NVIDIA

Perception with Confidence: A Conformal Prediction Perspective

Heng Yang
Research Scientist, NVIDIA
Incoming Assistant Professor, Harvard University

Oct. 23, 2022

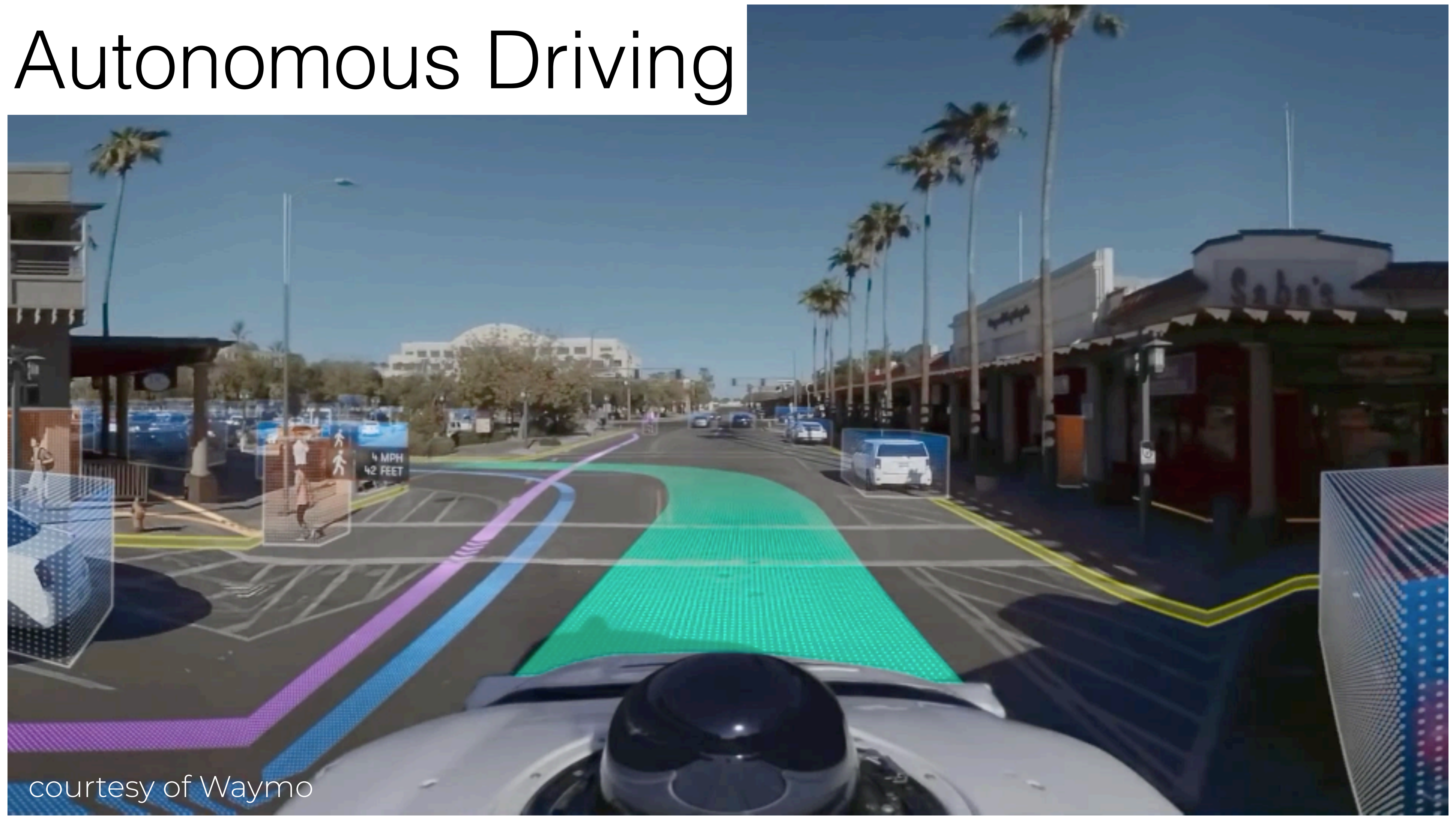
Joint work with Marco Pavone (NVIDIA/Stanford)

ECCV TEL AVIV 2022

Workshop on 3D Perception for Autonomous Driving

Robotics and Autonomy

Autonomous Driving

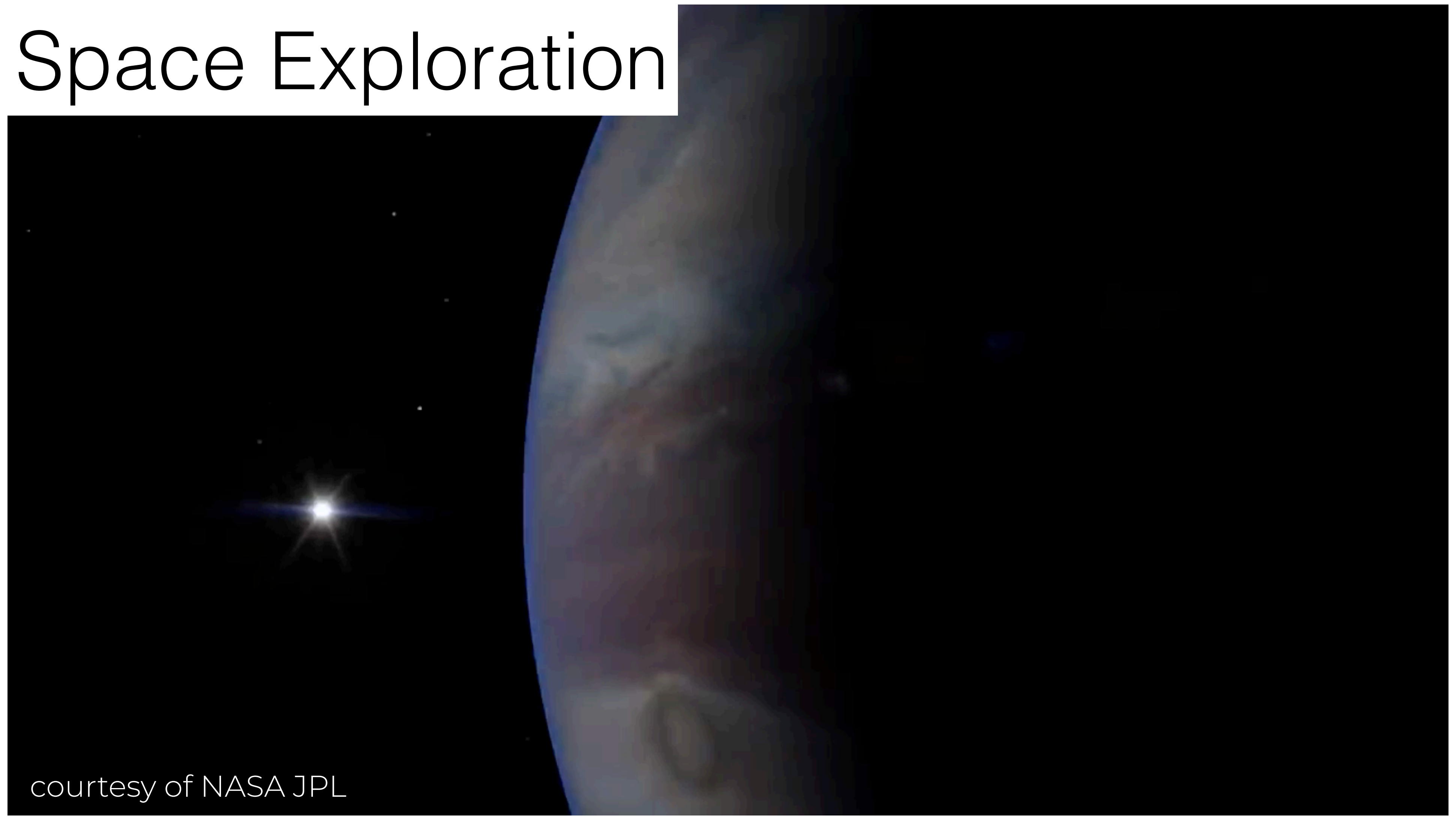


courtesy of Waymo

Intelligent Flight

courtesy of Skydio

Space Exploration



courtesy of NASA JPL

Search and Rescue

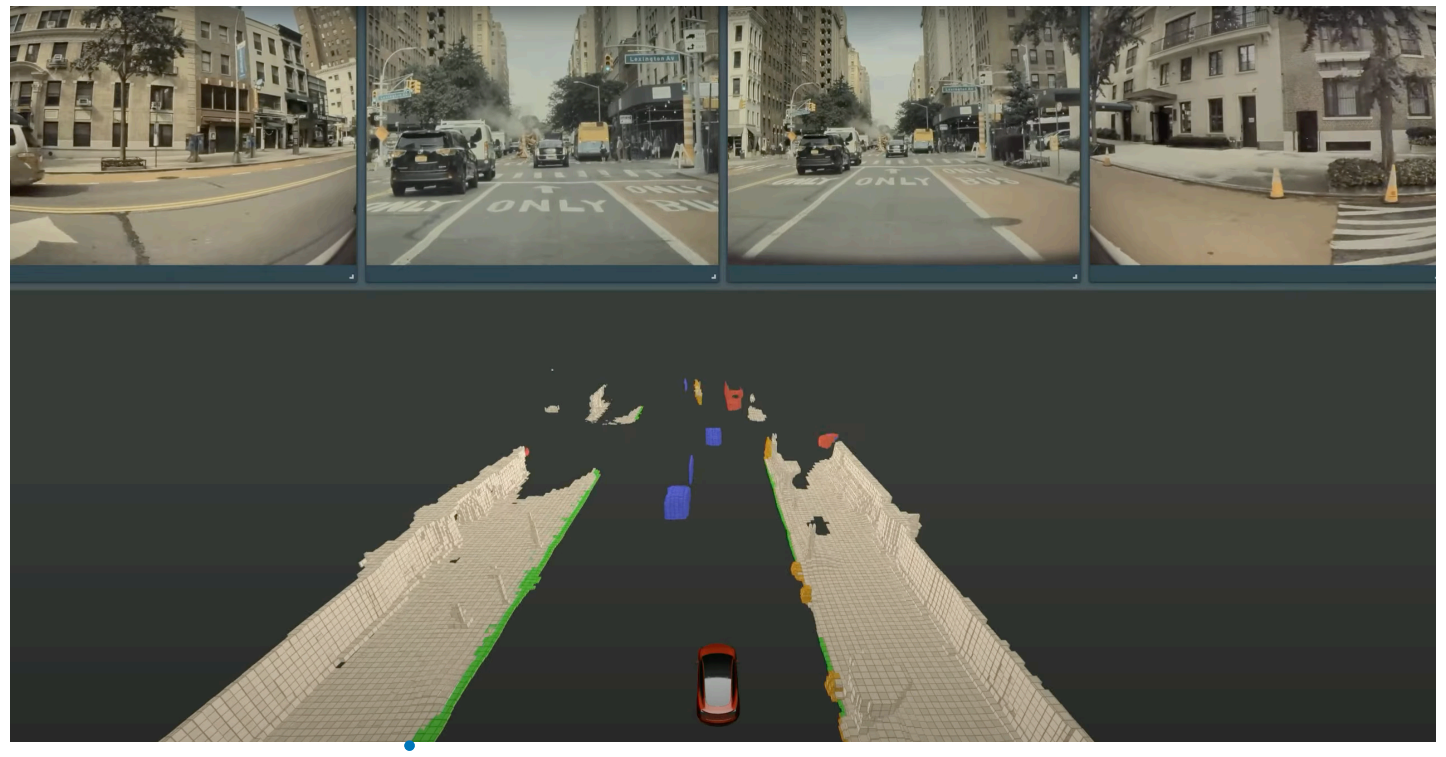
courtesy of NVIDIA

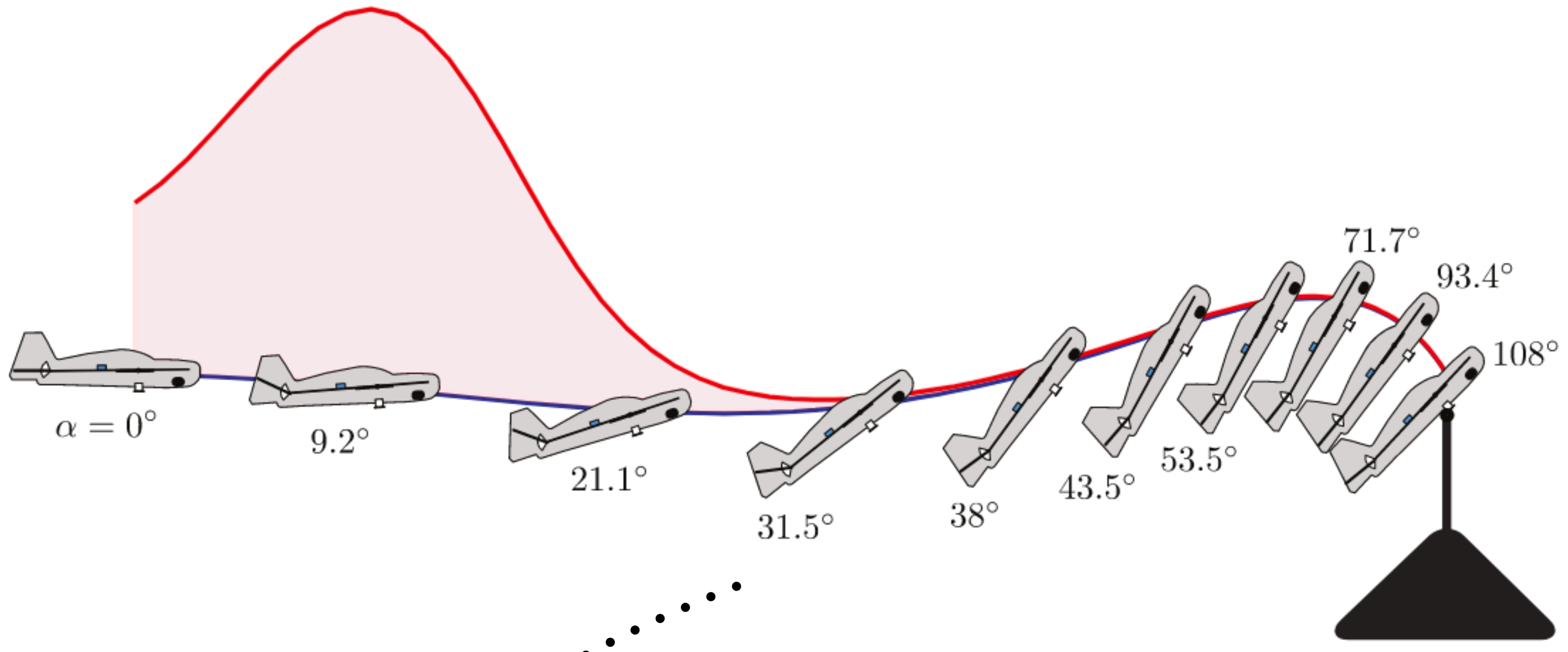
Robotics and Autonomy

Neural Representations — powerful but difficult to verify

Neural Vision

Classification
Segmentation
Detection
3D Vision
SLAM





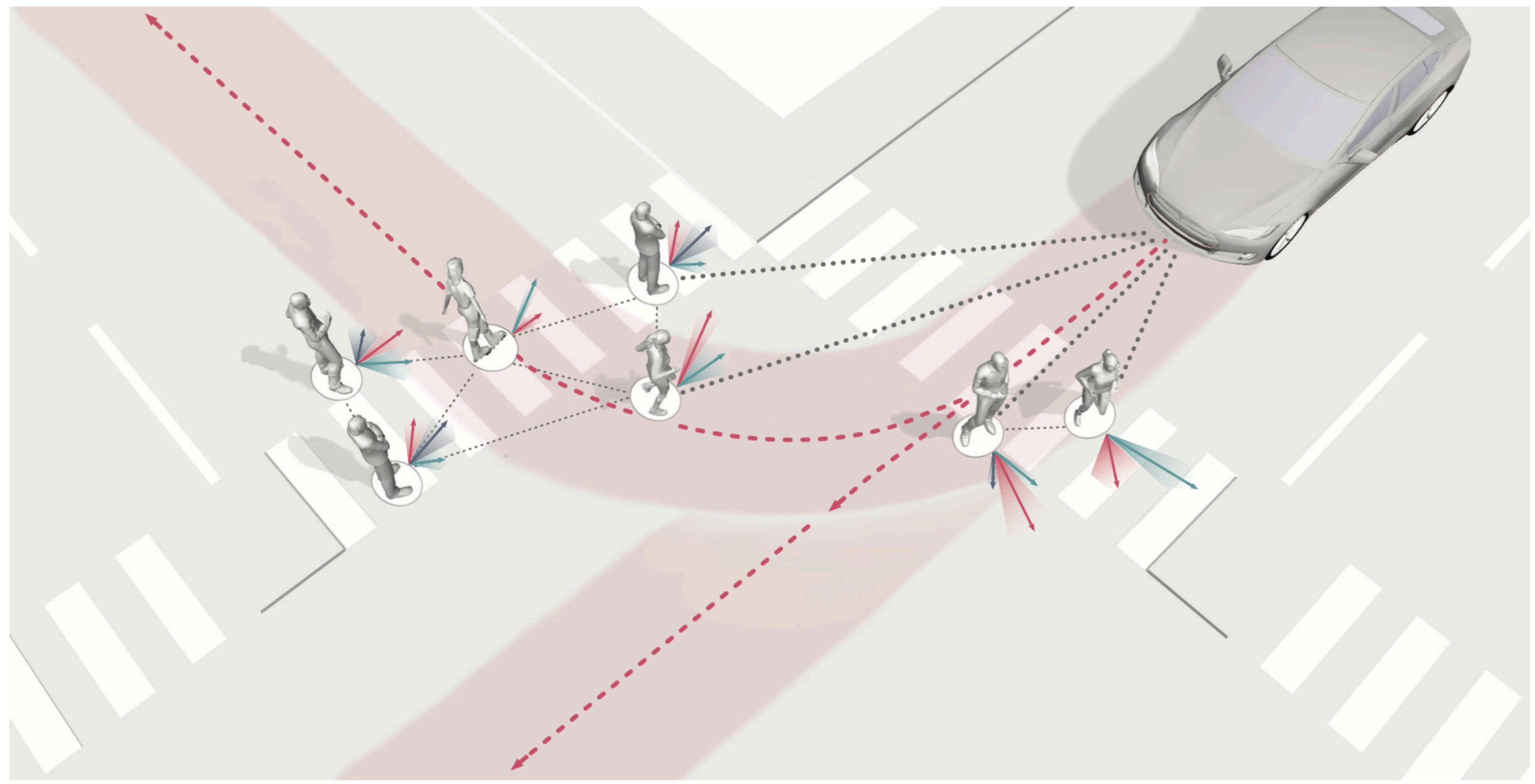
Neural Planning

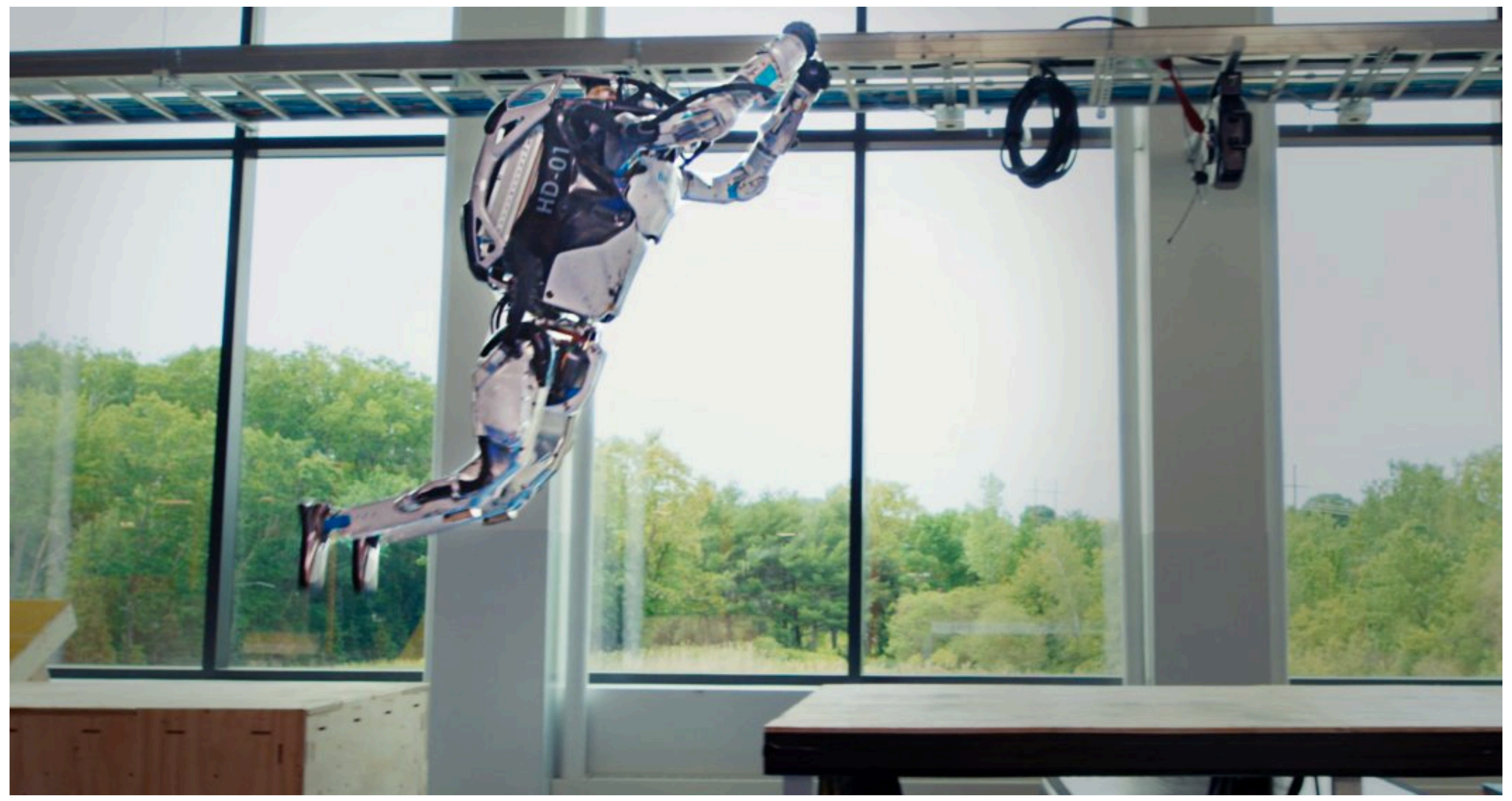
Differentiable planner
Learning to optimize

Sensing → Perception → Prediction → Planning → Control → Action

Neural Prediction

Behavior understanding
Relationship modeling





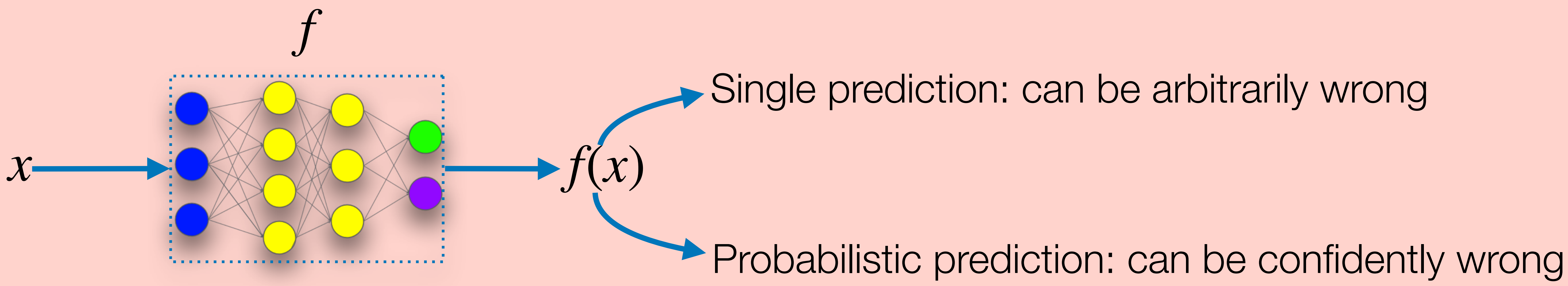
Neural Control

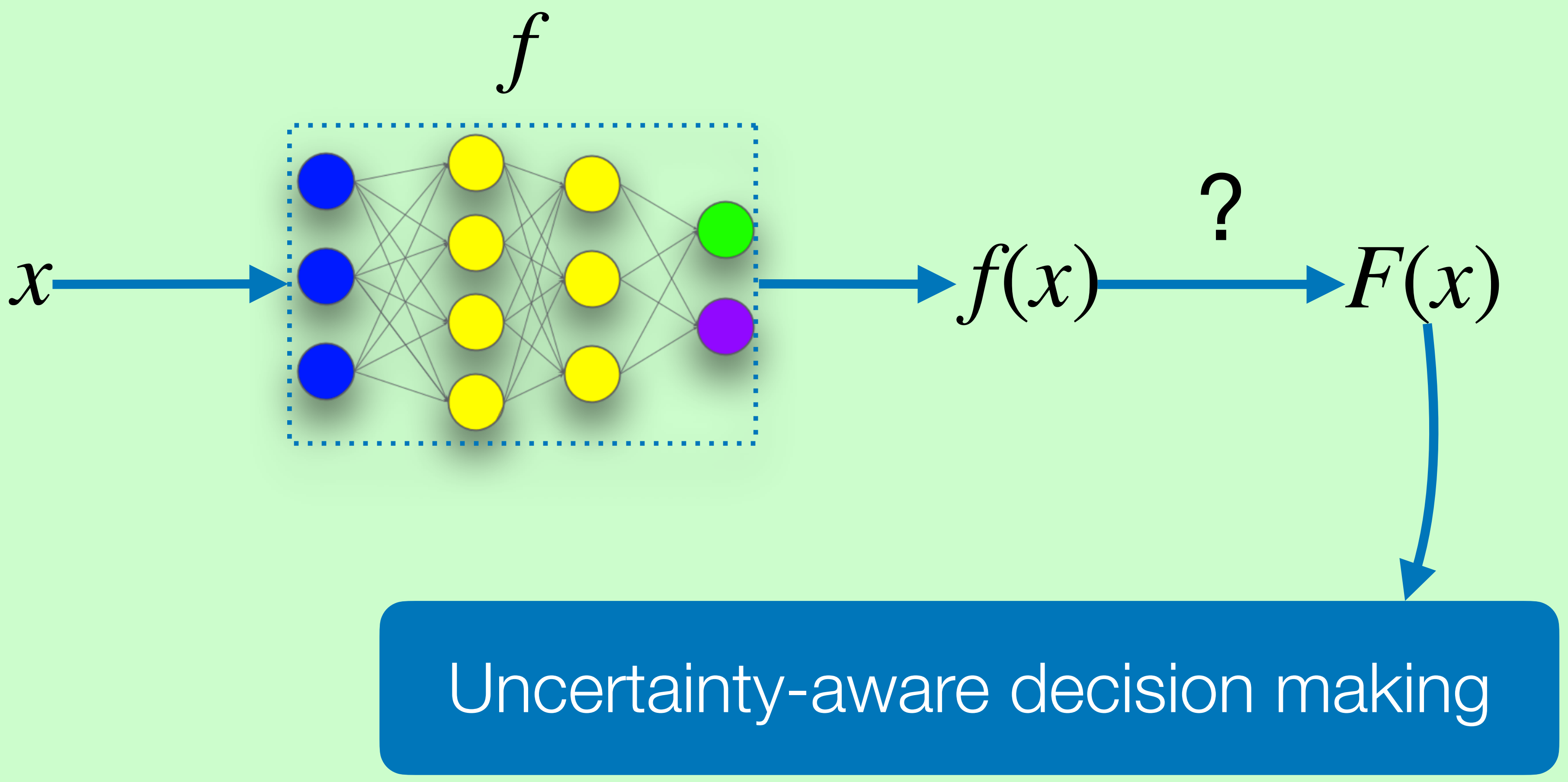
Complicated dynamics
High-dimensional certificates
Deep Policy

$$(x_i \leftrightarrow y_i), i = 1, \ldots, l$$

Observation	Label
Image	Bounding box
(State_t, Action_t)	State_{t+1}
Optimization	Solution

⋮



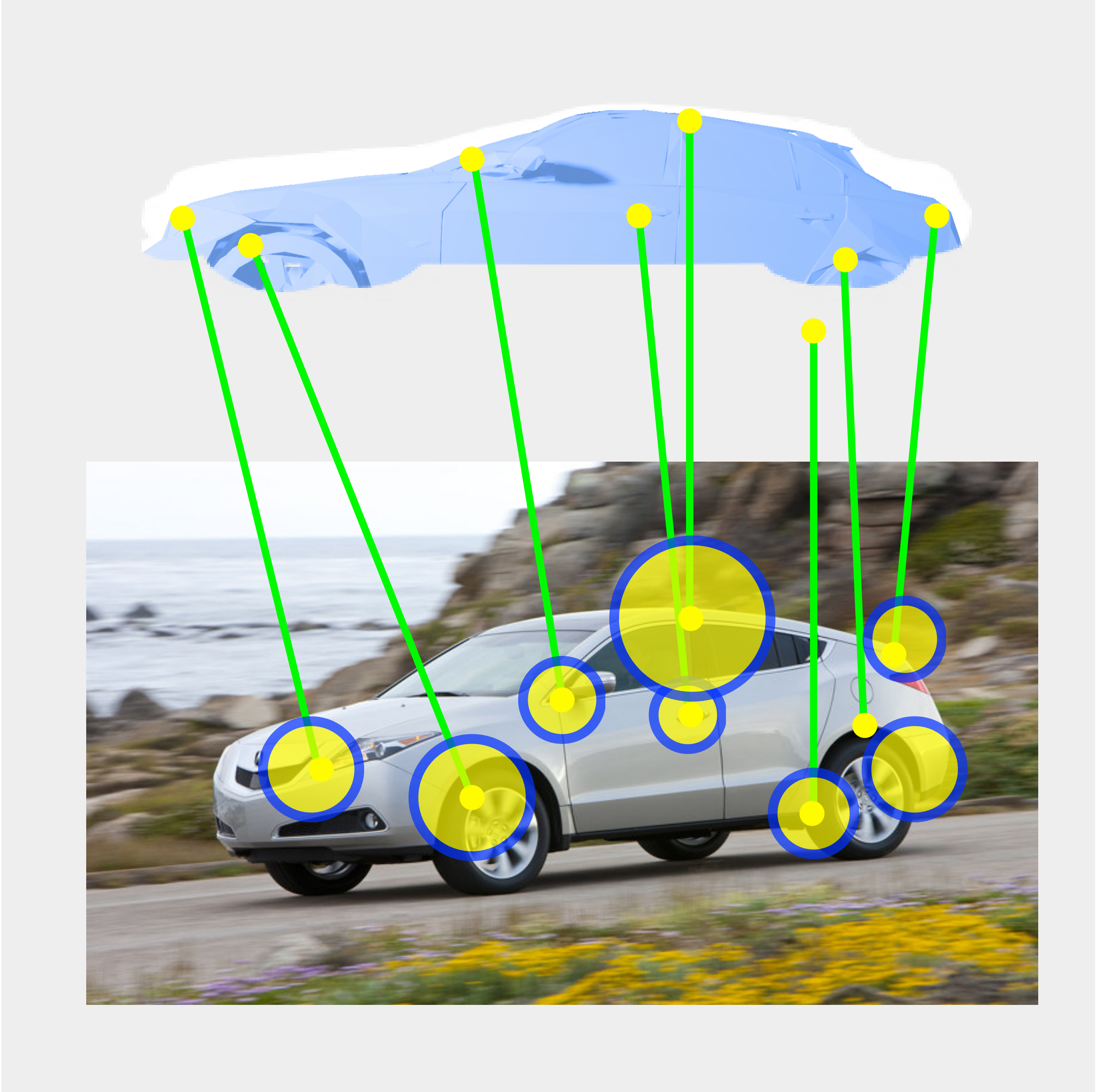


Provably correct prediction set

- Provable coverage: $y \in F(x)$
- Tight and adaptive
- Easy to compute
- Minimal assumption

This Talk: Provably Correct Conformal Prediction Set

Inductive Conformal Prediction



Probabilistically Correct Object Pose Estimation

Inductive Conformal Prediction (ICP)

Given a training set $\{z_i = (x_i, y_i)\}_{i=1}^{l}$ drawn i.i.d. from an unknown distribution Q on the space $\mathcal{Z} = \mathcal{X} \times \mathcal{Y}$

Training

- proper training set $\{z_i\}_{i=1}^{m}$, calibration set $\{z_i\}_{i=m+1}^{l}$ $(n := l - m)$
- Train any heuristic prediction f using the proper training set

Conformal Calibration

- Define a nonconformity function $S : \mathcal{Z}^m \times \mathcal{Z} \to \mathbb{R}$:

$$S(\{z_1, \ldots, z_m\}, z) = r(y, f(x))$$

- Compute the nonconformity scores in the calibration set:

$$\alpha_i = r(y_i, f(x_i)), i = m + 1, \ldots, l$$

- Sort the nonconformity scores: $\alpha_{\pi(1)} \geq \alpha_{\pi(2)} \geq \ldots \geq \alpha_{\pi(n)}$

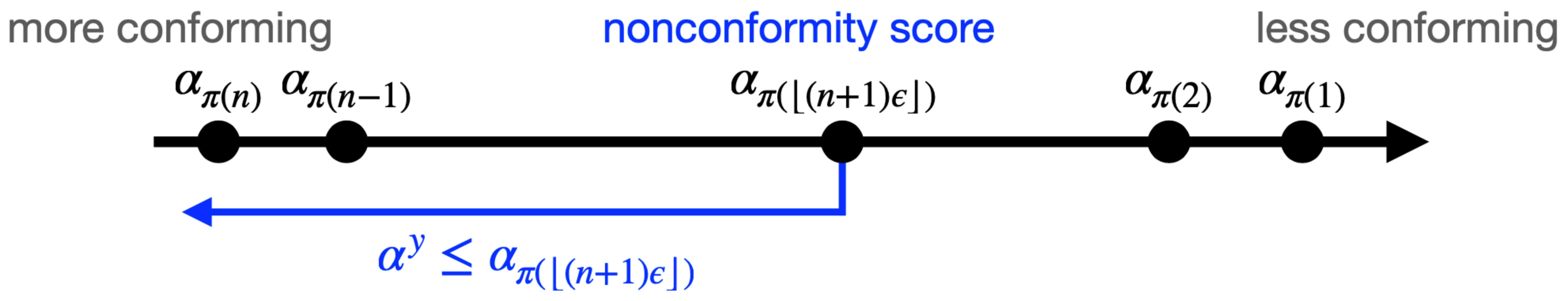
Conformal Prediction

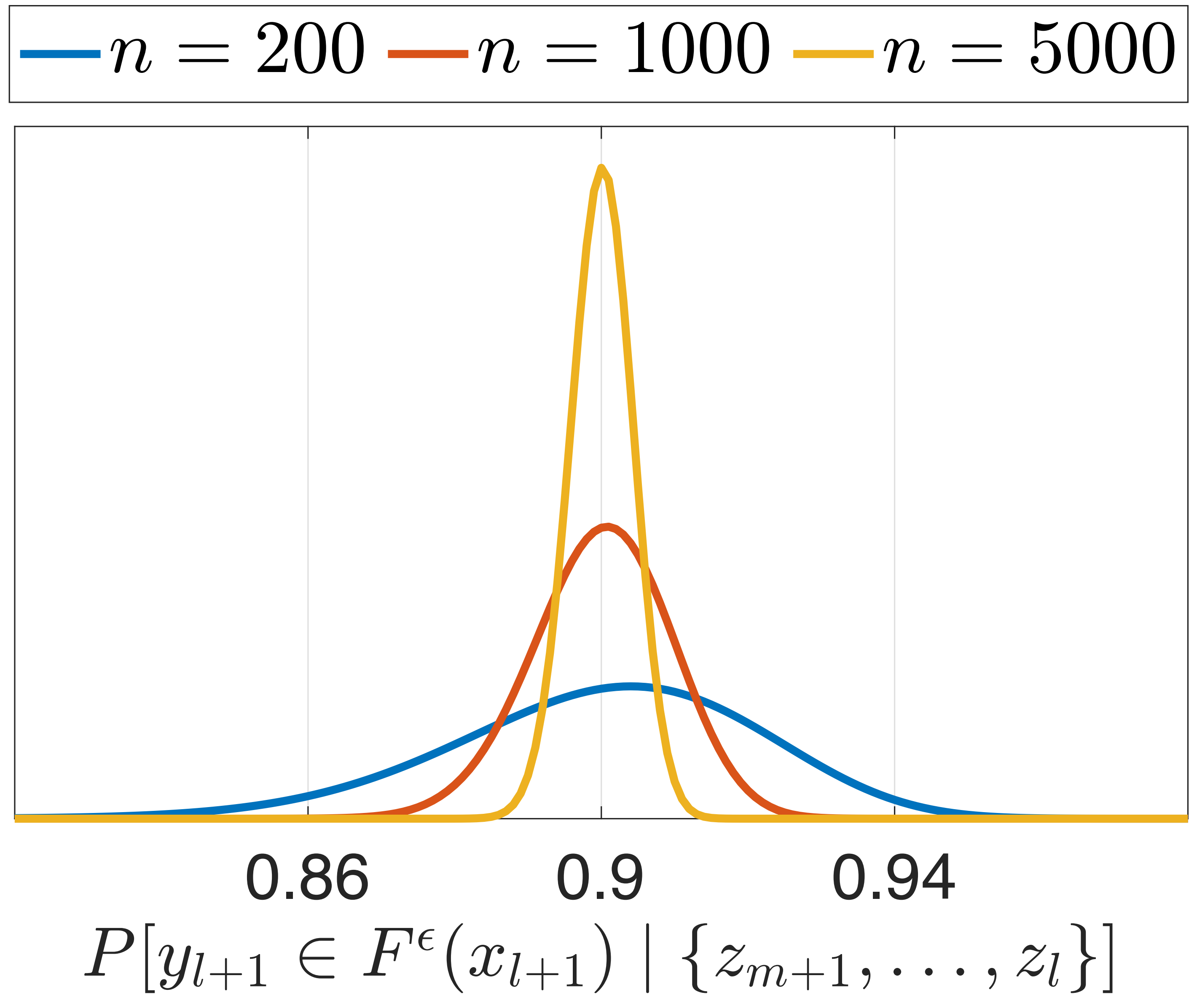
- Given x_{l+1}, and user-specified error rate $\epsilon \in (0,1)$, output

$$F^{\epsilon}(x) = \{y \in \mathcal{Y} \mid \alpha^y := r(y, f(x_{l+1})) \leq \alpha_{\pi(\lfloor (n+1)\epsilon \rfloor)}\}$$

Inductive Conformal Prediction (ICP)

$$F^{\epsilon}(x) = \{y \in \mathcal{Y} \mid \alpha^{y} := \boxed{r(y, f(x_{l+1}))} \leq \alpha_{\pi(\lfloor (n+1)\epsilon \rfloor)}\}$$





Theorem Validity of ICP

If $z_{l+1} = (x_{l+1}, y_{l+1})$ is exchangeable with $\{z_i\}_{i=m+1}^{l}$, then for any $0 < \epsilon < 1$

$$1 - \epsilon \leq \mathbb{P}[y_{l+1} \in F^{\epsilon}(x_{l+1})] \leq 1 - \epsilon + \frac{1}{n+1}$$

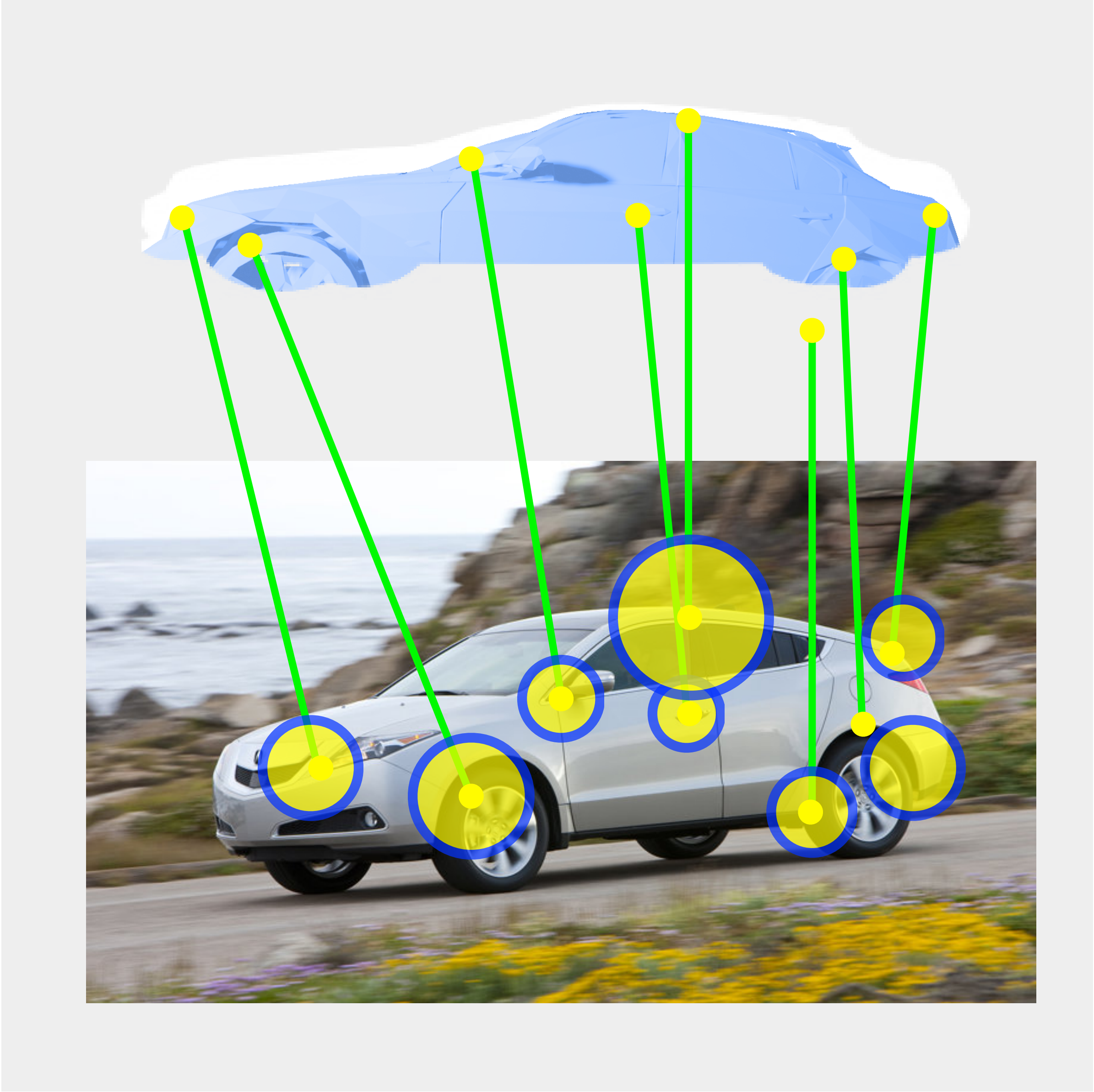
Moreover, conditioned on the calibration set, we have

$$\mathbb{P}\left[y_{l+1} \in F^{\epsilon}(x_{l+1}) \mid \{z_i\}_{i=m+1}^{l}\right] \sim \text{Beta}(n + 1 - t, t)$$

with $t = \lfloor (n+1)\epsilon \rfloor$.

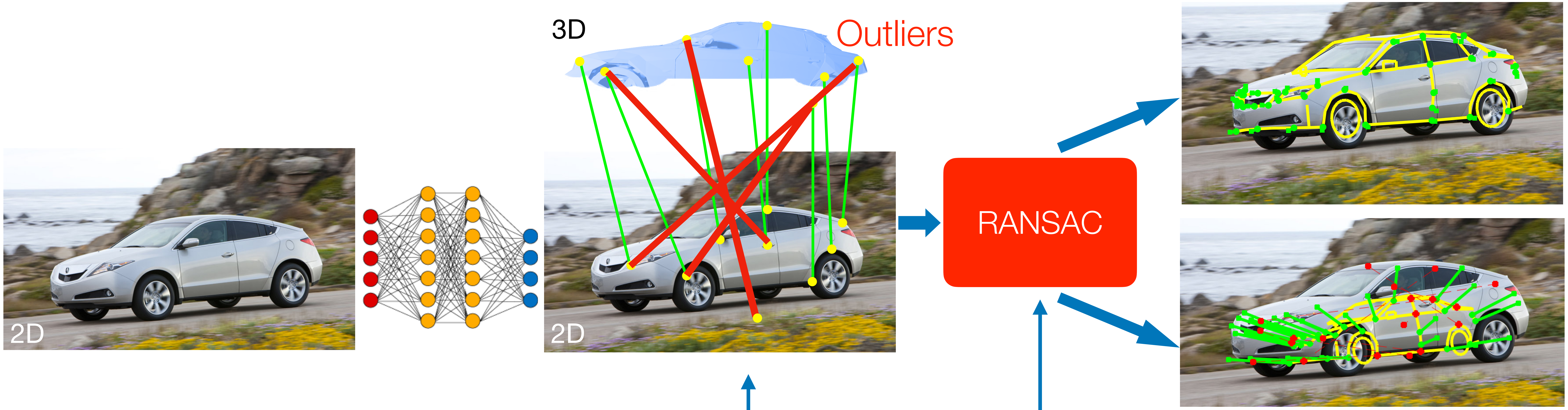
This Talk: Provably Correct Conformal Prediction Set

Inductive Conformal Prediction



Probabilistically Correct Object Pose Estimation

Object Pose Estimation

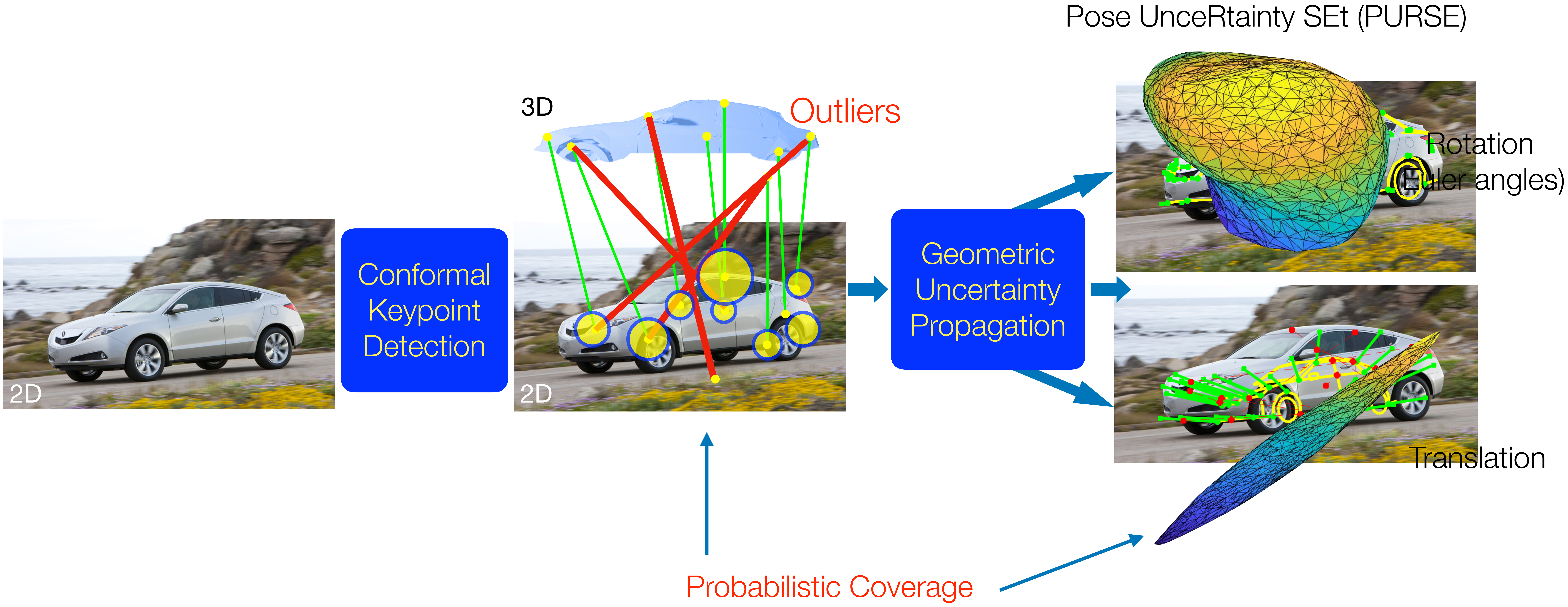


RANSAC can fail without notice

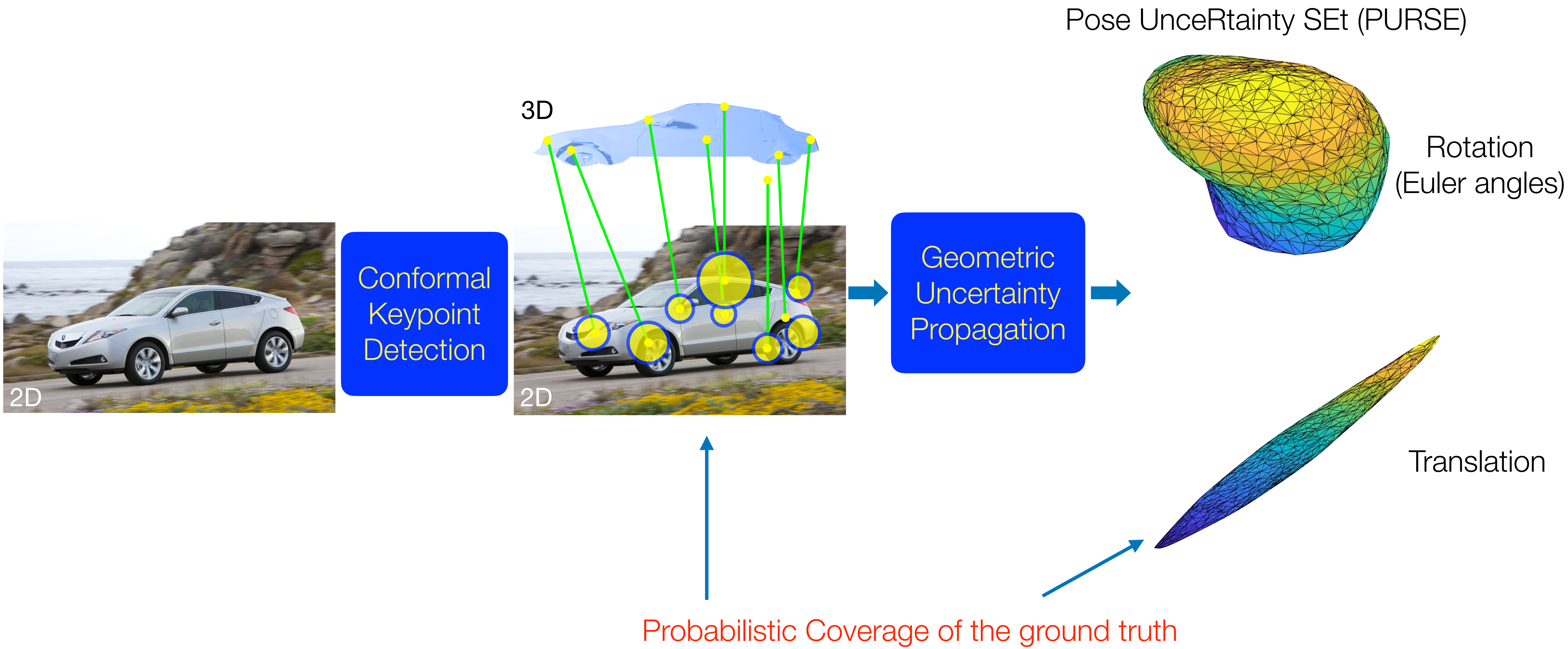
Keypoints can be arbitrarily wrong

No provably correct pose uncertainty

Probabilistically Correct Object Pose Estimation



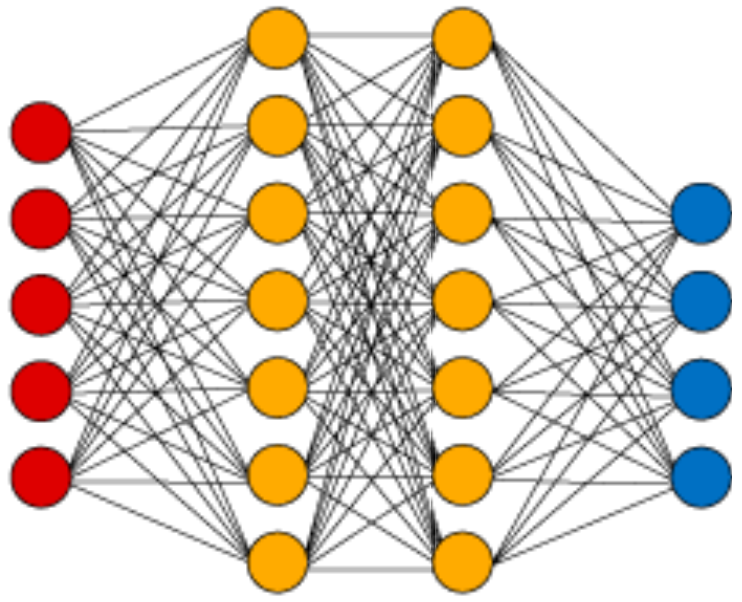
Probabilistically Correct Object Pose Estimation

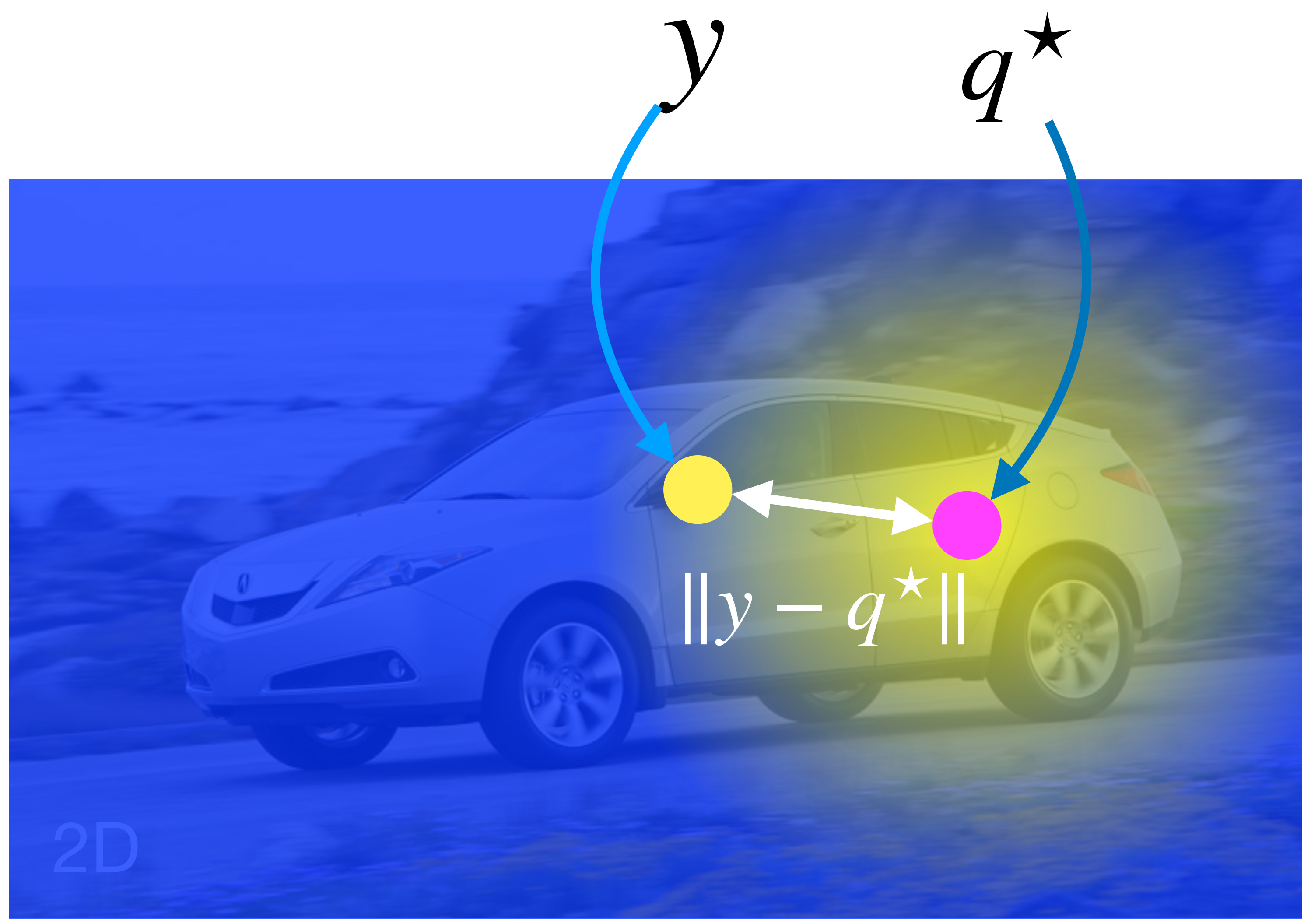


Conformal Keypoint Detection

x

f





Nonconformity function

$$r(y, f(x)) = p^\star \|y - q^\star\|$$

$$p^\star = \max\{f(x)\}$$

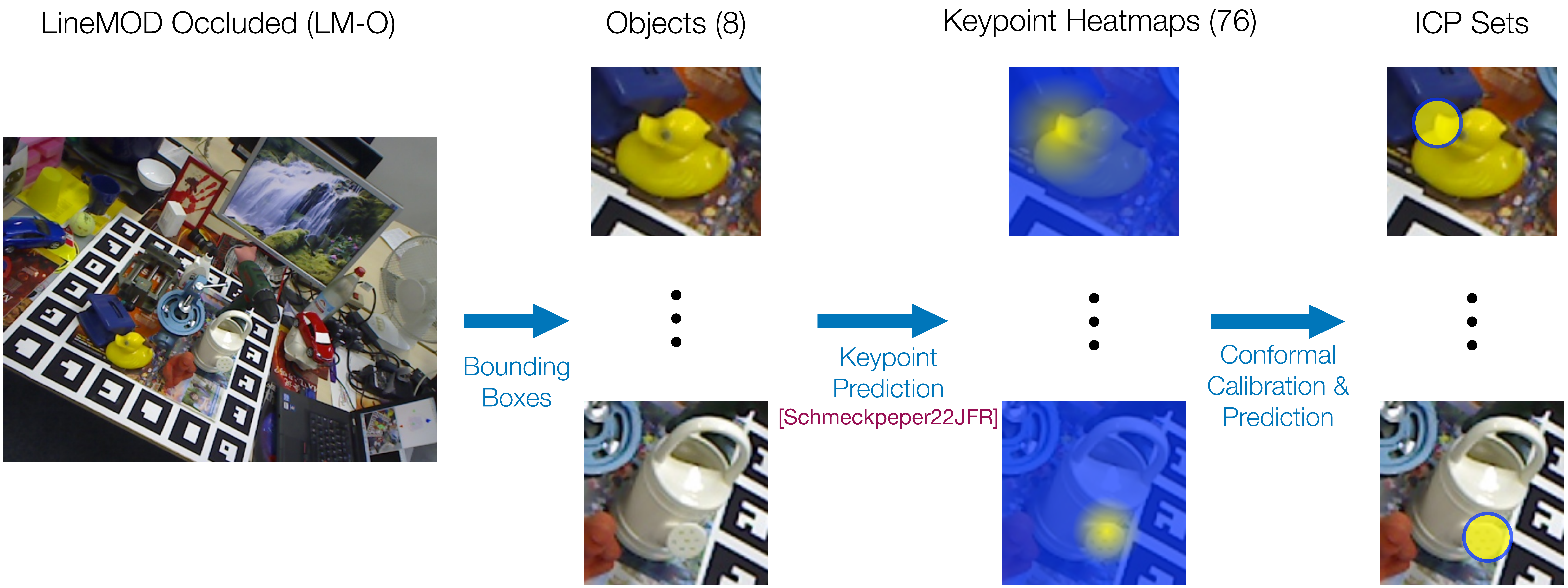
$$q^\star = \arg\max_{q \in x}\{f(x)\}$$

→ **Calibration** → Scores: $\{\alpha_{\pi(i)}\}_{i=1}^{n}$ → Quantile: $\alpha_{\pi(\lfloor (n+1)\epsilon \rfloor)}$ (User: ϵ) → **Prediction** (New sample: x_{l+1}) →

$$F^\epsilon = \left\{ \|y - q^\star_{l+1}\| \leq \frac{\alpha_{\pi(\lfloor (n+1)\epsilon \rfloor)}}{p^\star_{l+1}} \right\}$$

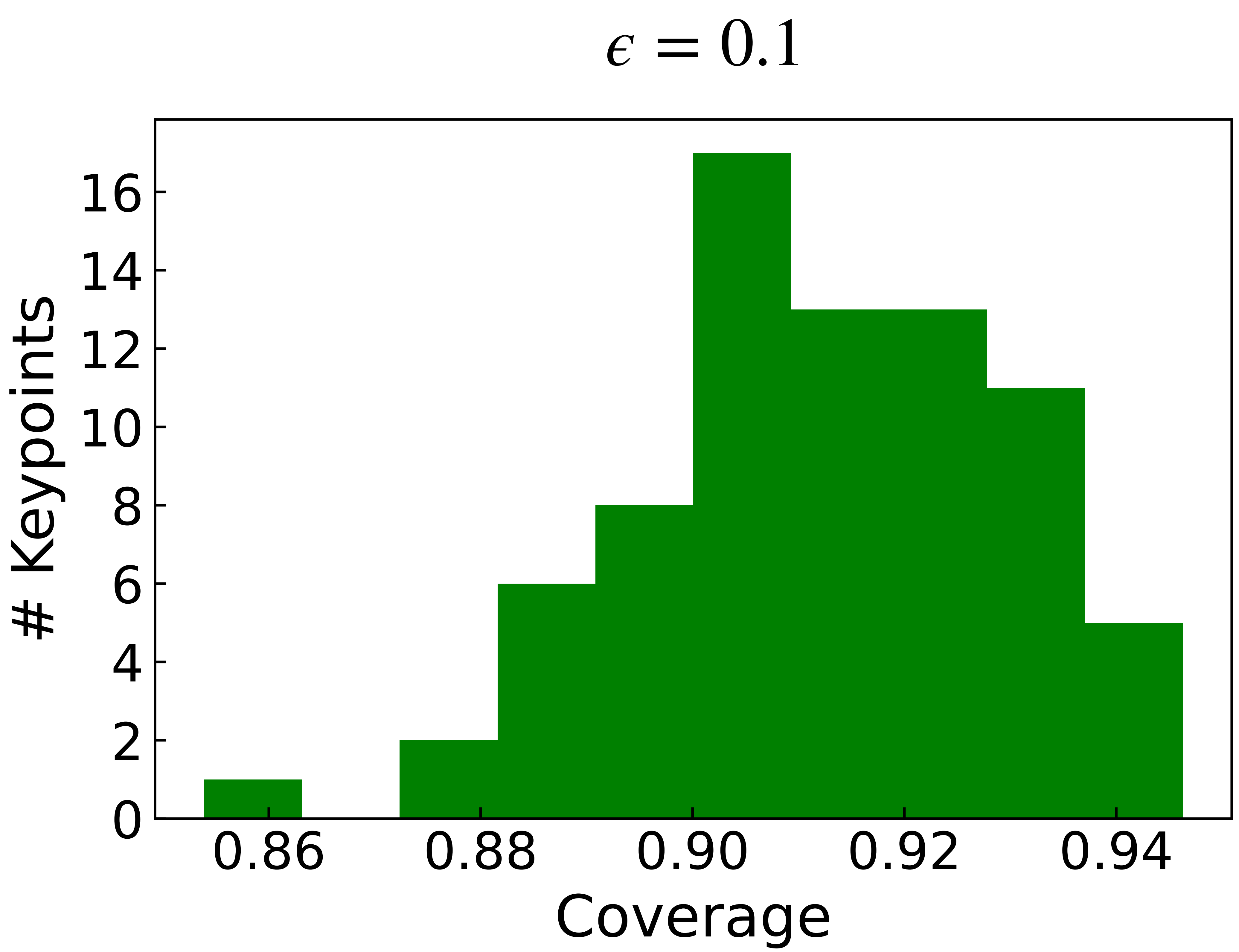
Center Radius

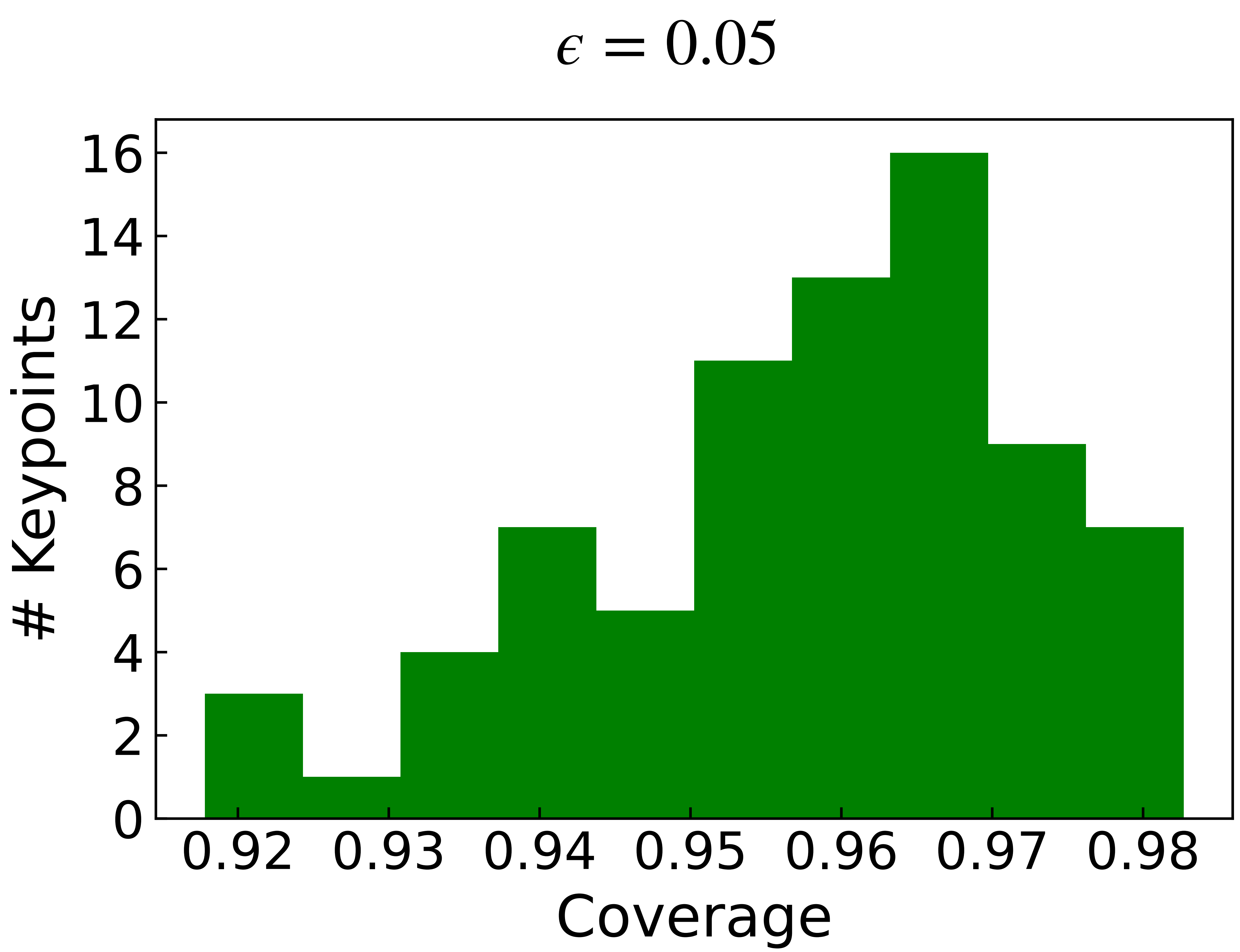
Does it work?



Calibration set size: 200 images; Test set size: 1214 images; $\epsilon = 0.1$ and $\epsilon = 0.05$

Valid Coverage

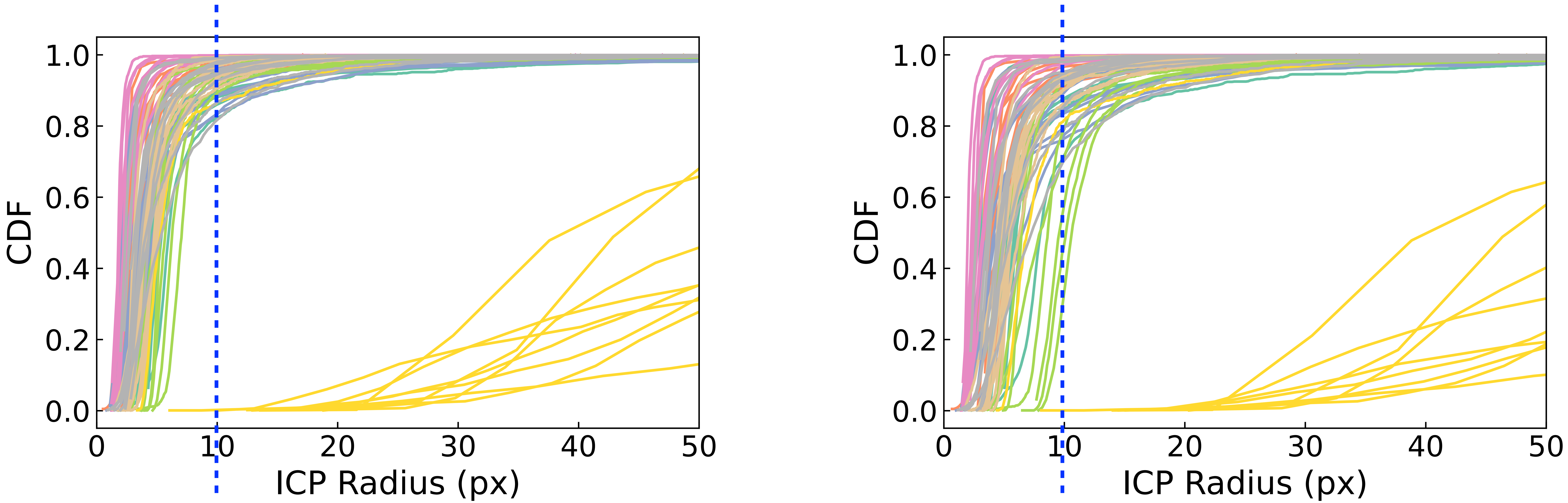




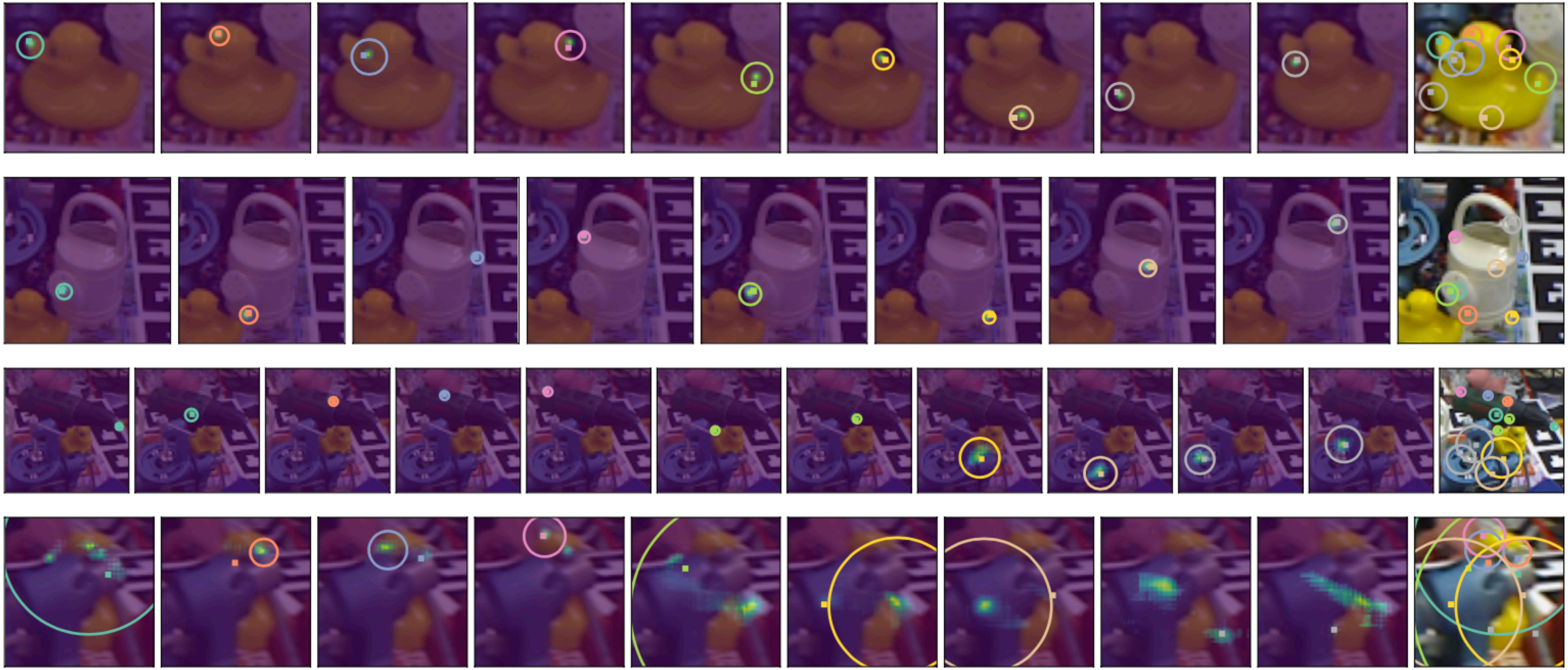
Tight

Heatmap size 64×64

$\epsilon = 0.1$ $\epsilon = 0.05$

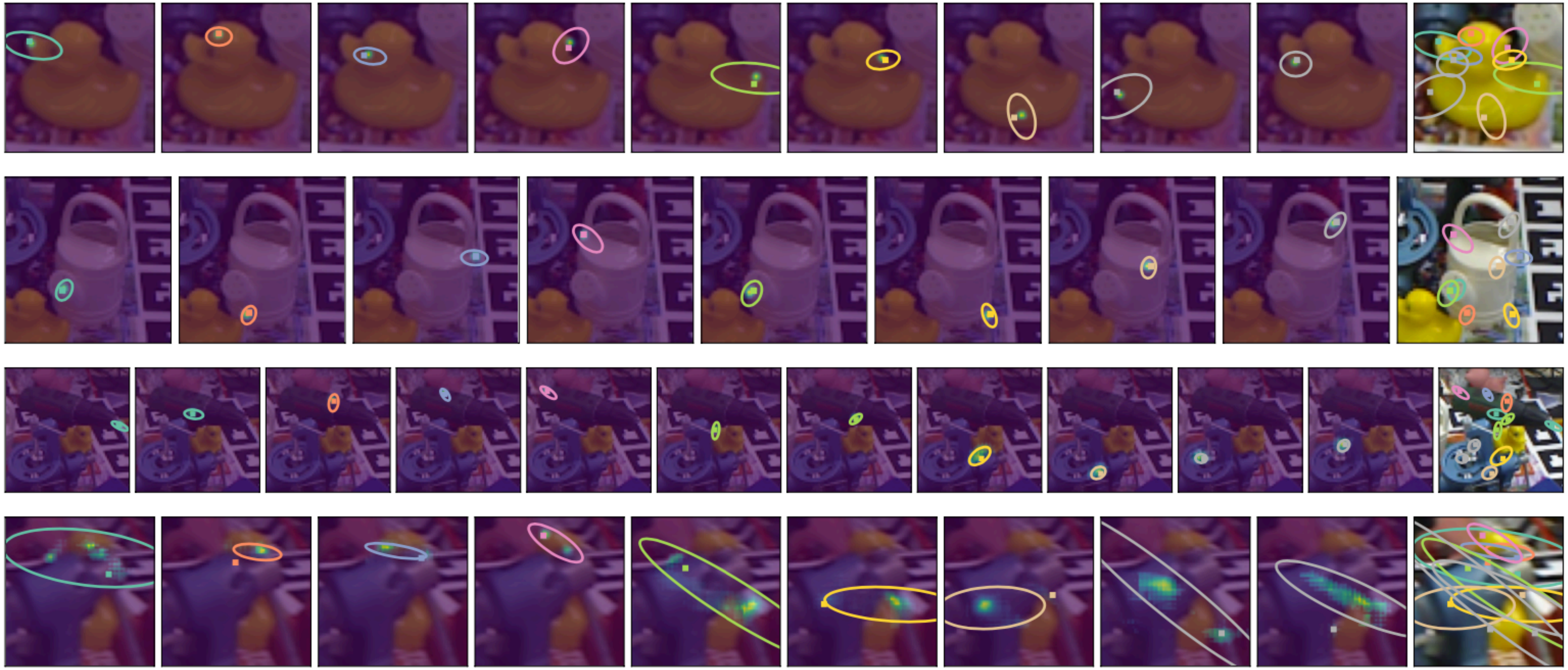


Adaptive

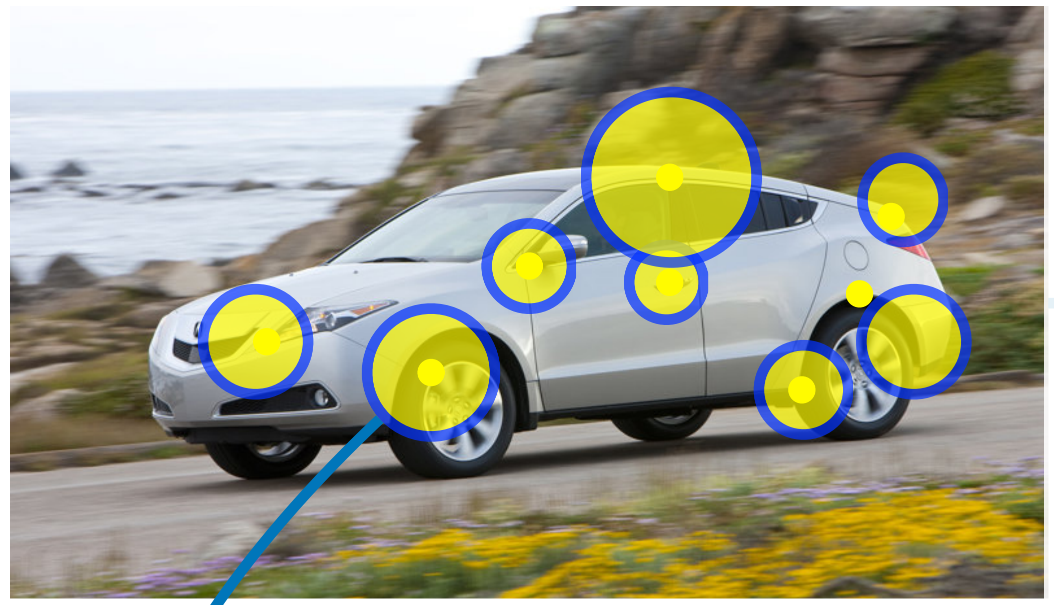


Variations

- We can design nonconformity functions to generate elliptical prediction sets [YP22NeurIPSWorkshop]
- We can "conformalize" the well-known Pixel Voting Network (PVNet)



Geometric Uncertainty Propagation



Uncertainty-aware Perspective-n-Points

Optimal pose $(R^\star, t^\star)$

- Difficult to compute
- No uncertainty quantification
- No bound to ground truth

$$\min_{(R,t)\in \mathrm{SE}(3)} \sum_{k=1}^{K} \|\Pi(RY_k + t) - q_k\|^2_{\Lambda_k}$$

Pose (to be estimated) | Camera projection | 3D Keypoint | Center of prediction set | "weight" depending on uncertainty of prediction set

k-th keypoint

With $(1-\epsilon)$ probability:

$$(y_k - q_k)^T \Lambda_k (y_k - q_k) \le 1$$

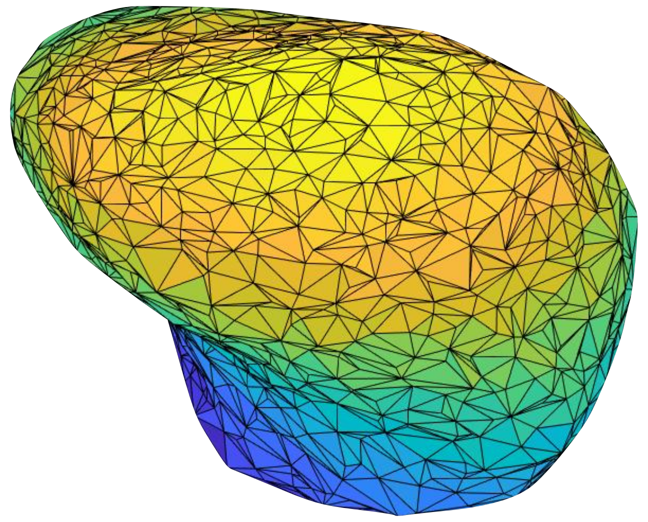
$$y_k = \Pi(RY_k + t)$$
$$p := [\mathrm{vec}(R)^T, t^T]^T$$

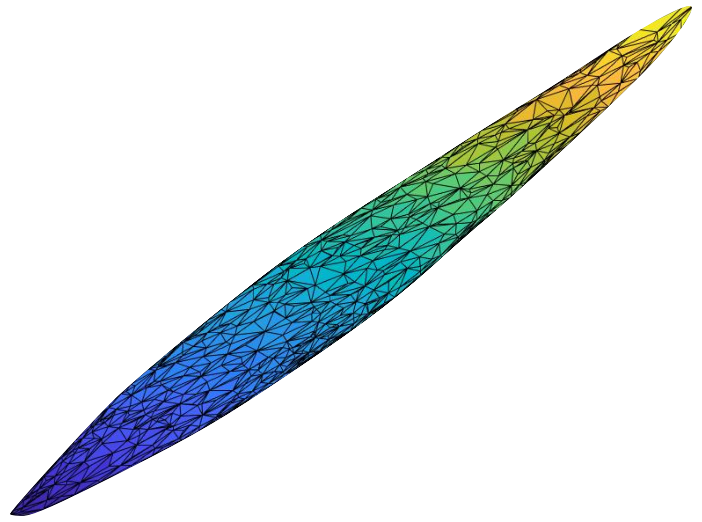
$$p^T A_k p \le 0$$

With $(1-\epsilon)^K$ probability:

$$p \in \mathscr{P} := \left\{ p \in \mathrm{SE}(3) \mid p^T A_k p \le 0, k = 1, \ldots, K \right\}$$

Pose UnceRtainty SEt (PURSE)





Characterizing Pose Uncertainty Set

Pose UnceRtainty SEt (PURSE)

With $(1-\epsilon)^K$ probability:

$$p \in \mathscr{P} := \left\{p \in \mathrm{SE}(3) \mid p^T A_k p \leq 0, k = 1, \ldots, K\right\}$$

Nonconvex set !!

Algorithm: Sample from PURSE

1. Random choose 3 (out of K) prediction sets (ball or ellipse)
2. Random sample 2D keypoints $\{\hat{y}_{k_i}\}_{i=1}^3$ inside the prediction sets
3. $\{(R_j, t_j)\}_{j=1}^4 =$ SolveP3P $\left(\{\hat{y}_{k_i} \leftrightarrow Y_{k_i}\}_{i=1}^3\right)$
4. If $\{(R_j, t_j)\}_{j=1}^4 \cap \mathscr{P} \neq \emptyset$: return success; else: go back to step 1

- Check membership
- Generate sample
- Approximate size ⟶ from samples
- Optimization ⟶ Semidefinite relaxation (Future work)

Does it work?

Coverage	Cat	Duck	Can	Ape	Driller	Eggbox	Glue	Holepuncher
$\epsilon = 0.1$ $(1-\epsilon)^K = 0.387$	0.761	0.772	0.686	0.826	0.771	0.741	0.748	0.672
$\epsilon = 0.05$ $(1-\epsilon)^K = 0.630$	0.880	0.855	0.819	0.932	0.867	0.885	0.883	0.824

PVNet Recall (2D Projection)	0.651	0.614	0.861	0.691	0.731	0.0843	0.554	0.698

No uncertainty quantification

Conclusions & Perspectives

Conclusions:

- Inductive conformal prediction
  - A simple, efficient, distribution-free statistical machinery for probabilistically correct prediction sets
  - Promising for a wide range of computer vision and robotics problems
- Probabilistically Correct Object Pose Estimation
  - Conformal keypoint detection: simple circular or elliptical prediction sets
  - Geometric uncertainty propagation: Pose Uncertainty Set (PURSE)

Perspectives:

- Assumption of exchangeability
- Use PURSE for downstream planning and control