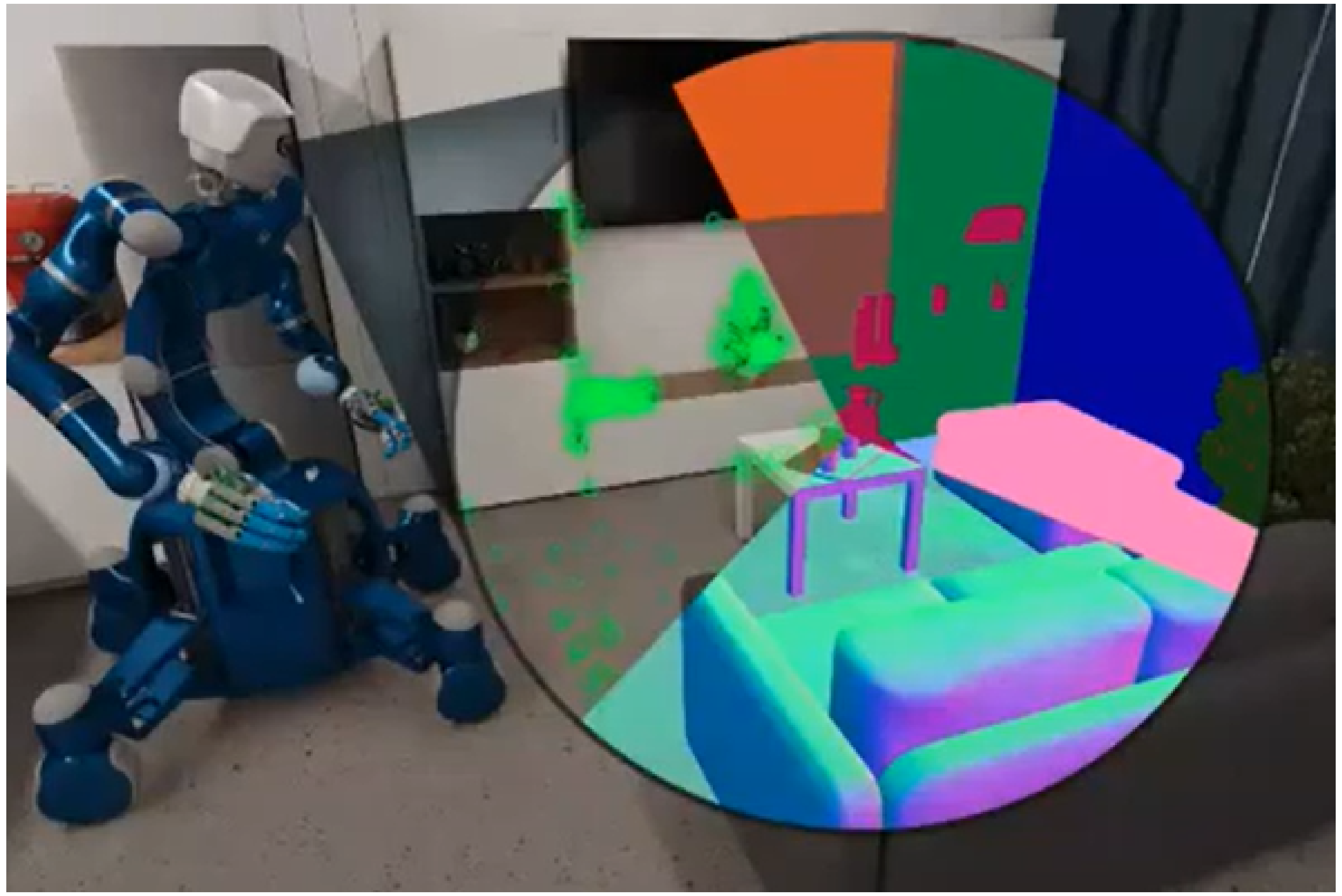
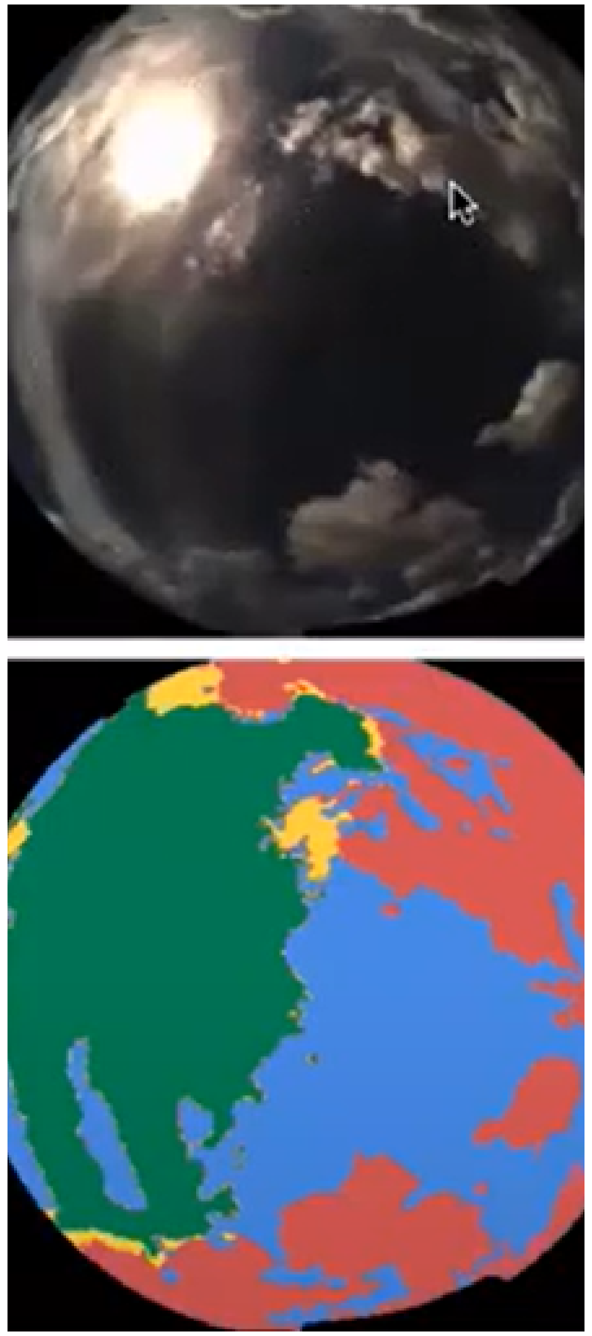
Data-driven Machine Vision for Fast and Reliable Predictions

Typical application of machine perception

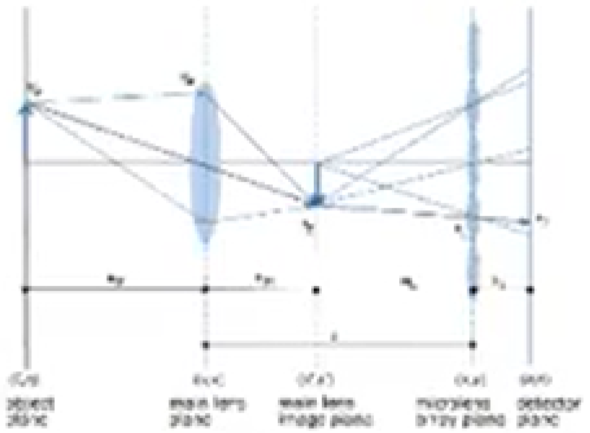




Model-based or Data-driven? Pros and Cons

Model-based

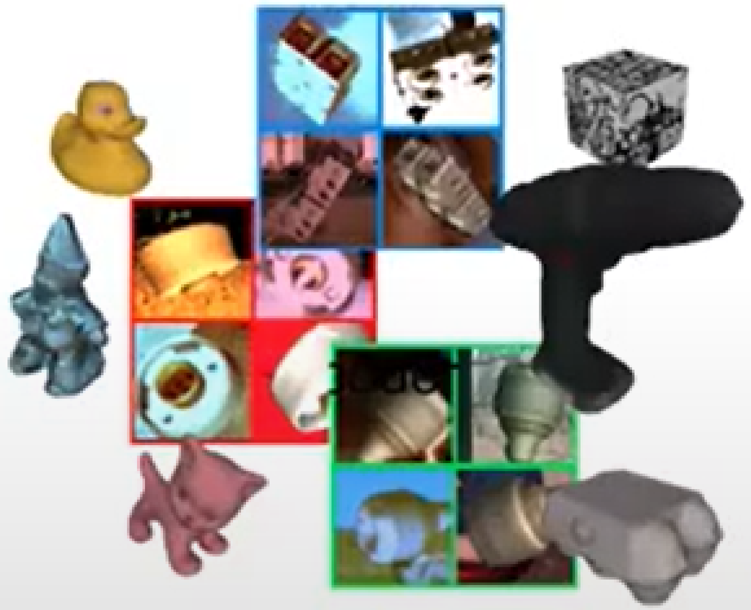
Small, specific data



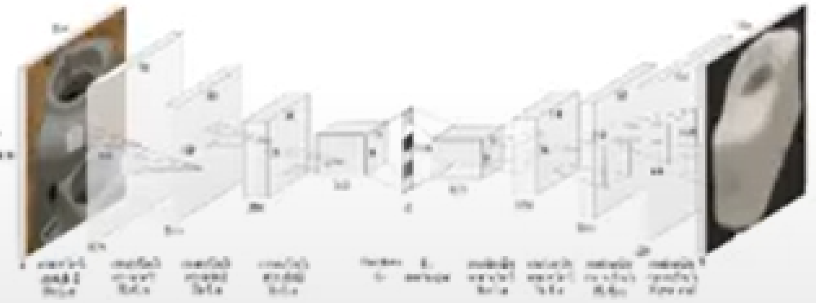
Precise, physical model

"Optimization"

Data-driven



Large, general data set



Generic, high-dimensional "model"

"Learning"

	model-based					**data-driven**
E.g.:	Camera calibration	Robot motion	Forward kinematics	Semantic segmentation	Speech recognition	Image classification

Three Major Challenges of Data-driven Perception Approaches

To leverage data-driven methods successfully, we need to tackle three major problems:

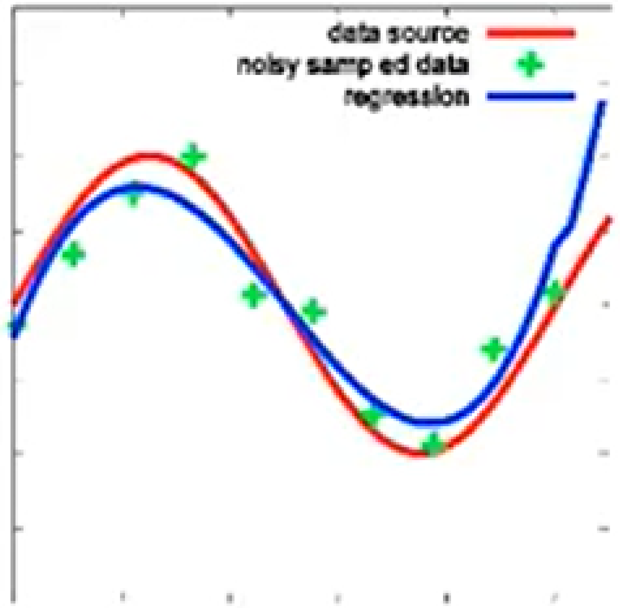
1. **Quality and amount of training data**
    - large networks require large training data sets
    - ground truth must be reliable
2. **Precise modelling of predictive uncertainty**
    - variations in the model and the data need to be taken into account
    - data-driven predictors should be neither under- nor over-confident
3. **Ressource-efficient inference**
    - often real-time requirements (e.g. robotics, autonomous driving)
    - limited hardware availability (e.g. embedded GPUs)

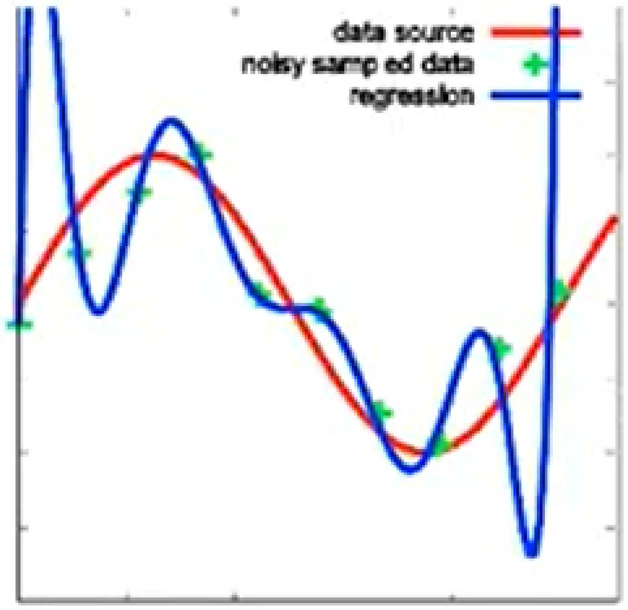
In the following, approaches to tackle these problems will be presented.

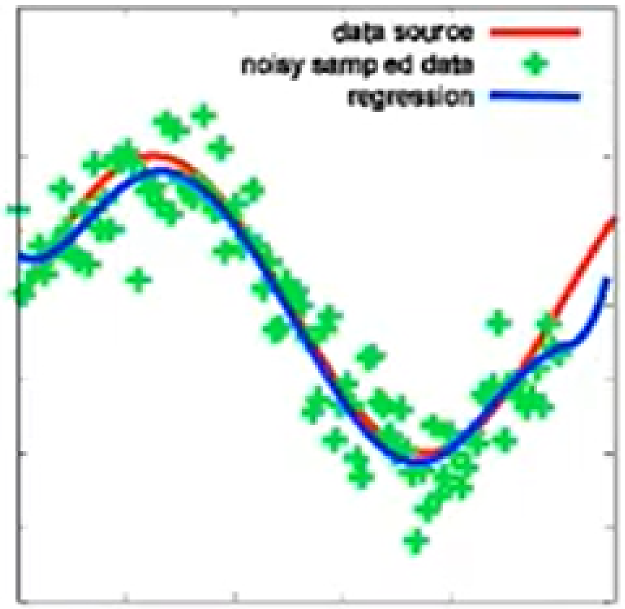
The Importance of Good and Large Training Data

Large models require large training data
→ overfitting

Many techniques exist, e.g.:
- regularisation
- data-augmentation







data p.:	**10**	**10**	**100**
model p.:	**5**	**10**	**10**

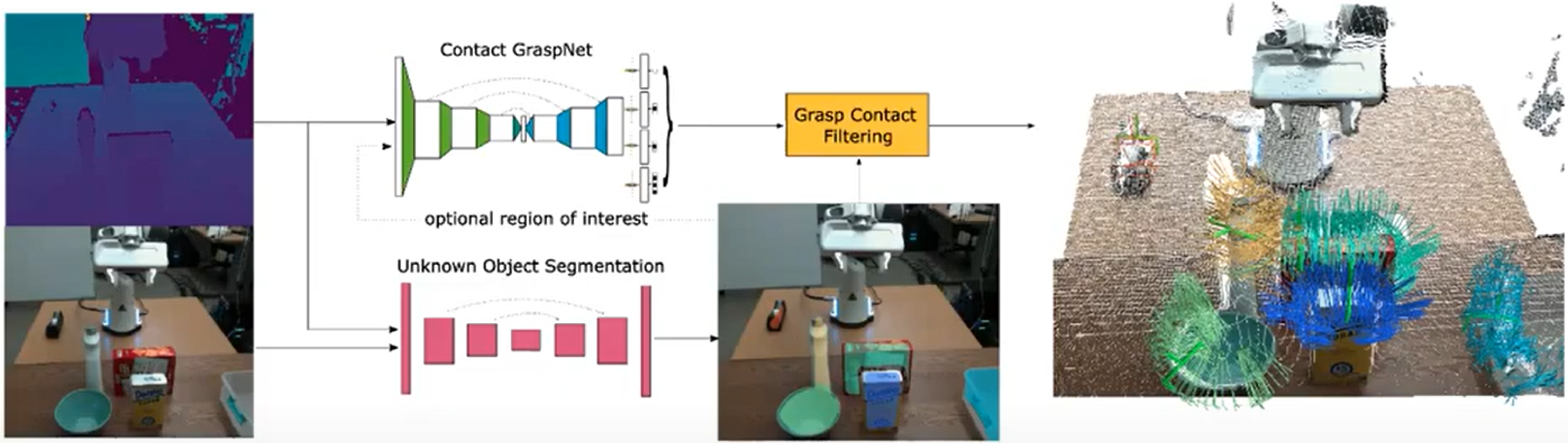
Recent Alternatives:
- Synthetic training data (e.g. CosyPose [1])
- self-supervised training

training data:
T-LESS, YCB +
1 million rendered images

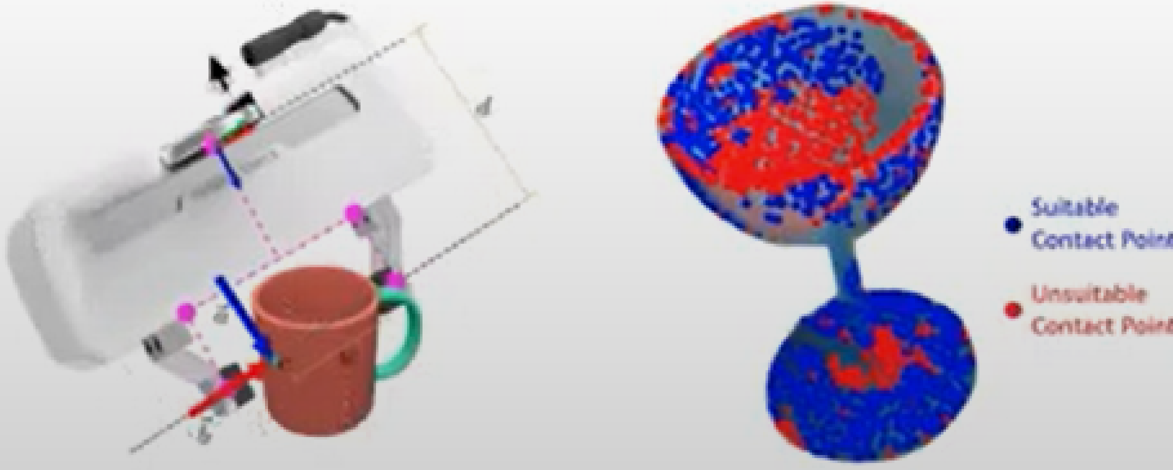
6D object pose estimation

from: [1] Labbé, Carpentier, Aubry, Sivic: “CosyPose: Consister multi-view multi-object 6D pose estimation”, ECCV 2020

Example: Synthetic Training Data for Grasp Detection



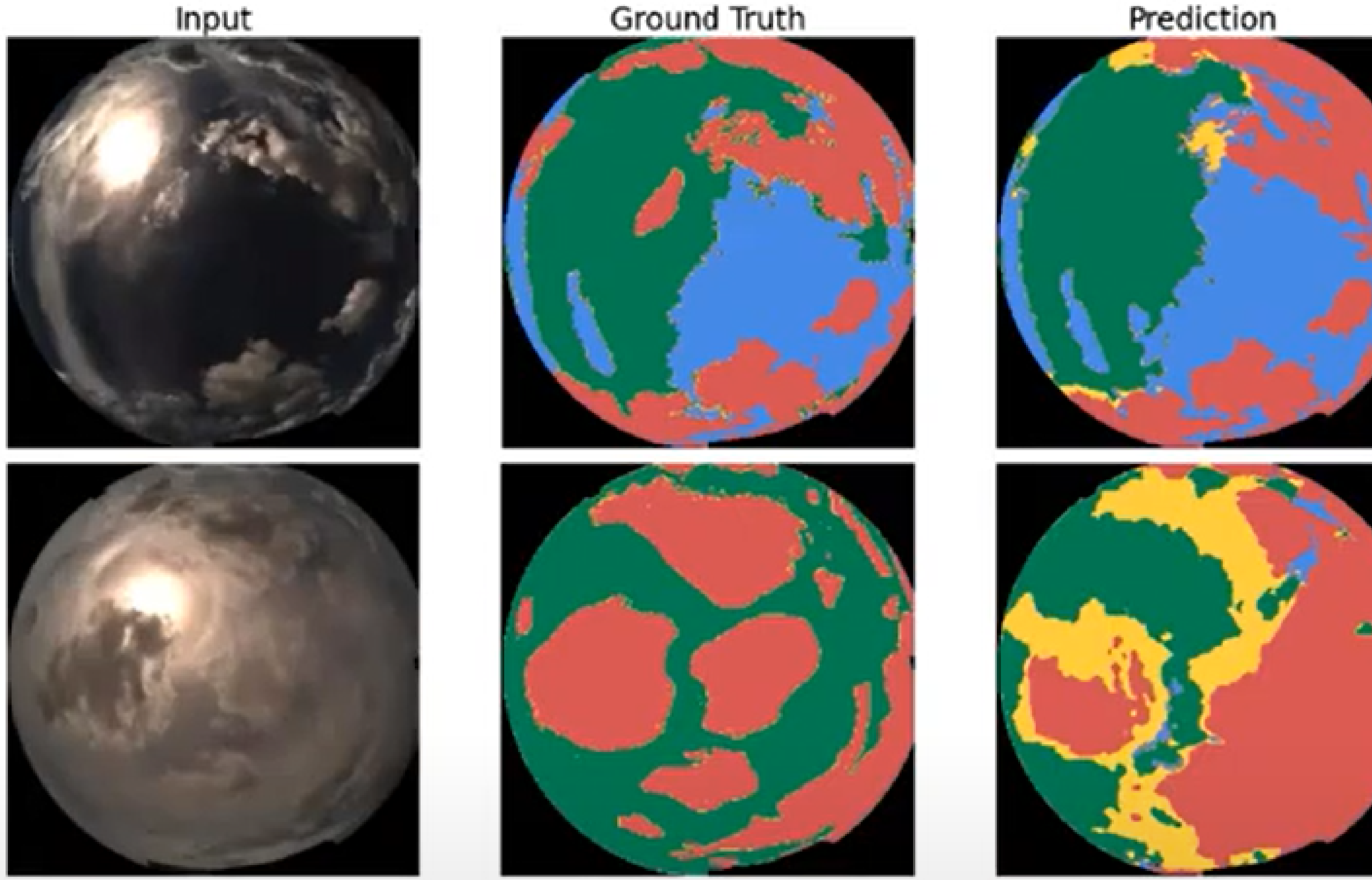
Main ideas:
- Use a representation for grasp contact points for 2-finger robotic grippers
- Train a network to predict feasible contact points from a large simulated training data
- Combine this with unknown-object segmentation to mask out objects



Sundermeyer, Mousavian, Triebel, Fox: "Contact-GraspNet: Efficient 6-DoF Grasp Generation in Cluttered Scenes"
IEEE Intern. Conf. *on Robotics and Automation (ICRA) 2021*

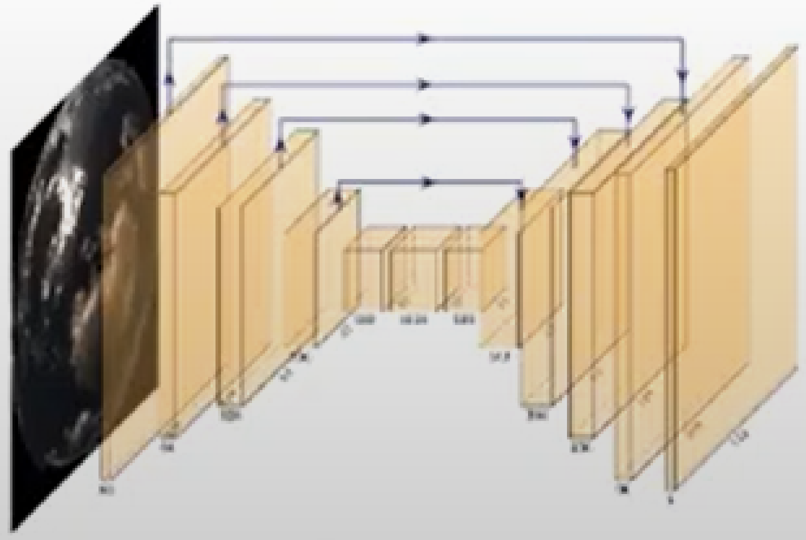
Example: Cloud Segmentation using Self-Supervised Learning

In-painting super-resolution



	Sky	Low-layer	Mid-layer	High-layer
IoU:	**91.73**	**69.52**	**40.82**	**49.83**

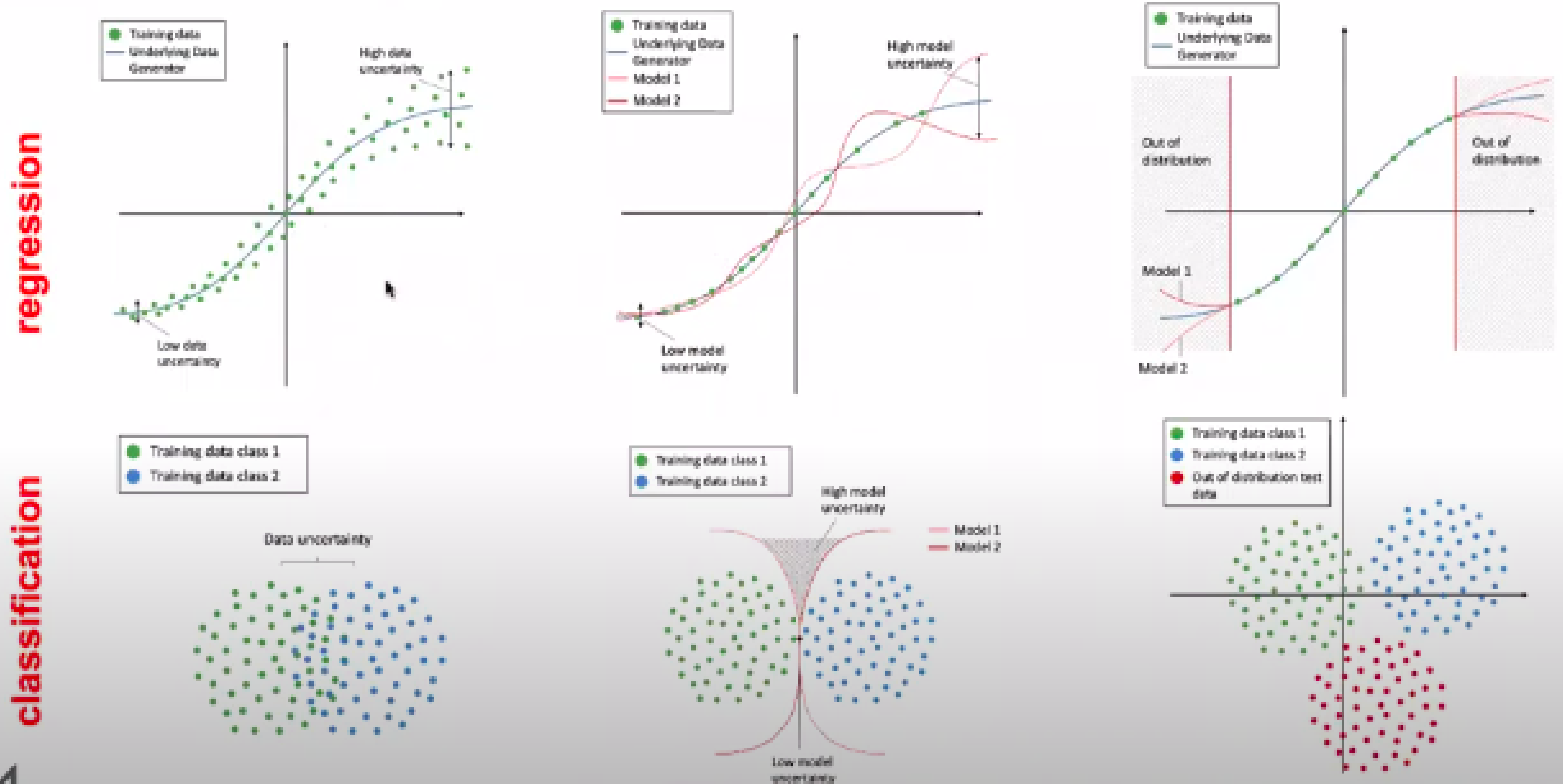
Instead of synthetic training data, we can also use **self-supervised learning**.



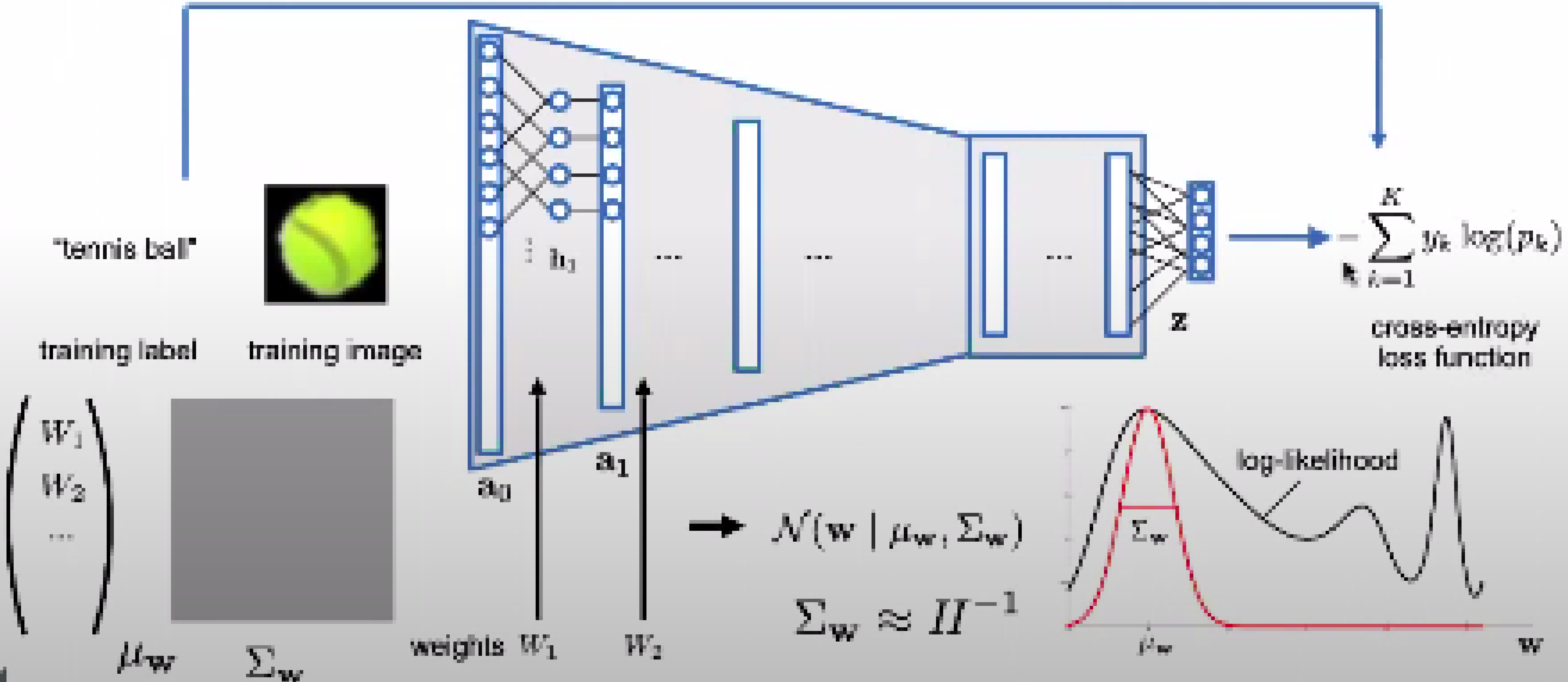
U-Net with ResNet34 backbone

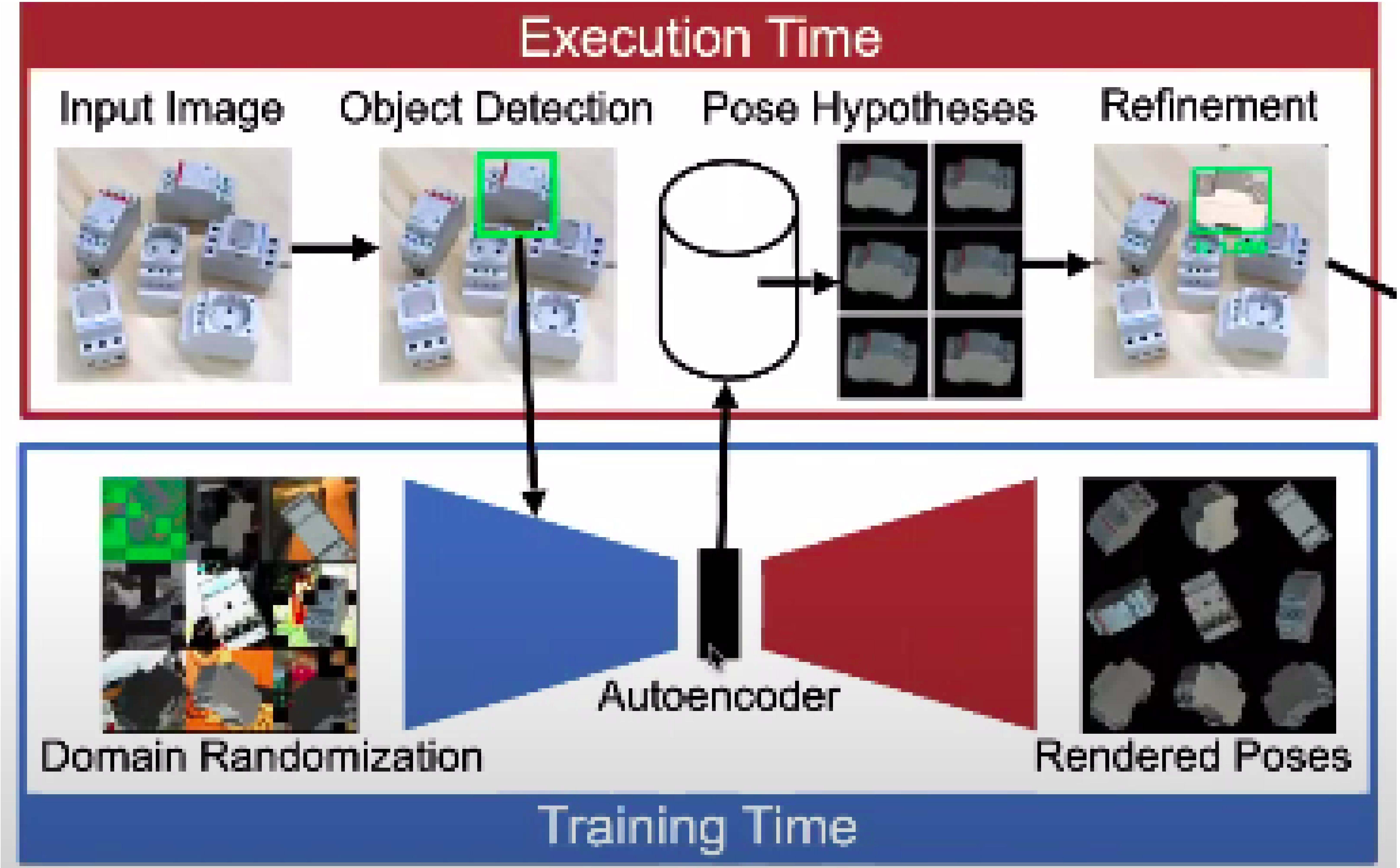
Y. Fabel, B. Nouri, S. Wilbert, N. Blum, R. Triebel, M. Hasenbalg, P. Kuhn, L. F. Zarzalejo, R. Pit
"Applying self-supervised learning for semantic cloud segmentation of all-sky images", Atmos.
Tech., 15, 797–809, 2022

Sources of Uncertainty



One way to address this are **Bayesian Neural Networks (BNN).**




Execution Time
Input Image
Object Detection
Pose Hypotheses
Refinement
Autoencoder
Domain Randomization
Rendered Poses
Training Time

Gaussian Process