

Improving Humanoid Grasp Success Rate based on Uncertainty-aware Metrics and Sensitivity Optimization

Woo-Jeong Baek*, Christoph Pohl*, Philipp Pelcz, Torsten Kröger, and Tamim Asfour

Abstract— We present an approach for the selection of robot grasp candidates by treating specified metrics in a probabilistic manner and maximizing the success rate through statistical optimization. Recently, progress has been made in grasping unknown objects in cluttered scenes by using deep neural networks or incorporating classifiers. Although existing methods deliver promising results, they either lack *explainability* or fail to account for *uncertainties* that accumulate over the entire system. To address this shortcoming, we optimize a ranking score based on the *sensitivities* of the grasp success with respect to a set of metrics. These sensitivities reflect each metric’s contribution to the success. To perform this optimization, we refer to a dataset of 932 randomly selected grasps recorded under *real-world* conditions with the humanoid robot ARMAR-6. By validating our approach on a separate data collection of 187 physical *real-world* grasps, we demonstrate that our approach yields a success rate of 73.8%, amounting to an improvement of more than 40% compared to a random grasp selection. The results exemplify that sensitivity optimization, scarcely applied in the context of robotic applications so far, can significantly enhance the grasp success by considering respective metrics in the face of *uncertainties*.

I. INTRODUCTION

Autonomous robotic grasping in unknown and unstructured environments is a complex task that is required for many real-world applications. While grasping in structured environments, e. g., assembly lines, can be handled by robots with very high accuracy and speed, the same process becomes very fragile, slow, and unreliable in unknown environments due to perceptual and systematic uncertainties. However, in everyday scenarios like helping in a household or elderly care, or even in inhospitable environments like nuclear power plants or landfills, such structured environments are rarely available. Therefore, autonomous manipulation abilities in unknown surroundings without prior knowledge become more and more important.

The selection of a suitable grasp to execute on an object or in a scene plays an essential role in many approaches to autonomous grasping (see e. g., [1], [2], [3]). For example, in Figure 1, the humanoid robot ARMAR-6 needs to select a suitable grasp in order to lift an object. Humans, on the other hand, can easily build on their scene understanding and prior knowledge to always select a promising grasp. However, this is not a trivial task for robots, as the large

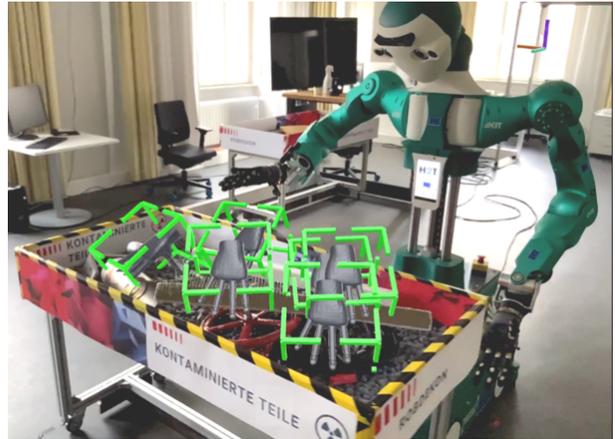


Fig. 1: Autonomously generated grasp candidates for execution on the humanoid robot ARMAR-6.

amount of raw visual data of the observed scene needs to be analyzed before any assumptions can be made. Additionally, this task becomes more complex due to uncertainties in the robot’s perception and execution. To address this issue, we propose a probabilistic approach for grasp selection built upon traditional statistical methods, that facilitates the derivation of a heuristic scoring function for grasp candidates extracted from visual perception based on previous grasp executions.

A. Previous Work

In our previous work [4], we investigated the influence of expert knowledge on the grasp selection process and showed that by selecting the correct autonomously-generated grasp candidate, the success rate of grasping with the humanoid robot ARMAR-6 [5] could be significantly increased. Therefore, we hypothesize that by improving only the quality of the grasp selection, we can significantly improve the autonomous grasping process. Additionally, in [6], we fuse multiple observations of manipulation actions in a scene using *Bayesian Recursive State Estimation*, to calculate the existence certainty, as well as the covariance of the pose, of a grasp action hypothesis. In real-world experiments on ARMAR-6, we showed that the probabilistic fusion of the grasp candidates improved the success rate of grasping compared to a non-probabilistic treatment. Therefore, our goal is to develop a probabilistic formulation of different metrics extracted from the visual perception and use statistical techniques to improve the grasp selection process. By doing so, we aim at enhancing the success rate of

* Both authors contributed equally to this paper.

The research leading to these results has received funding from the German Federal Ministry of Education and Research (BMBF) under the competence center ROBDEKON (13N14678).

The authors are with the Institute for Anthropomatics and Robotics (IAR), Karlsruhe Institute of Technology (KIT) {baek, pohl, torsten, asfour}@kit.edu.

autonomous grasping and gaining further insights into the influence of perceptual uncertainties.

B. Contributions

In this work, we propose an approach that employs traditional statistical tools to maximize the grasp success rate in a robot grasping application. To do so, we derive a scalar ranking score to rate grasp candidates. Specifically, we exploit the *sensitivities* of the grasp success with respect to selected grasp metrics, where sensitivities reflect to which extent each metric contributes to the success rate. In doing so, we distinguish ourselves from existing literature by presenting a method that is not only generalizable but also offers the possibility to analyze and interpret results in detail, for instance by studying the correlation and causality of different parameters. Specifically, we characterize each grasp by four Gaussian-distributed metrics, representing high-level scene understanding and its uncertainties. We collect data on the humanoid robot ARMAR-6 consisting of 932 grasps executed under real-world conditions in a random selection mode, providing the basis for the sensitivity optimization. For the optimized selection of grasp candidates, we employ the ranking score which is calculated on the basis of the randomly collected dataset. In particular, we assign weights to each metric according to their influence on the grasp success to rate each grasp candidate. We denote this approach with the term *Uncertainty-aware Sensitivity Optimization*. By validating our method in separate experiments, we show that explicitly considering uncertainties for the selection of grasp candidates according to our method yields a notable improvement of the grasp success rate. Additionally, we perform correlation studies and statistical tests to gain a deeper understanding of each metric’s contribution to the enhanced success rate. Overall, the scientific novelty of our contribution is twofold: We (i) present an explainable and generalizable uncertainty-aware grasp selection method incorporating Gaussian distributed uncertainties build upon sensitivity analyses. In addition, we (ii) complement obtained results by thoroughly studying the impact of the metrics’ uncertainties on the grasp success.

II. RELATED WORK

The ability of a robot to decide which action to execute in an unstructured environment is fundamental for autonomous manipulation. However, due to an incomplete knowledge and understanding of the scene, this is still a very challenging problem. A common approach for robotic grasping consists of generating several possible grasp candidates and subsequently selecting the best one for execution. State-of-the-Art methods in the field of discriminative grasping as [7], [8] employ neural networks to predict grasp poses based on point clouds or suggest the development of machine-learned classifiers as in [9], [10], [11] to distinguish highly promising grasp candidates. However, employing these black-box classifiers, the decisions are often non-comprehensible for human operators. Additionally, while some authors (e.g., [12] and [8]) propose to incorporate uncertainties by modeling metrics

as Gaussian distributions, their influence on the grasp success is neither quantified explicitly nor analyzed thoroughly. Motivated by this gap, we present an uncertainty-aware approach to grasp selection that makes transparent decisions based on previously executed grasps by exploiting the nature of statistical tools.

One possible approach for grasp selection is the success prediction of all candidates and the subsequent execution of the one with the highest probability of success. For example, in [9], a probabilistic framework for rating grasp hypotheses based on a semi-supervised *Kernel Logistic Regression* is developed. The authors use a smoothness assumption to train their classifier on partially unlabeled data and show that their approach can improve the success rate in grasping unknown objects. A *Gaussian Process*-based classifier is used in [10] to predict the success probability of grasps based on multiple metrics from literature. In [11], the performance of different learning methods is compared with respect to the prediction of a success probability for a manually annotated grasp dataset. Each classifier is trained with seven metrics as input. In contrast to these approaches that make use of data-driven methods to predict the success of one grasp candidate, the authors of [13] introduce four hand-crafted metrics combined with a probabilistic surface representation to predict the likelihood of the success. Since this work incorporates uncertainties, it shows similarities to our approach. However, as their scoring functions are hand-crafted, they automatically rely on the validity of their assumptions which prohibits learning from experience.

A second line of work suggests ranking grasp candidates with respect to certain criteria or databases. In [14], a segmented point cloud of an object of interest is used to calculate four metrics. Afterward, a pairwise ranking approach is applied to a set of grasp candidates, where a classifier is trained to predict whether one grasp candidate is preferred depending on object features and its corresponding metrics. In this way, a ranking of the grasp candidates is obtained from a set of training data and can be transferred to novel objects. In contrast, the authors of [15] build a library of successful grasps for shape patches and use the local features of patches of a queried shape to identify the best matching grasp in the library. They use the lowest distance of the queried shape template to the library shapes to rank grasp candidates. Even though this makes the decision process quite transparent, perceptual uncertainties are not accounted for in their approach.

A third line of work makes use of deep neural networks (DNNs) for robotic grasp selection applications. A physics simulation, in addition to crowdsourcing, is used in [7] to generate a large body of grasp data. The authors use this dataset as the basis for their analysis of stability metrics and the training of a DNN for grasp success prediction. In [8], a mixture density network is used to predict a prior distribution of feasible grasps specific to an object, which is used to sample grasp candidates that are rated using a voxel-based convolutional neural network for success prediction. For training, they use a large dataset generated in simulation

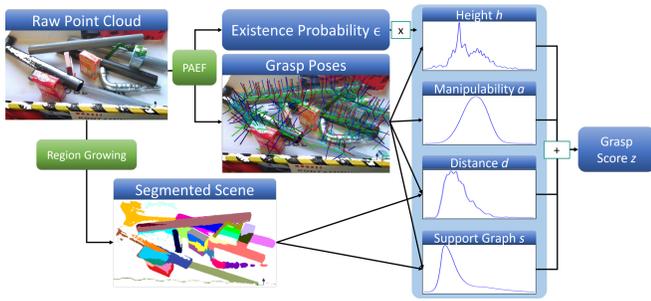


Fig. 2: Overview of our approach presented in Section III.

and evaluate their method on eight unseen objects on a real robot in very simple scenes.

One key characteristic of our contribution is the incorporation of a sensitivity optimization method to enhance the grasp success rate in a robotic application. While the authors of [16] and [17] investigate how the drawbacks of missing data can be alleviated for the assessment of clinical trials by incorporating sensitivity analyses, large-scale experiments as presented in [18] have provided experimental evidence that sensitivity optimization methods effectively contribute to discoveries of the lightest massive particles.

Inspired by these achievements and the flexible applicability of statistical techniques, we present a method to frame the maximization of the success rate in robotic grasping as a sensitivity optimization problem. In contrast to existing approaches, we place particular focus on both, the explicit consideration of uncertainties as well as the explainability of obtained results by carrying out in-depth analyses with respect to correlation and causality.

III. UNCERTAINTY-AWARE SENSITIVITY OPTIMIZATION

An overview of our approach for the autonomous selection of grasp candidates can be seen in Figure 2. Grasp candidates are generated based on the local surface geometry of a raw point cloud and probabilistically fused over multiple observations of the same scene to approximate the certainty in the grasp pose (see [6]). This pose and a segmented version of the point cloud are used to calculate four Gaussian-distributed grasp quality metrics. Based on a previously recorded dataset of random grasps, the sensitivities of the metrics towards the grasp success rate is estimated. Based on these sensitivities, the grasp score z is calculated and used for the selection of the most-likely grasp to succeed. In order to derive our methodology, we first provide an overview of the fundamentals, definitions, and assumptions we make throughout this work.

A. Preliminaries and Assumptions

As indicated above, we characterize each grasp g by n specified grasp metrics m_i , i. e., $g \mapsto m_1, \dots, m_n$. In particular, the key idea of our approach is to capture uncertainties of these metrics. However, as we are not provided with detailed knowledge about the specific behavior of these metrics, e. g., functional models, we argue that they can be represented by Gaussian distributions $m_i \propto N(\mu_i, \sigma_i)$ with mean value μ_i

and standard deviation σ_i . In addition, each grasp candidate g is classified as either succeeded g_s or failed g_f after execution. For optimization purposes, we make use of the *Kullback-Leibler divergence* D_{KL} – a measure that reflects to which extent a distribution P differs from a reference distribution Q ; i. e.,

$$D_{KL}(P||Q) := \int_{-\infty}^{\infty} p(x) \log \left(\frac{p(x)}{q(x)} \right) dx. \quad (1)$$

Here, $p(x)$ and $q(x)$ denote the probability density functions of P and Q , respectively.

Specifically, we aim to account for possible deviations due to perceptual inaccuracies. For instance, estimating the above metrics m_1, \dots, m_n depends on employed measurement devices such as sensors, cameras, and their relationships with each other. We argue that resulting inaccuracies are a consequence of measurement uncertainties. Generally, measurements are one method to estimate real-world parameters. To obtain reliable estimates, however, respective data must be evaluated thoroughly, taking into account limitations of sensory devices or the lack of available data which impede the achievement of a perfect accuracy of 100%. We thus rely on the *metrological* viewpoint of uncertainties presented in the *Guideline to the Expression of Uncertainties in Measurement (GUM)* [19]. Accordingly, the term *sensitivity* describes how the output of a system Y behaves with respect to changes in the input parameters x_i , that can be expressed by the local derivative $\frac{\partial Y}{\partial x_i}$. As the behavior of the metrics' uncertainties strongly depends on the considered application, we suggest to specify them based on the user's experience and system knowledge.

B. Problem Statement and Method

One central goal of our contribution is to analyze whether we can improve the rate of succeeded grasps defined by

$$r_s := \frac{g_s}{g_{tot}}, \quad (2)$$

where g_s is the number of succeeded grasps and g_{tot} represents the total number of executed grasps. Especially, our goal is to derive a scalar ranking score z which is aimed at maximizing r_s . Specifically, as each grasp is characterized by its metrics and respective uncertainties, the main challenge lies in the derivation of a functional model $y : \mathbb{R}^+ \rightarrow \mathbb{R}_0^+$ for the ranking score z given by

$$z := y(m_1, \dots, m_n) \\ = y(N(\mu_1, \sigma_1), \dots, N(\mu_n, \sigma_n), c), \quad (3)$$

where the constant c stands for application-specific biases or constant values which are included in the scoring function, but not subject to uncertainties. To develop a method that addresses the above problem, we assume that a dataset containing *randomly* collected grasps covers the respective range of all metrics m_i and reliably distinguishes between the group of successful grasps g_s and failed ones g_f . In other words, we treat the randomly gathered data as our

population, which serves as the basis for the sensitivity optimization. Given this information, and taking the uncertainties for all metrics σ_i into account, we obtain a probability density function (PDF) $p_i(x)$ for the population of each metric m_i . This function returns the probability p_i of the occurrence of any value x of metric m_i for each grasp candidate g , i. e., $p_i(m_i=x|o)$. Here, o denotes the binary outcome of a grasp $o \in \{s, f\}$, where s stands for the successful and f for the failed grasps. We thus obtain one PDF corresponding to the successful grasps $p_{i,s}(x)$ and one to the failed grasps $p_{i,f}(x)$ for each metric. Referring to these PDFs, we carry out a two-stage optimization by scoring each newly generated grasp candidate g as described in the following:

- *Global Weighting*: First, we rank all considered metrics according to their influence on the grasp success. Intuitively, the difference between the PDFs of the successful and failed grasps indicates how *sensitive* the grasp success behaves with respect to the considered metric. Therefore, we calculate the KL divergence defined in Equation 1 between the distributions $P_{i,s}$ and $P_{i,f}$ for each metric. We define the *global score* $f_{glob,i}$ for each metric m_i as follows

$$f_{glob,i} := D_{KL}(P_{i,s}||P_{i,f}), \quad (4)$$

where $P_{i,s}$ and $P_{i,f}$ are the distributions representing the PDFs $p_{i,s} = p_i(m_i|o=s)$ and $p_{i,f} = p_i(m_i|o=f)$, respectively.

- *Local Weighting*: In the second step, we calculate the likelihood of belonging to the set of successful grasps g_s for each candidate g , denoting this scalar quantity as the *local score* $f_{loc,i}$. To be specific, we calculate the fraction

$$f_{loc,i} := \frac{P_{i,s}}{P_{i,s} + P_{i,f}} \quad (5)$$

by again referring to the distributions $P_{i,f}$ and $P_{i,s}$.

Finally, we suggest a total score $f_{tot,i}$ for each metric i given by:

$$f_{tot,i} := f_{glob,i} \cdot f_{loc,i}. \quad (6)$$

By multiplying $f_{glob,i}$ with $f_{loc,i}$, we assign one weighting factor $f_{tot,i}$ to each grasp candidate. Particularly, $f_{tot,i}$ ensures that the influence of each metric on the grasp success (given by $f_{glob,i}$) and the likelihood for a successful outcome ($f_{loc,i}$) are both considered to rank each grasp candidate.

IV. SENSITIVITY OPTIMIZATION FOR GRASP SELECTION

In this section, we describe how the *Uncertainty-aware Sensitivity Optimization* can be used for grasp selection using several metrics derived from visual perception.

A. Metrics

For the characterization of a grasp candidate g , we chose four metrics, which are defined based on visual and proprioception sensor information and modeled as Gaussian distributions. These metrics are computed based on the *Probabilistic Action Extraction and Fusion* (PAEF) [6]. To

this end, the local surface geometry of the raw point cloud is analyzed and the principal curvatures and normals for each point extracted. To obtain averaged statistics of the surface information, the points are clustered in *Supervoxels*. Using the averaged information of a *Supervoxel*, affordances are extracted and a temporally-consistent coordinate frame is calculated, which can be used to track an action hypothesis over multiple observations of the scene. An *unscented Kalman filter* is employed to calculate the covariance of an action's execution pose. Additionally, a *Hidden Markov Model*, which tracks the number of scene observations in which a specific action hypothesis could be identified, is used to estimate the *Existence Certainty* ϵ of the action.

1) *Grasp Height (h)*: This metric reflects the height of a grasp candidate above the floor. It should favor objects, which lie on top of the clutter. The mean and variance (μ_h, σ_h^2) are computed from the covariance of the grasping pose, which is obtained through the computations of *PAEF*.

2) *Distance to Center (d)*: This metric describes the distance of the grasp position to the center of the object-oriented bounding box of the point cloud segment closest to the grasping pose. It is computed combining *PAEF* and a segmentation of the scene. The metric should favor grasps that are situated close to the center of mass of an object. The variance σ_d^2 is approximated to amount to 10% of the bounding box's length, while the mean value μ_d equals the distance to the center of the bounding box.

3) *Support Relations (s)*: The metric provides the number of objects that are supported by the point cloud segment closest to the grasping pose and is computed using *PAEF* and the segmented scene. This metric favors objects which are not covered by other objects. The mean and variance (μ_s, σ_s^2) are obtained from a probabilistic support graph described in [20]. Here, a large number of scene graphs are generated based on all RANSAC shape estimations of a segmented scene and a probabilistic representation is calculated based on the distribution of shapes and support relations between them.

4) *Manipulability (a)*: This metric represents the extended manipulability score proposed in [21], which is a quality measure for how freely an end-effector can move at a certain workspace position and is computed using only the grasping pose from *PAEF*. The mean and variance (μ_a, σ_a^2) are taken from a manipulability map. To this end, random joint configurations are sampled and the mean and variance of the corresponding manipulability in a voxelized workspace representation are updated. This metric favors grasps that are easily reachable by the robot.

Additionally, we take the grasp's *Existence Certainty* ϵ into account, which is a scalar value indicating the probability that a grasp affordance actually exists at a certain position. However, it is important to note that this value does not correspond to or is in any way related to a grasp quality measure. This value is not modeled as a Gaussian distribution, as it corresponds to a probability. Therefore, ϵ is not treated as a metric. The *Existence Certainty* is also computed by *PAEF*.



Fig. 3: Examples of the experimental setup with ARMAR-6.

B. Scoring Function

As our primary goal lies in validating the efficacy of the proposed *Uncertainty-aware Sensitivity Optimization*, we choose a straightforward scoring function. To account for the possibility of including further scalar parameters which might be used for scaling purposes of the ranking score by a factor α or stand for application-specific biases β , we propose a functional model y_{exp} given by

$$y_{exp}(\alpha, m_i, \beta) := \alpha \cdot \sum_{i=1}^n f_{tot,i} + \beta. \quad (7)$$

We emphasize that this scoring function is generalizable for other applications, where the purpose lies in the optimization of data with respect to one category. For example, it would be possible to exchange a grasp g for any other kind of sample with binary properties. In addition, we highlight that we applied deterministic methods to derive above scoring function and that the parameters for the Gaussian distributions of the metrics are explicitly defined such that our method is explainable, i. e., it is possible to perform more detailed studies on the obtained results. We assign the following application-specific parameters to the metrics m_1, \dots, m_4 as well as to the parameters α and β :

$$\begin{aligned} m_1 &\mapsto h; & m_2 &\mapsto d; & m_3 &\mapsto s; & m_4 &\mapsto a; \\ \alpha &\mapsto \epsilon; & \beta &= 0; \end{aligned} \quad (8)$$

According to these assignments, the resulting scoring function for rating grasp candidates becomes

$$\begin{aligned} z &:= y_{exp}(\epsilon, h, d, s, a) \\ &= \epsilon \cdot (f_{tot,h} + f_{tot,d} + f_{tot,s} + f_{tot,a}). \end{aligned} \quad (9)$$

As the *Existence Certainty* ϵ is related to the number of observations of a grasp candidate at a certain position and, therefore, the accuracy of the estimated parameters, the candidates that correspond to $\epsilon=0$ are neglected, while such with high ϵ are attributed to higher ranking scores.

V. EXPERIMENTS AND RESULTS

For the data collection, as well as the evaluation of our approach, multiple real-world grasping experiments have been conducted on the humanoid robot ARMAR-6 [5] resulting in more than 1100 grasp attempts on unknown objects. A video illustrating the approach and experiments can be found under <https://youtu.be/puJmGsK6hSE>.

A. Experimental Setup and Data Acquisition

To improve the grasp success rate using our approach described in Section III-B, a dataset consisting of *randomly* chosen grasp candidates and their outcomes is required. For this, we used a similar setup to the box emptying experiments in [6], which can be seen in Figure 3: 11 objects (five plastic pipes, four boxes, and two metal pipes) are randomly placed in a box and grasp candidates are extracted based on the PAEF approach. Additionally, the scene was segmented using a region-growing segmentation. Afterward, the inverse kinematics of each candidate, as well as the distance to the borders of the box, were checked to ensure reachability and prevent collisions with the box. Finally, a random grasp was chosen from among all reachable candidates and was executed by the robot. Together with the result of the grasp attempt, the metrics described in Section IV-A were calculated for the executed grasp in order to perform the global and local weighting steps described in Equation 4 and Equation 5. After grasping an object, the robot placed the object in the same box to introduce random changes in the scene. If the scene did not change for multiple attempts (either because no object was graspable or the same object was grasped repeatedly), the objects in the box were randomly rearranged by a human operator.

B. Evaluation and Results

To apply the *Uncertainty-aware Sensitivity Optimization*, 932 random grasps were executed during four days of experiments on ARMAR-6. The probability density functions of these grasps can be seen in Figure 4. Among them, 304 attempts were classified as successful while the remaining 628 were classified as failed. Therefore, the success rate of the randomly performed grasps amounts to 32.6%. The resulting values for the KL divergence are shown in Table I.

To rate grasp candidates according to the scoring function (Equation 9), the PDFs of the random dataset have been used to evaluate the grasp selection: The same experiments as described in Section V-A, including an identical grasp candidate generation process, were run again, with the difference that the scoring function z was used this time to select a grasp candidate. Specifically, the metrics were calculated for each candidate and the one with the highest score was chosen for execution. For the evaluation of our approach, 187 grasps were carried out with ARMAR-6 on the last day of experiments. Using our optimized grasp selection described in Section III and IV, the success rate could be increased to 73.8% with 138 successful and 49 failed grasp attempts.

TABLE I: Kullback-Leibler divergences of the metrics obtained from random grasping.

Metric	Kullback-Leibler Divergence
<i>Grasp Height</i> h	0.460
<i>Distance to Center</i> d	0.034
<i>Support Relations</i> s	0.014
<i>Manipulability</i> a	0.010

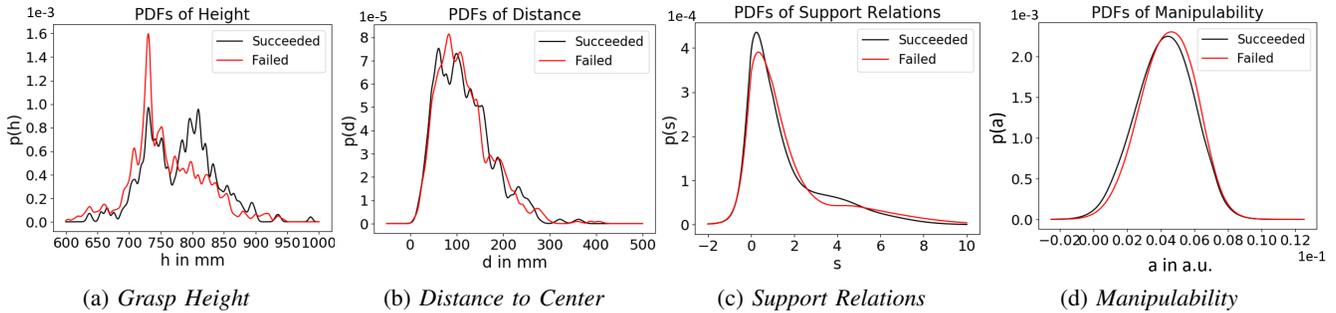


Fig. 4: Probability density functions for succeeded (black) and failed (red) grasp attempts of considered metrics for randomly selected grasps. These distributions provide the basis for the calculation of the ranking score according to Equation 9.

This large increase in performance indicates that our scoring function was able to rate the grasp candidates according to their success probability. However, as visible in Figure 4a, the PDFs of the *Grasp Height* h show the biggest difference between the black (succeeded grasps) and red (failed grasps) curve, which suggests that h had a dominant influence on the scoring function, and therefore on the grasp selection. The KL divergence of h in Table I, which is one order of magnitude greater than the KL divergences of the remaining metrics, further supports this observation. To investigate the interdependencies of the metrics, as well as demonstrate the explainability of our approach, additional analyses were performed.

VI. DATA ANALYSIS AND DATA INTERPRETATION

As shown in Table I, the KL divergence of the height h is one order of magnitude higher than the others and thus plays the most dominant role in the scoring function given by Equation 9. Accordingly, the highest grasp candidate is chosen for execution while the remaining metrics do not seem to contribute to the grasp selection due to their small KL divergence values. However, we hypothesize that small changes in the scoring function originating from the KL variations of the distance d , manipulability a , and support relations s become relevant for grasp candidates on similar heights. Apart from that, the score z obtained by applying Equation 9 depends on the specified uncertainties of each metric. In the following, we analyse the metrics' contributions and the influence of the uncertainties on the success rate.

A. Intra-Cluster Correlation Studies

To investigate the dependency of the score z on the height h , we refer to the correlation plot in Figure 5, which results from the data obtained in the second round of experiments (validation data set). It is apparent that z generally increases with h , which matches our expectations. On the other hand, it can be seen that clusters are formed. In particular, higher scores are attributed to grasp candidates with *lower* height values in these clusters, which indicates that the remaining metrics were decisive. In order to quantify the contribution of each metric to z , we calculate the so-called *intra-cluster correlations (IC)* in Table II, which describe the correlation with the score z within the circles depicted in Figure 5.

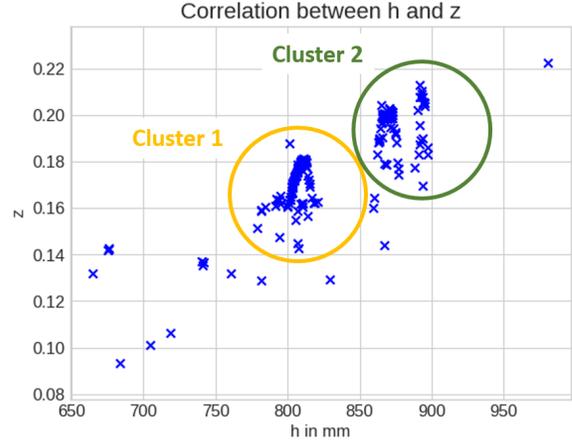


Fig. 5: Correlation between the final score z and the height h of the validation dataset. The circles in orange and green depict the clusters 1 and 2, respectively.

Obviously, the correlation values shown in the second column of this table consolidate the dominant character of the height h . Considering the entire dataset, its correlation with z is more than twice as much as the remaining three metrics. This, however, does not apply to the IC values: While the IC value for the height h decreases within the clusters, a higher absolute value of the IC values of the support relations s can be observed for cluster 2, indicating an increase in the influence of s in this cluster. In addition, the IC for the distance d changes its sign in cluster 1. These observations can be explained as follows: In the case of grasp candidates showing significant height differences, higher grasps are clearly preferred. The dominant character of the height h leads to the suppression of all remaining metrics on a large scale. On smaller scales of the height h ,

TABLE II: Correlation and intra-cluster correlation (IC) values between the final score and the metrics.

Metric	Corr. with z	IC Cluster 1	IC Cluster 2
<i>Grasp Height</i> h	0.8267	0.3592	0.2236
<i>Distance to Center</i> d	0.2998	-0.2967	0.1387
<i>Support Relations</i> s	-0.0982	-0.1424	-0.1902
<i>Manipulability</i> a	0.0345	-0.0749	-0.0166

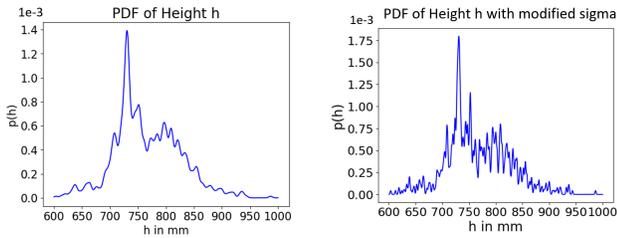


Fig. 6: PDFs of a population with the uncertainty from Section IV-A (left) and with a modified uncertainty (right).

however, namely inside clusters 1 and 2, the significance of the other three metrics becomes relevant. At first glance, the correlation values for the distance d and support relations s indicate that grasp candidates with larger distances to the object center are preferred. However, the IC values clearly demonstrate that the influence of these metrics on z increases drastically and that the IC values even invert their signs in some cases on small height scales. For example, the selection of candidates close to the object center is favored in cluster 1. Generally, the IC values lie in the same order of magnitudes for all considered metrics except for the manipulability a that, according to our results, does not significantly contribute to the final score. We clearly note that higher amounts of data are required to draw conclusions on the causality of our findings.

B. Influence of Uncertainties

The explainable nature of the *Uncertainty-aware Sensitivity Optimization* enables us to not only perform correlation studies but also to perform more thorough analyses of the obtained results. As described above, we suggested specifying the metrics and uncertainties individually for each application to facilitate the generalization to other scenarios. So far, although our results convincingly demonstrated that applying *Uncertainty-aware Sensitivity Optimization* yields a remarkable improvement of the grasp success rate, it is not yet obvious to which extent the specification of the uncertainties has contributed to the improved success rate. In particular, the PDFs in Figure 4 result from the total amount of random grasps, where each grasp is modeled as a Gaussian distribution with uncertainties σ specified in Section IV-A. Thus, different uncertainty specifications would lead to different PDFs. The right part in Figure 6 illustrates the PDF for a reduced uncertainty by a factor of 0.3. In the following, we perform studies on how the metrics' uncertainties influence the grasp success rate. We focus on the two metrics, which indicate the highest contributions to the score z : the height h and the distance to the object d .

1) *Modification of Uncertainties*: The optimization relies on the PDFs in Figure 4. Since these distributions belong to the uncertainty specifications in Section IV-A, their shapes will vary when modifying the uncertainties σ as illustrated in Figure 6. For analysis purposes, we consider three uncertainty settings for the height h and the distance d .

2) *Bootstrapping*: We bootstrap the distribution belonging to the group of succeeded grasps (black curve in Figure 4). To do so, we generate 10 000 simulated datasets with each 932

random samples from the modified distributions. According to the Central Limit Theorem, we obtain a Gaussian distribution by plotting the mean values of these 10 000 simulated distributions, which provide the information of the expected values for the succeeded and failed grasp attempts for each uncertainty constellation. We thereby consider a confidence level of 95 %.

3) *Hypothesis Testing*: We generate samples from the modified PDFs representing the distributions of the three uncertainty settings shown in Table III. We conduct a p-test (see e. g., [22]) to decide whether a sample belongs to the group of succeeded or failed grasp attempts. To do so, we refer to the distributions in the *Bootstrapping step*. We define our null hypothesis H_0 as follows: *The selected candidate does not belong to the group of succeeded grasp attempts*. As we deal with binary classifications of each grasp, the rejection of H_0 would automatically hint at a failed grasp attempt which equals the acceptance of the alternate hypothesis H_1 stating *the selected candidate does belong to the group of succeeded grasp attempts*. The hypothesis testing is conducted on a significance level of $\rho_H = 0.05$. We divide the number of simulated grasp samples, which reject the null hypothesis and the total number of 932 samples to obtain the success rate. Table III displays our results.

C. Discussion

Generally, the intra-cluster correlation analyses support our claim that the influence of the remaining metrics increases for objects with negligible height differences. In particular, observing the behavior of the IC values shows that a mere glimpse of the correlation on the entire dataset can be misleading. We conclude from the findings of our IC studies, that the effect of the distance to the object d and the support relations s should indeed be considered for scenes with a high number of objects. However, the validation dataset contains only 187 grasps, which is a rather small sample size from a statistical point of view. Specifically, these 187 selections represent only a small subset of the available choices, as an average of 100-200 candidates were suggested by PAEF per scene. Nevertheless, the results from the evaluation presented in Section V-B also demonstrate the ability of the *Uncertainty-aware Sensitivity Optimization*, as even a relatively small dataset resulted in such a large increase in performance over the randomly selected grasp candidates. The results of our statistical analyses presented in Table III imply that reducing the metrics' uncertainties does not necessarily yield higher success rates. In fact, Table III presents the expected success rates for three settings, where the uncertainties of the grasp metrics are reduced by factors 0.5, 0.25 and 0.1 compared to the PDFs in Figure 4.

TABLE III: Expected success rates for modified σ .

Metric	σ	0.50σ	0.25σ	0.10σ
Grasp Height h	73.80 %	64.05 %	57.63 %	60.19 %
Distance to Center d	73.80 %	74.01 %	76.04 %	70.96 %

Specifically for the height h , it can be seen that lower uncertainty values yield decreased success rates: According to the results of the hypothesis tests, reducing the uncertainty by a factor of 0.25 would lead to a grasp success rate of 57.63%, which corresponds to a decrease by more than 16%. Arguing that the accurate estimation of uncertainties is essential for the increase of the grasp success, we conclude that the specified uncertainties σ_h and σ_d in the experiments were appropriately chosen. In the case of the distance to the object d , however, slight improvements are found to be attainable by reducing the respective uncertainty. As the estimated success rate in the third constellation (0.1σ) in Table III becomes smaller, we infer that reducing the uncertainty by a factor of 0.25 would be desired to achieve an improvement of approximately 3% of the success rate.

VII. CONCLUSION AND OUTLOOK

In this work, we presented an explainable and generalizable approach for *Uncertainty-aware Sensitivity Optimization* and applied it to autonomous grasp selection. To do so, we recorded 932 randomly selected grasps under real-world conditions with the humanoid robot ARMAR-6, which were generated using our previous work on probabilistic action extraction and execution described in [6]. Referring to this dataset, we introduced a scoring function that takes into account four specified metrics to rate grasp candidates by applying a global and local weighting. In a second round of experiments, we used this scoring function to select the most promising grasp candidate for a given constellation of objects in the scene. Doing so, we were able to achieve a grasp success rate of 73.8% (in comparison to 32.6% using randomly selected grasps). This significant increase in performance demonstrates that the choice of metrics was indeed suitable for predicting successful grasp candidates. In addition to the real-world grasping experiments, we performed in-depth analyses on the influence of the grasp metrics, as well as their uncertainties, on the grasp success. It was found that the height of grasp candidates contributes to the improvement of the success rate the most. On the other hand, our intra-cluster correlation studies indicated that the remaining metrics become decisive for objects on similar heights. In addition, we analyzed the impact of the metrics' uncertainty specifications on the grasp success rate. Interestingly, our results signify that decreased uncertainties would not generally yield higher success rates. However, from a statistical point of view, we note that obtained results were based on a rather small amount of data of 932 grasps and thus must be augmented by further experiments and respective analyses in the future. To this end, using simulated grasp experiments to generate a large number of grasps could be a promising approach. Therefore, we plan to study the versatility of our approach by applying it to different scenarios, as it is not limited to an application in grasp selection. Moreover, we aim to investigate how the incorporation of additional metrics as well as different scoring functions would influence the grasp selection.

REFERENCES

- [1] A. Sahbani, S. El-Khoury, P. Bidaud, P. B. A. Overview, A. Sahbani, S. El-Khoury, and P. Bidaud, "3D Object Grasp Synthesis Algorithms," *Robotics and Autonomous Systems*, vol. 60, no. 3, pp. 326–336, 2012.
- [2] J. Bohg, A. Morales, T. Asfour, and D. Kragic, "Data-Driven Grasp Synthesis—A Survey," *IEEE Transactions on Robotics*, vol. 30, no. 2, pp. 289–309, 2014.
- [3] K. Kleeburger, R. Bormann, W. Kraus, and M. F. Huber, "A Survey on Learning-Based Robotic Grasping," *Current Robotics Reports*, vol. 1, no. 4, pp. 239–249, 2020.
- [4] C. Pohl, K. Hitzler, R. Grimm, A. Zea, U. D. Hanebeck, and T. Asfour, "Affordance-Based Grasping and Manipulation in Real World Applications," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2020, pp. 9569–9576.
- [5] T. Asfour, M. Wächter, L. Kaul, S. Rader, P. Weiner, S. Ottenhaus *et al.*, "ARMAR-6: A High-Performance Humanoid for Human-Robot Collaboration in Real World Scenarios," *IEEE Robotics & Automation Magazine*, vol. 26, no. 4, pp. 108–121, 2019.
- [6] C. Pohl and T. Asfour, "Probabilistic Spatio-Temporal Fusion of Affordances for Grasping and Manipulation," *IEEE Robotics & Automation Letters*, vol. 7, no. 2, pp. 3226–3233, 2022.
- [7] D. Kappler, J. Bohg, and S. Schaal, "Leveraging big data for grasp planning," in *IEEE International Conference on Robotics and Automation*, 2015, pp. 4304–4311.
- [8] Q. Lu, M. Van der Merwe, B. Sundaralingam, and T. Hermans, "Multifingered Grasp Planning via Inference in Deep Neural Networks: Outperforming Sampling by Learning Differentiable Models," *IEEE Robotics & Automation Magazine*, vol. 27, no. 2, pp. 55–65, 2020.
- [9] A. N. Erkan, O. Kroemer, R. Detry, Y. Altun, J. Piater, and J. Peters, "Learning probabilistic discriminative models of grasp affordances under limited supervision," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2010, pp. 1586–1591.
- [10] A. K. Goins, R. Carpenter, W.-K. Wong, and R. Balasubramanian, "Evaluating the efficacy of grasp metrics for utilization in a Gaussian Process-based grasp predictor," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2014, pp. 3353–3360.
- [11] C. Rubert, D. Kappler, A. Morales, S. Schaal, and J. Bohg, "On the relevance of grasp metrics for predicting grasp success," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2017, pp. 265–272.
- [12] P. Ni, W. Zhang, X. Zhu, and Q. Cao, "Pointnet++ grasping: learning an end-to-end spatial grasp generation algorithm from sparse point clouds," in *IEEE International Conference on Robotics and Automation*, 2020, pp. 3619–3625.
- [13] D. Chen, V. Dietrich, Z. Liu, and G. von Wichert, "A Probabilistic Framework for Uncertainty-Aware High-Accuracy Precision Grasping of Unknown Objects," *Journal of Intelligent & Robotic Systems*, vol. 90, no. 1–2, pp. 19–43, 2018.
- [14] D. Kent and R. Toris, "Adaptive Autonomous Grasp Selection via Pairwise Ranking," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2018, pp. 2971–2976.
- [15] A. Herzog, P. Pastor, M. Kalakrishnan, L. Righetti, J. Bohg, T. Asfour, and S. Schaal, "Learning of grasp selection based on shape-templates," *Autonomous Robots*, vol. 36, no. 1–2, pp. 51–65, 2014.
- [16] S. Cro, T. P. Morris, M. G. Kenward, and J. R. Carpenter, "Sensitivity analysis for clinical trials with missing continuous outcome data using controlled multiple imputation: a practical guide," *Statistics in medicine*, vol. 39, no. 21, pp. 2815–2842, 2020.
- [17] A.-D. Hazewinkel, J. Bowden, K. H. Wade, T. Palmer, N. J. Wiles, and K. Tilling, "Sensitivity to missing not at random dropout in clinical trials: use and interpretation of the trimmed means estimator," *Statistics in Medicine*, vol. 41, no. 8, pp. 1462–1481, 2022.
- [18] The KATRIN Collaboration, "Direct neutrino-mass measurement with sub-electronvolt sensitivity," *Nature Physics*, vol. 18, no. 2, pp. 160–166, 2022.
- [19] J. C. for Guides in Metrology (JCGM), "Evaluation of measurement data — guide to the expression of uncertainty in measurement," JCGM, Tech. Rep., 2009.
- [20] F. Paus and T. Asfour, "Probabilistic Representation of Objects and Their Support Relations," in *International Symposium on Experimental Robotics (ISER)*, 2020, pp. 510–519.
- [21] N. Vahrenkamp, T. Asfour, G. Metta, G. Sandini, and R. Dillmann, "Manipulability Analysis," in *IEEE-RAS International Conference on Humanoid Robots*, 2012, pp. 568–573.
- [22] G. Cowan, *Statistical data analysis*. Oxford university press, 1998.