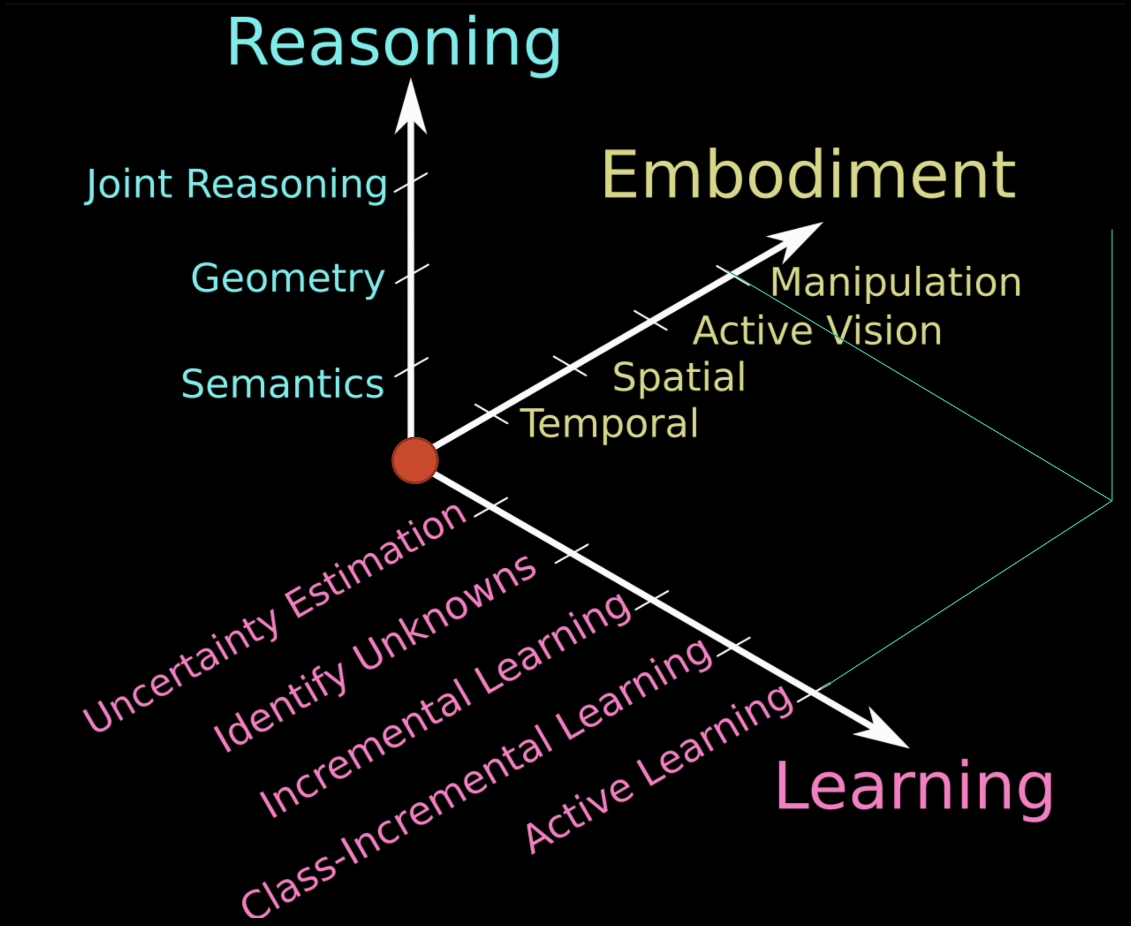
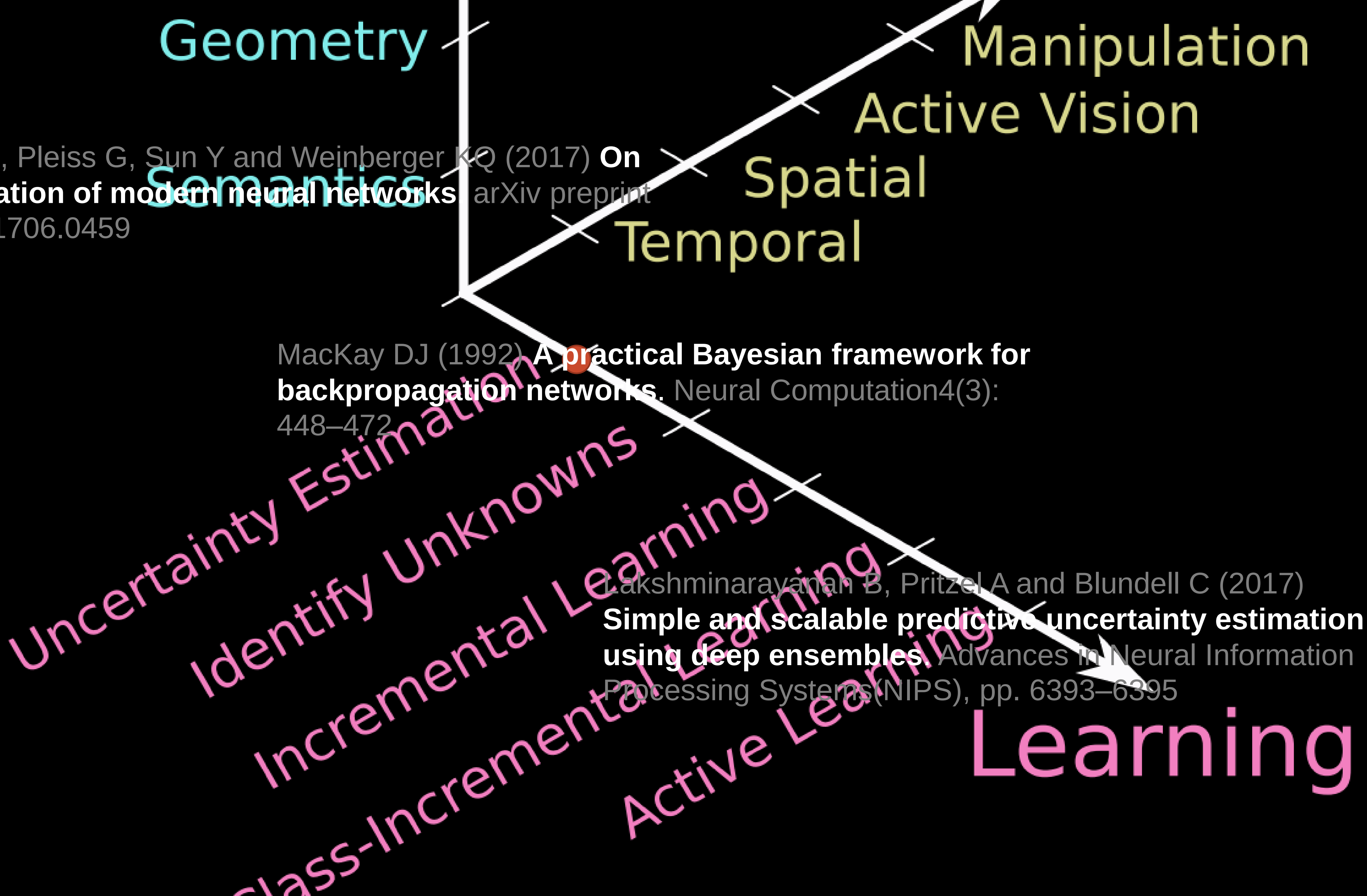
Limits and potentials of deep learning in robotics

Flexible learning reading group

Aravind Battaje

Challenges in robotic vision

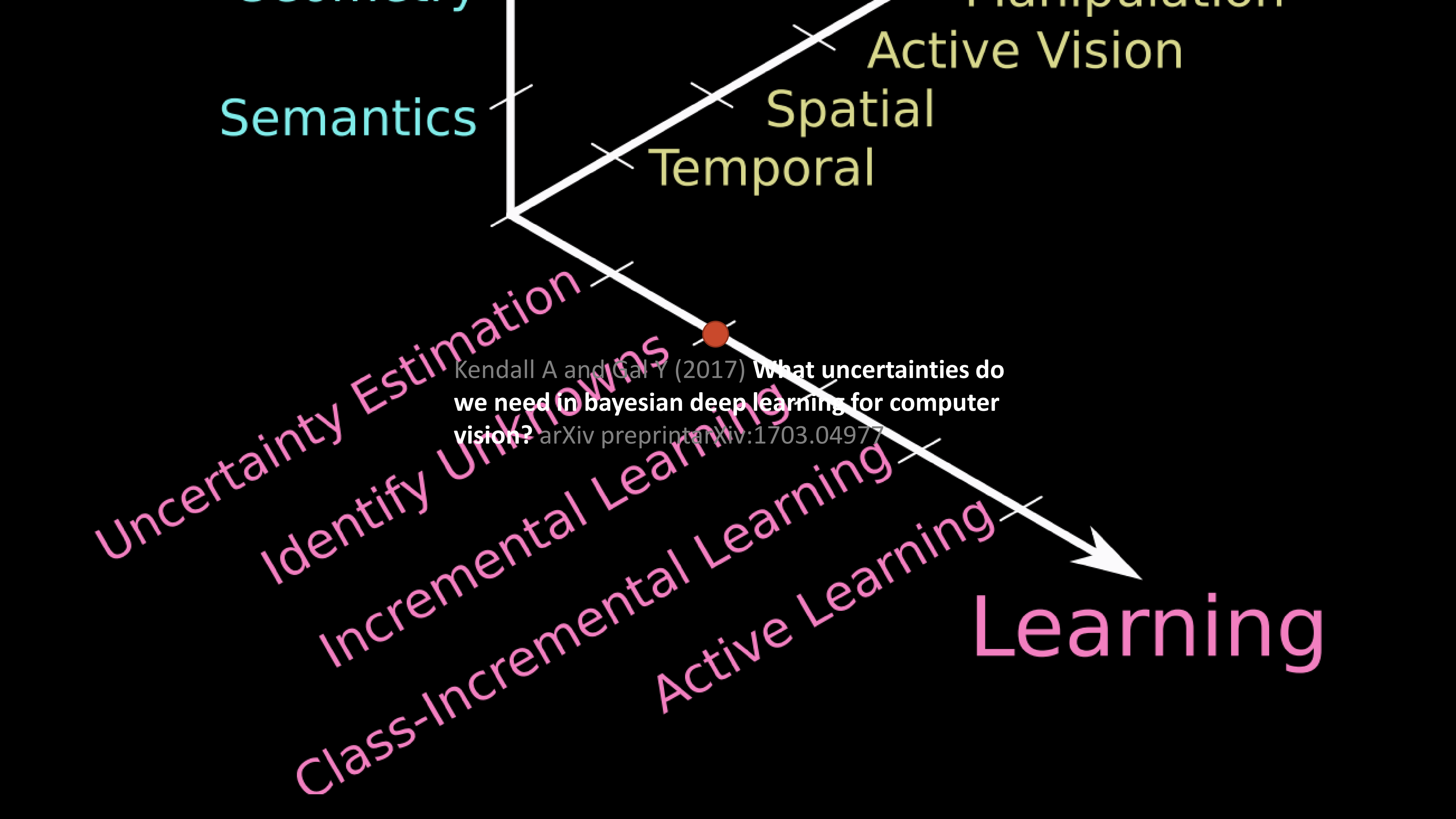




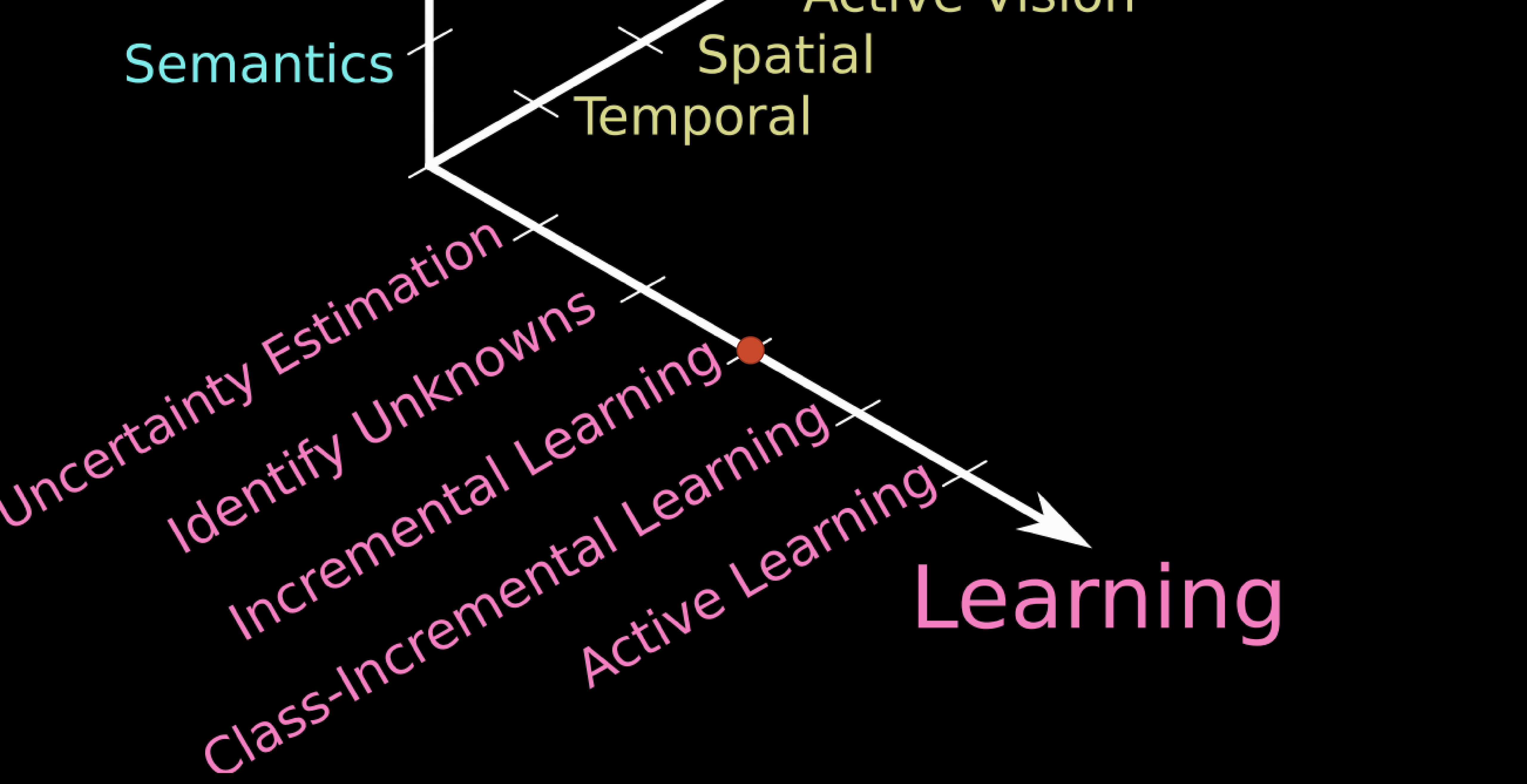
Guo C, Pleiss G, Sun Y and Weinberger KQ (2017) **On calibration of modern neural networks**. arXiv preprint arXiv:1706.0459

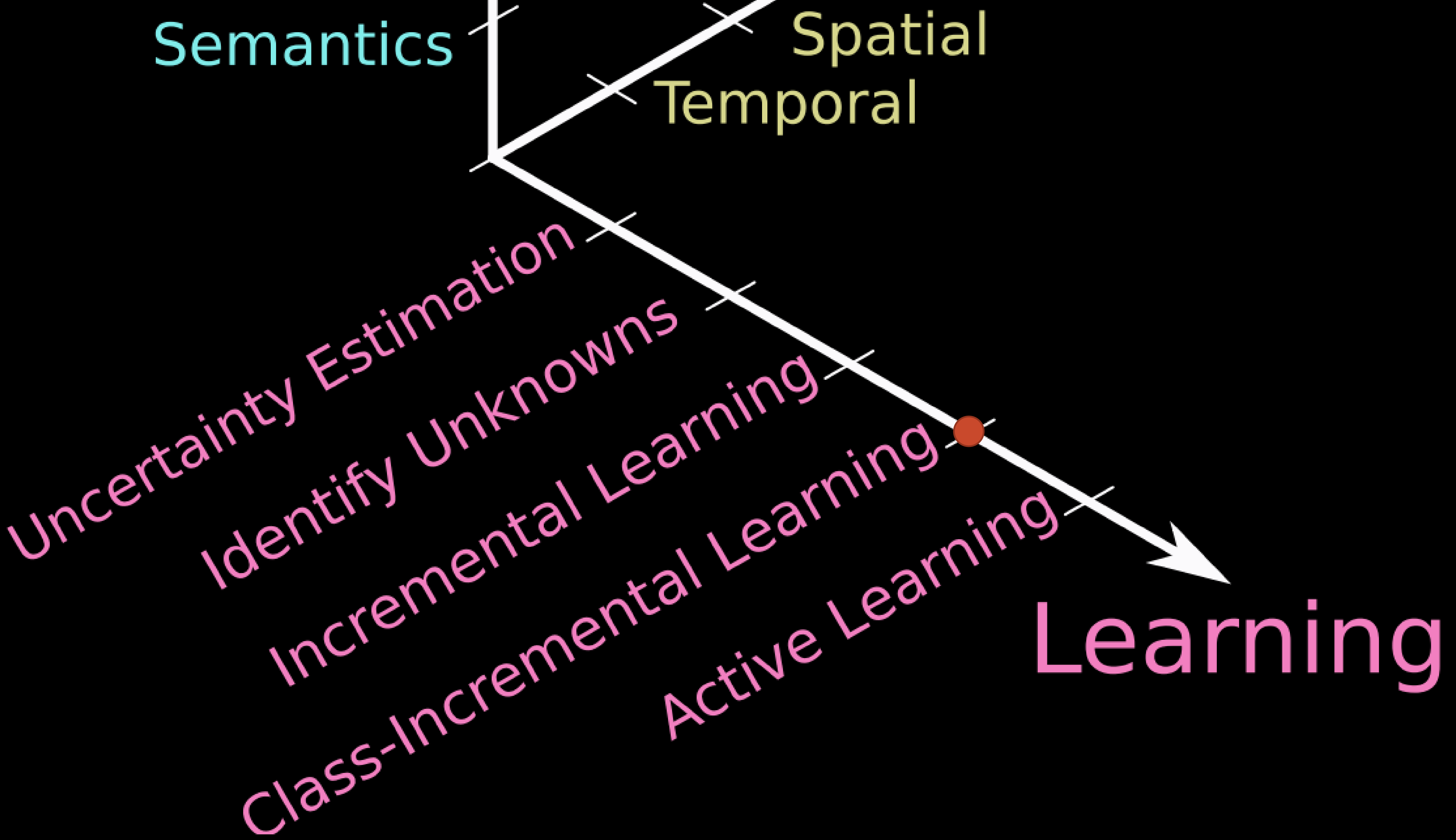
MacKay DJ (1992) **A practical Bayesian framework for backpropagation networks**. Neural Computation4(3): 448–472

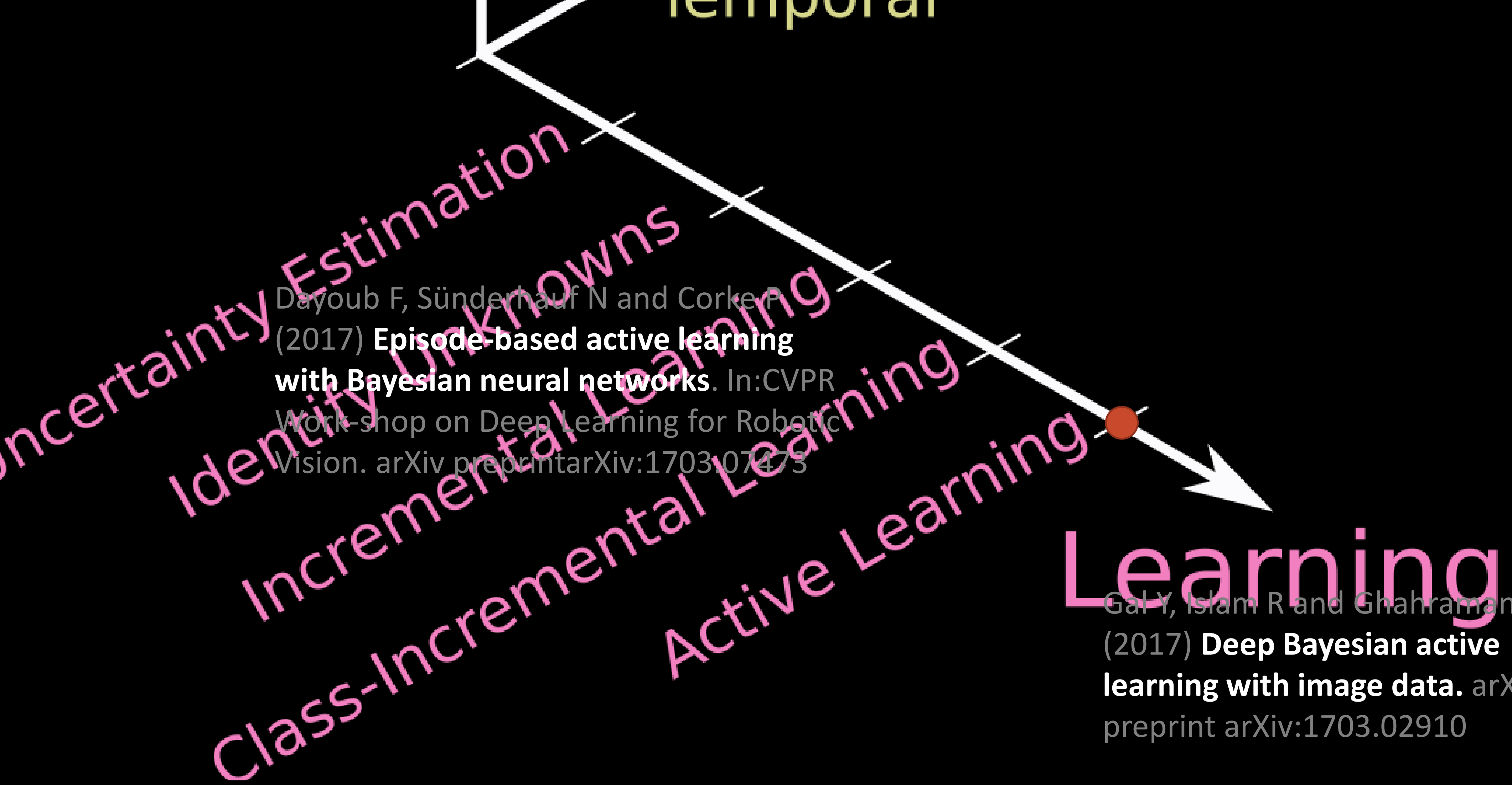
Lakshminarayanan B, Pritzel A and Blundell C (2017) **Simple and scalable predictive uncertainty estimation using deep ensembles**. Advances in Neural Information Processing Systems(NIPS), pp. 6393–6395



Kendall A and Gal Y (2017) **What uncertainties do we need in bayesian deep learning for computer vision?** arXiv preprintarXiv:1703.04977

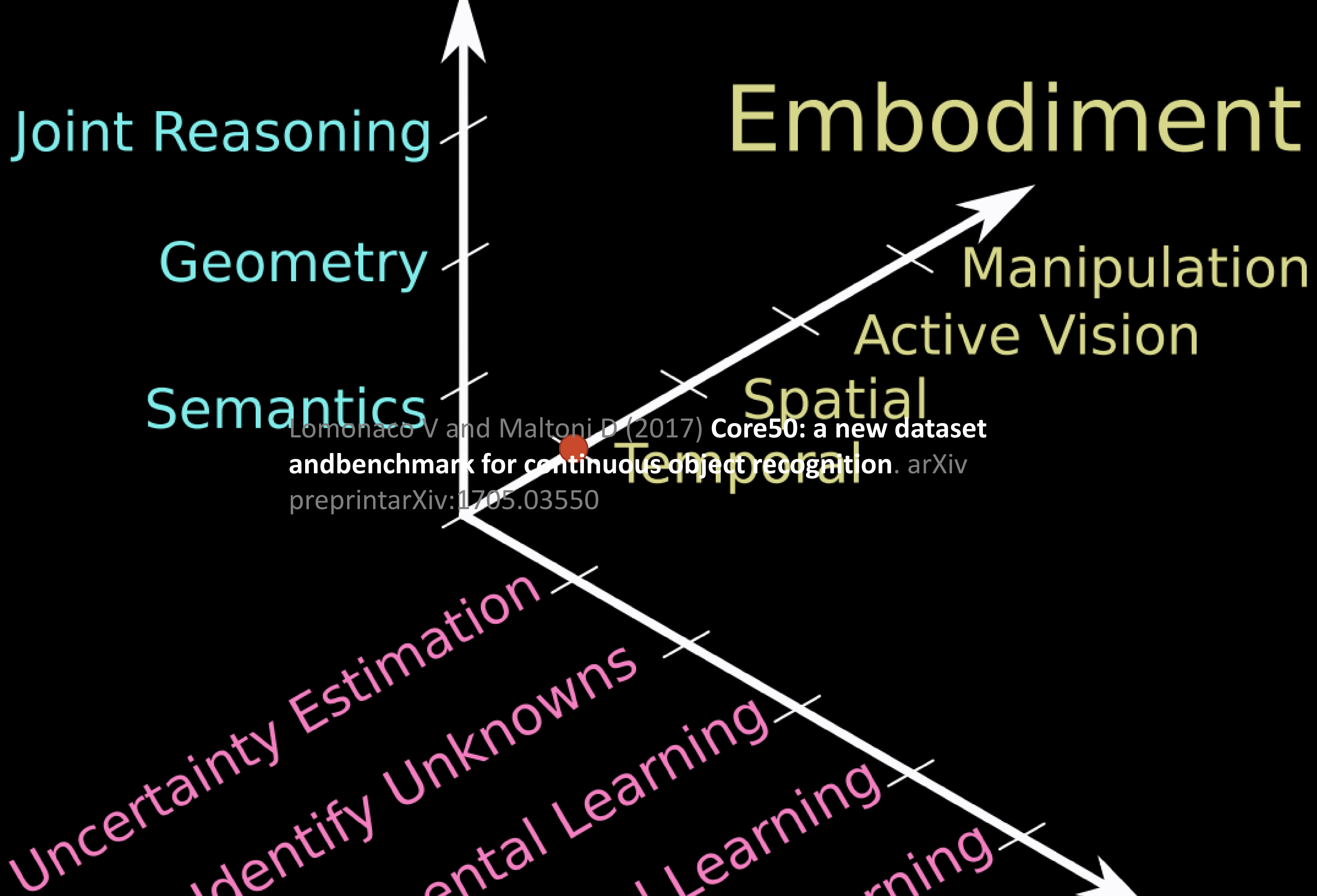

Semantics
Spatial
Temporal
Uncertainty Estimation
Identify Unknowns
Incremental Learning
Class-Incremental Learning
Active Learning
Learning


Semantics
Spatial
Temporal
Uncertainty Estimation
Identify Unknowns
Incremental Learning
Class-Incremental Learning
Active Learning
Learning

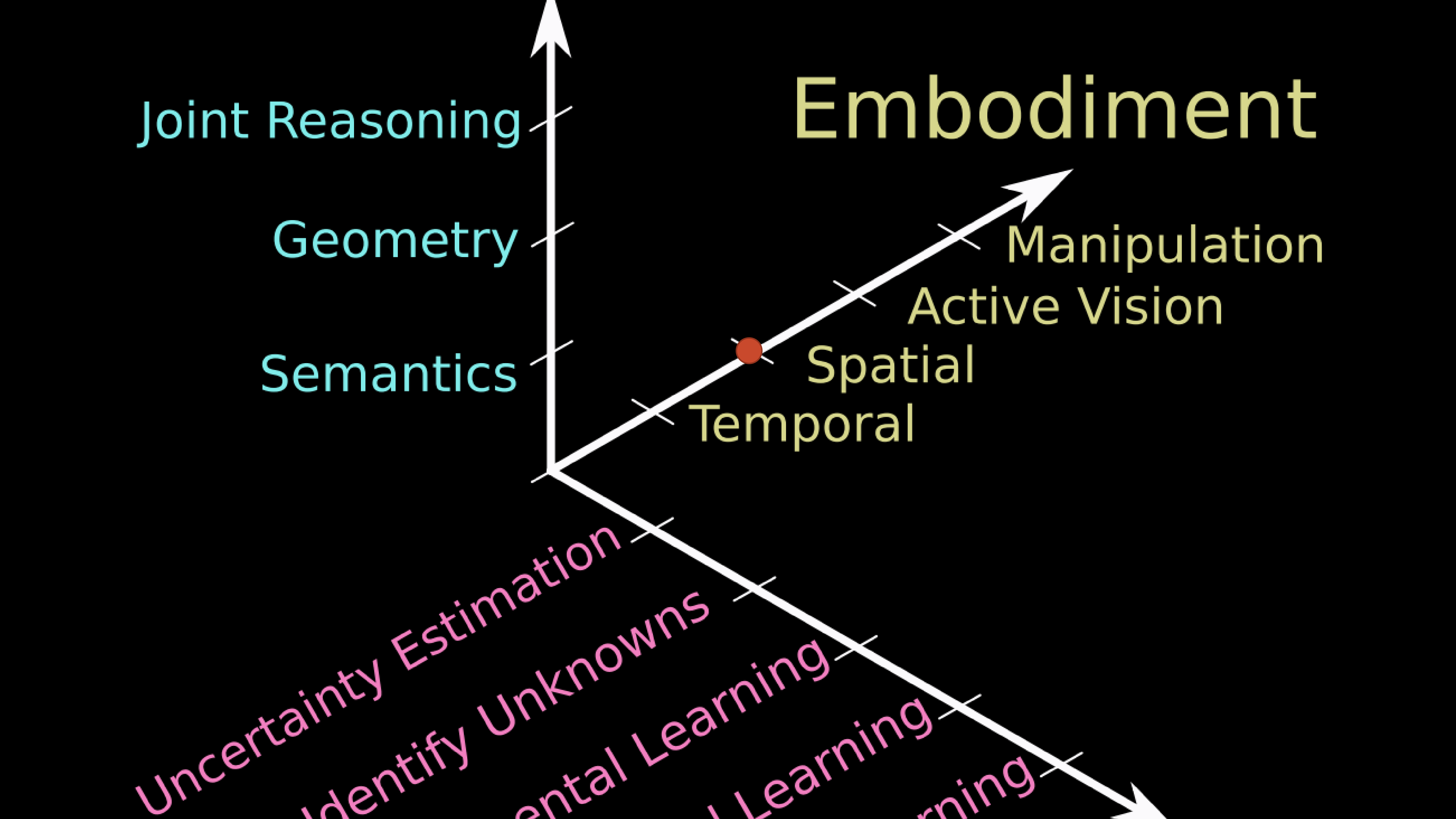


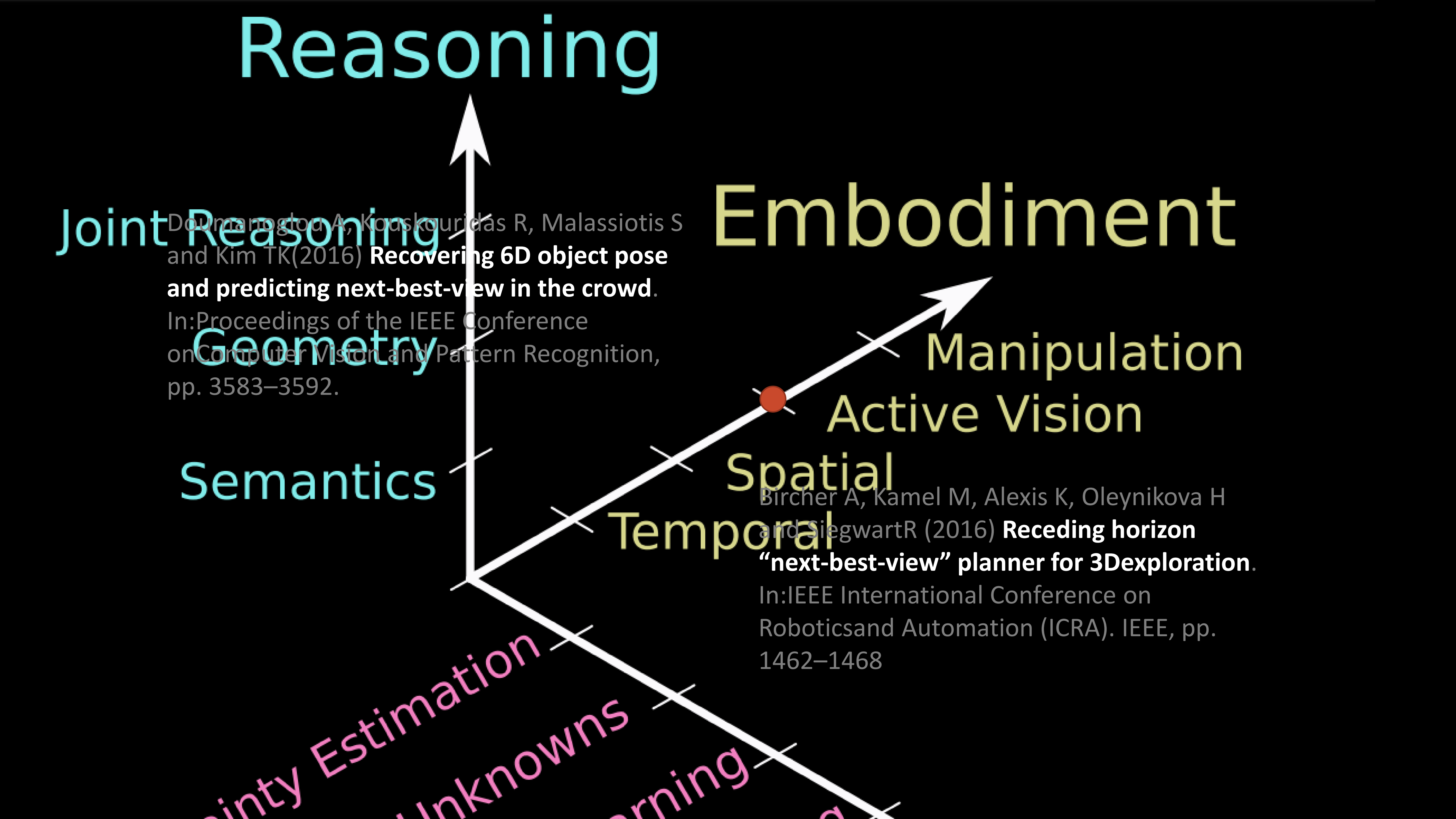
Dayoub F, Sünderhauf N and Corke P (2017) **Episode-based active learning with Bayesian neural networks**. In:CVPR Work-shop on Deep Learning for Robotic Vision. arXiv preprintarXiv:1703.07473

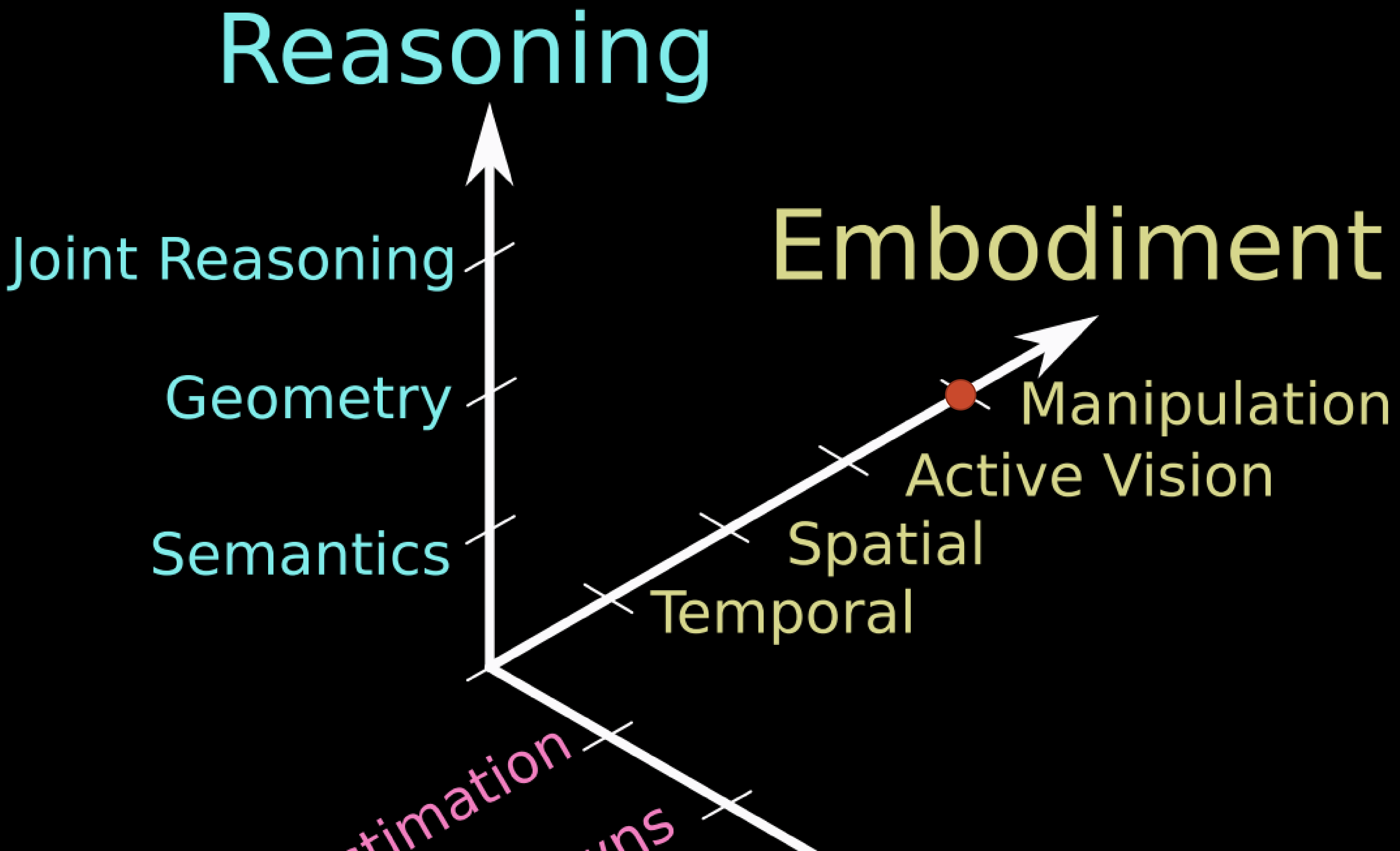
Gal Y, Islam R and Ghahramani Z (2017) **Deep Bayesian active learning with image data.** arXiv preprint arXiv:1703.02910



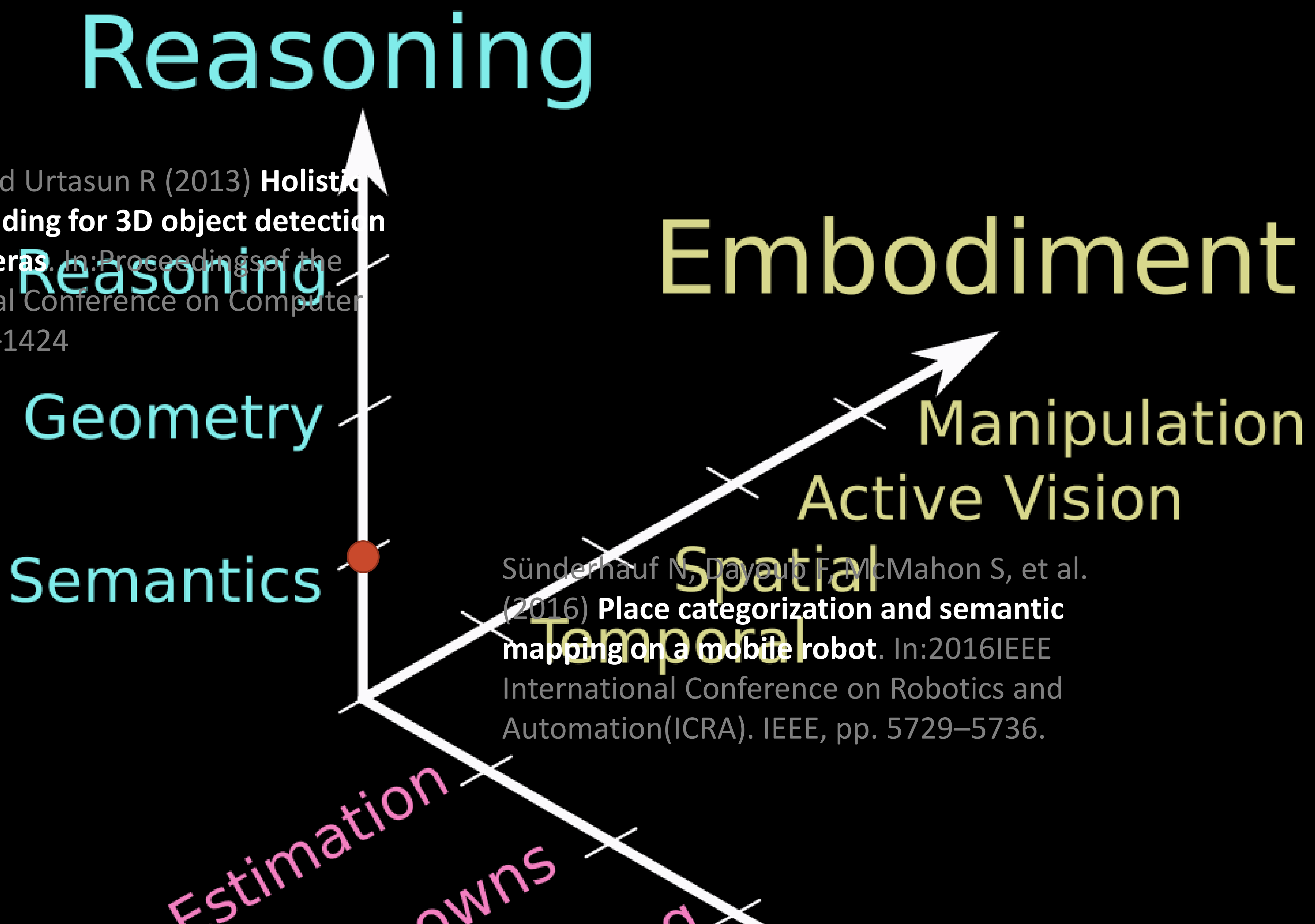
Lomonaco V and Maltoni D (2017) **Core50: a new dataset andbenchmark for continuous object recognition**. arXiv preprintarXiv:1705.03550


Embodiment
Joint Reasoning
Geometry
Semantics
Manipulation
Active Vision
Spatial
Temporal
Uncertainty Estimation
Identify Unknowns
Learning
Learning

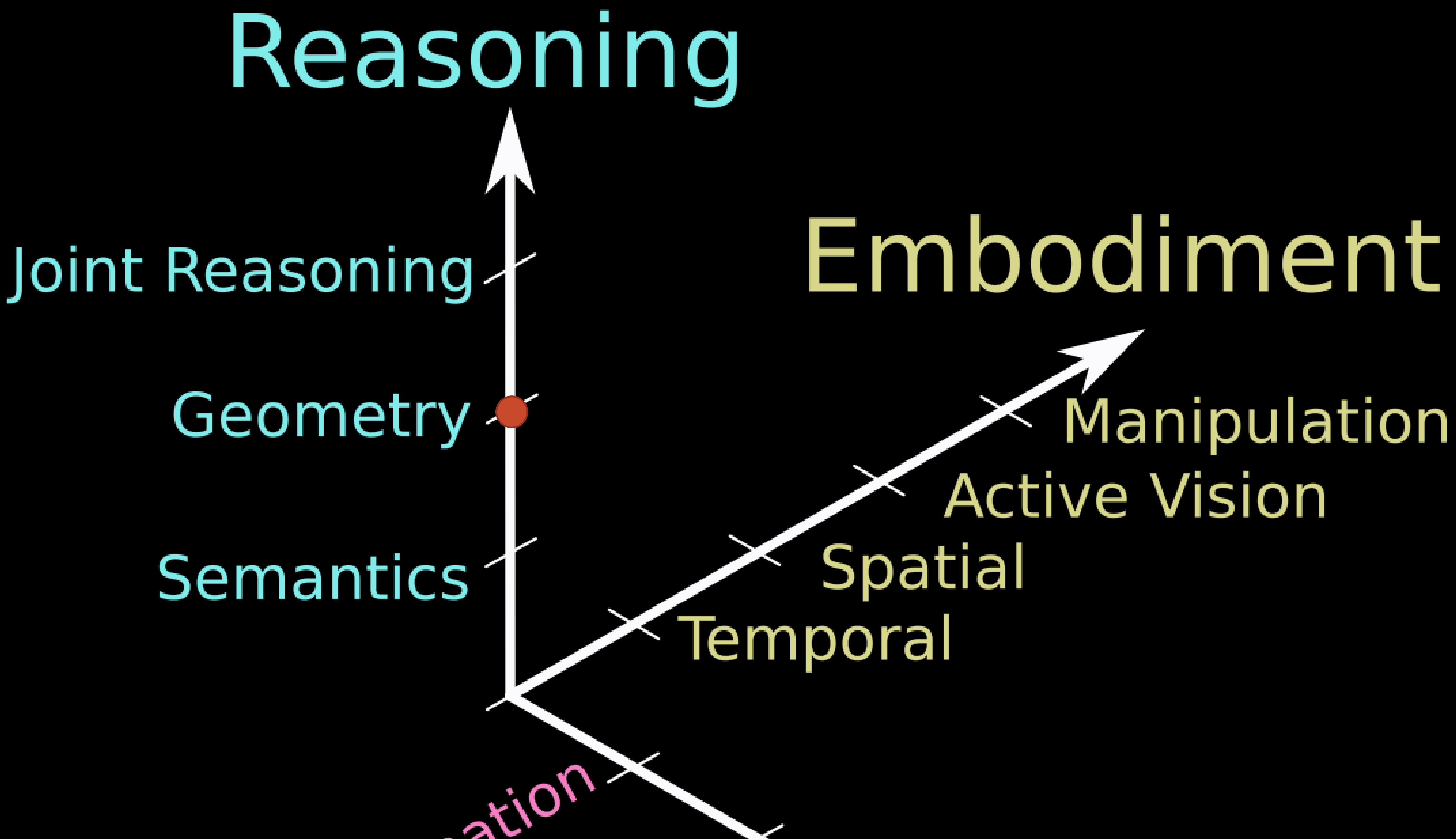

Reasoning
Joint Reasoning
Geometry
Semantics
Embodiment
Manipulation
Active Vision
Spatial
Temporal
inty Estimation
Unknowns
rning
Doumanoglou A, Kouskouridas R, Malassiotis S and Kim TK(2016) **Recovering 6D object pose and predicting next-best-view in the crowd**. In:Proceedings of the IEEE Conference onComputer Vision and Pattern Recognition, pp. 3583–3592.
Bircher A, Kamel M, Alexis K, Oleynikova H and SiegwartR (2016) **Receding horizon "next-best-view" planner for 3Dexploration**. In:IEEE International Conference on Roboticsand Automation (ICRA). IEEE, pp. 1462–1468

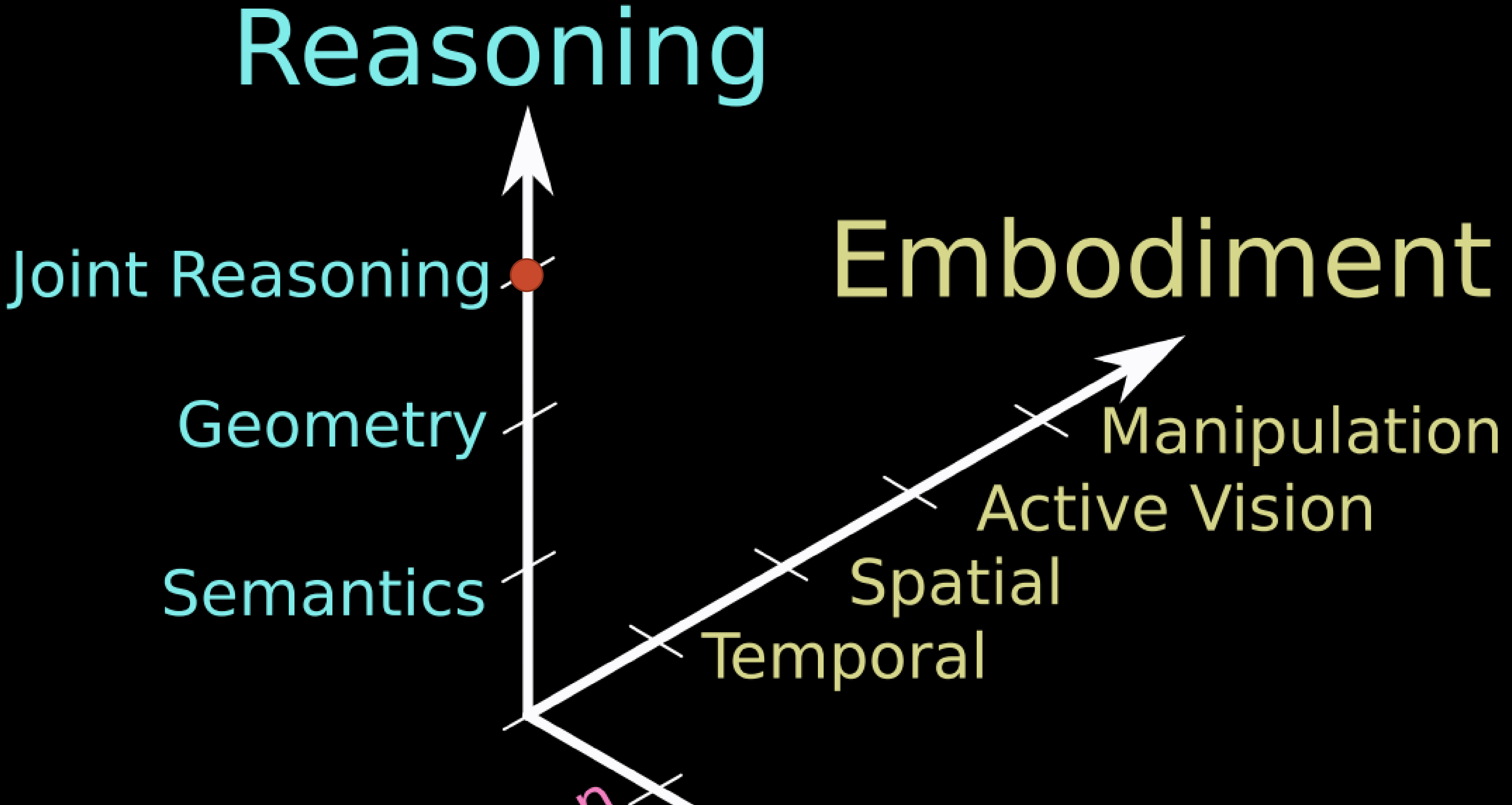

Reasoning
Joint Reasoning
Geometry
Semantics
Embodiment
Manipulation
Active Vision
Spatial
Temporal

Lin D, Fidler S and Urtasun R (2013) **Holistic scene understanding for 3D object detection with RGBD cameras**. In:Proceedingsof the IEEE International Conference on Computer Vision, pp.1417–1424

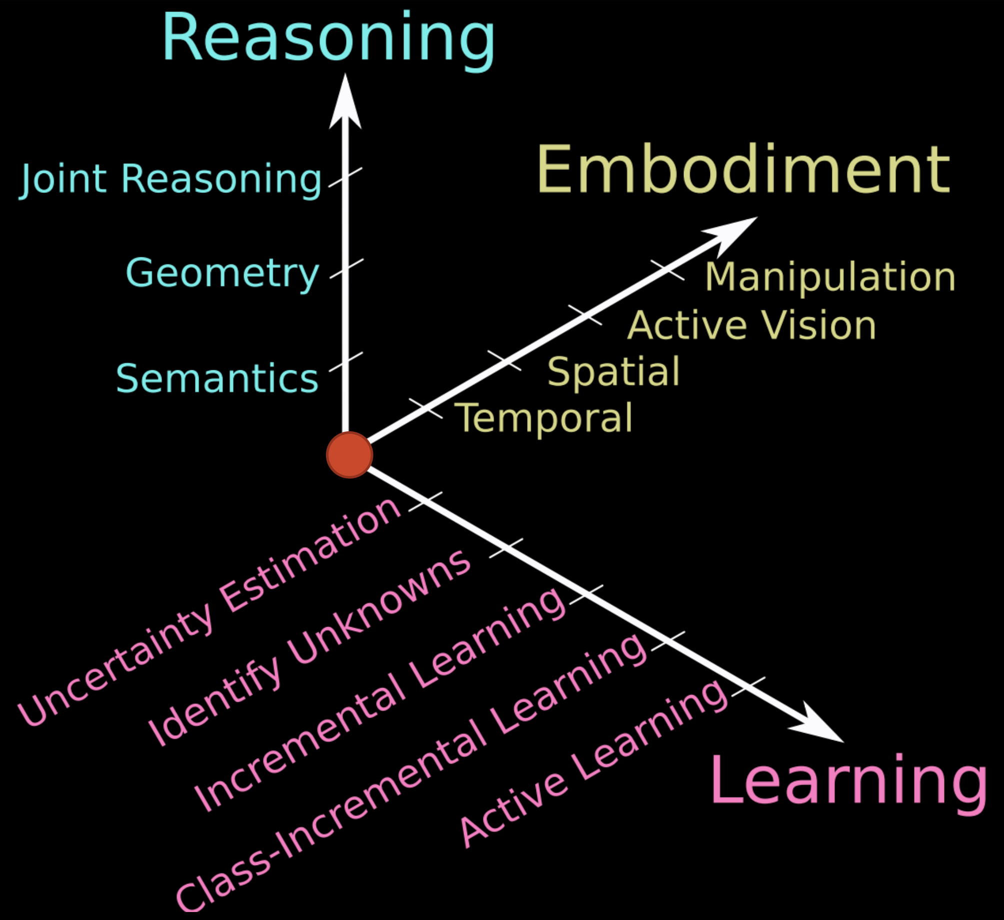


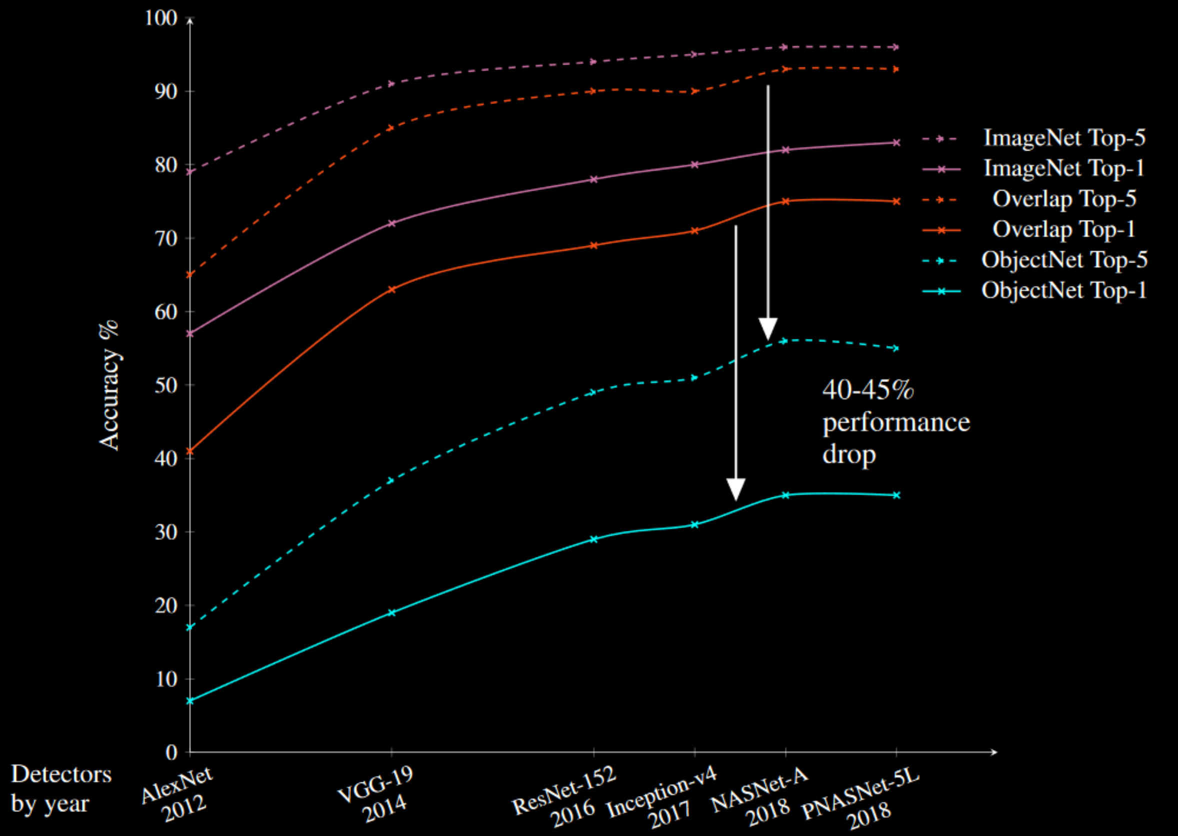
Sünderhauf N, Dayoub F, McMahon S, et al. (2016) **Place categorization and semantic mapping on a mobile robot**. In:2016IEEE International Conference on Robotics and Automation(ICRA). IEEE, pp. 5729–5736.


Reasoning
Joint Reasoning
Geometry
Semantics
Embodiment
Manipulation
Active Vision
Spatial
Temporal


Reasoning
Joint Reasoning
Geometry
Semantics
Embodiment
Manipulation
Active Vision
Spatial
Temporal

Challenges in robotic vision





Andrei Barbu, David Mayo, Julian Alverio, William Luo, Christopher Wang, Dan Gutfreund, Josh Tenenbaum, and Boris Katz. **Objectnet: A large-scale bias-controlled dataset for pushing the limits of object recognition models.** In Advances in Neural Information Processing Systems 32, pages 9448–9458. 2019.

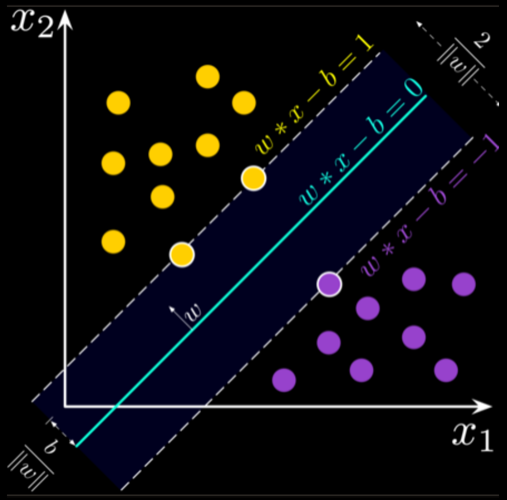
Importance of open-set recognition

- Known known classes
- Known unknown classes
- Unknown unknown classes

mitigate

Problem with **features**
or
readout layer?

Problem with readout layer



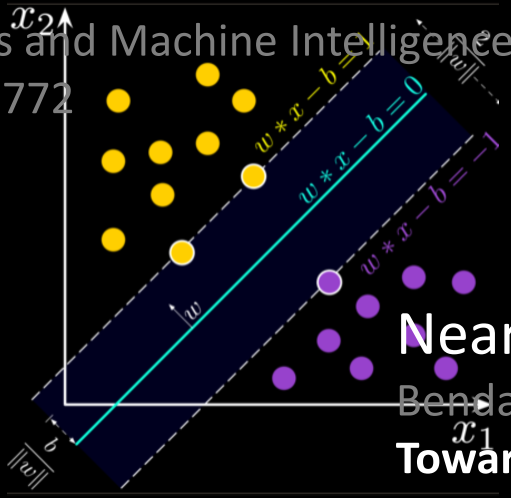
$$\sigma(\mathbf{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}}$$

$$\cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|}$$

Problem with readout layer

Slab-based linear classifier

Scheirer WJ, de Rezende Rocha A, Sapkota A and Boult TE(2013) **Toward open set recognition.** IEEE Transactions onPattern Analysis and Machine Intelligence35(7): 1757–1772



$$\sigma(\mathbf{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}}$$

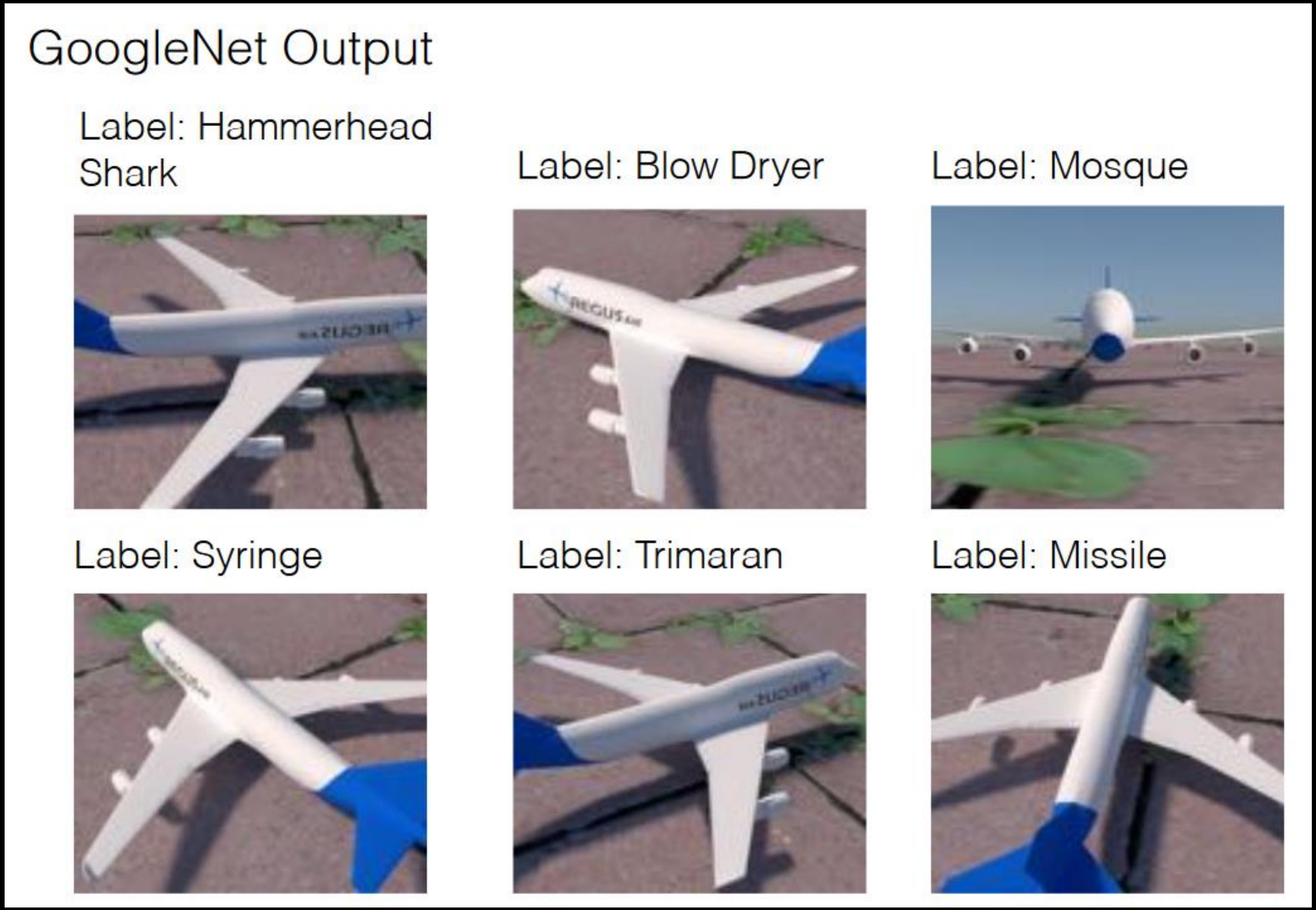
Nearest non-outlier models

Bendale A and Boult TE (2015) **Towards open world recognition.** In:Proceedings of the IEEE Conference on Computer Visionand Pattern Recognition, pp. 1893–1902.

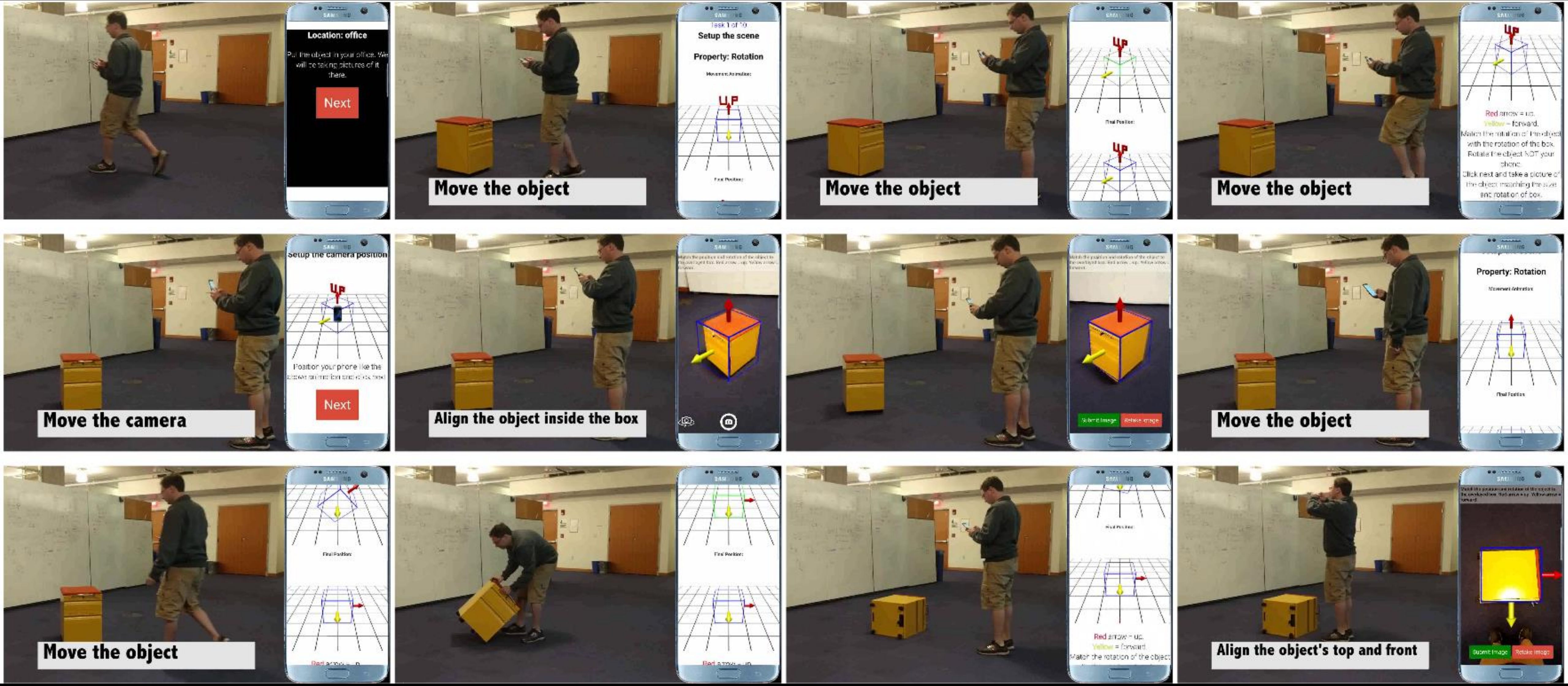
Extreme value-theory based calibration of decision boundaries

Bendale A and Boult TE (2016) **Towards open set deep networks.** In:Proceedings of the IEEE Conference on Computer Visionand Pattern Recognition, pp. 1563–1572

$$\cos(\theta) = \frac{\mathbf{A}\cdot\mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|}$$

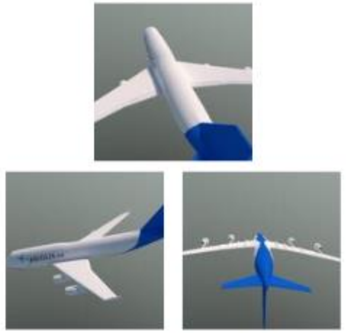


Source: Walter Scheirer slides in Deep Learning RSS 2016 workshop

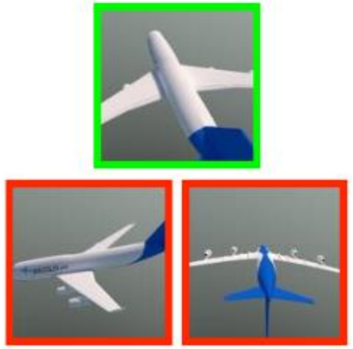

Location: office
Put the object in your office. We will be taking pictures of it there.
Next
Setup the scene
Property: Rotation
UP
Move the object
Move the object
Red arrow = up.
Yellow = forward.
Match the rotation of the object with the rotation of the box.
Rotate the object NOT your phone.
Click next and take a picture of the object matching the size and rotation of box.
Move the object
Setup the camera position
Position your phone like the above animation and click next
Next
Move the camera
Align the object inside the box
Submit Image
Retake image
Property: Rotation
Move the object
Move the object
Align the object's top and front

Psychophysics pipeline

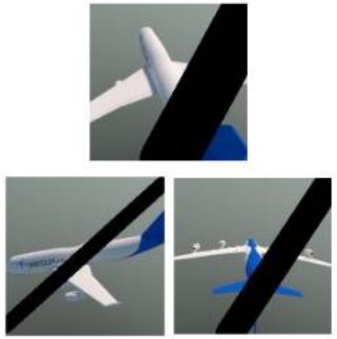
1. Render Class Canonical View (CCV) Candidates



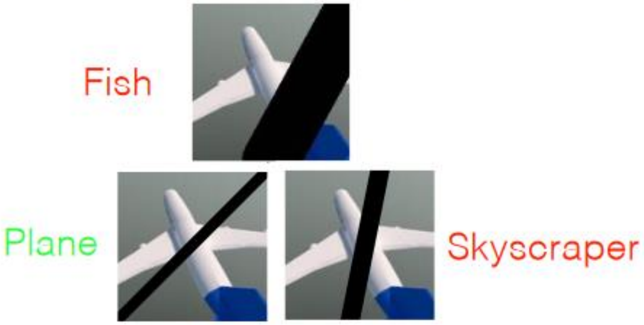
2. CCV Classifier



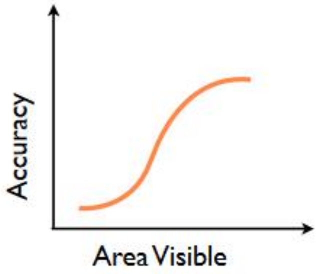
3. Manipulate Chosen Variable



4. Classify Images

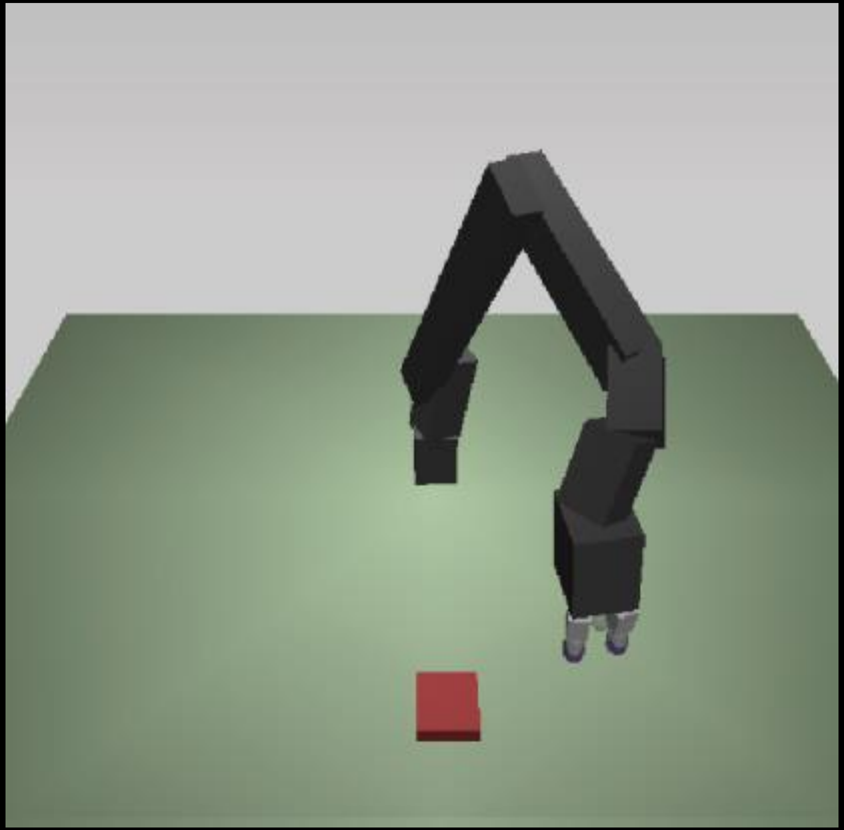


5. Generate Psychometric Curve



Brandon RichardWebster

Source: Walter Scheirer slides in Deep Learning RSS 2016 workshop



Jaco & MuJoCo

Purely real infeasible ⇒ Sim2Real

BUT, REALITY GAP

REALITY GAP

Transfer learning

REALITY GAP

Transfer learning

- Augment data

  Su H, Qi CR, Li Y and Guibas LJ (2015) **Render for CNN: view-point estimation in images using CNNs trained with rendered 3dmodel views**. In:2015 IEEE International Conference on Computer Vision (ICCV 2015), Santiago, Chile, 7–13 December2015, pp. 2686–2694

- Soft constraint on feature distribution

  Long M, Cao Y, Wang J and Jordan MI (2015) **Learning transferable features with deep adaptation networks**. In:Proceedingsof the 32nd International Conference on Machine Learning(ICML 2015), Lille, France, 6–11 July 2015, pp. 97–105.

- Aligned data (triple loss)

  Tzeng E, Devin C, Hoffman J, et al. (2015a) **Towards adapting deep visuomotor representations from simulated to real environments**. CoRR abs/1511.07111.

Transfer learning
- Augment data
- Soft constraint on feature distribution
- Aligned data (triple loss)

REALITY GAP

Domain randomization

Tobin J, Fong R, Ray A, Schneider J, Zaremba W and Abbeel P(2017) **Domain randomization for transferring deep neural net-works from simulation to the real world**. In:2017 IEEE/RSJInternational Conference on Intelligent Robots and Systems(IROS). IEEE, pp. 23–30

Aside: Automatic domain randomization

Transfer learning

- Augment data
- Soft constraint on feature distribution
- Aligned data (triple loss)

Domain randomization

REALITY GAP

Progressive nets

Rusu A, Rabinowitz N, Desjardins G, et al. (2016) **Progressive neural networks**. arXiv preprint arXiv:1606.04671

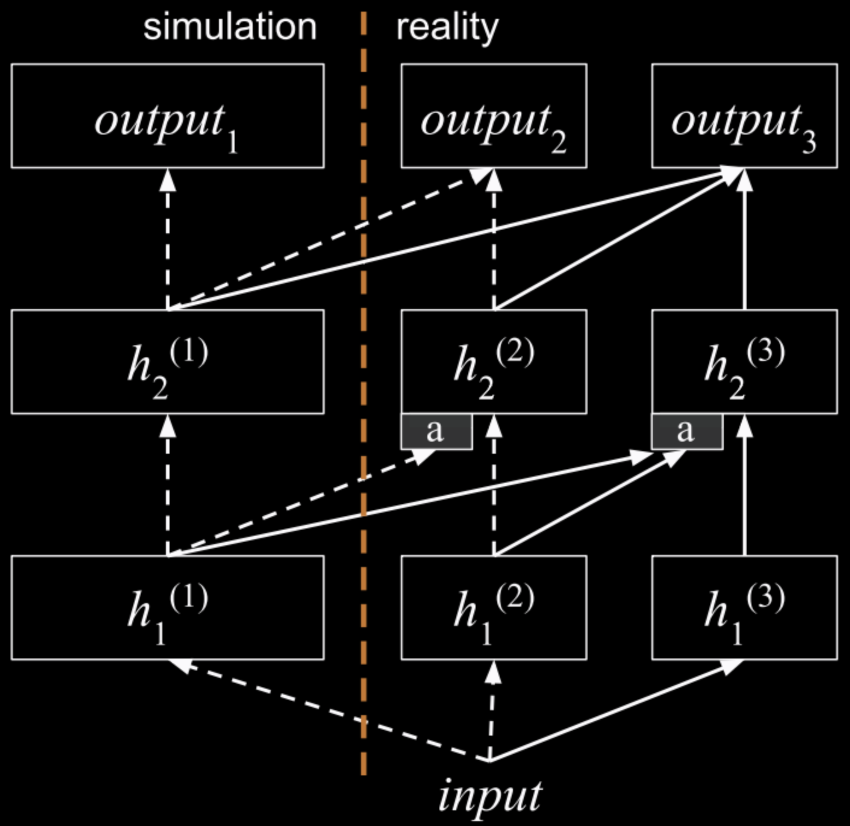
REALITY GAP

Transfer learning

- Augment data
- Soft constraint on feature distribution
- Aligned data (triple loss)

Domain randomization

Progressive nets



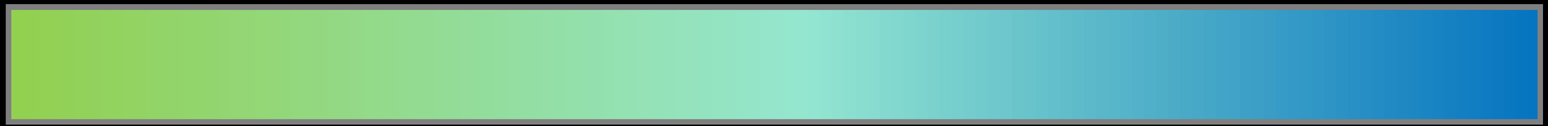
Rusu A, Rabinowitz N, Desjardins G, et al. (2016) **Progressive neural networks**. arXiv preprint arXiv:1606.04671

Perfect (physics) model
of the world

Why learn?

Perfect black-boxes
of the world

Why care about physics?



Physics

Blackbox

Sweet spot?
⇓
Robot overlords!

Perfect (physics) model
of the world

↓

Why learn?

Perfect black-boxes
of the world

↓

Why care about physics?

Sweet spot?
⇓
Robot overlords!

Perfect (p xes
of t

Wh ysics?

Physics

Blackbox

Sweet spot?
⇓
Robot overlords!

Perfect (physics) model
of the world

Why learn?

Perfect black-boxes
of the world

Why care about physics?

Physics

Blackbox

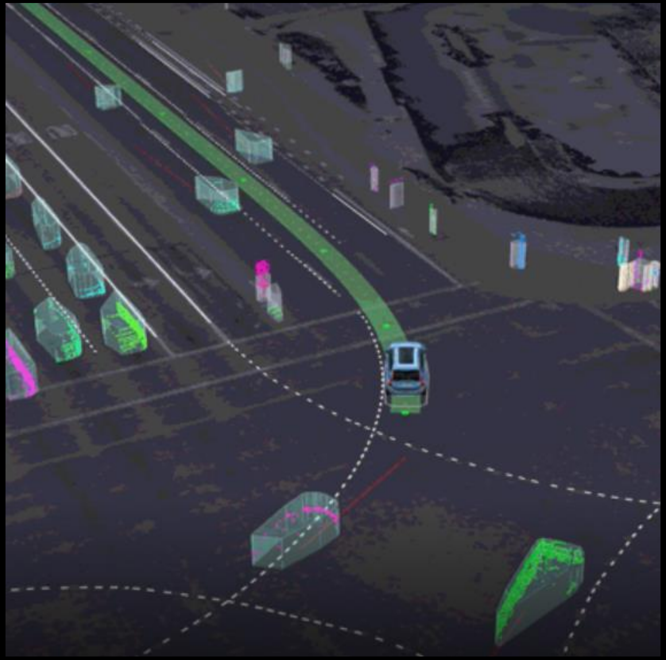
Sweet spot?

$\Downarrow$

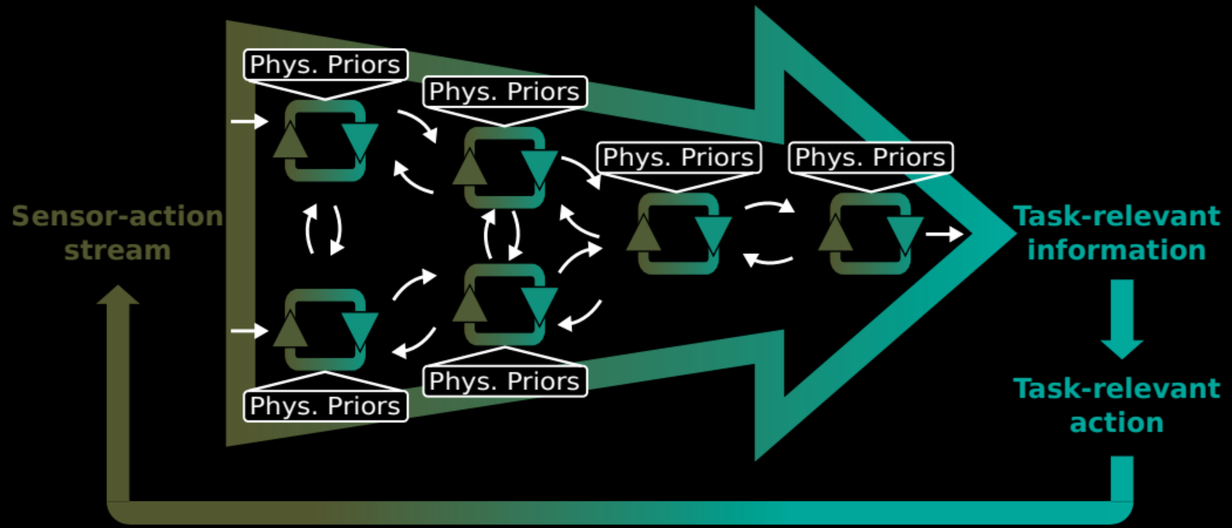
Robot overlords!

Physics-model-based		Deep-learning-based
Explicit	Representation	Implicit
Broadly applicable	Generality	Only in trained regime
Small basin of convergence	Robustness	Large basin of convergence
Very high	Data efficiency	Very low
Good in local regime(?)	Computational efficiency	Highly efficient once trained

Key challenges (or thoughts?)



Source: Uber



Physics-based models ⇒ less data

Physics-model-based

Deep-learning-based

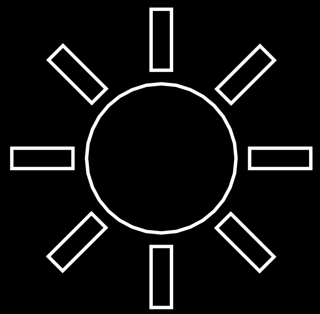
Programming

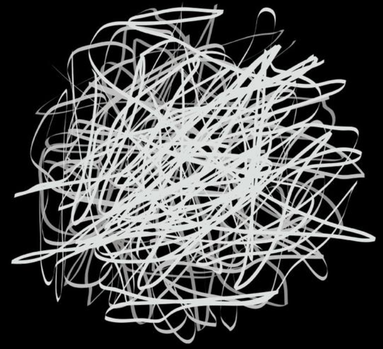
Data

Quicksort

Decision trees

Deep learning





Implement

C++

Learning Weights

Network architecture, loss function,
training procedure and data

Programming vs data, conclusions

1. Move towards programming
2. Use deep learning to discover structure
3. Divide problems: what we know and don’t know

Discover structure?

Hinton G, Vinyals O and Dean J (2015)
Distilling the knowledge in a neural network.
CoRRabs/1503.0253

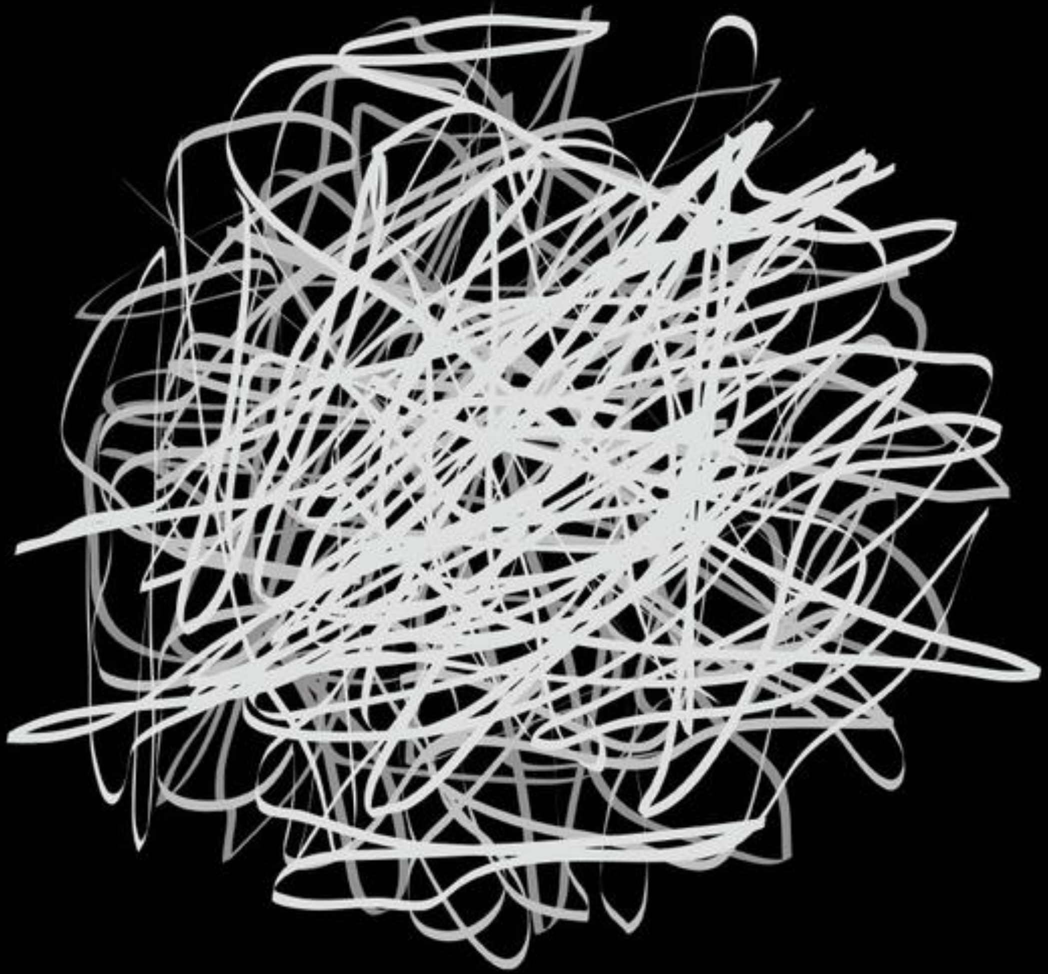
Lopez-Paz D, Bottou L, Schölkopf B and Vapnik V (2016)
Unifying distillation and privileged information.
CoRRabs/1511.03643

Divide problems?

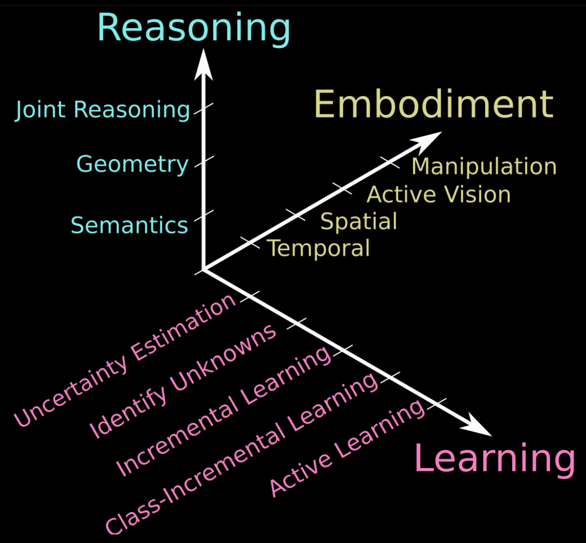
Byravan A and Fox D (2017)
SE3-nets: Learning rigid body motion using deep neural networks.
In:Proceedings of theIEEE International Conference on Robotics and Automation(ICRA).

Haarnoja T, Ajay A, Levine S and Abbeel P (2016)
Backprop KF: learning discriminative deterministic state estimators.
CoRRabs/1605.07148.

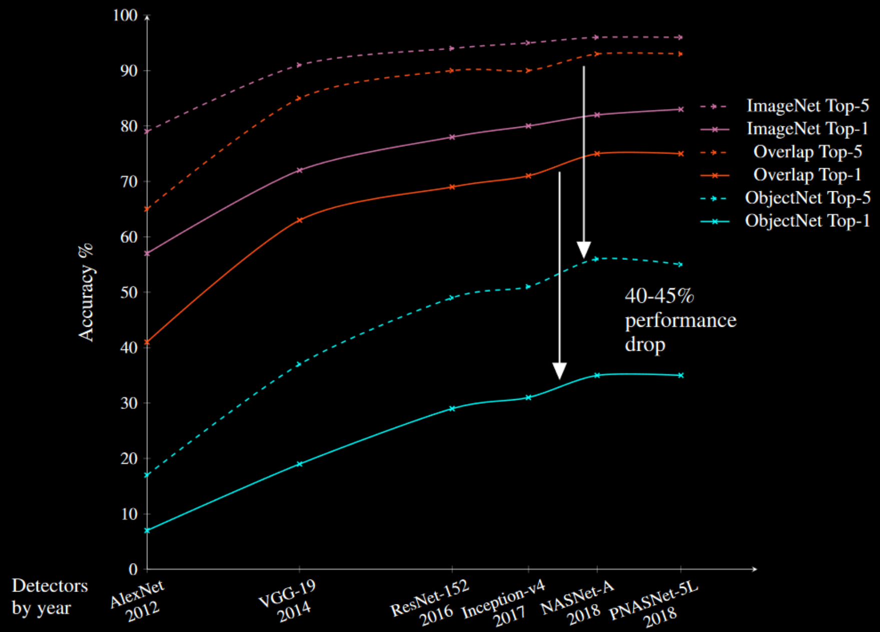
Jonschkowski R and Brock O (2016)
End-to-end learnable histogram filters.
In:Workshop on Deep Learning for Action andInteraction at the Conference on Neural Information Process-ing Systems (NIPS).



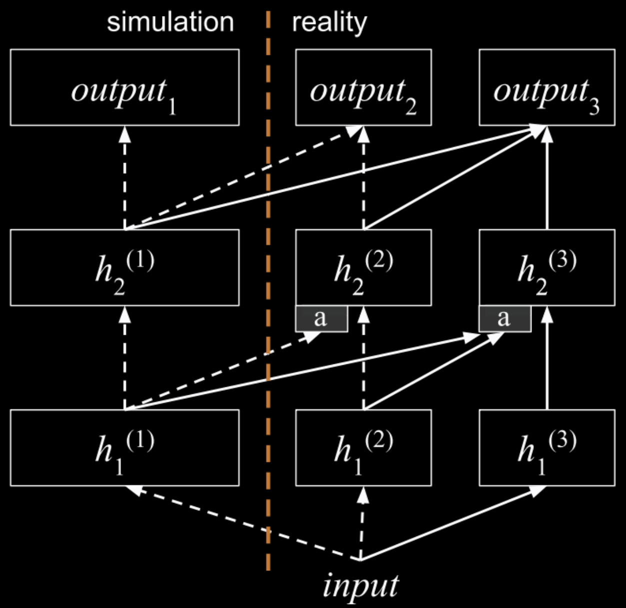
Summary



Reasoning
Joint Reasoning
Geometry
Semantics
Embodiment
Manipulation
Active Vision
Spatial
Temporal
Uncertainty Estimation
Identify Unknowns
Incremental Learning
Class-Incremental Learning
Active Learning
Learning



Accuracy %
100
90
80
70
60
50
40
30
20
10
0
ImageNet Top-5
ImageNet Top-1
Overlap Top-5
Overlap Top-1
ObjectNet Top-5
ObjectNet Top-1
40-45% performance drop
Detectors by year
AlexNet 2012
VGG-19 2014
ResNet-152 2016
Inception-v4 2017
NASNet-A 2018
PNASNet-5L 2018



simulation
reality
$output_1$
$output_2$
$output_3$
$h_2^{(1)}$
$h_2^{(2)}$
$h_2^{(3)}$
a
a
$h_1^{(1)}$
$h_1^{(2)}$
$h_1^{(3)}$
input

Physics-model-based
Deep-learning-based
Programming
Data