foroosh, M. Tappen and M. Penksy. Sparse Convolutional Neural Networks, 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, 2015, pp. 806-814.
89. S. Han, H. Mao, W.J. Dally: Deep Compression: Compressing Deep Neural Networks with Pruning, Trained Quantization and Huffman Coding (ICLR'16, best paper award)
90. S. Han, X. Liu, H. Mao, J. Pu, A. Pedram, M.A. Horowitz. and W.J. Dally. EIE: Efficient Inference Engine on Compressed Deep Neural Network. International Conference on Computer Architecture (ISCA), 2016.
91. M. Rastegari, V. Ordonez, J. Redmon, & A. Farhadi. XNOR-Net: ImageNet Classification Using Binary Convolutional Neural Networks. arXiv preprint arXiv:1603.05279., year 2016.
92. A. Eitel, J.T. Springenberg, L. Spinello, M. Riedmiller, W. Burgard. Multimodal Deep Learning for Robust RGB-D Object Recognition. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Hamburg, Germany, 2015.
93. D. Ciresan, U. Meier, J. Schmidhuber, Multi-column deep neural networks for image classification, in: Proceedings of the CVPR 2012.
94. O. Vinyals, A. Toshev, S. Bengio, D. Erhan. Show and tell: A neural image caption generator. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2015.
95. M. Malinowski, M. Rohrbach and M. Fritz. Ask Your Neurons: A Neural-based Approach to Answering Questions about Images. ICCV 2015.
96. Y. Zhu, R. Kiros, R. Zemel, R. Salakhutdinov, R. Urtasun, A. Torralba, S. Fidler. Aligning Books and Movies: Towards Story-Like Visual Explanations by Watching Movies and Reading Books. The IEEE International Conference on Computer Vision (ICCV), 2015, pp. 19-27.
97. Cesar Cadena. Anthony Dick and Ian D. Reid. Multi-modal Auto-Encoders as Joint Estimators for Robotics Scene Understanding. Proceedings of Robotics: Science and System Proceedings. June 2016
98. Bo Li, Tianlei Zhang and Tian Xia. Vehicle Detection from 3D Lidar Using Fully Convolutional Network. Proceedings of Robotics: Science and System Proceedings. June 2016
99. Pablo F. Alcantarilla, Simon Stent, German Ros, Roberto Arroyo and Riccardo Gherardi. Street-View Change Detection with Deconvolutional Networks. Proceedings of Robotics: Science and System Proceedings. June 2016
100. Niko Sunderhauf, Sareh Shirazi, Adam Jacobson, Feras Dayoub, Edward Pepperell, Ben Upcroft, and Michael Milford. Place Recognition with ConvNet Landmarks: Viewpoint-Robust, Condition-Robust, Training-Free. Proceedings of Robotics: Science and System Proceedings.July 2015
101. D. Albani, A. Youssef, V. Suriani, D. Nardi, and D.D. Bloisi. A Deep Learning Approach for Object Recognition with NAO Soccer Robots.RoboCup International Symposium. July 2016.
102. Daniel Speck, Pablo Barros, Cornelius Weber and Stefan Wermter. Ball Localization for Robocup Soccer using Convolutional Neural Networks. RoboCup International Symposium. July 2016.

103. Chelsea Finn, Xin Yu Tan, Yan Duan, Trevor Darrell, Sergey Levine and Pieter Abbeel. Deep Spatial Autoencoders for Visuomotor Learning. IEEE International Conference on Robotics and Automation (ICRA). May 2016.
104. Yang Gao, Lisa Anne Hendricks, Katherine J. Kuchenbecker and Trevor Darrell. Deep Learning for Tactile Understanding from Visual and Haptic Data. IEEE International Conference on Robotics and Automation (ICRA). May 2016.
105. Farzad Husain, Babette Dellen and Carme Torra. Action Recognition based on Efficient Deep Feature Learning in the Spatio-Temporal Domain. IEEE International Conference on Robotics and Automation (ICRA). May 2016.
106. Gabriel L. Oliveira, Abhinav Valada, Claas Bollen, Wolfram Burgard and Thomas Brox. Deep Learning for Human Part Discovery in Images. IEEE International Conference on Robotics and Automation (ICRA). May 2016.
107. Hasan F. M. Zaki, Faisal Shafait and Ajmal Mian. Convolutional Hypercube Pyramid for Accurate RGB-D Object Category and Instance Recognition. IEEE International Conference on Robotics and Automation (ICRA). May 2016.
108. Caio Cesar Teodoro Mendes, Vincent Fremont and Denis Fernando Wolf. Exploiting Fully Convolutional Neural Networks for Fast Road Detection. IEEE International Conference on Robotics and Automation (ICRA). May 2016.
109. Lerrel Pinto and Abhinav Gupta. Supersizing Self-supervision: Learning to Grasp from 50K Tries and 700 Robot Hours. IEEE InternationalConference on Robotics and Automation (ICRA). May 2016.
110. Ashesh Jain, Avi Singh, Hema S Koppula, Shane Soh, and Ashutosh Saxena. Recurrent Neural Networks for Driver Activity Anticipation via Sensory-Fusion Architecture. IEEE International Conference on Robotics and Automation (ICRA). May 2016.
111. Joel Schlosser, Christopher K. Chow, and Zsolt Kira. Fusing LIDAR and Images for Pedestrian Detection using Convolutional Neural Networks. IEEE International Conference on Robotics and Automation (ICRA). May 2016.
112. Gabriele Costante, Michele Mancini, Paolo Valigi and Thomas A. Ciarfuglia. Exploring Representation Learning with CNNs for Frame to Frame Ego-Motion Estimation. IEEE International Conference on Robotics and Automation (ICRA). May 2016.
113. Yiyi Liao, Sarath Kodagoda, Yue Wang, Lei Shi and Yong Liu. Understand Scene Categories by Objects: A Semantic Regularized Scene Classifier Using Convolutional Neural Networks. IEEE International Conference on Robotics and Automation (ICRA). May 2016.
114. Jeffrey Mahler, Florian T. Pokorny, Brian Hou, Melrose Roderick, Michael Laskey, Mathieu Aubry, Kai Kohlhoff, Torsten Kroger, James Kuffner and Ken Goldberg. Dex-Net 1.0: A Cloud-Based Network of 3D Objects for Robust Grasp Planning Using a Multi-Armed Bandit Model with Correlated Rewards. IEEE International Conference on Robotics and Automation (ICRA). May 2016.
115. Di Guo, Tao Kong, Fuchun Sun and Huaping Liu. Object Discovery and Grasp Detection with a Shared Convolutional Neural Network. IEEE International Conference on Robotics and Automation (ICRA). May 2016.
116. Xiaochuan Yin and Qijun Chen. Deep Metric Learning Autoencoder for Nonlinear Temporal Alignment of Human Motion. IEEE International Conference on Robotics and Automation (ICRA). May 2016.
117. Alessandro Giusti, Jérôme Guzzi Dan C. Ciresan, Fang-Lin He, Juan P. Rodríguez, Flavio Fontana, Matthias Faessler, Christian Forster, Jürgen Schmidhuber, Gianni Di Caro, Davide Scaramuzza and Luca M. Gambardella. A Machine Learning Approach to Visual Perception of Forest Trails for Mobile Robots. IEEE International Conference on Robotics and Automation (ICRA). May 2016.
118. David Held, Sebastian Thrun and Silvio Savarese. Robust Single-View Instance Recognition. IEEE International Conference on Robotics and Automation (ICRA). May 2016.
119. Shichao Yang, Daniel Maturana and Sebastian Scherer. Real-time 3D Scene Layout from a Single Image Using Convolutional Neural Networks. IEEE International Conference on Robotics and Automation (ICRA). May 2016.
120. Peter Uršic, Rok Mandeljc, Aleš Leonardis and Matej Kristan. Part-Based Room Categorization for Household Service Robots. IEEE International Conference on Robotics and Automation (ICRA). May 2016.
121. Rudy Bunel, Franck Davoine and Philippe Xu. Detection of Pedestrians at Far Distance. IEEE International Conference on Robotics and Automation (ICRA). May 2016.
122. Adithyavairavan Murali, Animesh Garg, Sanjay Krishnan, Florian T. Pokorny, Pieter Abbeel, Trevor Darrell and Ken Goldberg. TSC-DL: Unsupervised Trajectory Segmentation of Multi-Modal Surgical Demonstrations with Deep Learning. IEEE International Conference on Robotics and Automation (ICRA). May 2016.
123. Alex Kendall and Roberto Cipolla. Modelling Uncertainty in Deep Learning for Camera Relocalization. IEEE International Conference on Robotics and Automation (ICRA). May 2016.
124. Farzad Husain, Hannes Schulz, Babette Dellen, Carme Torras and Sven Behnke. Combining Semantic and Geometric Features for Object Class Segmentation of Indoor Scenes. IEEE International Conference on Robotics and Automation (ICRA). May 2016.
125. Judy Hoffman, Saurabh Gupta, Jian Leong, Sergio Guadarrama and Trevor Darrell. Cross-Modal Adaptation for RGB-D Detection. IEEE International Conference on Robotics and Automation (ICRA). May 2016.
126. Niko Sunderhauf, Feras Dayoub, Sean McMahon, Ben Talbot, Ruth Schulz, Peter Corke, Gordon Wyeth, Ben Upcroft, and Michael Milford. Place Categorization and Semantic Mapping on a Mobile Robot. IEEE International Conference on Robotics and Automation (ICRA). May 2016.
127. M. Szarvas, A. Yoshizawa, M. Yamamoto and J. Ogata, "Pedestrian detection with convolutional neural networks," IEEE Proceedings. Intelligent Vehicles Symposium, 2005., 2005, pp. 224-229. doi: 10.1109/IVS.2005.1505106

128. Klaus Greff, Rupesh Kumar Srivastava, Jan Koutník, Bas R. Steunebrink, Jürgen Schmidhuber. LSTM: A Search Space Odyssey.arXiv:1503.04069.
129. Robocup Standard Platform League: http://www.robocup.org/robocup-soccer/standard-platform/
130. Torch: A scientific computing framework for luaJIT. Official website. Available (July 2016) in: http://torch.ch/
131. Theano Development Team. Theano: A Python framework for fast computation of mathematical expressions. arXiv:1605.02688v1. May 2016.
132. James Bergstra, Olivier Breuleux, Frédéric Bastien, Pascal Lamblin, Razvan Pascanu, Guillaume Desjardins, Joseph Turian, David Warde- Farley and Yoshua Bengio. Theano: A CPU and GPU Math Compiler in Python. Proc. Of The 9th Python In Science Conf. (SCIPY 2010).
133. Ian J. Goodfellow, David Warde-Farley, Pascal Lamblin, Vincent Dumoulin, Mehdi Mirza, Razvan Pascanu, James Bergstra, Frédéric Bastien, and Yoshua Bengio. "Pylearn2: a machine learning research library". arXiv preprint arXiv:1308.4214.
134. Theano Lights. GitHub Repository. Available (July 2016) in: https://github.com/Ivaylo-Popov/Theano-Lights.
135. Bart van Merriënboer, Dzmitry Bahdanau, Vincent Dumoulin, Dmitriy Serdyuk, David Warde-Farley, Jan Chorowski, and Yoshua Bengio, "Blocks and Fuel: Frameworks for deep learning," arXiv preprint arXiv:1506.00619 [cs.LG], 2015.
136. Lasagne Library. GitHub Repository. Available (July 2016) in: https://github.com/Lasagne/Lasagne
137. TensorFlow. Official Website. Available (July 2016) in: https://www.tensorflow.org/
138. Tianqi Chen, Mu Li, Yutian Li, Min Lin, Naiyan Wang, Minjie Wang, Tianjun Xiao, Bing Xu, Chiyuan Zhang, and Zheng Zhang. MXNet: A Flexible and Efficient Machine Learning Library for Heterogeneous Distributed Systems. In Neural Information Processing Systems, Workshop on Machine Learning Systems, 2015
139. Deeplearning4j. Official website. Available (July 2016) in: http://deeplearning4j.org/
140. S. Tokui, K. Oono, S. Hido and J. Clayton, Chainer: a Next-Generation Open Source Framework for Deep Learning,Proceedings of Workshop on Machine Learning Systems(LearningSys) in The Twenty-ninth Annual Conference on Neural Information Processing Systems (NIPS), (2015)
141. Amit Agarwal, Eldar Akchurin, Chris Basoglu, Guoguo Chen, Scott Cyphers, Jasha Droppo, Adam Eversole, Brian Guenter, Mark Hillebrand, T. Ryan Hoens, Xuedong Huang, Zhiheng Huang, Vladimir Ivanov, Alexey Kamenev, Philipp Kranen, Oleksii Kuchaiev, Wolfgang Manousek, Avner May, Bhaskar Mitra, Olivier Nano, Gaizka Navarro, Alexey Orlov, Hari Parthasarathi, Baolin Peng, Marko Radmilac, Alexey Reznichenko, Frank Seide, Michael L. Seltzer, Malcolm Slaney, Andreas Stolcke, Huaming Wang, Yongqiang Wang, Kaisheng Yao, Dong Yu, Yu Zhang, Geoffrey Zweig (in alphabetical order), "An Introduction to Computational Networks and the Computational Network Toolkit", Microsoft Technical Report MSR-TR-2014-112, 2014.
142. Pierre Sermanet, David Eigen, Xiang Zhang, Michael Mathieu, Rob Fergus, Yann LeCun. OverFeat: Integrated Recognition, Localization and Detection using Convolutional Networks. arXiv:1312.6229. Feb 2014.
143. B. C. Ooi, K.-L. Tan, S. Wang, W. Wang, Q. Cai, G. Chen, J. Gao, Z. Luo, A. K. H. Tung, Y. Wang, Z. Xie, M. Zhang, and K. Zheng. SINGA: A distributed deep learning platform. ACM Multimedia (Open Source Software Competition) 2015.
144. W. Wang, G. Chen, T. T. A. Dinh, B. C. Ooi, K.-L.Tan, J. Gao, and S. Wang. SINGA: putting deep learning in the hands of multimedia users. ACM Multimedia 2015.
145. ConvNetJS: Deep Learning in your browser. Official Website. Available (July 2017) in: http://cs.stanford.edu/people/karpathy/convnetjs/.
146. Cuda-convnet2. GitHub Repository. Available (July 2016) in: https://github.com/akrizhevsky/cuda-convnet2
147. MatConvNet: CNNs for MATLAB. Official Website. Available (July 2016) in: http://www.vlfeat.org/matconvnet/
148. Neon: Nervana 's Python-based deep learning library. Official Website. Available (July 2016) in: http://neon.nervanasys.com/
149. Veles: Distributed platform for rapid Deep learning application development. Official Website. Avalaible (July 2016) in: https://velesnet.ml/
150. R.J. Williams, J. Peng. An efficient gradient-based algorithm for on-line training of recurrent network trajectories. Neural Computation, Volume 2, Issue 4, Winter 1990, Pages 490-501
151. J. Ngiam, Z. Chen, P.W. Koh , A.Y. Ng. Learning Deep Energy Models. in: Proceedings of the ICML, 2011
152. NVIDIA CUDA toolkit documentation: http://docs.nvidia.com/cuda/
153. NVIDIA CUDNN GPU Accelerated Deep Learning: https://developer.nvidia.com/cudnn
154. PASCAL VOC 2012: http://host.robots.ox.ac.uk:8080/leaderboard/main_bootstrap.php
155. NYU2: http://cs.nyu.edu/~silberman/datasets/nyu_depth_v2.html
156. MPII Human Pose Dataset: http://human-pose.mpi-inf.mpg.de/#results
157. KITTI: http://www.cvlibs.net/datasets/kitti/eval_object.php
158. MS COCO: http://mscoco.org/dataset/#detections-leaderboard
159. LabelMe: http://labelme.csail.mit.edu/Release3.0/browserTools/php/dataset.php
160. Cityscapes: https://www.cityscapes-dataset.com/benchmarks/
161. MNIST - CIFAR 10 - CIFAR 100: http://rodrigob.github.io/are_we_there_yet/build/#datasets
162. SUN dataset: http://vision.princeton.edu/projects/2010/SUN/

163. ImageNet: http://www.image-net.org/
164. Software links for Deep Learning. Available (July 2017) in: http://deeplearning.net/software_links/
165. Datasets for Deep Learning. Available (July 2017) in: http://deeplearning.net/datasets/
166. Guosheng Lin, Chunhua Shen, Anton van den Hengel and Ian Reid. Efficient Piecewise Training of Deep Structured Models for Semantic Segmentation. arXiv:1504.01013v4 2016
167. Alejandro Newell, Kaiyu Yang, and Jia Deng. Stacked Hourglass Networks for Human Pose Estimation.. arXiv:1603.06937v1. 2016.
168. Georgia Gkioxari, Ross Girshick and Jitendra Malik. Contextual Action Recognition with R*CNN. arXiv:1505.01197v3 2016
169. Google Translate Application: https://play.google.com/store/apps/details?id=com.google.android.apps.translate
170. Soheil Bahrampour, Naveen Ramakrishnan, Lukas Schott, Mohak Shah. Comparative Study of Deep Learning Software Frameworks. arXiv:1511.06435v3 2016.
171. MNIST: http://yann.lecun.com/exdb/mnist/
172. CIFAR-10 and CIFAR-100: https://www.cs.toronto.edu/~kriz/cifar.html
173. Li Wan, Matthew Zeiler, Sixin Zhang, Yann LeCun, Rob Fergus. Regularization of Neural Network using DropConnect. International Conference on Machine Learning 2013
174. Benjamin Graham. Fractional Max-Pooling. arXiv:1412.6071v4. 2015
175. Djork-Arne Clevert, Thomas Unterthiner and Sepp Hochreiter. Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs). arXiv:1511.07289v5. 2016
176. Yu Xiang, Wongun Choi, Yuanqing Lin, and Silvio Savarese. Subcategory-aware Convolutional Neural Networks for Object Proposals and Detection. arXiv:1604.04693v1 2016
177. S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In ICML, 2015
178. Caffe Model Zoo: http://caffe.berkeleyvision.org/model_zoo.html
179. M. Jaderberg, K. Simonyan, A. Zisserman, K. Kavukcuoglu: Spatial transformer networks. In: NIPS. (2015)
180. A. Handa, M. Bloesch, V. Patraucean, S. Stent, J. McCormac, A. Davison, gvnn: Neural Network Library for Geometric Computer Vision. In: ECCV 2016.
181. R. Verschae, J. Ruiz-del-Solar. Object Detection: Current and Future Directions, Frontiers in Robotics and AI, Article 29, Vol. 2, Nov. 2015.
182. Y. Bengio. Deep learning of representations for unsupervised and transfer learning," in ICML Unsupervised and Transfer Learning, Volume 27 of JMLR Proceedings, eds I. Guyon, G. Dror, V. Lemaire, G. W. Taylor, and D. L. Silver (Bellevue: JMLR.Org), 17–36, year 2012.
183. D. Kotzias, M. Denil, P. Blunsom,and N. de Freitas. Deep Multi-Instance Transfer Learning. CoRR, abs/1411.3128, year 2014.
184. M. Mathieu, M. Henaff and Y. LeCun. Fast training of convolutional networks through ffts, arXiv preprint arXiv:1312.5851, 2013
185. D.O. Hebb. The Organization of Behavior. New York: Wiley & Sons, 1949
186. T. Poggio, F. Girosi. Regularization Algorithms for Learning that are Equivalent to Multilayer Networks. Science, Volume 247, Issue 4945,pp. 978-982, 1990
187. K. Funahashi. On the approximate realization of continuous mappings by neural networks. Neural Networks, Volume 2 Issue 3, 1989. Pages183-192
188. Jia Deng, Alexander C. Berg, Kai Li, and Li Fei-Fei. 2010. What does classifying more than 10,000 image categories tell us?. In Proceedings of the 11th European conference on Computer vision: Part V (ECCV'10), Kostas Daniilidis, Petros Maragos, and Nikos Paragios (Eds.). Springer-Verlag, Berlin, Heidelberg, 71-84
189. A. Saxena, H. Pandya, G. Kumar, K. M. Krishna. Exploring Convolutional Networks for End-to-End Visual Servoing. http://robotics.iiit.ac.in/people/harit.pandya/servonets/
190. Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A. Rusu, Joel Veness, Marc G. Bellemare, Alex Graves, Martin Riedmiller, Andreas K. Fidjeland, Georg Ostrovski, Stig Petersen, Charles Beattie, Amir Sadik, Ioannis Antonoglou, Helen King, Dharshan Kumaran, Daan Wierstra, Shane Legg, Demis Hassabis. Human-level Control through Deep Reinforcement Learning. In Nature, 518: 529–533, 2015.
191. T Lei, L Ming. A robot exploration strategy based on q-learning network. Real-time Computing and Robotics (RCAR), IEEE International Conference on, 6-10 June 2016.
192. Yuke Zhu, Roozbeh Mottaghi, Eric Kolve, Joseph J. Lim, Abhinav Gupta, Li Fei-Fei, Ali Farhadi. Target-driven Visual Navigation in Indoor Scenes using Deep Reinforcement Learning. arXiv preprint arXiv:1609.05143.
193. Piotr Mirowski, Razvan Pascanu, Fabio Viola, Hubert Soyer, Andrew J. Ballard, Andrea Banino, Misha Denil, Ross Goroshin, Laurent Sifre, Koray Kavukcuoglu, Dharshan Kumaran, Raia Hadsell. Learning to Navigate in Complex Environments. Under

review as a conference paper at ICLR 2017.

194. Alborz Geramifard, Thomas J. Walsh, Stefanie Tellex, Girish Chowdhary, Nicholas Roy, and Jonathan P. How. 2013. A Tutorial on Linear Function Approximators for Dynamic Programming and Reinforcement Learning. Found. Trends Mach. Learn. 6, 4 (December 2013), 375- 451. DOI=http://dx.doi.org/10.1561/2200000042

195. Ziyu Wang, Victor Bapst, Nicolas Heess, Volodymyr Mnih, Remi Munos, Koray Kavukcuoglu, Nando de Freitas. Sample Efficient Actor- Critic with Experience Replay. arXiv:1611.01224 [cs.LG]. 2016

196. David Silver, Aja Huang, Chris J. Maddison, Arthur Guez, Laurent Sifre, George van den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, Sander Dieleman, Dominik Grewe, John Nham, Nal Kalchbrenner, Ilya Sutskever, Timothy Lillicrap, Madeleine Leach, Koray Kavukcuoglu, Thore Graepel, Demis Hassabis. Mastering the game of Go with deep neural networks and tree search. Nature, Vol. 529, No. 7587. (27 January 2016), pp. 484-489, doi:10.1038/nature16961.

197. GPU-Based Deep Learning Inference: A Performance and Power Analysis. NVIDIA whitepaper. https://www.nvidia.com/content/tegra/embedded-systems/pdf/jetson_tx1_whitepaper.pdf

198. Nicolás Cruz, Kenzo Lobos-Tsunekawa, Javier Ruiz-del-Solar. Using Convolutional Neural Networks in Robots with Limited Computational Resources: Detecting NAO Robots while Playing Soccer. arXiv:1706.06702 (2017)

199. Jimmy Ren, Xiaohao Chen, Jianbo Liu, Wenxiu Sun, Jiahao Pang, Qiong Yan, Yu-Wing Tai, and Li Xu. Accurate Single Stage Detector Using Recurrent Rolling Convolution. arXiv preprint arXiv:1704.05776 (2017).

200. Zhaowei Cai, Quanfu Fan, Rogerio S. Feris, and Nuno Vasconcelos. A unified multi-scale deep convolutional neural network for fast objectdetection. In European Conference on Computer Vision, pp. 354-370. Springer International Publishing, 2016.

201. Y. Xiang, W. Choi, Y. Lin and S. Savarese: Subcategory-aware Convolutional Neural Networks for Object Proposals and Detection. IEEE Winter Conference on Applications of Computer Vision (WACV) 2017.

202. Yousong Zhu, Jinqiao Wang, Chaoyang Zhao, Haiyun Guo, and Hanqing Lu. Scale-Adaptive Deconvolutional Regression Network for Pedestrian Detection. In Asian Conference on Computer Vision, pp. 416-430. Springer, Cham, 2016.

203. F. Yang, W. Choi and Y. Lin: Exploit All the Layers: Fast and Accurate CNN Object Detector with Scale Dependent Pooling and Cascaded Rejection Classifiers. Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition 2016.

204. Jianxiong Xiao, James Hays, Krista A. Ehinger, Aude Oliva, and Antonio Torralba. Sun database: Large-scale scene recognition from abbeyto zoo. In Computer vision and pattern recognition (CVPR), 2010 IEEE conference on, pp. 3485-3492. IEEE, 2010.

205. Bolei Zhou, Agata Lapedriza, Aditya Khosla, Aude Oliva, and Antonio Torralba. Places: A 10 million image database for scene recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence (2017).

206. J. Xiao, J. Hays, K. A. Ehinger, A. Oliva, and A. Torralba. SUN database: Large-scale scene recognition from abbey to zoo. In CVPR, 2010.

207. Shuran Song, Samuel P. Lichtenberg, and Jianxiong Xiao. Sun rgb-d: A rgb-d scene understanding benchmark suite. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 567-576. 2015.

208. Liang-Chieh, Chen George Papandreou, Florian Schroff, and Hartwig Adam. Rethinking atrous convolution for semantic image segmentation. arXiv preprint arXiv:1706.05587 (2017).

209. Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia. Pyramid scene parsing network. arXiv preprint arXiv:1612.01105 (2016).

210. Zifeng, Wu, Chunhua Shen, and Anton van den Hengel. Wider or deeper: Revisiting the resnet model for visual recognition. arXiv preprint arXiv:1611.10080 (2016).

211. Panqu Wang, Pengfei Chen, Ye Yuan, Ding Liu, Zehua Huang, Xiaodi Hou, and Garrison Cottrell. Understanding convolution for semantic segmentation. arXiv preprint arXiv:1702.08502 (2017).

212. Rui Zhang, Sheng Tang, Yongdong Zhang, Jintao Li, and Shuicheng Yan. Scale-Adaptive Convolutions for Scene Parsing. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2031-2039. 2017.

213. Ping Luo, Guangrun Wang, Liang Lin, and Xiaogang Wang. Deep Dual Learning for Semantic Image Segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2718-2726. 2017.

214. Jun Fu, Jing Liu, Yuhang Wang, and Hanqing Lu. Stacked Deconvolutional Network for Semantic Segmentation. arXiv preprint arXiv:1708.04943 (2017).

215. Guangrun Wang, Ping Luo, Liang Lin, and Xiaogang Wang. Learning object interactions and descriptions for semantic image segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5859-5867. 2017.

216. ILSVRC 2016 scene classification. http://image-net.org/challenges/LSVRC/2016/results

217. Places Challenge 2016: http://places2.csail.mit.edu/results2016.html

218. Jifeng Dai, Yi Li, Kaiming He, Jian Sun. R-FCN: Object Detection via Region-based Fully Convolutional Networks. arXiv:1605.06409

219. Jimmy Ren, Xiaohao Chen, Jianbo Liu, Wenxiu Sun, Jiahao Pang, Qiong Yan, Yu-Wing Tai, Li Xu. Accurate Single Stage Detector Using Recurrent Rolling Convolution. CVPR 2017.

220. Jonathan Huang, Vivek Rathod, Chen Sun, Menglong Zhu, Anoop Korattikara, Alireza Fathi, Ian Fischer, Zbigniew Wojna, Yang Song, Sergio Guadarrama, Kevin Murphy. Speed/accuracy trade-offs for modern convolutional object detectors. arXiv:1611.10012. 2017

221. Jie Hu, Li Shen, Gang Sun. Squeeze-and-Excitation Networks. arXiv:1709.01507

222. Wang F, Jiang M, Qian C, et al. Residual Attention Network for Image Classification. arXiv preprint arXiv:1704.06904, 2017.

223. Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, Alex Alemi. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. AAAI. 2017: 4278-4284.

224. Olaf Ronneberger, Philipp Fischer, Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. arXiv preprint arXiv:1505.04597, 2015.

225. Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, Serge Belongie. Feature pyramid networks for object detection.arXiv preprint arXiv:1612.03144, 2016.

226. Abhinav Shrivastava, Rahul Sukthankar, Jitendra Malik, Abhinav Gupta. Beyond skip connections: Top-down modulation for object detection. arXiv preprint arXiv:1612.06851, 2016.

227. Xingyu Zeng, Wanli Ouyang, Junjie Yan, Hongsheng Li, Tong Xiao, Kun Wang, Yu Liu, Yucong Zhou, Bin Yang, Zhe Wang, Hui Zhou, Xiaogang Wang. Crafting GBD-Net for Object Detection. arXiv preprint arXiv:1610.02579, 2016.

228. Yu Chen, Chunhua Shen, Xiu-Shen Wei, Lingqiao Liu, Jian Yang. Adversarial PoseNet: A Structure-aware Convolutional Network for Human Pose Estimation. arXiv:1705.00389

229. J. Wang, Z. Liu, Y. Wu, J. Yuan. Learning actionlet ensemble for 3d human action recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 36, no. 5, pp. 914-927, 2014.

230. Huiwen Guo, Xinyu Wu, Wei Fengab. Multi-stream deep networks for human action classification with sequential tensor decomposition. Signal Processing, Volume 140, November 2017, Pages 198-206

231. Y.H. Ng, M. Hausknecht, S. Vijayanarasimhan, O. Vinyals, R. Monga, G. Toderici. Beyond short snippets: deep networks for video classification. IEEE Conference on Computer Vision and Pattern Recognition(CVPR) (2015), pp. 4694-4702

232. Z. Wu, Y.G. Jiang, X. Wang, H. Ye, X. Xue. Multi-stream multi-class fusion of deep networks for video classification. ACM onMultimedia Conference (2016), pp. 791-800

233. G. Lev, G. Sadeh, B. Klein, L. Wolf. RNN fisher vectors for action recognition and image annotation. European Conference on Computer Vision(ECCV) (2016), pp. 833-850

234. Du Tran, Lubomir Bourdev, Rob Fergus, Lorenzo Torresani, Manohar Paluri. Learning Spatiotemporal Features with 3D Convolutional Networks. arXiv:1412.0767 (2014)

235. Gao Huang, Zhuang Liu, Kilian Q. Weinberger, Laurens van der Maaten. Densely Connected Convolutional Networks. arXiv:1608.06993, 2016.