Conformal Semantic Keypoint Detection with Statistical Guarantees

Heng Yang and **Marco Pavone**
NVIDIA Research
{hengy,mpavone}@nvidia.com

Abstract

Detecting semantic keypoints is a critical intermediate task for object detection and pose estimation from images. Existing approaches, albeit performing well on standard benchmarks, offer no provable guarantees on the quality of the detection. In this paper, we apply the statistical machinery of *inductive conformal prediction* that, given a *calibration* dataset (*e.g.,* 200 images) and a *nonconformity* function, converts a heuristic heatmap detection into a *prediction set* that *provably* covers the true keypoint location with a user-specified probability (*e.g.,* 90%). We design three different nonconformity functions leading to circular or elliptical prediction sets that are easy to compute. On the LINEMOD Occluded dataset we demonstrate that (i) the empirical coverage rate of the prediction sets is valid; (ii) the prediction sets are *tight*, *e.g.,* a ball with radius 10 pixels covers true keypoint locations on most test images; and (iii) the prediction sets are *adaptive*, *i.e.,* their sizes become larger for keypoints that are difficult to detect and smaller for easy instances.

1 Introduction

Image-based object detection and pose estimation is a longstanding problem in computer vision and robotics, finding extensive applications in augmented reality [7], autonomous driving [19], robotic manipulation [12], and space robotics [4]. One of the most popular and best-performing paradigms for object detection and pose estimation is a *two-stage* pipeline [15, 14, 17, 22, 27, 21, 20, 5], where the first stage (often called *feature matching* or *correspondence learning*) uses neural networks to detect (semantic) keypoints of the objects on the image, and the second stage (often called *model estimation*) computes the 6D rotation and translation of the objects by solving an optimization problem, known as *Perspective-n-Points* (PnP), that minimizes reprojection errors of the detected keypoints.

Safety-critical applications call for *provably correct* perception algorithms. For the two-stage pipeline in object pose estimation, classical [8, 18] and recent [26, 25] works have shown that it is possible to solve the PnP optimization in the second stage to provable *global optimality*. However, no existing works can provide performance guarantees for the neural network-based keypoint detector in the first stage. Therefore, it remains challenging to ensure the safety of the full two-stage pipeline.

Contributions. We contribute the first framework for detecting 2D semantic keypoints with *provable statistical guarantees*. Given a (heuristic) learned detector that outputs a *heatmap* of a target keypoint, we leverage the statistical machinery of *inductive conformal prediction* (ICP) [24, 2] to convert it into a *prediction set* that provably guarantees to contain the true keypoint location (a property called *coverage*) with a user-specified probability (*e.g.,* 90%). In order to apply the ICP framework, we need (i) a set of images (*e.g.,* 200) to *calibrate* the performance of the heuristic detector, which we show is straightforward to obtain in popular benchmarks [6]; and (ii) a *nonconformity* function that measures the performance of the heuristic detector, for which we instantiate three different designs that are intuitive, easy to compute, and lead to simple prediction sets such as balls and ellipses (*cf.* Fig. 4). We validate our framework on the LINEMOD Occluded dataset (LM-O) [3] using the keypoint detector in [14] and demonstrate that: (i) the user-specified probability of coverage is empirically valid for all three nonconformity functions; (ii) the prediction sets are *tight*, *e.g.,* a ball with radius 10 pixels covers the true keypoint location on most test images; (iii) the prediction sets are *adaptive*, *i.e.,* they become larger for keypoints that are difficult to detect and smaller for easy instances.

Paper organization. Section 2 introduces the ICP framework. Section 3 applies ICP to keypoint detection. Section 4 demonstrates experimental results and Section 5 provides concluding remarks.

2 Preliminaries: Inductive Conformal Prediction

Given a set $\{z_i = (x_i, y_i)\}_{i=1}^{l}$ with observation $x_i \in \mathcal{X}$ and label $y_i \in \mathcal{Y}$ such that each $z_i \in \mathcal{Z} := \mathcal{X} \times \mathcal{Y}$ is drawn i.i.d. from an *unknown* distribution on $\mathcal{Z}$, inductive conformal prediction (also called split conformal prediction) [13, 2] provides a simple yet powerful framework to learn a *set prediction* $F^\epsilon \subseteq \mathcal{Y}$, parameterized by a user-specified error rate $0 < \epsilon < 1$, such that given a new sample $z_{l+1} = (x_{l+1}, y_{l+1})$ satisfying an *exchangeability* condition (elaborated in Theorem 1), we have

$$\mathbb{P}\left[y_{l+1} \in F^\epsilon(x_{l+1})\right] \geq 1 - \epsilon, \tag{1}$$

i.e., the prediction set F^ϵ guarantees to contain the true label y_{l+1} with probability at least $1 - \epsilon$.

Training. We start by dividing $\{z_i\}_{i=1}^{l}$ into a *proper training set* $\{z_1, \ldots, z_m\}$ and a *calibration set* $\{z_{m+1}, \ldots, z_l\}$. We shorthand $n = l - m$ as the size of the calibration set. We learn a heuristic prediction function $f : \mathcal{X} \to \tilde{\mathcal{Y}}$ from the proper training set using *any* method. This flexibility allows us to fully exploit the power of modern deep learning. The prediction space $\tilde{\mathcal{Y}}$ can be the same as the label space $\mathcal{Y}$, or can contain auxiliary information such as a heuristic notion of uncertainty (*e.g.,* softmax scores in classification or a heatmap in the case of keypoint detection).

Conformal calibration. Leveraging the learned f, we define a *nonconformity* function $S : \mathcal{Z}^m \times \mathcal{Z} \to \mathbb{R}$ to measure how well a given sample $z = (x, y)$ *conforms* to the proper training set:

$$S\left(\{z_1, \ldots, z_m\}, (x, y)\right) = r(y, f(x)), \tag{2}$$

where $r : \mathcal{Y} \times \tilde{\mathcal{Y}} \to \mathbb{R}$ is a measure of disagreement between the label y and the prediction $f(x)$. For example, consider $\mathcal{Y} = \tilde{\mathcal{Y}} = \mathbb{R}$, one can design $r(y, f(x)) = |y - f(x)|$: if (x, y) poorly conforms to the training set, f will incur large errors. While the function S can be arbitrary, (2) is a typical and convenient definition since f is implicitly dependent on $\{z_i\}_{i=1}^{m}$. We then compute the nonconformity score for each sample in the calibration set as $\alpha_i = r(y_i, f(x_i)), i = m + 1, \ldots, l$, and sort them in *nonincreasing* order $\alpha_{\pi(1)} \geq \ldots \geq \alpha_{\pi(n)}$, where $\pi(i) \in \{m + 1, \ldots, l\}$ is an index permutation.

Conformal prediction. Given a new observation x_{l+1} (with an unknown label y_{l+1}) and a user-specified error rate $\epsilon \in (0, 1)$, we compute the inductive conformal prediction (ICP) set as

$$F^\epsilon(x_{l+1}) = \left\{y \in \mathcal{Y} \mid \alpha^y \leq \alpha_{\pi(\lfloor(n+1)\epsilon\rfloor)}\right\}, \tag{3}$$

where $\alpha^y = r(y, f(x_{l+1}))$ is the nonconformity score of the new sample when fixing the label to be y. Fig. 1 illustrates the conformal calibration and prediction procedure, and we have the following result stating that the ICP set (3) provides a valid statistical coverage of the true label y_{l+1}.

Theorem 1 (Validity of ICP Coverage [24, 9, 23]) *If* $z_{m+1}, \ldots, z_l, z_{l+1} = (x_{l+1}, y_{l+1})$ *are exchangeable, i.e., their distribution is invariant under permutation, then*

$$1 - \epsilon \leq \mathbb{P}\left[y_{l+1} \in F^\epsilon(x_{l+1})\right] \leq 1 - \epsilon + 1/(n + 1) \tag{4}$$

for any $\epsilon \in (0, 1)$. *Conditioned on the calibration set, we have*

$$\mathbb{P}\left[y_{l+1} \in F^\epsilon(x_{l+1}) \mid \{z_{m+1}, \ldots, z_l\}\right] \sim \text{Beta}(n + 1 - t, t) \tag{5}$$

with $t = \lfloor(n+1)\epsilon\rfloor$, *i.e., the conditional coverage satisfies a Beta distribution.*

A few remarks are in order. First, asking $z_{m+1}, \ldots, z_l, z_{l+1}$ to be exchangeable is weaker than asking them to be independent. This weaker assumption can be beneficial in handling temporal correlations [11], though for our experiments on LM-O, the calibration and test images are fully i.i.d. Second, under the exchangeability assumption, $\alpha_{l+1} := r(y_{l+1}, f(x_{l+1}))$ is *equally likely* to fall in anywhere between the calibration scores $\{\alpha_{\pi(i)}\}_{i=1}^{n}$ in Fig. 1. Hence, $\mathbb{P}\left[y_{l+1} \in F^\epsilon(x_{l+1})\right] = \mathbb{P}\left[\alpha_{l+1} \leq \alpha_{\pi(\lfloor(n+1)\epsilon\rfloor)}\right] = 1 - \lfloor(n+1)\epsilon\rfloor/(n+1) \geq 1 - \epsilon$ and the lower bound in (4) can be intuitively proved. The upper bound in (4) states that $1 - \epsilon$ is not overly conservative (indeed tight if n is large) and there exist test samples for which the ICP set fails to contain the true label. Lastly, and importantly, the probabilistic guarantee in (4) is *marginal* over the randomness of the calibration set, which means the *empirical coverage*, given different calibration sets, will fluctuate as predicted by the *conditional* distribution in (5). Fig. 2 plots the Beta distribution at $\epsilon = 0.1$ with different sizes of the calibration set. We observe that as n increases the empirical coverage becomes more concentrated at $1 - \epsilon$. In Section 4, our experiments show that even with a small ($n = 200$) calibration set, the empirical coverage is close to, and mostly higher than, $1 - \epsilon$.
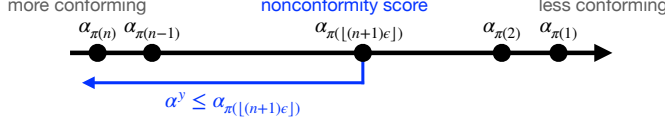
Figure 1: Given a learned prediction function and a calibration set of n samples, conformal calibration uses a nonconformity function (2) to compute and sort nonconformity scores $\{\alpha_{\pi(i)}\}_{i=1}^{n}$. Given a new observation and an error rate ϵ, conformal prediction (3) outputs a prediction set of all labels under which the nonconformity score of the new sample is no larger than $\alpha_{\pi(\lfloor (n+1)\epsilon \rfloor)}$.
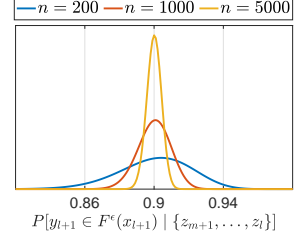
Figure 2: Beta distribution of the conditional coverage with different n in (5).

3 Conformal Semantic Keypoint Detection

To apply ICP to keypoint detection, we need a heuristic detector f and a nonconformity function r.

Heatmap-based keypoint detector. Given an RGB image $x \in \mathbb{R}^{H \times W \times 3}$, heatmap-based methods [14, 17, 10] output $f(x) \in \Delta^{HW} := \{v \in \mathbb{R}^{HW} \mid v_i \geq 0, \sum_i^{HW} v_i = 1\}$ where $f(x)_i \geq 0$ indicates the probability of the keypoint on the i-th pixel (we vertically concatenate all pixels). Let $\pi(\cdot)$ denote an index permutation that sorts $f(x)$ as $f(x)_{\pi(1)} \geq \ldots \geq f(x)_{\pi(HW)}$.

Nonconformity function. It is tempting to treat keypoint detection as a classification problem and adopt popular nonconformity functions designed for classification (*e.g.,* the one in [16, 1]). However, in the Supplementary Material we show the resulting prediction sets are loose and hard to interpret. This motivates us to design the following three nonconformity functions.

(I) Peak. Let $q^{\star} \in \mathbb{R}^2$ be the pixel location with maximum probability $p^{\star} := f(x)_{\pi(1)}$. We design

$$r(y, f(x)) = p^{\star} \|y - q^{\star}\| . \tag{Peak}$$

According to (3), we have the following inductive conformal prediction set

$$F^{\epsilon}(x_{l+1}) = \left\{ y \in \mathcal{Y} \mid \|y - q^{\star}_{l+1}\| \leq \alpha_{\pi(\lfloor (n+1)\epsilon \rfloor)}/p^{\star}_{l+1} \right\} . \tag{ICP-Peak}$$

This ICP set is a ball centered at $q^{\star}_{l+1}$ with a radius *inversely* proportional to $p^{\star}_{l+1}$. Intuitively, when $p^{\star}_{l+1}$ is small, *i.e.,* $f(x)$ is uncertain, the ball is enlarged to account for higher uncertainty.

(II) Variance. Let $q_i \in \mathbb{R}^2$ be the i-th pixel location. Compute $\bar{q} = \sum_{i=1}^{K} f(x)_{\pi(i)} \cdot q_{\pi(i)}$ as the expected location of the top-K keypoints, and $\gamma^2 = \sum_{i=1}^{K} f(x)_{\pi(i)} \cdot \|q_{\pi(i)} - \bar{q}\|^2$ as the "variance" (we use $K = 100$ pixels because the heatmap can be quite noisy). We design

$$r(y, f(x)) = \|y - \bar{q}\|/\gamma. \tag{Var}$$

According to (3), we have the inductive conformal prediction set

$$F^{\epsilon}(x_{l+1}) = \left\{ y \in \mathcal{Y} \mid \|y - \bar{q}_{l+1}\| \leq \gamma_{l+1}\alpha_{\pi(\lfloor (n+1)\epsilon \rfloor)} \right\} . \tag{ICP-Var}$$

This ICP set is a ball centered at $\bar{q}_{l+1}$ with a radius proportional to the "standard deviation" γ_{l+1}. Intuitively, when the heatmap is spread out and $f(x)$ has higher uncertainty, the ball becomes larger.

(III) Covariance. Compute the expected top-K keypoint location $\bar{q}$ as before. Then compute the *covariance matrix* $\Sigma = \sum_{i=1}^{K} f(x)_{\pi(i)} \cdot (q_{\pi(i)} - \bar{q})(q_{\pi(i)} - \bar{q})^{\mathsf{T}}$. We design

$$r(y, f(x)) = (y - \bar{q})^{\mathsf{T}} \Sigma^{-1} (y - \bar{q}). \tag{Cov}$$

According to (3), we have the inductive conformal prediction set

$$F^{\epsilon}(x_{l+1}) = \left\{ y \in \mathcal{Y} \mid (y - \bar{q}_{l+1})^{\mathsf{T}} \Sigma^{-1}_{l+1} (y - \bar{q}_{l+1}) \leq \alpha_{\pi(\lfloor (n+1)\epsilon \rfloor)} \right\} . \tag{ICP-Cov}$$

This ICP set is an ellipse centered at $\bar{q}_{l+1}$. The eigenvectors of $\Lambda_{l+1} := \Sigma^{-1}_{l+1}/\alpha_{\pi(\lfloor (n+1)\epsilon \rfloor)}$ point in the directions of the principal axes, while the eigenvalues are $1/a^2$ and $1/b^2$, with $a \leq b$ the lengths of the semi-axes. Similar to (ICP-Var), the ellipse gets larger if the heatmap has higher uncertainty. Different from (ICP-Var), the ellipse better captures nonuniform uncertainty, as we show in Section 4.
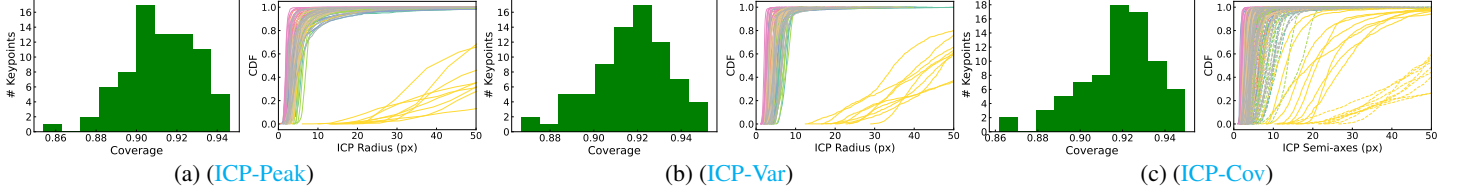
3

(a) (ICP-Peak) (b) (ICP-Var) (c) (ICP-Cov)

Figure 3: ICP statistics on the LM-O dataset [3] with $\epsilon = 0.1$. In (c), solid and dashed lines correspond to CDFs of a and b ($a \leq b$), respectively. The histograms show that, regardless of the nonconformity function, the empircal converage is close to, and mostly above $1 - \epsilon = 90\%$. The CDFs show that the ICP sets are tight, *e.g.,* a ball with radius 10 pixels covers true keypoints for most images.

4 Experiments

We test our approach on the LINEMOD Occluded test set (LM-O) [3] with 1214 RGB images each containing 8 possible objects. There are in total 76 semantic keypoints labeled by the authors of [17].

Heatmap-based detector. We use a state-of-the-art heatmap-based detector in [17, 14]. Given a 2D image, we use the groundtruth bounding boxes to generate image patches of single objects, and run the heatmap detector to get 76 heatmaps (one per keypoint) each with a dimension 64×64.

Conformal calibration and prediction. We calibrate the heuristic detector on 200 images taken from LM-O according to the BOP'19/20 challenge [6]. For each keypoint, we compute nonconformity scores according to (Peak), (Var), or (Cov), and calculate $\alpha_{\pi(\lfloor (n+1)\epsilon \rfloor)}$ with $\epsilon = 0.1$. We then compute (ICP-Peak), (ICP-Var), or (ICP-Cov), on the entire LM-O test set with 1214 images.

We demonstrate the performance of the three ICP sets with $\epsilon = 0.1$. We provide additional results with $\epsilon = 0.2$ and $\epsilon = 0.05$ in the Supplementary Material.

Validity. The ICP sets, regardless of the underlying nonconformity functions, provide *statistically guaranteed* coverage as promised by Theorem 1. Fig. 3 plots the histogram of the empirical coverage of 76 keypoints, *i.e.,* we compute $\mathbb{P}[y_{l+1} \in F^\epsilon(x_{l+1})]$ in LM-O for each keypoint and histogram 76 numbers. We see the empirical coverage is close to, and mostly above, 90%.

Tightness. The ICP sets are tight. Fig. 3 plots the empirical cumulative distribution functions (CDF) of the sizes of the ICP sets: radii for balls and semi-axes for ellipses (each colored curve plots the CDF for one of the 76 keypoints). We observe that for most keypoints, an ICP set with radius (or semi-axis) 10 pixels covers over 80% of the test images.

Adaptiveness. Fig. 4 shows qualitative examples of the ICP sets. In the left example, the "duck" is observable and the heatmap is well predicted, leading to small ICP sets. In the right example, the "duck" is occluded and the heatmap is spread out, leading to larger ICP sets. The elliptical (ICP-Cov) sets better capture nonuniform uncertainty. More examples are provided in Supplementary Material.
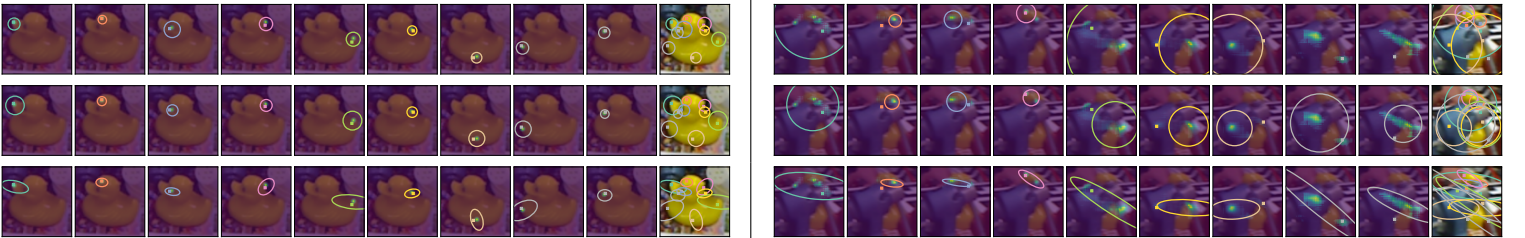

Figure 4: Qualitative examples of ICP sets on LM-O [3] with $\epsilon = 0.1$. Squares: true keypoints. Balls & ellipses: ICP sets. Last image of each row overlays true keypoints, ICP sets of *all* keypoints, and the original image. Other images overlay heatmap, true keypoint, ICP set of a *single* keypoint, and the original image. Top to bottom: (ICP-Peak), (ICP-Var), (ICP-Cov). Left: easy, right: challenging.

5 Conclusions

We applied inductive conformal prediction for statistically guaranteed keypoint detection. Given a heuristic heatmap, we design three nonconformity functions and calibrate the heatmap into a circular or elliptical prediction set that guarantees coverage of the true keypoint location with a user-specified probability. We demonstrate validity, tightness, and adaptiveness of the prediction sets on LM-O.

References

[1] Anastasios Angelopoulos, Stephen Bates, Jitendra Malik, and Michael I Jordan. Uncertainty sets for image classifiers using conformal prediction. In *Intl. Conf. on Learning Representations (ICLR)*, 2020. 3

[2] Anastasios N Angelopoulos and Stephen Bates. A gentle introduction to conformal prediction and distribution-free uncertainty quantification. *arXiv preprint arXiv:2107.07511*, 2021. 1, 2

[3] Eric Brachmann, Alexander Krull, Frank Michel, Stefan Gumhold, Jamie Shotton, and Carsten Rother. Learning 6d object pose estimation using 3d object coordinates. In *European Conf. on Computer Vision (ECCV)*, pages 536–551. Springer, 2014. 1, 4

[4] Bo Chen, Jiewei Cao, Alvaro Parra, and Tat-Jun Chin. Satellite pose estimation with deep landmark regression and nonlinear pose refinement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, pages 0–0, 2019. 1

[5] Bo Chen, Alvaro Parra, Jiewei Cao, Nan Li, and Tat-Jun Chin. End-to-end learnable geometric vision by backpropagating pnp optimization. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 8100–8109, 2020. 1

[6] Tomas Hodan, Frank Michel, Eric Brachmann, Wadim Kehl, Anders GlentBuch, Dirk Kraft, Bertram Drost, Joel Vidal, Stephan Ihrke, Xenophon Zabulis, et al. Bop: Benchmark for 6d object pose estimation. In *European Conf. on Computer Vision (ECCV)*, pages 19–34, 2018. 1, 4

[7] Georg Klein and David Murray. Parallel tracking and mapping for small ar workspaces. In *2007 6th IEEE and ACM international symposium on mixed and augmented reality*, pages 225–234. IEEE, 2007. 1

[8] Laurent Kneip, Hongdong Li, and Yongduek Seo. Upnp: An optimal o(n) solution to the absolute pose problem with universal applicability. In *European Conf. on Computer Vision (ECCV)*, pages 127–142. Springer, 2014. 1

[9] Jing Lei, Max G'Sell, Alessandro Rinaldo, Ryan J Tibshirani, and Larry Wasserman. Distribution-free predictive inference for regression. *Journal of the American Statistical Association*, 113(523):1094–1111, 2018. 2

[10] Yunzhi Lin, Jonathan Tremblay, Stephen Tyree, Patricio A Vela, and Stan Birchfield. Single-stage keypoint-based category-level object pose estimation from an RGB image. In *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, pages 1547–1553. IEEE, 2022. 3

[11] Rachel Luo, Shengjia Zhao, Jonathan Kuck, Boris Ivanovic, Silvio Savarese, Edward Schmerling, and Marco Pavone. Sample-efficient safety assurances using conformal prediction. *arXiv preprint arXiv:2109.14082*, 2021. 2

[12] Lucas Manuelli, Wei Gao, Peter Florence, and Russ Tedrake. kpam: Keypoint affordances for category-level robotic manipulation. In *The International Symposium of Robotics Research*, pages 132–157. Springer, 2019. 1

[13] Harris Papadopoulos. Inductive conformal prediction: Theory and application to neural networks. In *Tools in artificial intelligence*. Citeseer, 2008. 2

[14] Georgios Pavlakos, Xiaowei Zhou, Aaron Chan, Konstantinos G Derpanis, and Kostas Daniilidis. 6-DoF object pose from semantic keypoints. In *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, pages 2011–2018. IEEE, 2017. 1, 3, 4

[15] Sida Peng, Yuan Liu, Qixing Huang, Xiaowei Zhou, and Hujun Bao. Pvnet: Pixel-wise voting network for 6dof pose estimation. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 4561–4570, 2019. 1

[16] Yaniv Romano, Matteo Sesia, and Emmanuel Candes. Classification with valid and adaptive coverage. In *Advances in Neural Information Processing Systems (NIPS)*, volume 33, pages 3581–3591, 2020. 3

[17] Karl Schmeckpeper, Philip R Osteen, Yufu Wang, Georgios Pavlakos, Kenneth Chaney, Wyatt Jordan, Xiaowei Zhou, Konstantinos G Derpanis, and Kostas Daniilidis. Semantic keypoint-based pose estimation from single RGB frames. *J. of Field Robotics*, 2022. 1, 3, 4

[18] Gerald Schweighofer and Axel Pinz. Globally optimal o(n) solution to the pnp problem for general camera models. In *British Machine Vision Conf. (BMVC)*, pages 1–10, 2008. 1

[19] Jingnan Shi, Heng Yang, and Luca Carlone. Optimal pose and shape estimation for category-level 3d object perception. In *Robotics: Science and Systems (RSS)*, 2021. 1

[20] Jingnan Shi, Heng Yang, and Luca Carlone. Optimal and robust category-level perception: Object pose and shape estimation from 2d and 3d semantic keypoints. *arXiv preprint arXiv:2206.12498*, 2022. 1

[21] Jiaming Sun, Zihao Wang, Siyu Zhang, Xingyi He, Hongcheng Zhao, Guofeng Zhang, and Xiaowei Zhou. Onepose: One-shot object pose estimation without cad models. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 6825–6834, 2022. 1

[22] Bugra Tekin, Sudipta N Sinha, and Pascal Fua. Real-time seamless single shot 6d object pose prediction. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 292–301, 2018. 1

[23] Vladimir Vovk. Conditional validity of inductive conformal predictors. In *Asian conference on machine learning*, pages 475–490. PMLR, 2012. 2

[24] Vladimir Vovk, Alexander Gammerman, and Glenn Shafer. *Algorithmic learning in a random world*. Springer Science & Business Media, 2005. 1, 2

[25] Heng Yang and Luca Carlone. In perfect shape: Certifiably optimal 3d shape reconstruction from 2d landmarks. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 621–630, 2020. 1

[26] Heng Yang and Luca Carlone. Certifiably optimal outlier-robust geometric perception: Semidefinite relaxations and scalable global optimization. *IEEE Trans. Pattern Anal. Machine Intell.*, 2022. 1

[27] Sergey Zakharov, Ivan Shugurov, and Slobodan Ilic. Dpod: 6d pose object detector and refiner. In *Intl. Conf. on Computer Vision (ICCV)*, pages 1941–1950, 2019. 1