Risk-Aware Grasping in Cluttered Environments via Modeling Aleatoric Uncertainty

Myungjoon Son[1], Seungjun Kim[2], Yoonseon Oh[1]
[1]Dept. of Electronic Eng., Hanyang University, Seoul, Republic of Korea
[2]Dept. of Artificial Intelligence, Hanyang University, Seoul, Republic of Korea
{choming03, rlatmdwnseo, yoh21}@hanyang.ac.kr

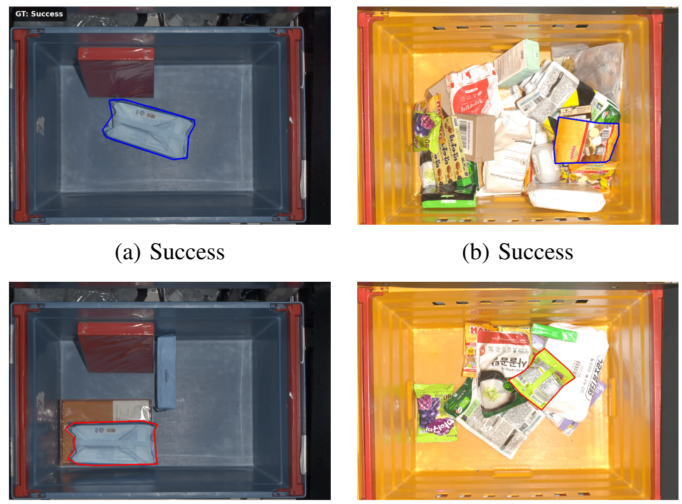
***Abstract*—Robotic grasping in unstructured environments suffers from intrinsic ambiguity (aleatoric uncertainty) due to sensor noise and occlusions, which conventional deterministic methods often overlook. In this paper, we propose a Risk-Aware Grasping strategy utilizing the Picking Success Prediction with Uncertainty (PSP-U) model. Instead of relying on computationally intensive 3D backbones, we employ a lightweight statistical feature abstraction and a dual-head network to simultaneously estimate grasp success probability and data-dependent uncertainty. Experimental results on a large-scale real-world dataset demonstrate that our approach achieves superior performance compared to heavy 3D point-based models by effectively filtering out high-risk candidates. Significantly, filtering out the top 50% of uncertain grasps improves the success rate from 84.1% to over 94%, validating the proposed method as a robust solution for safety-critical industrial applications.**

***Index Terms*—Robotic Grasping, Aleatoric Uncertainty, Risk-Awareness, Deep Learning in Robotics**

I. Introduction

Robotic grasping in cluttered environments suffers from sensor noise and occlusions. In these scenarios, identical grasping attempts on visually similar objects can yield divergent outcomes (see Fig. 1)—a phenomenon we refer to as *intrinsic ambiguity*, formally characterized as *aleatoric uncertainty* [1]. While conventional deterministic approaches often overlook this stochasticity—leading to overconfident failures—robust operation requires explicitly quantifying predictive confidence. To address this, the robotic system must autonomously identify and refrain from executing high-risk attempts based on uncertainty estimation.

In this work, we propose the PSP-U model to explicitly quantify this uncertainty. Our main contributions are summarized as follows: (1) **Uncertainty-Aware Modeling:** We propose the PSP-U model utilizing a dual-head architecture and Laplacian likelihood to explicitly quantify aleatoric uncertainty, enabling the detection of intrinsic ambiguities in cluttered environments. (2) **Efficient Feature Abstraction:** We demonstrate that lightweight statistical feature abstraction effectively filters sensor noise and outperforms computationally heavy 3D point-based networks (e.g., PointNet++) in real-world industrial settings. (3) **Risk-Aware Grasping Strategy:** We validate a robust filtering policy based on predicted uncertainty, which significantly enhances the grasping success rate (from 84.1% to 94%) by avoiding high-risk candidates. (4) **Large-Scale Real-World Validation:** We validate our approach on a large-scale dataset comprising 11,210 physical grasp attempts, demonstrating superior reliability compared to deterministic baselines.



(a) Success (b) Success

(c) Failure (d) Failure

Fig. 1: **Illustrative examples of aleatoric uncertainty in bin-picking.** (a)-(b) Successful grasps in ideal scenarios where objects are spatially isolated. (c)-(d) Failed grasps in deceptively clear scenarios subject to latent physical constraints. Note the contrast between (a) and (c): despite being visually similar and occlusion-free, the outcome diverges due to intrinsic ambiguity. This necessitates uncertainty modeling to predict risks that are not explicitly observable.

II. Related Work

A. Data-Driven Grasping

Deep learning approaches using 3D point clouds, such as PointNet++ [2], have demonstrated high performance in grasping tasks. However, their complex 3D backbones often incur high computational costs. Recently, Li et al. [3] demonstrated that lightweight models using hand-crafted features are highly effective for large-scale industrial applications. Following this insight, we adopt a feature-based MLP architecture to ensure real-time capability while focusing on uncertainty quantification.

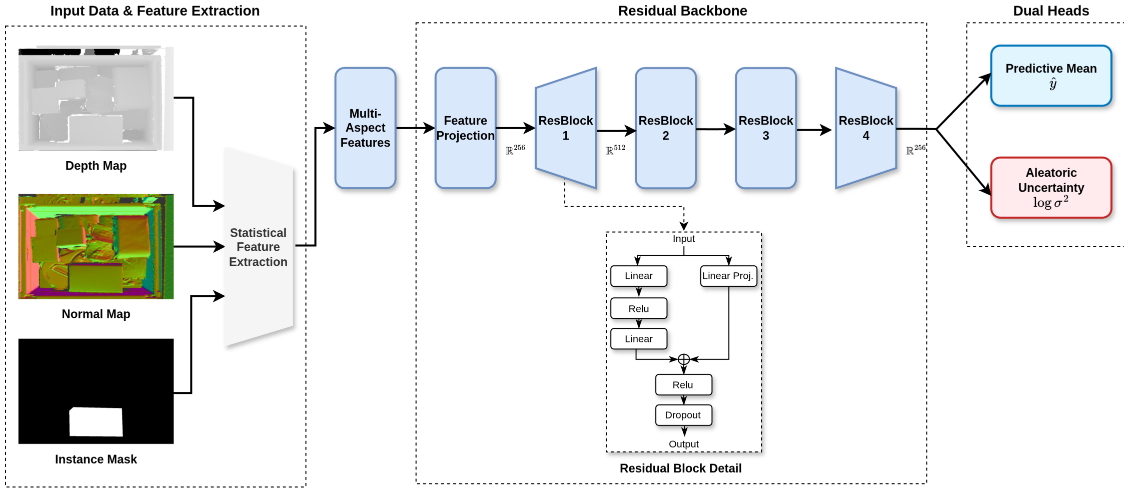


Fig. 2: **Architecture of the proposed PSP-U.** Taking a 13-dimensional statistical feature vector as input, the network predicts grasp success probability ($\hat{y}$) and aleatoric uncertainty ($\log \sigma^2$) via a dual-head mechanism. The architecture employs an *Expansion-Compression* strategy: it first expands the manifold ($\mathbb{R}^{256} \to \mathbb{R}^{512}$) to disentangle non-linear interactions, followed by a bottleneck compression ($\mathbb{R}^{512} \to \mathbb{R}^{256}$) to filter task-irrelevant noise. The residual blocks incorporate linear projection shortcuts to preserve feature identity across varying dimensions.

B. Uncertainty in Robotics

Modeling uncertainty is essential for safety-critical applications. Kendall and Gal [1] highlighted the importance of modeling aleatoric uncertainty to capture inherent data noise. While significant benchmarks like SuctionNet-1Billion [4] have advanced grasp quality scoring, they often focus on deterministic predictions. Our work distinguishes itself by leveraging *aleatoric uncertainty* as a dynamic risk indicator to proactively filter out unreliable grasp candidates.

III. Proposed Method

A. Statistical Feature Abstraction

Instead of processing raw 3D point clouds directly, we extract a compact set of 13 *multi-aspect features* from depth maps, surface normals, and instance masks, as detailed in Table I. While deep learning typically relies on raw inputs, industrial depth sensors exhibit significant high-frequency noise. We empirically found that utilizing statistical feature abstraction acts as a robust low-pass filter, preventing the model from overfitting to sensor artifacts compared to point-based networks.

TABLE I: List of 13 Statistical Input Features

Category	**Feature Names**
Geometric (4)	Normal Std (σ_n), Normal Std (X, Y) (σ_{nx}, σ_{ny}), Depth Std (σ_d)
Positional (4)	Center (c_x, c_y), Center Depth Mean/Std (μ_{cd}, σ_{cd})
Morphological (5)	Mask Ratio (r_m), BBox Cov. (c_b), Occlusion (o), Aspect Ratio (α), Rotated Box Ratio (r_r)

B. Aleatoric Uncertainty Modeling via Laplace Prior (PSP-U)

We propose the PSP-U model based on a Residual MLP architecture. To quantify uncertainty, the network employs a Dual-Head Output: one for the predictive mean ($\hat{y}$) and another for the log-variance ($s = \log \sigma^2$) representing aleatoric uncertainty. The architecture is illustrated in Fig. 2.

To ensure robustness against outliers common in industrial data, we assume a Laplacian distribution for the aleatoric noise (rather than Gaussian) and train the model using a Laplacian Heteroscedastic Loss based on the L1 norm:

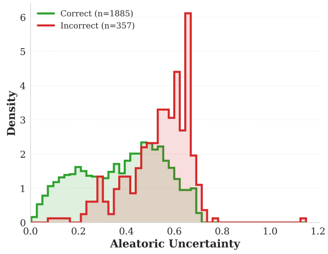
$$\mathcal{L} = \frac{1}{2} \exp(-s)|y - \hat{y}| + \frac{1}{2} s \quad (1)$$

where $y \in \{0, 1\}$ is the ground truth. We adopt the L1 norm (Laplacian prior) over the standard L2 norm (Gaussian prior) because bin-picking data contains frequent outliers due to sensor noise and occlusions. The heavy-tailed Laplacian distribution is more robust to these outliers, preventing the gradient from exploding. Furthermore, the learned variance s acts as a heteroscedastic weight; it down-scales the loss contribution of uncertain samples (where $|y - \hat{y}|$ is large), effectively allowing the model to suppress the influence of confusing data during training.

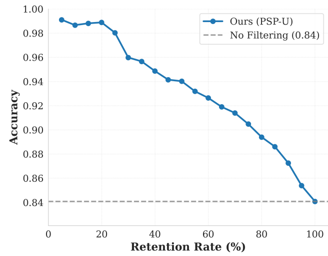
IV. Experiments and Results

A. Experimental Setup

We evaluated our PSP-U model using a large-scale dataset of 11,210 grasp attempts collected from a real-world logistics center. For fair comparison, we employed a deterministic MLP and PointNet++ [2] as baselines under the same experimental settings. *Implementation Details:* The dataset was split into 80% training and 20% validation sets. The backbone comprises four residual blocks with channel sizes [256, 512, 512,



(a) Uncertainty Density



(b) Retention Curve

Fig. 3: **Quantitative Analysis of Uncertainty Utility.** (a) Normalized distributions reveal a clear separation: **correct predictions** (green) cluster at low uncertainty, while **incorrect predictions** (red) shift significantly rightward. *(Histograms are normalized to account for class imbalance.)* (b) The retention curve demonstrates that progressively filtering out data with the highest uncertainty (keeping the top $X\%$ samples) monotonically improves grasping accuracy from 84% to over 94%.

512, 256] and a dropout rate of 0.1. We trained the model on a single NVIDIA RTX 3060 GPU for 50 epochs using the AdamW optimizer ($lr = 10^{-4}, wd = 10^{-4}, b = 64$) with a Cosine Annealing scheduler.

B. Quantitative Analysis & Efficiency

Table II compares prediction accuracy and input representation complexity. Our PSP-U achieves 84.1% accuracy, significantly outperforming PointNet++ (75.0%). This confirms that compact statistical features are more robust to high-frequency sensor noise than raw point clouds in cluttered environments. Furthermore, compared to the deterministic MLP (82.8%), explicitly modeling uncertainty acts as a regularizer, improving generalization.

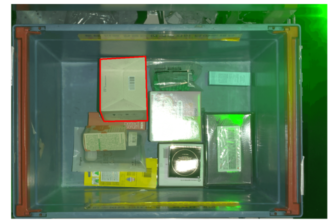
TABLE II: Comparison of Accuracy and Input Representation

Model	**Uncertainty**	**Input Representation**	**Accuracy**
PointNet++ [2]	$\times$	Raw Point Cloud ($N \times 3$)	75.0%
Det. MLP	$\times$	13 Statistical Features	82.8%
PSP-U (Ours)	$\circ$	**13 Statistical Features**	**84.1%**

C. Risk-Aware Filtering Effect

To validate the utility of uncertainty, we sorted test samples by their predicted uncertainty (σ) and analyzed the retention curve. As shown in Fig. 3, filtering leads to a monotonic increase in performance. By excluding the top 50% of uncertain predictions, the success rate improves from 84.1% to over 94%, confirming that uncertainty serves as a reliable metric for identifying and avoiding high-risk grasps. Crucially, high uncertainty serves as a trigger for a fallback strategy (e.g., selecting the next-best candidate) rather than causing a system deadlock, thereby ensuring continuous operation.

(a) $\sigma = 0.68$



(b) $\sigma = 0.69$

(c) $\sigma = 0.71$

(d) $\sigma = 0.69$

Fig. 4: **Qualitative Analysis.** High aleatoric uncertainty ($\sigma > 0.6$) effectively flags False Positives (deterministic errors), acting as a safety net against latent physical constraints.

D. Qualitative Analysis

Since False Positives (incorrect success predictions) cause physical failures and potential hardware damage, we focus on validating if our module flags these errors. Fig. 4 visualizes the predicted uncertainty for false-positive cases. Even when the model incorrectly predicts success, the uncertainty remains high ($\sigma > 0.6$). This demonstrates the model's capability to detect physical risks such as deformability, concave geometry, and occlusion, which are difficult to capture with deterministic predictions alone.

V. Conclusion

We presented the PSP-U model to address aleatoric uncertainty in robotic bin-picking. By explicitly modeling data-dependent uncertainty with a lightweight dual-head network, we enabled a Risk-Aware Grasping strategy. Our results show that using uncertainty as a risk filter allows the system to achieve near-perfect success rates by avoiding ambiguous grasp attempts.

Acknowledgment

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. RS-2024-00409492, Beyond-G Global Innovation Center) and by CMES Co., Ltd.

References

[1] A. Kendall and Y. Gal, "What uncertainties do we need in bayesian deep learning for computer vision?" in *NeurIPS*, 2017.
[2] C. R. Qi et al., "PointNet++: Deep hierarchical feature learning on point sets in a metric space," in *NeurIPS*, 2017.
[3] S. Li et al., "Demonstrating Large-Scale Package Manipulation via Learned Metrics of Pick Success," in *RSS*, 2023.
[4] H. Cao et al., "SuctionNet-1Billion: A large-scale benchmark for suction grasping," *IEEE RA-L*, vol. 6, no. 4, 2021.