Machine Learning in Robotics

Applications, Challenges & Potentials

Field and Service Robots

M.Eng Jasper Hatilima, MSc. Nasser Gyagenda, MSc. Omar Gamal
Prof. Dr.-Ing. Dr. h. c. Hubert Roth

Contents

- Definitions
- Example application areas for field and service (F&S) robots
- Capabilities of F&S robots with respect to the AI components: With a specific emphasis on learning.
- Machine learning for perception:
  - field robots
  - service robots
- General challenges

FIELD AND SERVICE ROBOTS

Field robots mostly used in unstructured or uncontrolled environments, including underwater, in mines, in forests, on farms and in the air [1]. Mostly outdoor applications.

Examples: autonomous vehicles, weeding and harvesting robots, powerline inspection aerial robots, search and rescue robots, robots for information gathering in disaster/dangerous areas.

Service robots are mainly for indoor or limited outdoor applications and usually target common people as main users.

Asimo, Pepper, Savioke Relay.

EXAMPLES & APPLICATIONS OF F&S ROBOTS

Applications for Field Robots – Facility Inspection

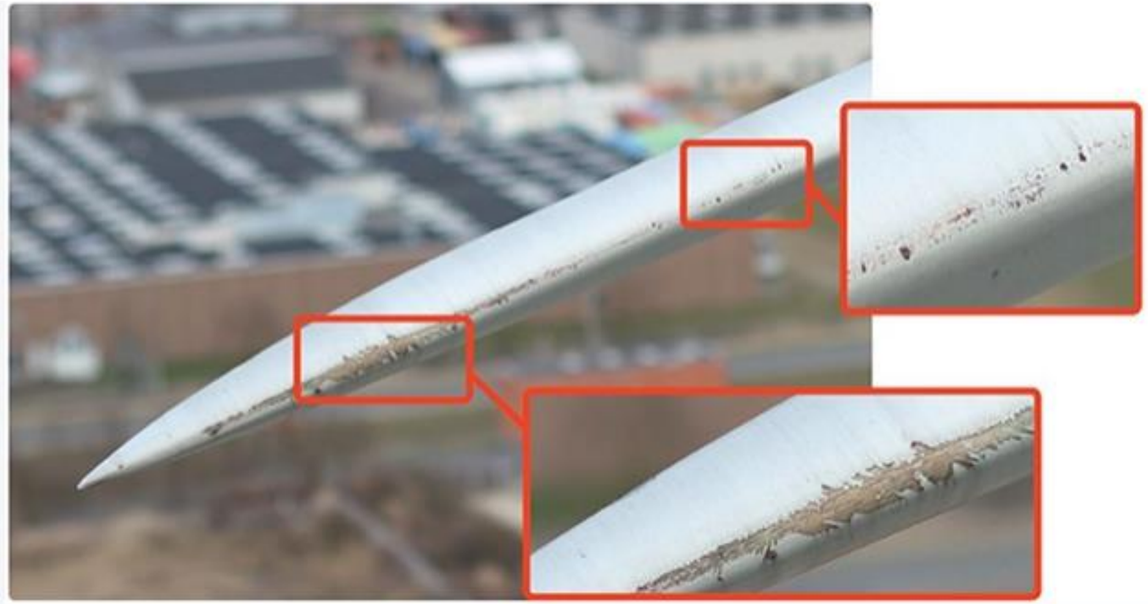
https://environm14.wordpress.com/2014/10/03/wind-turbines-2/

http://web.cornis.fr/en/pages/panoblade

Fig 1: Wind Turbine Inspection

Applications for Field Robots (Cont'd)

http://www.asctec.de/en/renewable-uk-cyberhawk-drone-wind-turbine-service/



https://forcetechnology.com/en/services/drone-inspection-of-wind-turbines-onshore-and-offshore

Fig 2: Wind Turbine Inspection using Aerial Robots

Applications for Field Robots (Cont’d)

https://www.chpinspection.com/drone

http://www.pacetechnologies.com/infrared-corona-electrical-scanning

Fig 3: Powerline Inspection (thermal imaging)

Examples of Service Robots

https://www.omnifi.co.uk/monitor/hotels-savioke-robot/

https://www.bettshow.com/bett-products-list/pepper-humanoid-robot

https://robots.nu/de/roboter/Asimo

Fig 4: Service Robots

CAPABILITIES OF FIELD & SERVICE ROBOTS

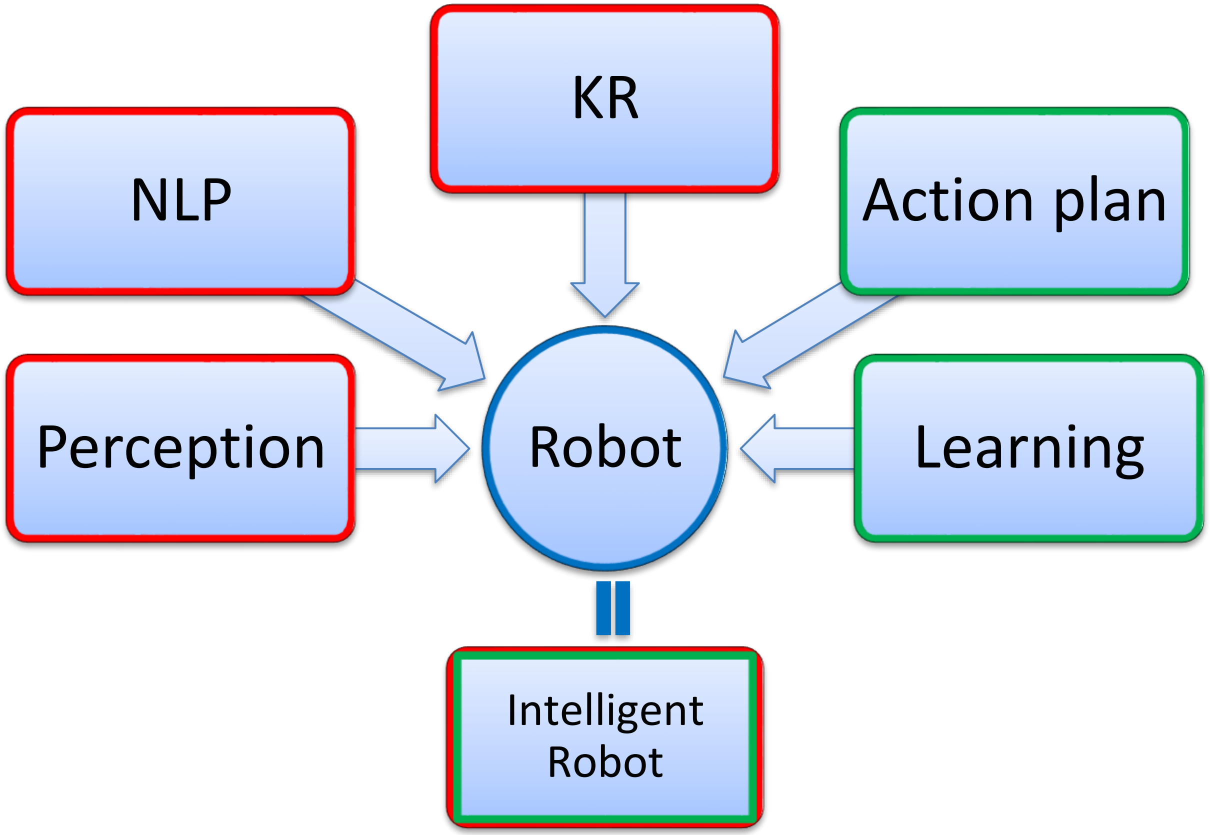
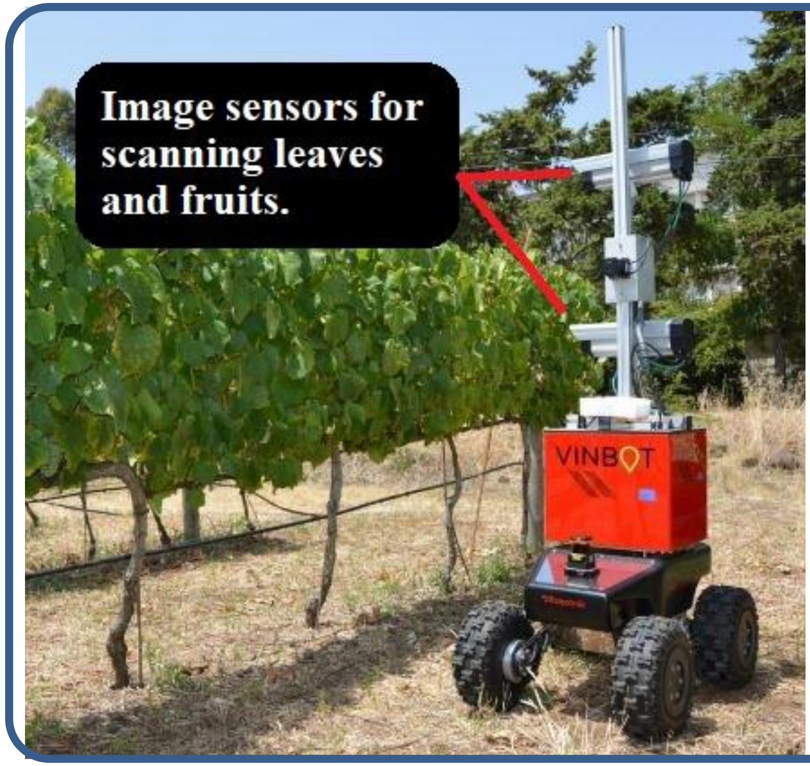


Fig 5: Capabilities of an Intelligent Robot

Field & Service Robot Capabilities based on Machine Learning

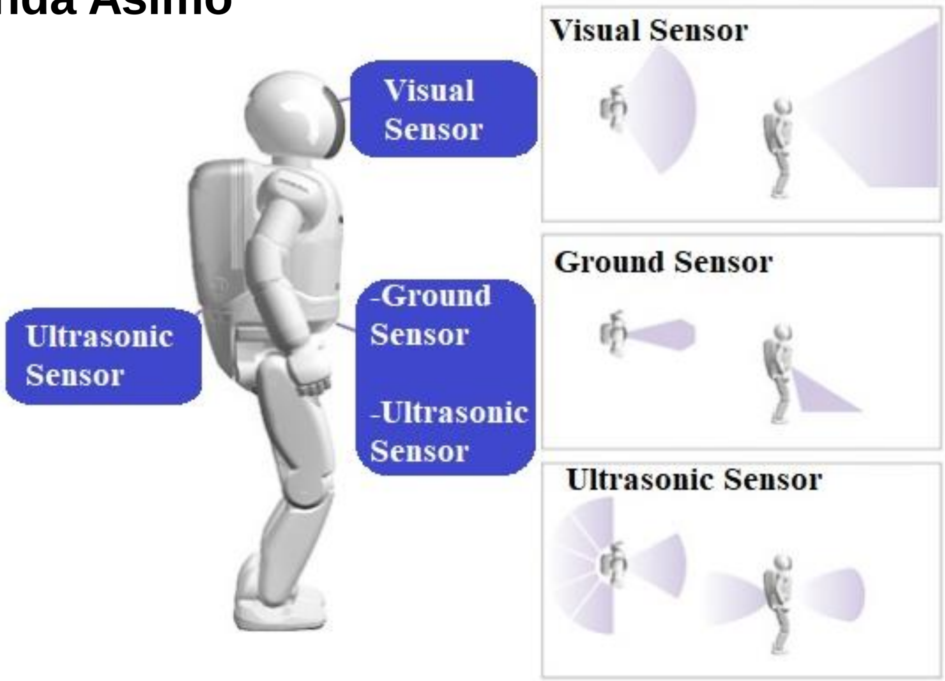
- **Service robots** must be able to understand natural language commands, as well as generate responses. (**NLP**)
- It is common for **field robots** to be able to process input and perform intelligent actions based on onboard intelligence. This means that AI models should be deployable on the agent (**edge device problem solving**).
- Additionally, both field and service robots must be able to **perceive** the environment. So **machine perception** is key.

MACHINE LEARNING FOR PERCEPTION IN F&S ROBOTS



Vinbot Captures:
- the amount of leaves,
- Amount of grapes,
- And estimates the leaf-to-fruit ratio, thereby predicting yields

Honda Asimo



https://robohub.org/protecting-european-wine-vinbot-rover-optimises-harvest-and-quality/

https://electroviees.wordpress.com/2013/07/04/humanoid-robots/

Fig 6: Perception in Field and Service Robots

Machine Learning Perception in Field Robots

Robot vision has traditionally been implemented using classical image processing techniques that use Hand-Crafted Features (HCF).

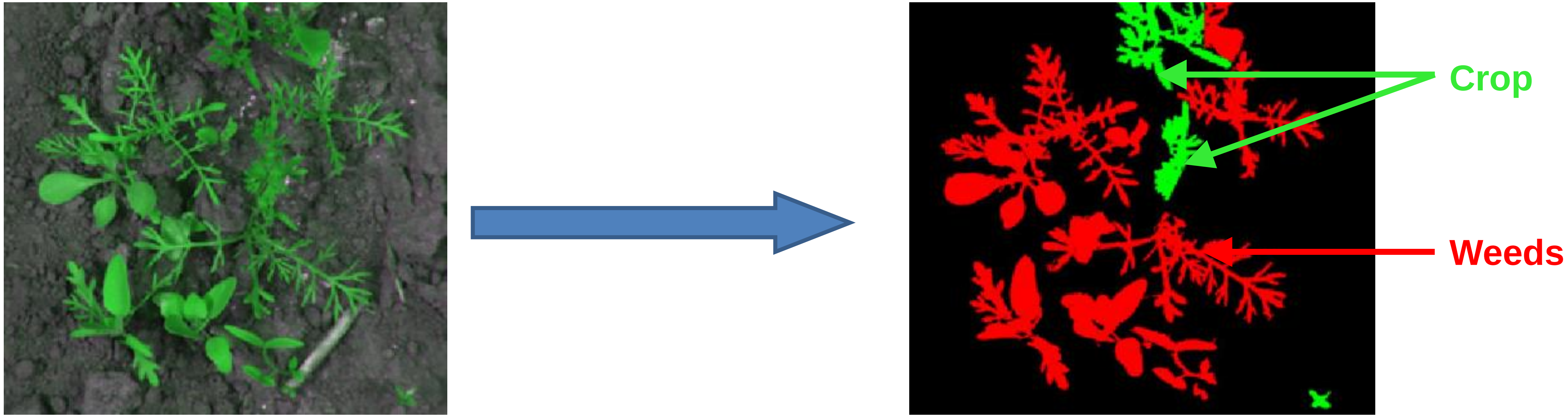


Fig 7: Weeds vs Crop Segmentation [4]

Some notable agricultural robot platforms using the perception of crops/weeds are AgBot II and Harvey from QUT.

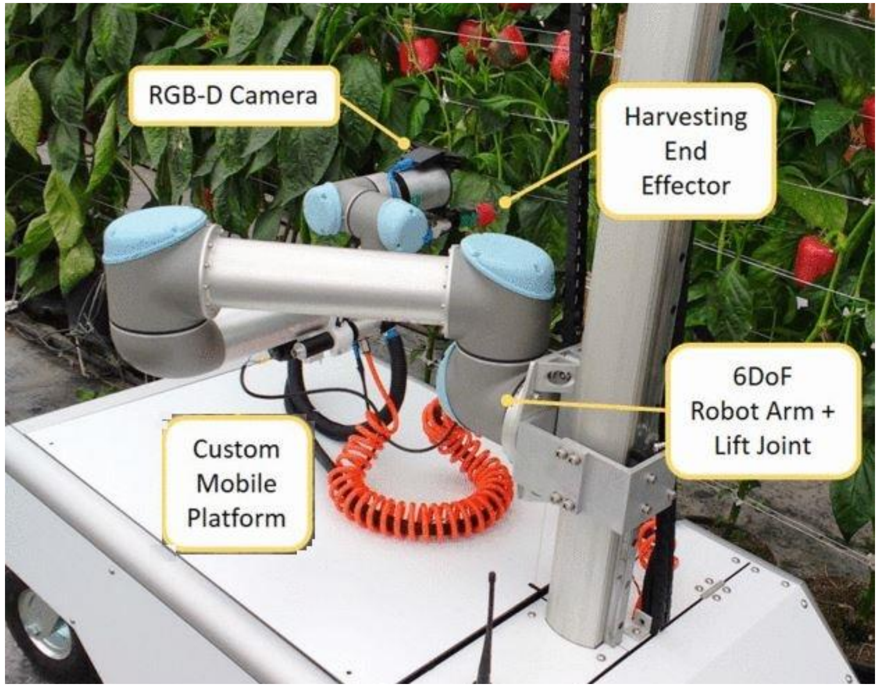


Fig 8: AgBot II (weed management robot) and Harvey (harvesting robot) respectively [4]

Limitation of HCF in Robot Perception

- Classical techniques use pixel level statistics [4] or shape features (area, major axis and compactness) in a Naïve Bayesian classifier to distinguish between crops and weeds [5].

- The methods above use hand-crafted features (HCF).

- The dependency on manual extraction of features places a limitation on the scalability of the methods, as well as the robustness to variations in image conditions.
- *Enter deep learning for perception in field and service robots!*

So how suitable is deep learning for field robot tasks?

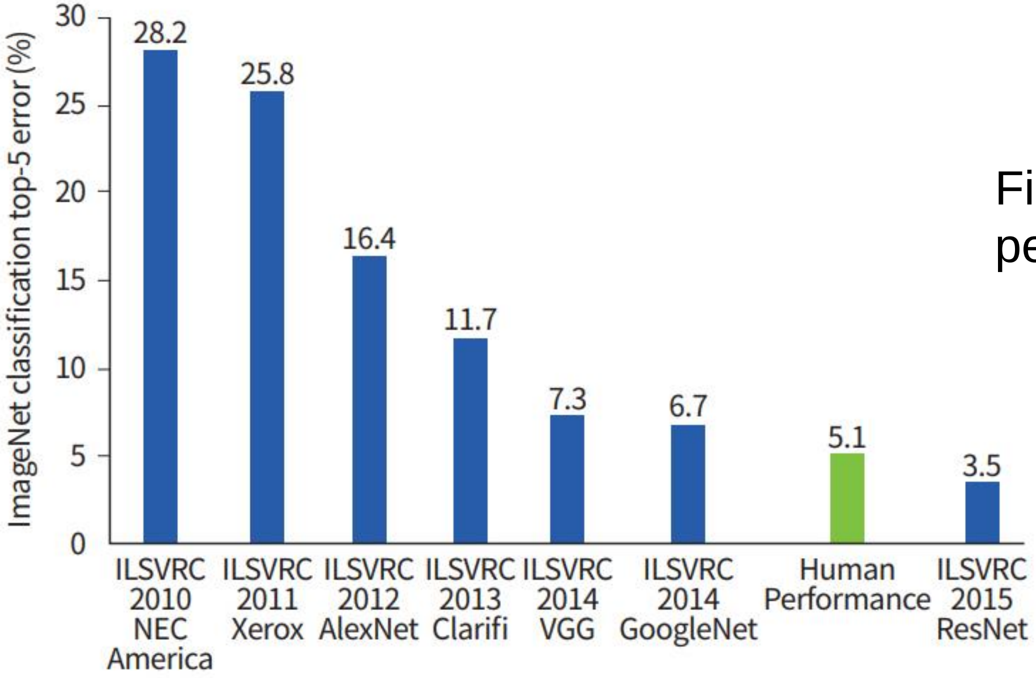


Fig 9: ILSVRC* Classification performance from 2010 to 2015 [5]

For object detection, gains have been made: R-CNN, YOLO, SSD, etc.

So can we readily replace HCF with these models in robotics?

*ILSVRC – ImageNet Large Scale Visual Recognition Competition

SOLUTIONS: Some Field Robots using DCNN

A DCNN (Adapted AlexNet) was used in[6] to extract features for use in a Random Forest (RF) classifier. This work also combined the CNN classifier output with a HCF classifier by using the sum rule[7].

The hybrid of DCNN and HCF showed better classification performance than each one of them singly applied.

Also robust to variation in scaling, occlusion, illumination.

A combination of CNN (Adapted AlexNet) and HCF

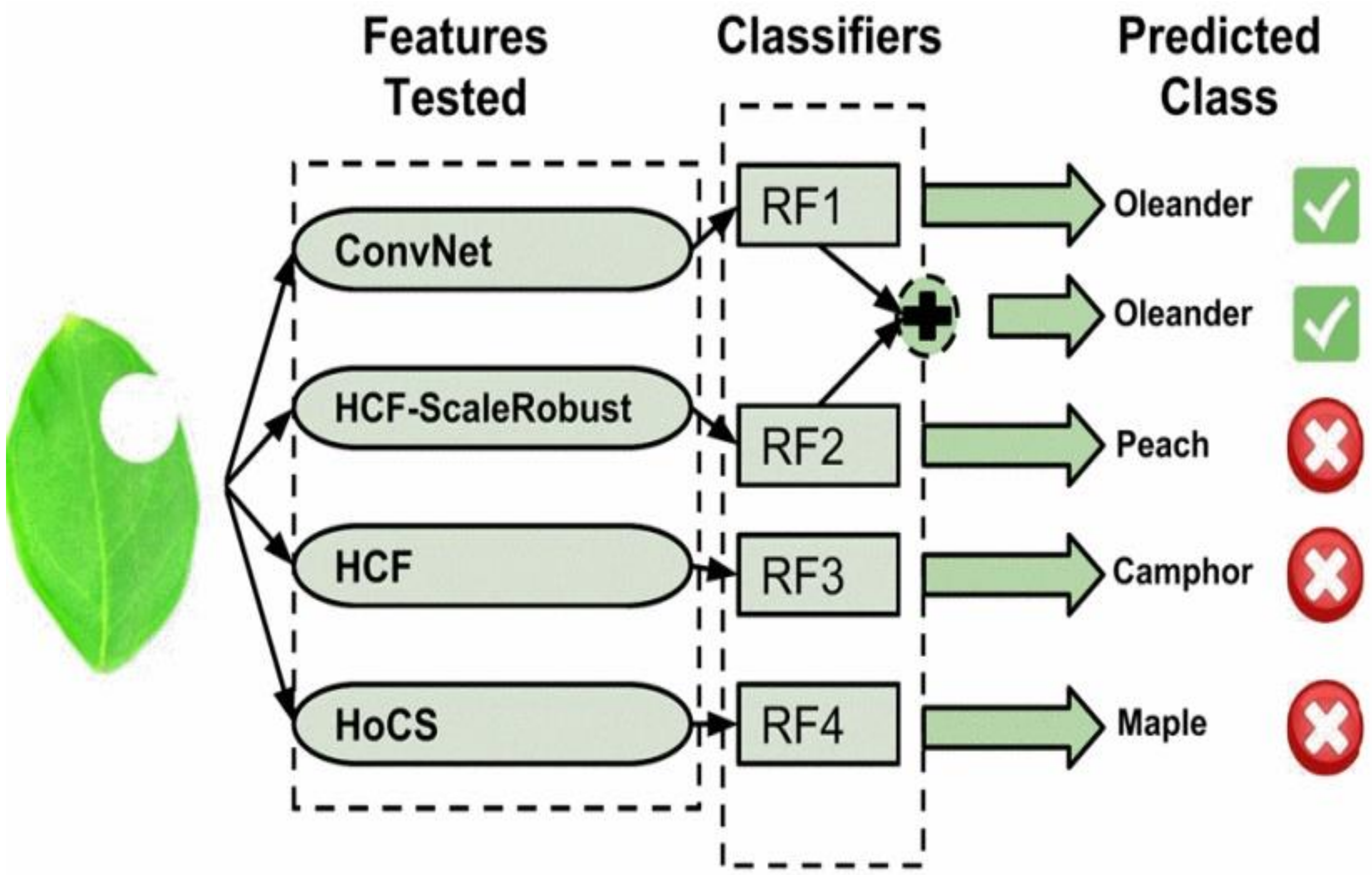


Fig 10: Classification based on various feature extraction techniques [6]

Sample results from adapted AlexNet on agricultural classification:

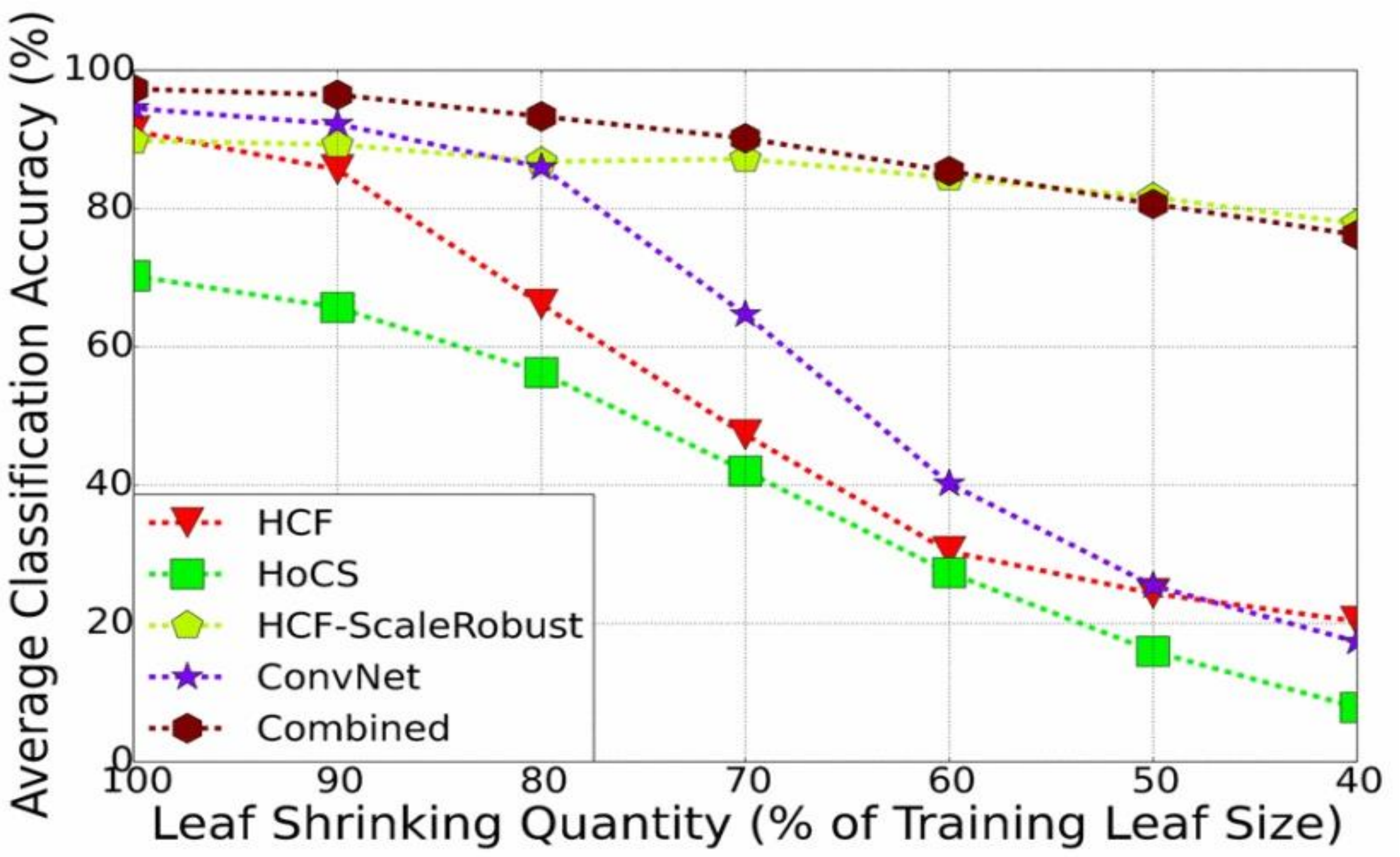


Fig 11: Classification accuracy w.r.t object shrinkage: CNN vs Image Processing vs Combination [6]

So why don't we fully go for CNNs in field & service robot perception?

- For most research in DCNNs, the focus has been to improve accuracy.
- Results: very deep models with millions of parameters and sometimes requiring training across GPU clusters: 128 GPUs in [8].
- But field and service robots do not have such high processing capabilities and memory capacities.
- Let us consider some of these aspects quantitatively…

Deep Learning Models and Hardware Platforms

Size of most networks(examples): Computation/Memory

Network	Number of Parameters	Size of Model
AlexNet	62 Million	240 MB
VGG-16	138 Million	533 MB [9]
ResNet	25 Million	102 MB [9]
Inception-v3	25 Million	96 MB [9]

Example target hardware: Xilinx Virtex-7 FPGA on-chip memory = 8.5 MB

Computational resources(examples): IoT vs Edge Computing

GPU	CUDA Cores	Memory (GB)	Mem. B/W (GB/s)	Power (W)
GTX TITAN X	3072	12	336.5	250
Jetson TX2	256	8	59.7	7.5
Jetson AGX Xavier	512	16	137	<=30

Desktop/Server [10]

Mobile/Edge [11]

Works to Optimise DCNN for Field Robot Usage

Several approaches have been taken to create lightweight models for use on mobile platforms. This offers advantages in terms of:

- Feasible embedded/FPGA/ASIC deployment – This will allow model deployment on memory-limited field and service robots.
- More efficient training without need for huge GPU clusters.
- Less overhead when exporting new models to clients – Over-the-air updates from servers to agents.

Some example works in optimizing DCNNs for field robots…

(1) SqueezeNet

Reduced the 240MB AlexNet to 0.47MB and maintain accuracy[12] by:

- **Replacing majority of 3x3 filters with 1x1 filters.** A 1x1 filter has 9X less parameters.
- **Decreasing number of input channels to 3x3 filters.** The total number of parameters in a layer that as 3x3 filters is equal to num. of input channels x num. of filters x (3x3). Besides decreasing filter sizes, SqueezeNet also reduced num. of input channels.
- **Downsample late in the network to maintain accuracy.**

The above SqueezeNet techniques are implemented using the SqueezeNet “Fire module” as indicated below:

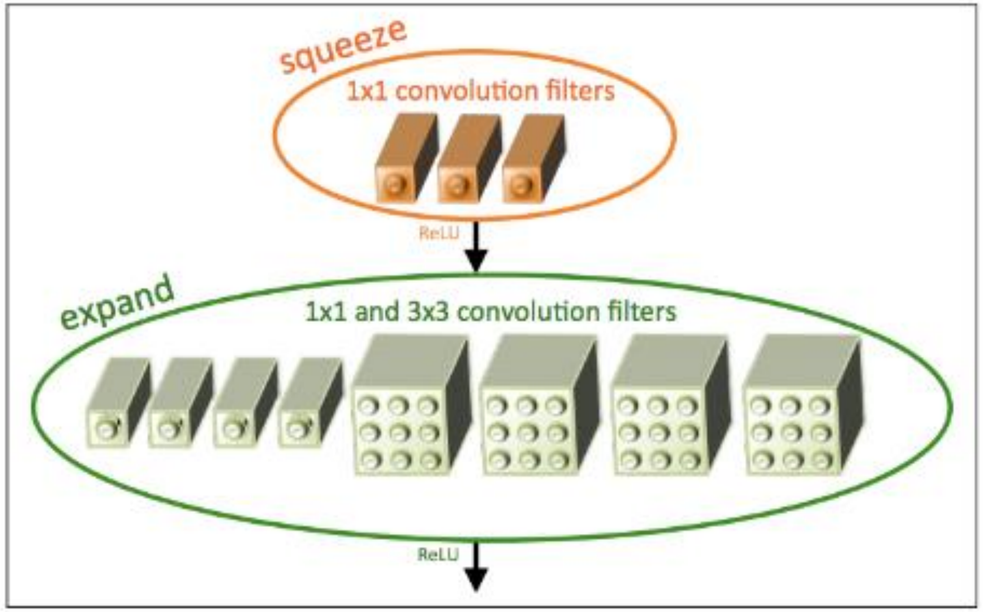


Fig 12: Organisation of convolution filters in the **Fire module** [12]

Reduced AlexNet by over 510X in size, 50X fewer parameters, same accuracy.

(2) Adapted Inception-v3 Ensemble

Another interesting model reduction technique was implemented by McCool et al. in [4] for a weeding robot. This involved three main steps:

- Adapt a pre-trained Inception-v3 model by using Agro-data for **transfer learning.**
- The adapted Inception-v3 model is then used as a **teacher** for a much smaller (lightweight) DCNN. The **student** network uses both logit output (before Softmax) to compare with teacher and classification loss to compare with ground truth.
- Create **ensemble** of lightweight models and find average decision.

Key outcomes of Adapted Inception-v3 ensemble:

- Trade-off between complexity and accuracy: The Inception-v3 after transfer learning had accuracy of 93.9% and 25 million parameters. The ensemble of **four lightweight networks** had accuracy of 90.3% but with only 1 million parameters.

- Reduced frame rate: The original Inception-v3 had over 5 fps while the adapted ensemble dropped to as low as 1.07 fps.

- Overall the lightweight model was small enough for use on field robot (AgBot).

Application of DL in Field and Service Robots - Search & Rescue

MAV trail following

In search and rescue missions, field robots must be able to follow man-made trails (such as those left be hickers).

A. Guisti et al. [13] proposed using CNNs to perceive where the trail is and then select a control signal for the MAV in order to remain on the trail. [**Pixel-to-Action Control**]

[13] was cast as a classification task by getting input from MAV front camera and then CNN outputs direction LEFT, STRAIGHT, RIGHT.

Application of DL in Field and Service Robots - Search & Rescue

MAV trail following

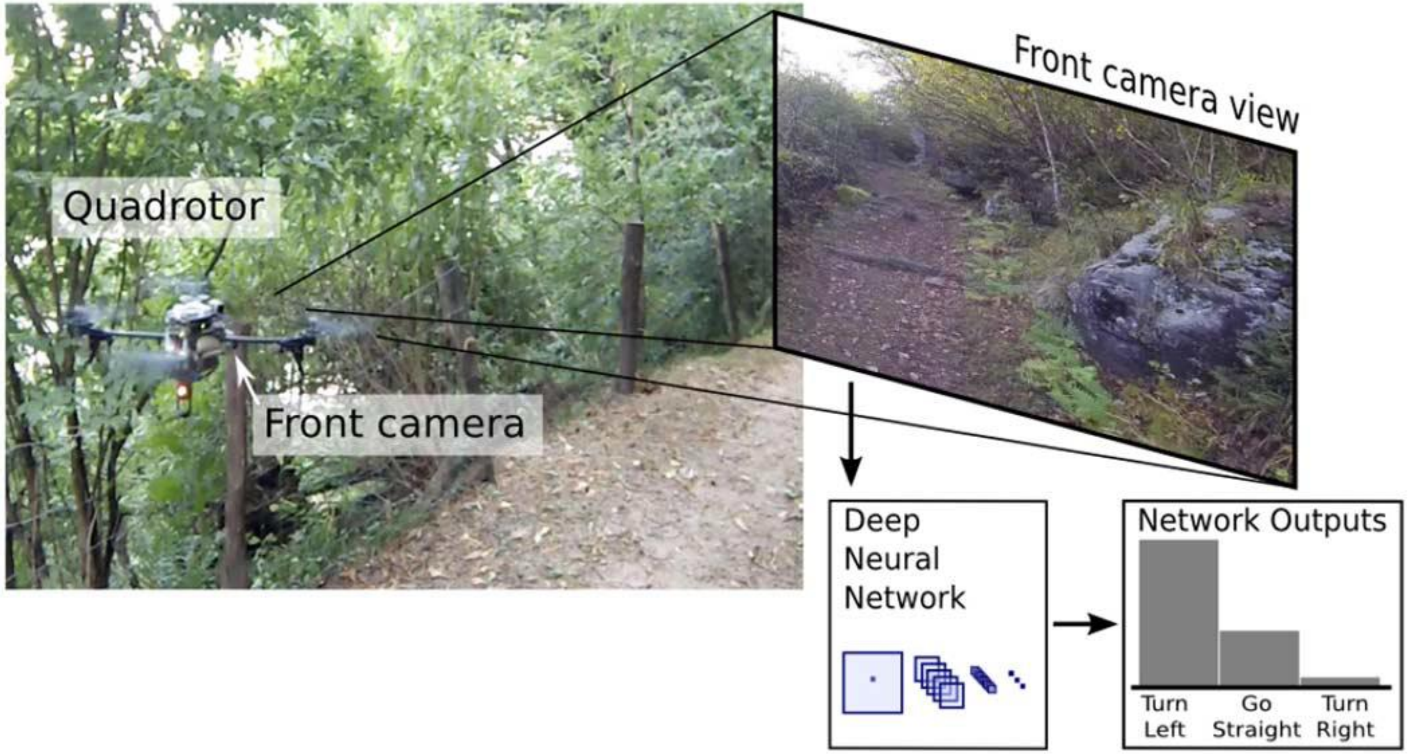


Fig 13: MAV Trail following [13]

Data collection was done by equipping a hiker with three cameras as indicated below. The two side cameras were placed 30º from the forward facing camera.

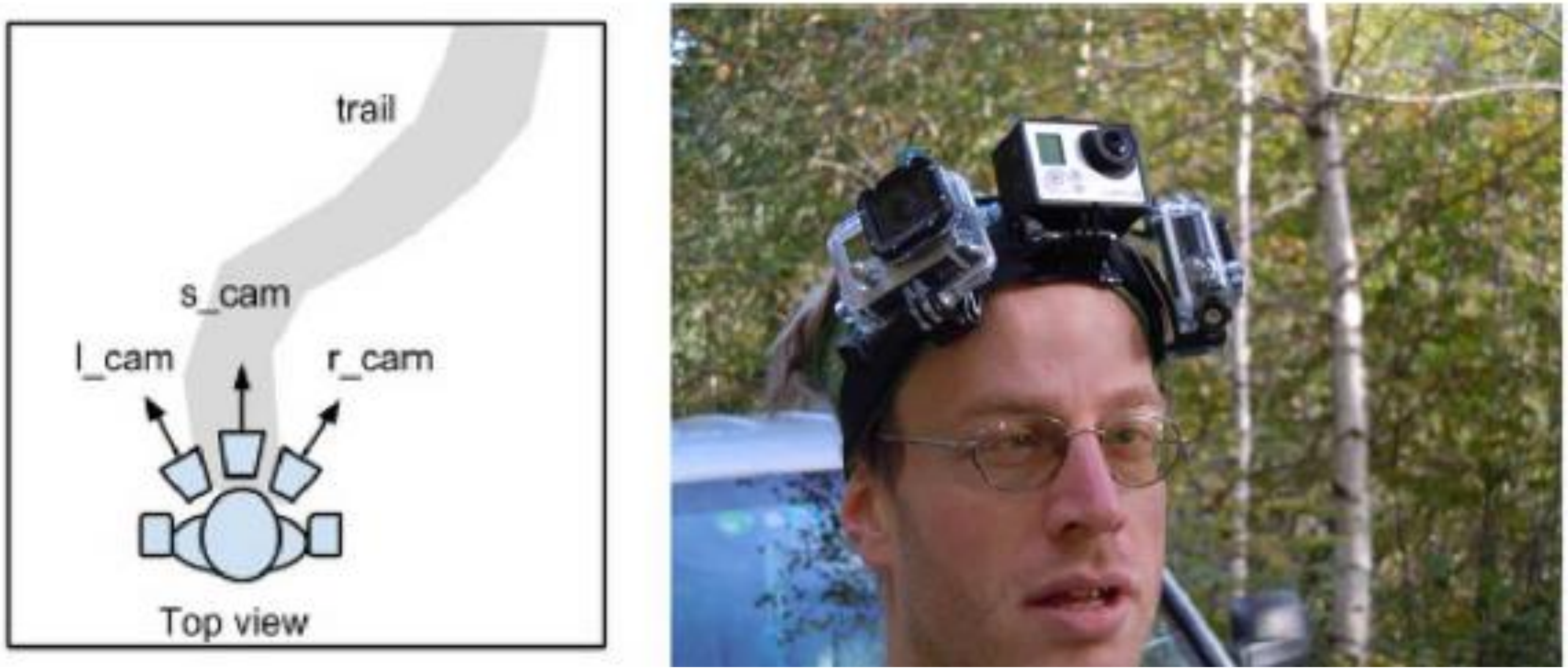


Fig 14: Creating the data set [13]

Output commands:

Fig 15: CNN output commands for MAV [13]

Application of DL in Service Robots

Perception-Action Learning in Pepper[14]

System capable of providing services to people in dynamic environments.

Integrated multi-module approach(Vision, speech, mapping).

System implemented on server with GTX Titan 12GB GPU. ROS communication between robot & server over WiFi (5GHz).

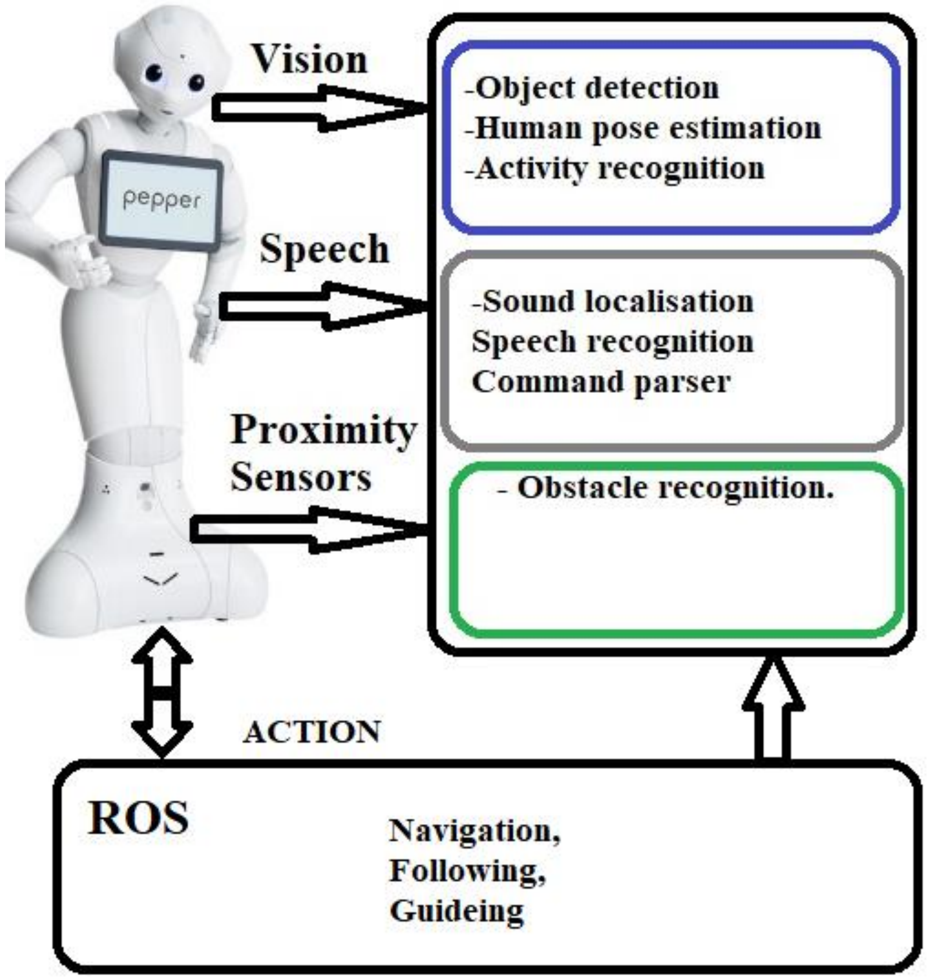


Fig 16: Service robot - Pepper

Sub modules:

Table 1: Some Perception Modules

Module	Description
Object Detection and Recognition	YOLOv2 (2016)
Scene Description	DenseCap (2016): Fully convolutional localization networks for captioning.
Speech Recognition	Google Speech API
Command Parser	Python Natural Language Toolkit (NLTK).

Table 2: Some Action Modules

Module	Description
Navigation	ROS, SLAM
Following	Control for following target person

Table 3: Some Learning Modules

Module	Description
Human Identification	Save person’s image for identification
Object and Person Position	In reference to map build in ROS Gmapping, save objects and person related information

General Challenges in learning for Field and Service Robots

- Big and accurate models vs edge computing power.
- Data sets: This challenge is even more when it comes to field robots and very specific field tasks. **Simulation data?**
- Metrics in deep learning for computer vision sometimes are not as stringent as requirements in robotics. For example, 90% accuracy in the top 5% would be very irrelevant in most robotics applications as there is no room for such an interval accuracy.

References

[1] C. Thorpe, H. Durant-Whyte, “Field Robots”, in *Proc. the 10th International Symposium of Robotics Research*, Lorne, Australia, 2001.

[2A] E&T: https://eandt.theiet.org/content/articles/2015/12/smartphone-based-object-recognition-paves-way-for-driverless-cars/

[2] University of Sydney: https://sydney.edu.au/engineering/our-research/robotics-and-intelligent-systems/australian-centre-for-field-robotics.html

[3] The Verge: https://www.theverge.com/circuitbreaker/2018/12/26/18156600/doomba-roomba-cleaning-maps-doom-levels-rich-whitehouse

[4] C. McCool, T. Perez and B. Upcroft, "Mixtures of Lightweight Deep Convolutional Neural Networks: Applied to Agricultural Robotics," *IEEE Robotics and Automotion Letters*, vol. 2, no. 3, pp. 1344-1351, 2017.

[5] https://www.pfmjournal.org/journal/Figure.php?xn=pfm-2018-00030.xml&id=f1-pfm-2018-00030&number=32&p_name=0752_32

[6] D. Hall, C. McCool, F. Dayoub, N. Sunderhauf and B. Upcroft, "Evaluation of Features for Leaf Classification in Challenging Conditions," in *Proc. IEEE Winter Conf. on Appl. of Computer Vision*, Waikoloa, HI, USA, 2015.

[7] J. Kittler, M. Hatef, R. P. W. Duin and J. Matas, "On Combining Classifiers," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 3, pp. 226-239, 1998.

[8] F. N. Iandola et al. “FireCaffe: Near-linear Acceleration of Deep Neural Network Training on Compute Clusters” in *Proc. IEEE Conf. on Comp. Vision and Patt. Recogn. (CVPR)*, June, 2016.

[9] https://www.pyimagesearch.com/2017/03/20/imagenet-vggnet-resnet-inception-xception-keras/

[10] https://www.geforce.com/hardware/desktop-gpus/geforce-gtx-titan-x/specifications

[11] https://devblogs.nvidia.com/jetson-tx2-delivers-twice-intelligence-edge/

[12] F. N. Iandola et al. “SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size” CoRR, 2016.

[13] A. G. Giusti et al. “A machine learning approach to visual perception of forest trails for mobile robots”, *IEEE Robotics and Automation Letters*, July, 2016

[14] Beom-Jin Lee et al. “Perception-Action-Learning System for Mobile Social-Service Robots using Deep Learning”, in *Association for Advancement of Artificial Intelligence (AAAI)*, 2018