Force-Aware 3D Contact Modeling for Stable Grasp Generation

Zhuo Chen[1], **Zhongqun Zhang**[1,2*], **Yihua Cheng**[1], **Aleš Leonardis**[1], **Hyung Jin Chang**[1]

[1]University of Birmingham,
[2]College of Software, Nankai University
zxc417@student.bham.ac.uk, zhangzhongqun@nankai.edu.cn, {y.cheng.2,a.leonardis,h.j.chang}@bham.ac.uk,

Abstract

Contact-based grasp generation plays a crucial role in various applications. Recent methods typically focus on the geometric structure of objects, producing grasps with diverse hand poses and plausible contact points. However, these approaches often overlook the physical attributes of the grasp, specifically the contact force, leading to reduced stability of the grasp. In this paper, we focus on stable grasp generation using explicit contact force predictions. First, we define a force-aware contact representation by transforming the normal force value into discrete levels and encoding it using a one-hot vector. Next, we introduce force-aware stability constraints. We define the stability problem as an acceleration minimization task and explicitly relate stability with contact geometry by formulating the underlying physical constraints. Finally, we present a pose optimizer that systematically integrates our contact representation and stability constraints to enable stable grasp generation. We show that these constraints can help identify key contact points for stability which provide effective initialization and guidance for optimization towards a stable grasp. Experiments are carried out on two public benchmarks, showing that our method brings about 20% improvement in stability metrics and adapts well to novel objects.

Code — https://github.com/chzh9311/force-aware-grasp.git

Introduction

Hand grasp synthesis aims at generating plausible grasping pose of the human hand with an object template (Jiang et al. 2021; Karunratanakul et al. 2021; Liu et al. 2023; Shimada et al. 2024; Lee et al. 2024; Zuo et al. 2024; Xu et al. 2024; Zhang et al. 2025). It has recently drawn increasing attention due to its wide applications in AR/VR (Grady et al. 2024), robotics (Li et al. 2024; Zhong et al. 2025), and human-computer interaction (Guo, Lu, and Yao 2021). Previous works have already extensively studied the geometric properties of Hand-Object Interaction (HOI), where contact modeling plays an important role (Jiang et al. 2021; Zhou et al. 2022; Grady et al. 2021; Zhang et al. 2024). Since hand-object contact is very sensitive to noise in hand pose, existing studies prefer to use contacts as geometric constraints to generate hand pose and perform well in plausibility, controllability (Zhang et al. 2024) and diversity (Liu et al. 2023).

However, the stability of grasp synthesis has been much less explored. Although some recent studies try to model physical properties in HOI, they are not yet properly related to contact geometry. On the one hand, force-agnostic stability modeling is hard to integrate with contacts. Some existing methods (Liu et al. 2022; Luo, Liu, and Yi 2024) address stability by setting equilibrium constraints with both forces and contact points as unknowns, which requires strong assumptions like zero-friction to obtain solvable constraints, leading to suboptimal results. Others directly use a simulator (Zhao et al. 2022; Yuan et al. 2023) in inference, but iterative simulations are too complicated to directly incorporate with contact-based pose refinement. On the other hand, existing studies the relation between forces and visual clues or object mass instead of contact geometry, and are only limited to specific object shapes, *e.g.*, boards (Grady et al. 2022, 2024; Zhao et al. 2025) or cubes (Murtaza et al. 2023). The relationship between grasping geometry and contact forces on general objects remains unexplored.

Meanwhile, we find that if the contact forces can be predicted, then modeling stability in contact-based grasp synthesis will be much easier. This is since stability can be measured as the object acceleration and is related to contact forces and points via physical laws. With the forces known, it is possible to identify the most important contact keypoints for static equilibrium without ignoring gravity and frictions. These keypoints provide valuable guides in grasp synthesis. They are simple geometric targets that a hand model (Romero, Tzionas, and Black 2022) can easily fit, but they also contain important physics information so that successfully contacting these points largely ensures stability, as shown in Fig. 1.

Thus, we propose to explicitly integrate forces into 3D contact modeling. Specifically, we present a novel force-aware contact representation, based on which we construct a pipeline with a contact generator and a pose optimizer following previous contact-based pose synthesis methods (Grady et al. 2021; Liu et al. 2023; Zuo et al. 2024). In the framework, we first extend the contact representation to include contact likelihood, hand part label, and contact normal force value for each sampled point on the object surface. The generator is then trained to generate all three maps. In the optimization stage, we model stability as the

*Zhongqun Zhang is the corresponding author.

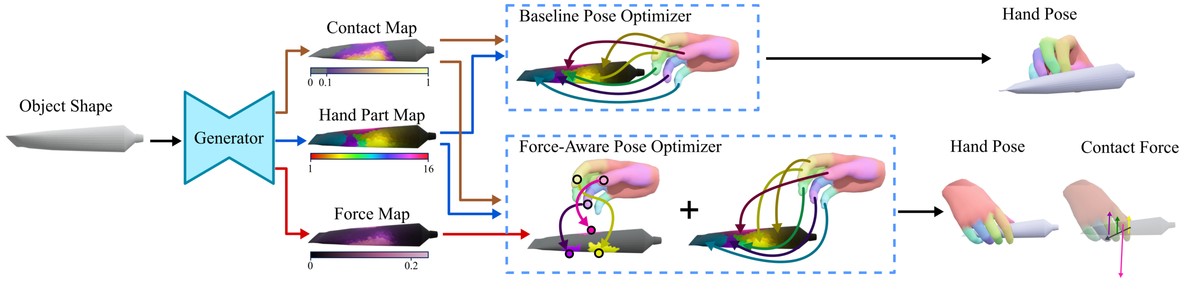


Figure 1: The difference between previous methods (without force predictions) and ours (with force predictions). Force information guides the optimization by identifying a few keypoints to ensure stability.

minimization of accelerations based on which a few key contact points are recognized. Then a strong initialization is performed by fitting the hand model (Romero, Tzionas, and Black 2022) to the keypoints. The pose optimizer then tries to fit the hand model to the keypoints and contact maps, leading to a stable grasp.

To train our generator, we need ground-truth force labels. However, existing HOI datasets with force labels (Grady et al. 2022, 2024; Zhao et al. 2025; Murtaza et al. 2023) are all limited to specific object shapes. Thus, we develop an automatic labeling pipeline using a simulator (Todorov, Erez, and Tassa 2012) to obtain the force label on general objects from HOI datasets (Taheri et al. 2020). The procedure searches for the optimal simulation parameters to restore the stability of ground truth data and output faithful force labels.

Finally, we test the generated grasp quality on two public benchmarks: GRAB (Taheri et al. 2020) and HO3Dv2 (Hampali et al. 2020). The results show that our method brings over 30% improvement in stability criteria on both in-domain and out-of-domain objects without affecting plausibility and diversity. It shows that our method effectively addresses stable grasp synthesis and adapts well to unseen objects.

Our contributions can be summarized as follows:

1. We formally model the relation between stability and contact points using force. We show that by optimizing accelerations, stability can be formulated as a solvable optimization of contact points and predicted contact forces;
2. We propose a novel pipeline for stable grasp synthesis. The pipeline is built upon force-aware contact representation, with a generator and optimizer. The former relation is utilized to identify keypoints to guide grasp generation;
3. We conduct experiments on two public benchmarks, showing that our method achieves state-of-the-art stability and excellent generality to novel objects.

Related Work

Contact-Based Grasp Synthesis. As physical plausibility can be affected by slight changes in hand pose (Grady et al. 2021), modeling hand-object contact has been a key to synthesizing plausible and stable grasps. Some methods propose to generate hand poses directly and refine the poses using predicted contact maps (Jiang et al. 2021; Taheri et al. 2020; Karunratanakul et al. 2021; Taheri et al. 2022). Usually, the generator is a cVAE (Sohn, Lee, and Yan 2015) and the refiner tries to fit MANO (Romero, Tzionas, and Black 2022) model to the contacts by training a separate model (Taheri et al. 2020) or optimizing iteratively (Jiang et al. 2021; Taheri et al. 2022). Other methods generally follow a two-stage pipeline to generate contact maps then optimize the hand pose to fit the contact (Grady et al. 2021). ContactGen (Liu et al. 2023) extends the contact representation to include hand part labels and orientations of the hand parts for better diversity, and GraspDiff (Zuo et al. 2024) integrates the optimization process into Diffusion (Ho, Jain, and Abbeel 2020) inference. Great improvements in plausibility and diversity have been observed, but stability is largely ignored because of the indirect relations between physical properties and geometry. In contrast, our research provides a novel perspective to bridge the gap via contact forces, greatly improving the stability while keeping the advantages of contact-based synthesis.

Modeling Grasp Stability. Grasping stability has long been an important criterion of grasp generation (Jiang et al. 2021). It offers guidance to increase grasp synthesis quality (Liu et al. 2022; Yuan et al. 2023). For stability-aware pose generation, physics engines (Coumans and Bai 2016–2021; Todorov, Erez, and Tassa 2012; Makoviychuk et al. 2021) are commonly used in inference by projecting the hand or human pose to physically plausible space via a mimic agent learnt by RL (Yuan et al. 2023). Though being more stable, the relations are not modeled explicitly, and the process is hard to integrate with contact modeling. So some other studies try to model stability according to equilibrium-related theories, like friction cone constraints (Luo, Liu, and Yi 2024) or simplified force closure (Nguyen 1988) conditions (Liu et al. 2022). One major challenge is the high dimensionality with both forces and contact points as unknowns, thus strong assumptions like zero-friction are applied, resulting in suboptimal results. Therefore, we propose to predict forces as part of the known variables, which directly solves this issue, allowing more precise equilibrium relations to be modeled in the synthesis process.

Method

In this section, we first describe our strategy for acquiring contact force labels for model training. Then we define our force-aware contact representation and the stability constraints, which form the core of our approach. Finally, we describe our model, which incorporates the representation and con-

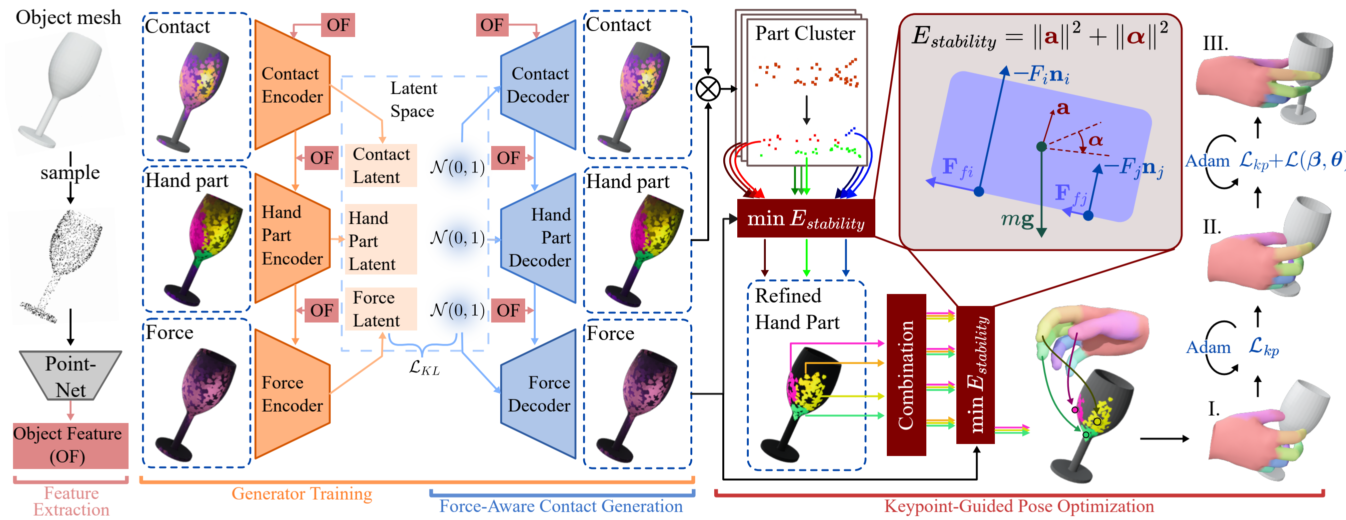


Figure 2: Our framework consists of a generator and an optimizer. The object features (OF) are obtained by processing sampled object points using PointNet++. Then the generator consisting of 3 successive cVAEs takes OF as an input condition. It generates force-aware contacts including the contact normal force value, which are then used to recognize important keypoints for equilibrium. In optimization stage, I. and II. are two-stage initialization based only on keypoints, and III. optimizes MANO parameters $\boldsymbol{\beta}, \boldsymbol{\theta}$ while also being guided by the keypoints to get the stable grasp pose.

straint to enable stable grasp synthesis. The framework is shown in Fig. 2.

Automatic Contact Force Labeling

Existing datasets typically rely on physical sensors to acquire force labels, which limits the data to plane-related objects, such as board-based (Grady et al. 2022, 2024; Zhao et al. 2025) or cube-based (Murtaza et al. 2023) interactions. In this paper, we obtain contact force labels using a physics simulator (Todorov, Erez, and Tassa 2012), based on hand and object meshes that can be derived from existing hand-object interaction datasets such as GRAB (Taheri et al. 2020).

To allow enough tolerance to the unavoidable errors in GT data and approximately modeling soft tissues, we propose to do *dynamic parameter searching*. First, we set the object free and take the contact forces when the object achieves the least acceleration within a certain displacement threshold. Then, we search for the hyperparameter setting resulting in the least displacement as the optimal parameters for the current sample, increasing the stable rate (displacement $< 5\,\mathrm{cm}$) to 65.4%. Please refer to supplementary for details in labeling procedure and data statistics.

The final label is in point-contact form. One point from the j^{th} contact area is determined by the simulator as the contact point $\mathbf{c}_j \in \mathbb{R}^3$. The label for this point also includes normal force value N_j (supportive force), lateral force (friction), and which hand part provides the force (hand part label).

Force-Aware Contact Representation

We use three hierarchical feature maps for force-aware contact representation. For each sampled point from the object surface, three scalars are assigned to respectively indicate the contact likelihood $C \in [0, 1]$, the hand part label in contact $P \in [1, 16]$, and the normal contact force value $F \in \mathbb{R}^+$.

With the hand and object meshes, the contact likelihood of each point is defined using the capsule distance d in (Grady et al. 2021) between the object point and the nearest vertex of the hand mesh: $C = \min\{1/d, 1\}$. The part label is defined as the hand part of the nearest vertex, where the hand is partitioned according to blend-skinning weights of MANO. Hands in Fig. 2 show the hand parts with different colors.

For contact force, we propose to only use the normal force values in our representation. Including frictions brings two extra dimensions for each contact point and makes it harder to predict. Moreover, the friction can be inferred from the equilibrium constraints using the normal force and is therefore unnecessary to predict. Finally, the force prediction is a scalar $F_i \geq 0$ for the i^{th} sampled point. And because the simulator labels forces on discrete points not continuous areas, we need to spread the force uniformly on the affinity object point set A_j, which includes all valid contact points whose nearest label point is $\mathbf{c}_j$, *i.e.*, $\forall i \in A_j, F_i = N_j/|A_j|$.

We then propose to represent the force values using one-hot vector. This is since classification-based methods empirically outperforms regression-based ones (Li et al. 2021). To build the discrete force distribution, we define a binning scheme over F using s bins. The first bin captures the probability of zero force ($F = 0$), while the remaining $s - 1$ bins cover the positive force range. To define the bin boundaries over the non-zero range, we assume the logarithm of the force magnitude, $\log F$, follows a normal distribution with mean $\mu_{\log F}$ and standard deviation $\sigma_{\log F}$. We uniformly divide the interval $[\mu_{\log F} - 3\sigma_{\log F}, \mu_{\log F} + 3\sigma_{\log F}]$ into $s - 2$

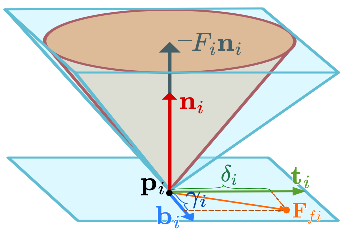


Figure 3: The friction cone (red) and its approximation (blue) in the stability energy function

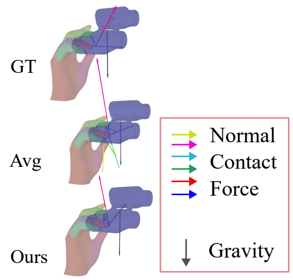


Figure 4: Normal contact forces by our method, avg. predictor, and GT

intervals. Set $l_{F1} = 0, l_{F(s+1)} = \infty$, and

$$\forall\, 2 \le i \le s,\ l_{Fi} = \exp\left\{\mu_{\log F} + \left(\frac{6(i-2)}{s-2} - 3\right)\sigma_{\log F}\right\}.$$

This construction ensures that non-zero force range is covered by $s-2$ bins uniformly distributed in log-space, accommodating the heavy-tailed nature of the force distribution.

The vectorized force is defined by Eq.(1), and the transformation from the one-hot vector back to a scalar force value is performed using the soft-argmax operation with a temperature parameter $t = 0.02$, as shown in Eq.(2).

$$\mathbf{v}_{Fi} = \begin{cases} 1, & \text{if } l_{Fi} \le F_i < l_{Fi+1}; \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

$$\hat{F} = \frac{\sum_{i=1}^{s} \frac{l_{Fi}+l_{Fi+1}}{2} \exp\{\mathbf{v}_{Fi}/t\}}{\sum_{i=1}^{s} \exp\{\mathbf{v}_{Fi}/t\}}. \tag{2}$$

Modeling Stability with Forces

The definition of stability can vary with applications, yet the ultimate goal is always to keep the object still. Formally, it means both the translational acceleration $\mathbf{a} \in \mathbb{R}^3$ and rotational acceleration $\boldsymbol{\alpha} \in \mathbb{R}^3$ are 0. Then our basic idea is to measure stability using the norm of accelerations.

With the predicted contact forces, it is possible to calculate the exact accelerations via physical laws. Denote the center of mass of the object as $\mathbf{p}_{CoM}$. Then formally, according to *Newton's Second Law* and *the Theory of Angular Momentum*:

$$\begin{aligned} \textstyle\sum_{i=1}^{n}(-F_i\mathbf{n}_i + \mathbf{F}_{fi}) + m\mathbf{g} &= m\mathbf{a}, & (3) \\ \textstyle\sum_{i=1}^{n}(\mathbf{p}_i - \mathbf{p}_{CoM}) \times (-F_i\mathbf{n}_i + \mathbf{F}_{fi}) &= I\boldsymbol{\alpha}, & (4) \end{aligned}$$

where $\mathbf{g}$ is the gravitational acceleration, $\mathbf{F}_{fi} = \mu F_i(\gamma_i\mathbf{b}_i + \delta_i\mathbf{t}_i)$ is the friction, $-1 \le \gamma_i, \delta_i \le 1$. $(\mathbf{b}_i, \mathbf{t}_i, \mathbf{n}_i)$ forms a unit orthogonal basis, and $\mathbf{n}_i$ is the surface normal. μ is the friction coefficient empirically set to 1. It is a common linear approximation for the friction cone (Hu et al. 2022) as shown in Fig. 3. The object mass m is set to 1 kg for normalization, and the moment of inertia I is approximated by regarding the object as a ball: $I = 0.4m \max_i \|\mathbf{p}_i - \mathbf{p}_{CoM}\|^2$.

Eq. (3) and (4) can be further formulated in a bilinear form as

$$N\mathbf{F} + \mu B(\boldsymbol{\gamma} \circ \mathbf{F}) + \mu T(\boldsymbol{\delta} \circ \mathbf{F}) + [\mathbf{g} \quad 0]^T = [\mathbf{a} \quad \boldsymbol{\alpha}]^T, \tag{5}$$

where $N = [N_m/m, N_t/I]^T$ and $N_m = [\mathbf{n}_1, \mathbf{n}_2, \ldots, \mathbf{n}_n]^T$, $N_t = [(\mathbf{p}_1 - \mathbf{p}_{CoM}) \times \mathbf{n}_1, (\mathbf{p}_2 - \mathbf{p}_{CoM}) \times \mathbf{n}_2, \ldots, (\mathbf{p}_n - \mathbf{p}_{CoM}) \times \mathbf{n}_n]^T$. B and T follow this definition. "$\circ$" is the Hadamard product. n is the point number and $\mathbf{F} = [F_1, F_2, \ldots, F_n]^T$ represents the force map in representation.

In Eq. (5), only if $\mathbf{F}$ is known can the acceleration vector $[\mathbf{a}, \boldsymbol{\alpha}]^T$ be linearly related to $\boldsymbol{\gamma}$ and $\boldsymbol{\delta}$. Then we define the stability energy as the norm of accelerations, which is formulated as a constrained quadratic optimization problem and can be solved by quadratic programming:

$$\begin{aligned} E_{stability}(\{\mathbf{F}_i\}_{i=1}^n, \{\mathbf{p}_i\}_{i=1}^n) &= \min_{\boldsymbol{\gamma},\boldsymbol{\delta}} \left\{\|\mathbf{a}\|^2 + \|\boldsymbol{\alpha}\|^2\right\}, \\ \text{s.t. } \forall 1 \le i \le n, -1 &\le \gamma_i, \delta_i \le 1. \end{aligned} \tag{6}$$

Since quadratic programming is non-differentiable, to integrate stability loss into training, we check the bounds of each element in $\mathbf{a}$ and $\boldsymbol{\alpha}$, and punish the cases when 0 is not inside. Use $\text{abs}(\cdot)$ to represent the element-wise absolute operator, and $B_{i\cdot}$ is the i^{th} row of B, then $-\text{abs}(B_{i\cdot})\mathbf{F} \le B_{i\cdot}(\boldsymbol{\gamma} \circ \mathbf{F}) \le \text{abs}(B_{i\cdot})\mathbf{F}$. The same goes for T. Thus, we can derive a differentiable loss term by comparing the upper or lower bounds of each element in Eq. (5) to 0 if 0 is outside the range. Using the notation $F_{fric} = \mu\, \text{abs}(B) + \mu\, \text{abs}(T)$, the loss term can be written as

$$\begin{aligned} \mathcal{L}_{stability} = \mathbf{1}^T \max&\left\{(N - F_{fric})\,\mathbf{F} + [\mathbf{g} \quad 0]^T, 0\right\} \\ - \mathbf{1}^T \min&\left\{(N + F_{fric})\,\mathbf{F} + [\mathbf{g} \quad 0]^T, 0\right\} \end{aligned} \tag{7}$$

Force-Aware Stable Grasp Synthesis

Generator Structure and Training Losses. The generator consists of 3 successive cVAEs, similar to (Liu et al. 2023). The positions and normals of all sampled object points are processed by PointNet++ (Qi et al. 2017), and the features are fed to all encoders and decoders as a part of the condition. The latent output of the former forms the other part.

The model is trained using reconstruction loss, KL-divergence loss, and stability loss. The reconstruction loss $\mathcal{L}_{rec} = \mathcal{L}_c + \mathcal{L}_{part} + \mathcal{L}_{force}$. The hand part map loss $\mathcal{L}_{part} = CE(\mathbf{P}, \hat{\mathbf{P}})$ is cross entropy loss, while contact loss $\mathcal{L}_c$ and force loss $\mathcal{L}_{force}$ are both L1 loss. KL-divergence $\mathcal{L}_{KL}$ on latent codes is also enforced following the training process of cVAE (Sohn, Lee, and Yan 2015). Note that in Eq. (7) we substitute $\mathbf{F}$ with $\hat{\mathbf{F}} \circ \hat{\mathbf{C}}$ to also regularize contact. Considering not all samples are stable, and we prefer stable samples to unstable ones, the total loss is weighted according to the displacement d in simulative labeling by $w = \min\{d_{th}/d, 1\}$ where $d_{th} = 5\,\text{cm}$. With preset weights, the final loss is

$$\mathcal{L}_{cVAE} = w\left(w_{rec}\mathcal{L}_{rec} + w_{KL}\mathcal{L}_{KL} + w_{stability}\mathcal{L}_{stability}\right). \tag{8}$$

Keypoint-Guided Stable Pose Optimization. The contact forces can provide guidance on the importance of contact points, but simply using the force values as weights results in suboptimal results (See Tab. 3 for details), since the optimization based on MANO (Romero, Tzionas, and Black 2022) is still non-convex. However, a better pose initialization is still possible, so the optimizer is less likely to get stuck in local minima. Previous studies apply random restarts (Grady et al. 2021) or iterative optimization (Liu et al. 2023), which are either non-robust or too complex to avoid local minima.

According to the theory of force closure (Nguyen 1988), as few as 3 contact points are enough for a stable grasp in 3D. The energy function in Eq. (6) is well suited for selecting the optimal contact points. Therefore, we propose to fit the MANO model to only a few keypoints that are important for object equilibrium as a strong initialization and stability-aware guidance through optimization.

To prevent the keypoints from falling inside the object (e.g., when the object is slim), the contact points of each part are first clustered and only one cluster is selected to produce the keypoint. We aggregate the supportive forces and contact points of cluster c of hand part h via force-weighted average over all the points as $\mathbf{F}_c^h$ and $\mathbf{p}_c^h$. Suppose H is the set of hand parts that are in touch with the object, then the optimal cluster label is the one with the least stability energy, *i.e.*,

$$c_0 = \underset{c}{\operatorname{argmin}} E_{stability}\left(\{\mathbf{F}_c^h, \mathbf{p}_c^h\} \cup \{\mathbf{F}_i, \mathbf{p}_i\}_{i\in H, i\neq h}\right). \tag{9}$$

Then we assign $\mathbf{F}_h = \mathbf{F}_{c_0}^h, \mathbf{p}_h = \mathbf{p}_{c_0}^h$ as the normal force and contact center of part p and iterate Eq. (9) over all contact parts.

Next, we want to select n_{kp} keypoints from the contact centers. Similarly, the optimal keypoint combination is the one with the least $E_{stability}$. In experiments, we set $n_{kp} = 3$ because it is the least possible number of contact points that forms a force closure, and a smaller number of target points means more traceable optimization. Thus, the optimal combination part labels is

$$H_0 = \underset{\substack{H_c\subset H\\ |H_c|=n_{kp}}}{\operatorname{argmin}} E_{stability}\left(\{\mathbf{F}_h\}_{h\in H_c}, \{\mathbf{p}_h\}_{h\in H_c}\right). \tag{10}$$

Then the target point of the hand part is obtained by moving the contact points along the surface normal for $r = 5\,\mathrm{mm}$, *i.e.*, $\mathbf{q}_i = \mathbf{p}_i + r\mathbf{n}_i$, roughly the finger radius.

Two-stage Initialization. Although fitting the hand joints to only 3 keypoints seems simple, the problem is still hard to solve analytically due to the non-linearity of the MANO model. Iterative optimization is an intuitive solution, but it is still possible to get stuck in local minima, and the initialization problem remains unsolved. Instead, in the first stage, we fit the average pose to the keypoints. Thus, the hand part centers are relatively fixed, and the problem then becomes point cloud registration. Suppose $\{\mathbf{p}_h^{hand}\}_{h\in H}$ are the hand part centers defined as the average position of connected joints, then the problem is registering $\{\mathbf{p}_h^{hand}\}_{h\in H_0}$ to $\{\mathbf{p}_h\}_{h\in H_0}$, which can be solved analytically via corresponding point set registration (Besl and McKay 1992), where $\mathcal{L}_{kp} = \sum_{h\in H_0}\|\mathbf{p}_h - \mathbf{p}_h^{hand}\|^2$ is minimized.

In the second stage, we optimize the hand pose parameters $\boldsymbol{\theta}$ together with the global pose to further minimize $\mathcal{L}_{kp}$ via iterative optimization. With the former global registration, such optimization is more likely to find the optimal solution and provides a stable initialization.

Keypoint-Guided Optimization. Based on the initialization, the next step would be to search for the optimal grasping pose around, where more contacts and less penetration should also be considered alongside stability. Inspired by (Liu et al. 2023), we use partitioned SDF to calculate the hand-object distance, and calculate the current contact map $\hat{\mathbf{C}}$ following (Grady et al. 2021). The penetration loss $\mathcal{L}_{pene}$ follows the definition of (Liu et al. 2023). The keypoint loss is kept in this stage as $\mathcal{L}_{kp}$. Together, regularization losses $\mathcal{L}_{reg} = \|\boldsymbol{\beta}\|^2 + \|\boldsymbol{\theta}\|^2$ are added to reduce implausible poses. Then the target of the optimization is as follows and is solved by an optimizer (we use Adam (Kingma and Ba 2015) in our experiment).

$$\min_{\boldsymbol{\beta},\boldsymbol{\theta}} \mathcal{L}_{optim} = w_{kp}\mathcal{L}_{kp} + w_c\mathcal{L}_c + w_{pene}\mathcal{L}_{pene} + w_{reg}\mathcal{L}_{reg}. \tag{11}$$

Experiment

Datasets and Metrics

Datasets and Baselines. GRAB (Taheri et al. 2020) is used as our training and test set due to its precise hand shapes resulting from motion captures. The 10 objects from HO3Dv2 (Hampali et al. 2020) are used to test the out-of-domain adaptivity of the method. During test, the input object is randomly rotated while the gravity direction keeps unchanged. Please refer to the supplementary for training details.

For fair comparison, we compare our method to those trained only with GRAB on GRAB benchmark. On HO3D benchmark, however, all methods are *not* trained on HO3D dataset, including methods from GRAB benchmark and GF (Karunratanakul et al. 2020) and GraspTTA (Jiang et al. 2021) which are trained on ObMan (Hasson et al. 2019) dataset.

Stability Assessment. Following previous research (Karunratanakul et al. 2020; Jiang et al. 2021; Liu et al. 2023; Zuo et al. 2024), we use PyBullet (Coumans and Bai 2016–2021) to simulate the object and the grasping hand and measure the displacement of the object center of mass. *Simulation Disp.* is the average displacement while *Stable Rate* is the proportion of samples with Simulation Disp. less than $2\,\mathrm{cm}$. To more comprehensively measure stability considering both displacements and penetrations, we adopt distinction rate from dexterous robotic grasping research (Wang et al. 2023; Lu et al. 2024). It is the proportion of samples where the grasp can withstand the gravity of the object in IsaacGym (Makoviychuk et al. 2021) simulation and the maximum penetration depth is less than $5\,\mathrm{mm}$. Since human hand labels usually represent deformations with slight penetrations (Grady et al. 2021), this criterion is rather strict for human hands, but it is reasonable by showing the proportion of top-quality grasps. Therefore, we rename it to *distinction rate*.

Plausibility and Diversity Assessment. We use contact ratio and penetration volume to assess the geometric plausibility. *Contact ratio* is the proportional of samples where the hand and object are in contact (nearest distance $< 5\,\mathrm{mm}$). *Penetration Vol.* measures the intersecting volume of the hand and the object. Diversity is assessed by the *Entropy* and *Cluster Size* of the clustering result of all generated hand poses following (Liu et al. 2023; Wu et al. 2025).

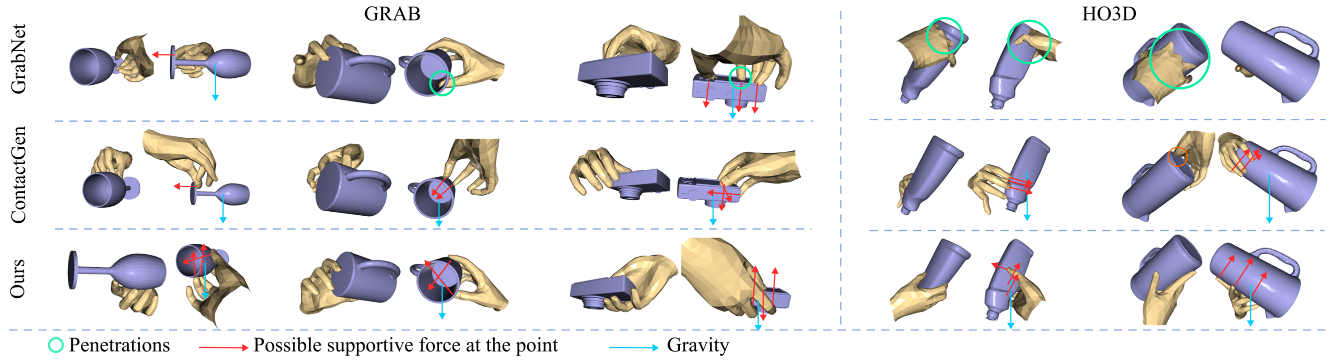


Figure 5: Grasp Samples from both datasets. Each sample is shown in two views with highlights of penetrations, possible supportive forces at the contact points, and gravity labels.

Methods	Simulation Disp. (cm)↓	Stable (<2 cm) Rate (%) ↑	Distinction Rate (%) ↑
Results on GRAB dataset (Taheri et al. 2020)			
GrabNet (2020)	0.90	91.6	9.2
ContactGen (2023)	2.82	72.5	12.5
FastGrasp (2025)	2.04	74.2	14.2
Ours	**0.68** ↓ 24.4%	**96.7** ↑ 5.5%	**17.5** ↑ 23.2%
Results on out-of-domain HO3D dataset (Hampali et al. 2020)			
GF (2020)	4.47	45.0	1.0
GrabNet (2020)	1.86	74.0	3.0
GraspTTA (2021)	1.91	74.0	7.5
ContactGen (2023)	4.62	51.0	9.0
FastGrasp (2025)	3.59	51.0	1.5
Ours	**1.50** ↓ 19.4%	**80.5** ↑ 8.5%	**17.0** ↑ 88.9%

Table 1: Quantitative results of stability performance on GRAB and HO3Dv2. '↑' after a criterion means the higher the better, while '↓' means the opposite. Our method outperforms previous ones by a large margin.

Grasp Synthesis Analysis

Quantitative Analysis on Stability. The stability performance on both datasets are shown in Tab. 1. On both benchmarks, our method achieves the best performance on all criteria, with the displacements decreased amazingly by 24.4% and 19.4%. The stable rate and distinction rate also increase obviously. Since the distinction rate considers both penetrations and stability, it is safe to conclude that our method finds a good balance between the two.

On the HO3D benchmark, since no training samples are seen, previous methods all suffer from an obvious performance drop on all metrics. In contrast, the stable rate and distinction rate of our method do not drop obviously. Since the keypoints are selected with physics-based analytical analysis instead of data-driven methods, our method shows strong adaptivity to general out-of-domain objects.

Analysis on Geometric Plausibility and Diversity. We report the relative criteria in Tab. 2. Our method shows the best contact ratio and comparable diversity. Although the penetration volumes are increased compared to previous methods, the state-of-the-art distinction rate in Tab. 1 indicates that the increase is acceptable. The good diversity partly results from the change of object pose relative to the gravity direction. With Eq. (7), our model favors more supportive forces in the anti-gravity direction, thus leading to a gravity-direction-related generation. By changing the gravity, our method can generate diverse and stable grasps.

Qualitative Analysis. To study the reasons for improvements, we visualize some generated samples from our method and two previous ones in Fig. 5.

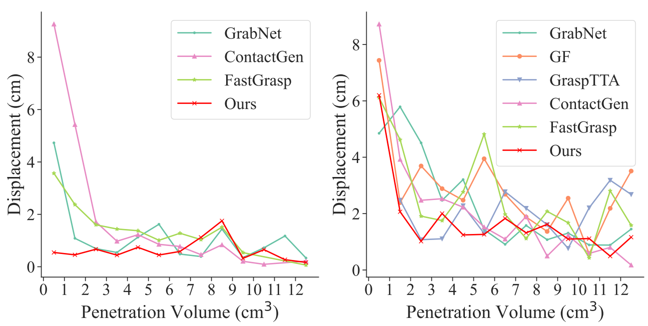
First, there is a trend that our method favors hands supporting the object from the bottom, which is preferable as they provide supportive forces against gravity. Fingers tend to touch the object from different directions due to the force-aware choice of keypoints. In contrast, GrabNet tends to produce penetrations (also shown in Tab. 1), and ContactGen prefers less contact, usually unstable grasps. Further, we can observe a common failure pattern from previous methods is that the hand ends up in touch with only one side of the object, like the 'camera' of GrabNet and 'mug' of ContactGen in Fig. 5. These cases mostly initialize the hand on one. Later optimization or refinement tries to lead the fingers through the object but is prevented by penetration losses. This indicates the importance of a stability-aware pose initialization.

Discussion on Stability and Penetration. Previous research has observed that larger penetrations are usually accompanied by smaller displacements (Jiang et al. 2021). We visualize the curve of penetration - displacement by plotting the average displacements from samples with penetration volumes in $[0, 1\text{cm}^3)$, $[1\text{cm}^3, 2\text{cm}^3)$, ..., $[12\text{cm}^3, \infty)$ in Fig. 6.

It is obvious that penetrations and displacements are negatively correlated, which indicates the trade-off between the two criteria: more stable grasps are usually tighter and of more penetrations. Therefore, more penetrations does not necessarily lead to implausible grasps, and simulation displacements reflect stability better if compared under similar penetrations. In Fig. 6, our penetration curve is generally the lowest considering all penetration levels. Our displacements are especially low for small penetrations ($<3\,\text{cm}^3$), which

Methods	Contact Ratio (%) ↑	Penetration Vol. (cm^3)↓	Entropy ↑	Cluster Size ↑
Results on GRAB dataset (Taheri et al. 2020)				
GrabNet (2020)	**100**	8.50	2.80	**4.40**
ContactGen (2023)	91.7	4.02	2.75	4.17
FastGrasp (2025)	98.3	**3.03**	2.81	3.21
Ours	**100**	5.13	**2.83**	4.37
Results on out-of-domain HO3D dataset (Hampali et al. 2020)				
GF (2020)	73.0	9.68	2.85	2.16
GrabNet (2020)	**100**	12.89	**2.88**	4.10
GraspTTA (2021)	**100**	5.34	2.76	3.39
ContactGen (2023)	85.0	**2.84**	2.84	**4.58**
FastGrasp (2025)	92.5	4.79	2.81	3.30
Ours	**100**	6.62	2.80	4.47

Table 2: Quantitative results of geometric plausibility and diversity criteria on both GRAB and HO3D Benchmarks. Note that although our method produces more penetrations, it is unavoidable due to more steady grasps.



(a) GRAB (b) HO3D

Figure 6: Simulation displacements in different penetration intervals. The curve of our method is in bright red and is the lowest especially for small penetrations.

explains the performance improvement in distinction rate.

Force Prediction Analysis

We test the precision of force prediction by comparing the force values in contact regions with simulated results. While there is no ground truth for synthesized poses, we can obtain the force by running the labeling pipeline for it. We also construct an average predictor by predicting all forces as the average force from the training set as the baseline method. The results of our method vs. average predictor are $0.21\,\mathrm{N}$ vs. $0.51\,\mathrm{N}$ for point-level forces and $4.13\,\mathrm{N}$ vs. $17.6\,\mathrm{N}$, which indicates that capturing the force distribution using a simple model like VAE is practical. Fig. 4 show a sample of force prediction, where we conclude that using average force can lead to very large force for all contacts but our method can judge where the force should be large and vice versa, thus producing results closer to the ground truth.

No.	Force Branch	Key point	Force value	Simulation Disp. (cm)↓	Penetration Vol. (cm^3)↓	Distinction Rate (%)↑
1	✗	✗	✗	2.26	7.27	5.0
2	✓	✗	✗	2.02	6.31	5.0
3	✓	✗	✓	1.60	7.70	6.7
4	✓	✓	✗	0.75	6.54	15.0
5	✓	✓	✓	**0.68**	**5.13**	**17.5**

Table 3: Ablation study using GRAB. The numbers of test means: 1. No force prediction in training and inference; 2. Use force map in training, but not in generation; 3. Use force map only as weights in optimization 4. Replace force predictions with average value; 5. Our full method.

n_{kp}	0	1	2	3	4	5	16
Simu. Disp. (cm)↓	2.02	1.78	1.48	0.68	0.66	0.67	**0.56**
Pene. Vol. (cm^3)↓	6.31	**5.00**	5.29	5.13	7.95	8.49	9.02
Distinction Rate (%)↑	5.0	10.0	7.5	**17.5**	6.7	7.5	5.8

Table 4: Ablation study on the number of keypoints n_{kp} using GRAB. $n_{kp} = 3$ is chosen for its highest distinction rate.

Ablation Study

In this section, we first study the effect of force prediction. We set up 5 different experiment configurations and report the settings and results in Tab. 3. Generally, using keypoint-based optimization and integrating exact force values are the components with the largest contributions. By comparing the increase from 2 to 3 and from 2 to 5, we can conclude that using force simply as weights is not as effective as using keypoints. The obvious increase from 2 to 4 demonstrates the effectiveness of keypoints in optimization. The last two rows show that using predicted force values effectively reduces unreasonable penetrations thus improving distinction rate. It indicates the predicted forces can mimic real forces and help to make keypoints easier to fit.

Moreover, we conduct an ablation to study the optimal number of keypoints and report the result in Tab. 4. The trend is clear: with the increase of keypoint number, the displacement decreases and the penetration increases. The distinction rate, considering both stability and penetrations, achieves the peak value at $n_{kp} = 3$. Essentially, although more keypoints lead to smaller displacements, it also comes with a higher risk of unreachable combinations, thus causing both inter- and self-penetrations. Therefore, we choose $n_{kp} = 3$ as the optimal keypoint number.

Conclusion

In this paper, we propose a novel contact-based grasp generation pipeline to improve stability. We extend the contact representation to include contact forces and derive an energy function to model stability by accelerations. Then, stability-aware key contact points are identified using the energy to guide the pose generation. The method achieves state-of-the-art stability in grasp synthesis, suggesting the significance of modeling contact forces and keypoint-guided optimization.

Acknowledgements

This research was funded by the China Scholarship Council - University of Birmingham PhD Scholarship programme (No. 202406230091), the MSIT(Ministry of Science and ICT), Korea, under the ITRC(Information Technology Research Center) support program(IITP-2025-RS-2020-II201789) supervised by the IITP(Institute for Information & Communications Technology Planning & Evaluation), and National Natural Science Foundation of China (Grant No. 62406163).

We would like to thank Yuming Chen for discussions on simulation strategies. We also thank Jungchan Cho and Jeongho Lee for their suggestions on paper writing.

References

Besl, P.; and McKay, N. D. 1992. A method for registration of 3-D shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2): 239–256.

Coumans, E.; and Bai, Y. 2016–2021. PyBullet, a Python module for physics simulation for games, robotics and machine learning. http://pybullet.org.

Grady, P.; Collins, J. A.; Tang, C.; Twigg, C. D.; Aneja, K.; Hays, J.; and Kemp, C. C. 2024. Pressurevision++: Estimating fingertip pressure from diverse rgb images. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 8698–8708.

Grady, P.; Tang, C.; Brahmbhatt, S.; Twigg, C. D.; Wan, C.; Hays, J.; and Kemp, C. C. 2022. PressureVision: estimating hand pressure from a single RGB image. In *European Conference on Computer Vision (ECCV)*, 328–345. Springer.

Grady, P.; Tang, C.; Twigg, C. D.; Vo, M.; Brahmbhatt, S.; and Kemp, C. C. 2021. ContactOpt: Optimizing Contact to Improve Grasps. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Guo, L.; Lu, Z.; and Yao, L. 2021. Human-machine interaction sensing technology based on hand gesture recognition: A review. *IEEE Transactions on Human-Machine Systems*, 300–309.

Hampali, S.; Rad, M.; Oberweger, M.; and Lepetit, V. 2020. Honnotate: A method for 3d annotation of hand and object poses. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, 3196–3206.

Hasson, Y.; Varol, G.; Tzionas, D.; Kalevatykh, I.; Black, M. J.; Laptev, I.; and Schmid, C. 2019. Learning joint reconstruction of hands and manipulated objects. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Ho, J.; Jain, A.; and Abbeel, P. 2020. Denoising Diffusion Probabilistic Models. In Larochelle, H.; Ranzato, M.; Hadsell, R.; Balcan, M.; and Lin, H., eds., *Advances in Neural Information Processing Systems (NeurIPS)*, 6840–6851. Curran Associates, Inc.

Hu, H.; Yi, X.; Zhang, H.; Yong, J.-H.; and Xu, F. 2022. Physical interaction: Reconstructing hand-object interactions with physics. In *SIGGRAPH Asia*, 1–9.

Jiang, H.; Liu, S.; Wang, J.; and Wang, X. 2021. Hand-Object Contact Consistency Reasoning for Human Grasps Generation. In *Proceedings of the International Conference on Computer Vision (ICCV)*.

Karunratanakul, K.; Spurr, A.; Fan, Z.; Hilliges, O.; and Tang, S. 2021. A Skeleton-Driven Neural Occupancy Representation for Articulated Hands. In *International Conference on 3D Vision (3DV)*.

Karunratanakul, K.; Yang, J.; Zhang, Y.; Black, M. J.; Muandet, K.; and Tang, S. 2020. Grasping field: Learning implicit representations for human grasps. In *International Conference on 3D Vision (3DV)*, 333–344. IEEE.

Kingma, D. P.; and Ba, J. 2015. Adam: A Method for Stochastic Optimization. In Bengio, Y.; and LeCun, Y., eds., *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*.

Lee, J.; Saito, S.; Nam, G.; Sung, M.; and Kim, T.-K. 2024. InterHandGen: Two-Hand Interaction Generation via Cascaded Reverse Diffusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 527–537.

Li, J.; Bian, S.; Zeng, A.; Wang, C.; Pang, B.; Liu, W.; and Lu, C. 2021. Human pose regression with residual log-likelihood estimation. In *Proceedings of the IEEE/CVF international conference on computer vision (CVPR)*, 11025–11034.

Li, S.; Bhagat, S.; Campbell, J.; Xie, Y.; Kim, W.; Sycara, K. P.; and Stepputtis, S. 2024. ShapeGrasp: Zero-Shot Task-Oriented Grasping with Large Language Models through Geometric Decomposition. *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 10527–10534.

Liu, S.; Zhou, Y.; Yang, J.; Gupta, S.; and Wang, S. 2023. Contactgen: Generative contact modeling for grasp generation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 20609–20620.

Liu, T.; Liu, Z.; Jiao, Z.; Zhu, Y.; and Zhu, S.-C. 2022. Synthesizing Diverse and Physically Stable Grasps With Arbitrary Hand Structures Using Differentiable Force Closure Estimator. *IEEE Robotics and Automation Letters (IR-AL)*, 470–477.

Lu, J.; Kang, H.; Li, H.; Liu, B.; Yang, Y.; Huang, Q.; and Hua, G. 2024. UGG: Unified Generative Grasping. In *Proceedings of the 2024 European Conference on Computer Vision*, volume 15125, 414–433.

Luo, H.; Liu, Y.; and Yi, L. 2024. Physics-aware Hand-object Interaction Denoising. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2341–2350.

Makoviychuk, V.; Wawrzyniak, L.; Guo, Y.; Lu, M.; Storey, K.; Macklin, M.; Hoeller, D.; Rudin, N.; Allshire, A.; Handa, A.; and State, G. 2021. Isaac Gym: High Performance GPU-Based Physics Simulation For Robot Learning.

Murtaza, Z.; Bonzini, A. A.; Althoefer, K.; and Jamone, L. 2023. QMCube - A Tactile Cube to Explore Hand Interaction Forces in Human Manipulation. In *2023 IEEE International Conference on Development and Learning (ICDL)*.

Nguyen, V.-D. 1988. Constructing force-closure grasps. *The International Journal of Robotics Research*, 7(3): 3–16.

Qi, C. R.; Yi, L.; Su, H.; and Guibas, L. J. 2017. PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc.

Romero, J.; Tzionas, D.; and Black, M. J. 2022. Embodied hands: Modeling and capturing hands and bodies together. *arXiv preprint arXiv:2201.02610*.

Shimada, S.; Mueller, F.; Bednarik, J.; Doosti, B.; Bickel, B.; Tang, D.; Golyanik, V.; Taylor, J.; Theobalt, C.; and Beeler, T. 2024. MACS: Mass Conditioned 3D Hand and Object Motion Synthesis. In *International Conference on 3D Vision (3DV)*.

Sohn, K.; Lee, H.; and Yan, X. 2015. Learning structured output representation using deep conditional generative models. *Advances in neural information processing systems (NeurIPS)*, 28.

Taheri, O.; Choutas, V.; Black, M. J.; and Tzionas, D. 2022. Goal: Generating 4d whole-body motion for hand-object grasping. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR))*, 13263–13273.

Taheri, O.; Ghorbani, N.; Black, M. J.; and Tzionas, D. 2020. GRAB: A Dataset of Whole-Body Human Grasping of Objects. In *European Conference on Computer Vision (ECCV)*.

Todorov, E.; Erez, T.; and Tassa, Y. 2012. MuJoCo: A physics engine for model-based control. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 5026–5033. IEEE.

Wang, R.; Zhang, J.; Chen, J.; Xu, Y.; Li, P.; Liu, T.; and Wang, H. 2023. DexGraspNet: A Large-Scale Robotic Dexterous Grasp Dataset for General Objects Based on Simulation. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, 11359–11366.

Wu, X.; Liu, T.; Li, C.; Ma, Y.; Shi, Y.; and He, X. 2025. FastGrasp: Efficient Grasp Synthesis with Diffusion. *International Conference on 3D Vision (3DV)*.

Xu, G.-H.; Wei, Y.-L.; Zheng, D.; Wu, X.-M.; and Zheng, W.-S. 2024. Dexterous grasp transformer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 17933–17942.

Yuan, Y.; Song, J.; Iqbal, U.; Vahdat, A.; and Kautz, J. 2023. Physdiff: Physics-guided human motion diffusion model. In *Proceedings of the IEEE/CVF international conference on computer vision (ICCV)*, 16010–16021.

Zhang, Z.; Cheng, Y.; Pérez-Pellitero, E.; Zhou, Y.; Deng, J.; Chang, H. J.; and Song, J. 2025. Single-view Image to Novel-view Generation for Hand-Object Interactions. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 10394–10402.

Zhang, Z.; Wang, H.; Yu, Z.; Cheng, Y.; Yao, A.; and Chang, H. J. 2024. NL2Contact: Natural Language Guided 3D Hand-Object Contact Modeling with Diffusion Model. In *European Conference on Computer Vision (ECCV)*, 284–300. Springer.

Zhao, Y.; Kwon, T.; Streli, P.; Pollefeys, M.; and Holz, C. 2025. EgoPressure: A Dataset for Hand Pressure and Pose Estimation in Egocentric Vision. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Zhao, Z.; Zuo, B.; Xie, W.; and Wang, Y. 2022. Stability-driven contact reconstruction from monocular color images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR))*, 1643–1653.

Zhong, Y.; Jiang, Q.; Yu, J.; and Ma, Y. 2025. DexGrasp Anything: Towards Universal Robotic Dexterous Grasping with Physics Awareness. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Zhou, K.; Bhatnagar, B. L.; Lenssen, J. E.; and Pons-Moll, G. 2022. TOCH: Spatio-Temporal Object Correspondence to Hand for Motion Refinement. In *European Conference on Computer Vision (ECCV)*. Springer.

Zuo, B.; Zhao, Z.; Sun, W.; Yuan, X.; Yu, Z.; and Wang, Y. 2024. GraspDiff: Grasping Generation for Hand-Object Interaction With Multimodal Guided Diffusion. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 1–13.