Supplementary Material for "An Economic Framework for 6-DoF Grasp Detection"

Xiao-Ming Wu[1,3,†,&], Jia-Feng Cai[1,3,&], Jian-Jian Jiang[1,3], Dian Zheng[1,3], Yi-Lin Wei[1,3], and Wei-Shi Zheng[1,2,3,4,*]

[1] School of Computer Science and Engineering, Sun Yat-sen University, China;
[2] Peng Cheng Laboratory, Shenzhen, China; [3] Key Laboratory of Machine Intelligence and Advanced Computing, Ministry of Education, China
[4] Pazhou Laboratory (Huangpu), Guangzhou, Guangdong, China
{wuxm65, caijf23, jiangjj35, zhengd35, weiylin5}@mail2.sysu.edu.cn, wszheng@ieee.org

S1 Ablation Study of Resource Costs

We test the resource costs of the main components of our framework. The costs are tested in an empty machine with one NVIDIA RTX3090 GPU for fair comparison. All the experiments are trained on the GraspNet-1Billion dataset [2] with the Kinect data. The results are shown in Table S1. We can observe that 1) economic supervision greatly enhances the model performance, only with little increase in the resource costs, which strongly proves the effectiveness of our supervision selection strategy. 2) the interactive head and the composite score estimation also cost little resource to further improve the model performance, making it surpass the SOTA grasping method under dense supervision.

Table S1: Ablation study of the resource costs of the main components.

economic supervision	interactive head	composite score	time(h)	memory(G)	GPUs(G)	mAP
			5.45	3.9	4.26	30.34
✓			8.16	4.2	4.78	42.31
✓	✓		8.25	4.2	5.11	44.14
✓	✓	✓	8.30	4.2	5.81	44.62

S2 Ablation Study of Different Supervision Types

We further try different types of sparse supervision to evaluate the effectiveness of our economic supervision. The experiments are conducted on the GraspNet-1Billion dataset [1] with the Kinect data. The implementation is based on our

† : project lead, &: equal key contributions, *: corresponding author.

EconomicGrasp framework and revising the output shape to fit different supervisions. As shown in Table S2, it shows that our economic supervision strategy is nontrivial, being the only one that surpassing the dense method.

Table S2: Comparison of different types of supervision.

Order	Supervision	Types	Seen ↑	Similar ↑	Novel ↑
1	GSNet [4]	Dense	61.19	47.39	19.01
2	Good demonstrations from natural	Sparse	34.47	31.70	12.99
3	Top-1 in each point	Sparse	43.59	34.09	13.36
4	Top-1 & random-99 in each point	Sparse	27.47	21.52	8.87
5	Keep angles in each point	Sparse	38.93	33.81	16.76
6	Keep depths in each point	Sparse	28.93	19.72	7.84
7	Keep partial views in each point	Sparse	59.19	47.69	18.27
8	EconomicGrasp (keep views in each point)	Sparse	**62.59**	**51.73**	**19.54**

S3 Plug-and-Play Analysis

Our EconomicGrasp framework is plug-and-play for current 6-DoF grasp methods, including the economic supervision and the focal representation module to suit our supervision. We test them in different 6-DoF grasp methods with different backbones on the GraspNet-1Billion dataset [1] with the Kinect data. To be noted that GSNet [4] is similar to our vanilla grasp framework except the grasp head to fit the supervision. Thus, we directly use our EconomicGrasp framework to be the GSNet [4] + EconomicGrasp. As shown in Table S3, it proofs our claim and shows great potential for our economic supervision.

Table S3: Plug-and-play tests of our supervision and method.

Models	Architecture	Seen ↑	Similar ↑	Novel ↑	Time (h) ↓
GraspNet Baseline [2]	PointNet++	27.56	26.11	10.55	116.7
GraspNet Baseline [2] + EconomicGrasp	PointNet++	**46.01**	**36.08**	**13.75**	**9.5**
TransGrasp [3]	Transformer	39.81	29.32	13.83	41.2
TransGrasp [3] + EconomicGrasp	Transformer	**56.41**	**45.26**	**17.84**	**14.9**
GSNet [4]	3DConv	61.19	47.39	19.01	37.8
GSNet [4] + EconomicGrasp	3DConv	**62.59**	**51.73**	**19.54**	**8.3**

S4 Diversity Analysis

In addition, we test the diversity of our method following a classical diversity metric [5], which evaluates the standard variance of translations and rotations

of the good grasps on each object. The experiments are tested in the GraspNet-1Billion dataset [2], Kinect, seen data. As soon in Table S4, we surprisingly find that our EconomicGrasp has more diversified high-quality grasps than GSNet [4], which is due to the fact that we keep views for each points to maintain diversity and meanwhile erase redundancy to make the model easy to be optimized. More analysis and potential of economic supervision are waiting for future exploration.

Table S4: Diversity of the generated good grasps in object level.

Method	Translation (m) ↑	Rotation (degree) ↑	Ratio of objects with less than 1 grasp ↓
GSNet	0.030	26.58	43.08%
EconomicGrasp	**0.064**	**56.95**	**12.03%**

S5 Video Demo

A video demo is attached. This video shows several real-world grasping scenarios and the training speed of our model. The real-world grasping experiments are conducted by the Franka Emika robot hand with a two-finger gripper and a RealSense camera. The training speed are tested in the same environment as Sec. S1. Moreover, we capture some high DoF examples from the video shown in Fig. S1, demonstrating the effectiveness and the flexibility of 6-DoF grasping.

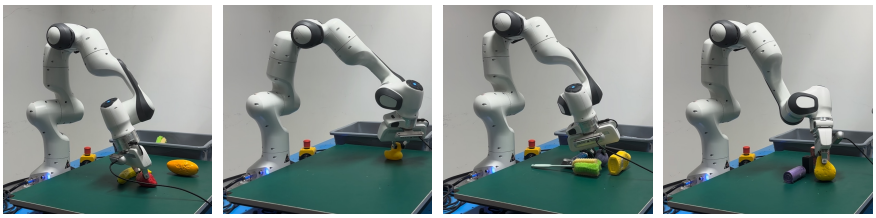

Fig. S1: The real-world grasping examples with flexible grasps.

References

1. Asif, U., Tang, J., Harrer, S.: Graspnet: An efficient convolutional neural network for real-time grasp detection for low-powered devices. In: International Joint Conferences on Artificial Intelligence (2018) 1, 2
2. Fang, H.S., Wang, C., Gou, M., Lu, C.: Graspnet-1billion: A large-scale benchmark for general object grasping. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (2020) 1, 2, 3

3. Liu, Z., Chen, Z., Xie, S., Zheng, W.S.: Transgrasp: A multi-scale hierarchical point transformer for 7-dof grasp detection. In: International Conference on Robotics and Automation (2022) 2
4. Wang, C., Fang, H.S., Gou, M., Fang, H., Gao, J., Lu, C.: Graspness discovery in clutters for fast and accurate grasp detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (2021) 2, 3
5. Xu, Y., Wan, W., Zhang, J., Liu, H., Shan, Z., Shen, H., Wang, R., Geng, H., Weng, Y., Chen, J., et al.: Unidexgrasp: Universal robotic dexterous grasping via learning diverse proposal generation and goal-conditioned policy. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (2023) 2