55HFF6D:实例级方法的论文 2022 年。针对问题：当前方法在具有挑战性的场景中通常表现不佳，例如不正确的初始姿势、突然重新定向和严重遮挡。所做工作：提出了一种鲁棒的 6D 物体姿态跟踪方法，具有新颖的分层特征融合网络，将其称为 HFF6D，旨在预测相邻帧之间物体的相对姿态。提出了一种具有注意机制的新型减法特征融合（SFF）模块，以在特征融合期间利用特征减法。利用数据增强技术，通过仅使用合成数据进行训练，使 HFF6D 在现实世界中更有效地使用，从而减少数据注释中的手动工作。

56NVRNET: 实例级方法的论文 2023 年。针对问题：提升旋转精度，消除了间接方法和端到端方法之间的差距。所做工作：提出了一种新颖的法线向量引导回归网络（NVR-Net），可在 3D 法线向量的指导下直接从单个 RGB 图像回归 6D 姿态。设计了一种新颖的方向感知特征（OAF）用于姿势估计。它由两组相应的 3D 法向量组成，以彻底将旋转与平移估计分开。然后，我们引入 CNN 来预测 OAF 的密集像素表示，而不会出现视点模糊。为了从 OAF 中单独估计旋转和平移，我们在可微封闭形式解决方案的指导下提出了一种来自法线向量 (PNV) 头部网络的新颖姿势。

57DGECN++:实例级方法的论文 2023 年。针对问题：大多数在执行可微 PnP 算法时几乎没有考虑 3D 空间中的几何特征，并且忽略了拓扑线索。提出了一种改进的端到端单目 6D 位姿估计方法（DGECN++），该方法结合了深度估计和几何感知的可学习 PnP 网络。我们的方法基于关键点。首先，我们检测与 3D 模型相对应的 2D 关

键点。然后，我们集成可微分 PnP/RANSAC 算法来创建用于 6D 位姿估计的端到端管道。

58GPT-COPE：类别级方法的论文 2023 年。针对问题：当前的方法仍然无法充分利用单个实例的内在几何特征，从而限制了它们处理具有复杂结构的对象（即相机）的能力。所做工作：提出 GPT-COPE，它利用图形引导的点变换器从观察到的点云中探索独特的几何特征。

TABLE I

QUANTITATIVE COMPARISON WITH STATE-OF-THE-ART METHODS ON THE NOCS-REAL DATASET [5]. THE FIRST AND SECOND HIGHEST VALUES ARE HIGHLIGHTED BY **BOLD** AND UNDERLINE. WE USE THE RESULTS OF FS-NET [11] PROVIDED BY GPV-NET [13] TO ENSURE A FAIRER COMPARISON

Method	Training Data	Prior	NOCS-REAL (↑)							#Params (M) (↓)
			IoU_{50}	IoU_{75}	5°2cm	5°5cm	10°2cm	10°5cm	10°10cm	
NOCS[CVPR'19] [5]	RGB	✗	80.5	30.1	7.2	9.5	13.8	26.7	26.7	-
SPD [ECCV'20] [6]	RGB-D	✓	77.3	53.2	19.3	21.4	43.2	54.1	-	18.3
DPN [ICCV'21] [28]	RGB-D	✗	79.8	62.2	29.3	35.9	50.0	66.8	-	67.9
SS-Conv [NeurIPS'21] [29]	RGB-D	✗	79.8	65.6	36.6	43.4	52.6	63.5	-	23.3
CR-Net [IROS'21] [7]	RGB-D	✓	79.3	55.9	27.8	34.3	47.2	60.8	-	21.4
SGPA [ICCV'21] [8]	RGB-D	✓	80.1	61.9	35.9	39.6	61.3	70.7	-	23.3
SSC-6D [AAAI'22] [31]	RGB-D	✗	73.0	-	16.8	19.6	44.1	54.5	56.2	-
CenterSnap [ICRA'22] [30]	RGB-D	✗	80.2	-	-	27.2	-	58.8	64.4	-
UDA-COPE [CVPR'22] [27]	RGB-D	✗	82.6	62.5	30.4	34.8	56.9	66.0	-	-
ShAPO [ECCV'22] [33]	RGB-D	✗	79.0	-	-	48.8	-	66.8	78.0	-
FS-Net [CVPR'21] [11]	D	✗	81.1	52.0	19.9	33.9	-	69.1	71.0	41.2
SAR-Net [CVPR'22] [26]	D	✗	79.3	62.4	31.6	42.3	50.3	68.3	-	6.3
GPV-Pose [CVPR'22] [13]	D	✗	**83.0**	64.4	32.0	42.9	-	73.3	74.6	8.6
UDA-COPE [CVPR'22] [27]	D	✗	79.6	57.8	21.2	29.1	48.7	65.9	-	-
GPT-COPE (Ours)	D	✓	82.0	**70.4**	**45.9**	**53.8**	**63.1**	**77.7**	**79.8**	7.1

TABLE II

QUANTITATIVE COMPARISON WITH STATE-OF-THE-ART METHODS ON THE NOCS-CAMERA DATASET [5]. THE FIRST AND SECOND HIGHEST VALUES ARE HIGHLIGHTED BY **BOLD** AND UNDERLINE

Method	Training Data	Prior	NOCS-CAMERA (↑)							#Params (M) (↓)
			IoU_{50}	IoU_{75}	5°2cm	5°5cm	10°2cm	10°5cm	10°10cm	
NOCS [CVPR'19] [5]	RGB	✗	83.9	69.5	32.3	40.9	48.2	64.6	65.1	-
SPD [ECCV'20] [6]	RGB-D	✓	93.2	83.1	54.3	59.0	73.3	81.5	-	18.3
CR-Net [IROS'21] [7]	RGB-D	✓	93.8	88.0	**72.0**	76.4	81.0	87.7	-	21.4
SGPA [ICCV'21] [8]	RGB-D	✓	93.2	**88.1**	70.7	74.5	**82.7**	88.4	-	23.3
DPN [ICCV'21] [28]	RGB-D	✗	92.4	86.4	64.7	70.7	77.2	84.7	-	67.9
CenterSnap [ICRA'22] [30]	RGB-D	✗	92.5	-	-	66.2	-	81.3	87.9	-
ShAPO [ECCV'22] [33]	RGB-D	✗	93.5	-	-	66.6	-	81.9	89.2	-
FS-Net[CVPR'21] [11]	D	✗	-	85.2	-	62.0	-	60.8	-	41.2
SAR-Net [CVPR'22] [26]	D	✗	86.8	79.0	66.7	70.9	75.3	80.3	-	6.3
GPV-Pose [CVPR'22] [13]	D	✗	92.9	86.6	67.4	76.2	-	87.4	-	8.6
GPT-COPE (Ours)	D	✓	92.5	86.9	70.4	**76.5**	81.3	**88.7**	**89.9**	7.1

59CLIPose：类别级方法的论文 2024 年。针对问题：从有限的点云样本中提取的类别特征可能并不全面。所做工作：研究是否可以利用其他方式的知识来获取类别信息。受此动机的启发，提出了 CLIPose，它采用预先训练的视觉语言模型来更好地学习对象类别信息，可以充分利用图像和文本模态中丰富的语义知识。为了使 3D 编码器更有效地学习特定类别的特征，我们通过多模态对比学习在特征空间中对齐三种模态的表示。除了利用 CLIP 模型的预训练知识之外，我们还期

望它对姿态参数更加敏感。因此，我们引入了一种即时调整方法来微调图像编码器，同时将旋转和平移信息合并到文本描述中。

TABLE I
COMPARISONS WITH STATE-OF-THE-ART METHODS ON REAL275 DATASET. OVERALL BEST RESULTS ARE IN BOLD AND THE SECOND BEST RESULTS ARE UNDERLINED. INPUT DENOTES THE FORMAT OF THE INPUT DATA USED BY THE METHODS, AND "S" DENOTES SHAPE PRIORS, "I/T" REPRESENTS IMAGE AND TEXT PRE-TRAINED KNOWLEDGE. '-' INDICATES PRIOR-FREE METHODS OR NO RESULTS REPORTED UNDER THIS METRIC IN ORIGINAL PAPER. (↑) REPRESENTS A HIGHER VALUE INDICATING BETTER PERFORMANCE.

Methods	Input	Prior	IoU_{50}↑	IoU_{75}↑	5°2cm↑	5°5cm↑	10°2cm↑	10°5cm↑	10°10cm↑	Speed(FPS)↑
NOCS [14]	RGB-D	-	80.5	30.1	-	9.5	13.8	26.7	26.7	9.9
CASS [40]	RGB-D	-	77.7	-	19.5	23.5	50.8	58.0	58.3	-
DualPoseNet [60]	RGB-D	-	79.8	62.2	29.3	35.9	50.0	66.8	-	2.6
SPD [15]	RGB-D	S	77.3	53.2	19.3	21.4	43.2	54.1	-	7.9
CR-Net [34]	RGB-D	S	79.3	55.9	27.8	34.3	47.2	60.8	-	-
SGPA [36]	RGB-D	S	80.1	61.9	35.9	39.6	61.3	70.7	-	6.6
DPDN [33]	RGB-D	S	83.4	76.0	46.0	50.7	70.4	78.4	80.4	31.6
IST-Net [18]	RGB-D	-	82.5	**76.6**	47.5	53.4	**72.1**	80.5	82.6	33.7
FS-Net [16]	D	-	81.1	63.5	19.9	33.9	-	69.1	71.0	16.1
GPV-Pose [17]	D	-	83.1	65.4	32.0	42.9	55.0	73.3	74.6	29.6
SSP-Pose [61]	D	S	82.3	66.3	34.7	44.6	-	77.8	79.7	37.0
RBP-Pose [62]	D	S	-	67.8	38.2	48.1	63.1	79.2	-	37.8
GPT-COPE [19]	D	S	82.0	70.4	45.9	53.8	63.1	77.7	79.8	-
DR-Pose [63]	D	S	78.9	68.2	41.7	46.0	67.7	76.3	-	-
HS-Pose [37]	D	-	82.7	75.3	45.3	54.9	68.6	83.6	84.6	39.8
Ours	D	I/T	**83.6**	**76.6**	**48.7**	**58.3**	70.4	**85.2**	**86.2**	**40.0**

TABLE II
COMPARISONS WITH STATE-OF-THE-ART METHODS ON CAMERA25 DATASET. OVERALL BEST RESULTS ARE IN BOLD AND THE SECOND BEST RESULTS ARE UNDERLINED. "S" DENOTES SHAPE PRIORS, "I/T" REPRESENTS IMAGE AND TEXT PRE-TRAINED KNOWLEDGE. '-' INDICATES PRIOR-FREE METHODS OR NO RESULTS REPORTED UNDER THIS METRIC IN ORIGINAL PAPER. (↑) REPRESENTS A HIGHER VALUE INDICATING BETTER PERFORMANCE.

Methods	Prior	IoU_{75}↑	5°2cm↑	5°5cm↑	10°2cm↑	10°5cm↑
NOCS [14]	-	85.3	32.3	40.9	48.2	64.6
DualPoseNet [60]	-	62.2	29.3	35.9	50.0	66.8
GPV-Pose [37]	-	88.3	72.1	79.1	-	89.0
IST-Net [18]	-	90.8	71.3	79.9	79.4	89.9
HS-Pose [37]	-	89.4	73.3	80.5	80.4	89.4
SPD [15]	S	83.1	54.3	59.0	73.3	81.5
SAR-Net [64]	S	70.9	66.7	70.9	75.3	80.3
CR-Net [34]	S	88.0	72.0	76.4	81.0	87.7
SGPA [36]	S	88.1	70.7	74.5	**82.7**	88.4
GPT-COPE [19]	S	86.9	70.4	76.5	81.3	88.7
RBP-Pose [62]	S	89.0	73.5	79.6	82.1	89.5
Ours	I/T	**91.0**	**74.8**	**82.2**	82.0	**90.6**

60SERA：类别级方法的论文 2022 年。本文提出了一种新的带有结构编码器和推理注意的类别级 6D 目标姿态估计框架。引入结构自编码器，通过一种不同的学习策略来挖掘同一类别内彩色图像中的共

享结构特征，该策略恢复另一个实例的图像，但与输入的图像具有最相似的姿态。在此基础上，堆叠一个推理注意解码器和全连接层，形成一个旋转预测网络，将结构特征和三维形状特征集成并投影到语义空间。语义空间包括观察模式和可学习模式，可学习模式通过在推理注意解码器上添加一个与之并行的快捷连接分支来学习。基于这些模式的进一步推理使解码器具有强大的特征表示。该方法在没有三维对象模型的情况下，在语义空间中隐式地对类别属性进行建模，并在该空间上进行推理，保证了 6D 对象姿态估计的良好性能。