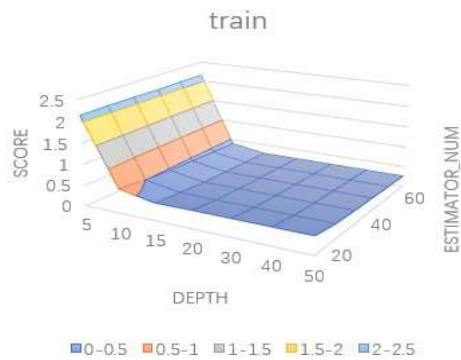


6.2 随机森林实验结果与分析

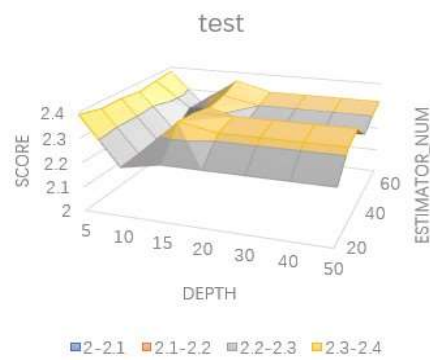
用随机森林回归 q 和 r 时，随着随机森林深度的加深，训练集 score 越来越小，而验证集 score 在短暂下降后立即上升（表 1），说明在 $depth$ 大于 15 已经出现过拟合了。而 $estimator$ 的数量无论是在验证集还是训练集中与 score 的相关性都不明显。因此在配置参数时将 $depth$ 配置为 10，可得到比较好的结果，取 $estimator_num=30,depth=10$ 作为随机森林的参数，其训练集 score 为 0.50，验证集 score 为 2.19。

表 1：随机森林回归 q 和 r 在训练集和验证集上的表现:

estimator num	20		30		40		50		60		70	
depth	train	test	train	test	train	test	train	test	train	test	train	test
5	2.14	2.39	2.15	2.37	2.15	2.37	2.14	2.36	2.15	2.37	2.15	2.38
10	0.58	2.20	0.50	2.19	0.46	2.20	0.43	2.20	0.41	2.26	0.39	2.23
15	0.42	2.23	0.36	2.35	0.32	2.34	0.30	2.32	0.28	2.33	0.26	2.36
20	0.43	2.23	0.35	2.32	0.32	2.34	0.30	2.27	0.28	2.30	0.26	2.32
30	0.43	2.23	0.35	2.32	0.32	2.34	0.30	2.27	0.28	2.30	0.26	2.32
40	0.43	2.23	0.35	2.32	0.32	2.34	0.30	2.27	0.28	2.30	0.26	2.32
50	0.43	2.23	0.35	2.32	0.32	2.34	0.30	2.27	0.28	2.30	0.26	2.32

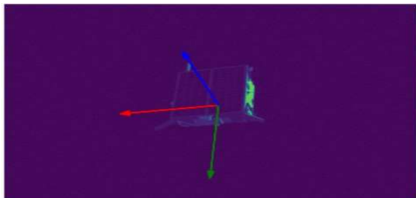


(d) 训练集 score 与两个参数的关系

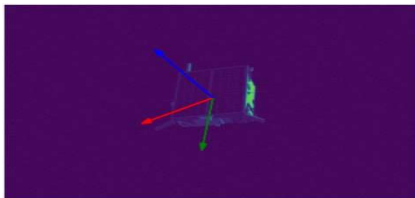


(e) 验证集 score 与两个参数的关系

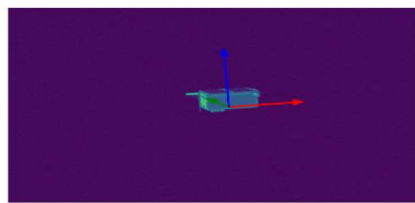
随机森林直接回归 q, r 的可视化效果图:



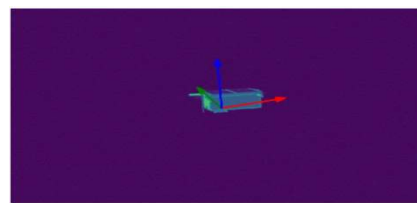
(f) 真实标签可视化



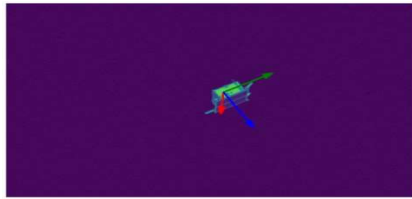
(g) 预测标签可视化



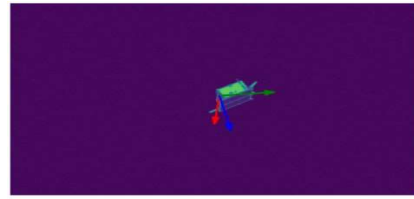
(h) 真实标签可视化



(i) 预测标签可视化



(j) 真实标签可视化

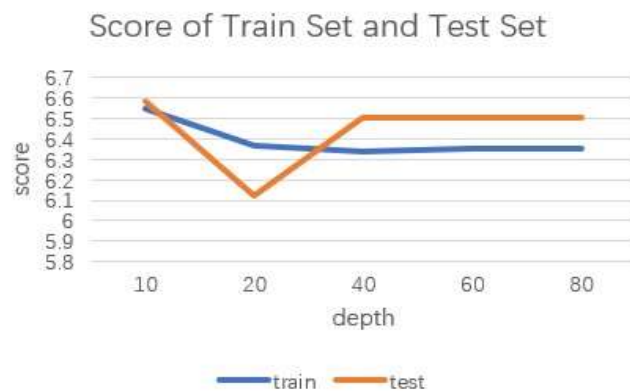


(k) 预测标签可视化

用随机森林回归关键点时,由上一个实验结果可知 $n_estimator$ 对结果影响不大,所以将 $n_estimator$ 固定为 20,调整 $depth$ 使算法从欠拟合到过拟合,整个过程当中 $score$ 都特别大(表 2),取 $estimator_num=20,depth=20$ 作为随机森林的参数,其训练集 $score$ 为 6.37,验证集 $score$ 为 6.12。

表 2: 随机森林回归关键点在训练集和验证集上的表现:

Depth	10	20	40	60	80
Train	6.55	6.37	6.34	6.35	6.34
Test	6.58	6.12	6.50	6.50	6.50



(l) 训练集和测试集 $score$ 与 $depth$ 的关系