HYPOTHESIS TESTING

First, taking the **3D object classification** experiment as an example, we prove that the proposed method is significantly different from other methods through the hypothesis testing.

Table 1. ModelNet shape classification. Comparison of the accuracy of the proposed model with the state-of-the-art. Our network has better performance on ModelNet10. Our network achieves state-of-the-art performance on ModelNet40 among deep nets on 3D input. The top 2 ranked values are highlighted in bold and in the colors red and blue, respectively.

Network	Representation	Input	Accuracy			
			ModelNet10		ModelNet40	
			(avg. class)	(overall)	(avg. class)	(overall)
MVCNN [40]	2D images	$80\times(164^2)$	-	-	89.7	92.0
3DShapeNets [47]	3D volumetric grid	30^3	83.5	-	77.3	-
OctNet [37]	3D volumetric grid	128^3	90.1	90.9	83.8	86.5
VoxNet [30]	3D volumetric grid	32^3	92.0	-	83.0	-
O-CNN [45]	3D volumetric grid	64^3	-	-	-	90.6
RGCNN [44]	points + normal	$1024\times(3+3)$	-	-	87.3	90.5
PointNet++ [33]	points + normal	$5000\times(3+3)$	-	-	-	91.9
So-Net [29]	points + normal	$5000\times(3+3)$	95.5	95.7	90.8	93.4
ECC [39]	points	1000×3	90.0	90.8	83.2	87.4
PointNet [9]	points	1024×3	-	-	86.2	89.2
DeepSets [53]	points	5000×3	-	-	-	90.0
PointNet++ [33]	points	$5000\times(3+3)$	-	-	-	90.7
Kd-network(depth 10)	points	$2^{15}\times3$	92.8	93.3	86.3	90.6
Kd-network(depth 15)	points	$2^{15}\times3$	93.5	94.0	**88.5**	**91.8**
So-Net [29]	points	2048×3	**93.9**	**94.1**	87.3	90.9
PointwiseNet	points	1024×3	**94.1**	**94.5**	**89.1**	**91.3**

As shown in Table 1, we did a comparison about the overall classification accuracy on ModelNet40 dataset. `-' indicates that the author of the corresponding method did not report the accuracy on that dataset. Considering that the main advantage of this method is the ability to directly process point cloud data, we only select the methods (ECC, PointNet, DeepSets, Kd-network (depth 10), Kd-network (depth 15), So-Net) that can directly process point cloud data for comparison.

Table 2.

Network	Representation	Input	Accuracy (overall)
ECC [39]	points	1000×3	87.4
PointNet [9]	points	1024×3	89.2
DeepSets [53]	points	5000×3	90.0
PointNet++ [33]	points	5000×(3 + 3)	90.7
Kd-network(depth 10) [24]	points	2^{15}×3	90.6
Kd-network(depth 15) [24]	points	2^{15}×3	**91.8**
So-Net [29]	points	2048×3	90.9
PointwiseNet	points	1024×3	**91.3**

We have performed 20 independent experiments (including the complete training process and validation process) for the presented method. The experimental results are shown as follows:

Table 3. The results of the proposed method in 20 independent experiments.

1	91.1525	**6**	90.9091	**11**	90.7873	**16**	90.8685
2	91.1932	**7**	90.9497	**12**	91.0308	**17**	90.9091
3	91.1120	**8**	90.7468	**13**	90.9903	**18**	90.7468
4	90.8685	**9**	91.1120	**14**	90.8656	**19**	91.3120
5	90.9844	**10**	90.8685	**15**	90.8279	**20**	90.9497

The mean of the seven other methods (ECC, PointNet, DeepSets, Kd-network (depth 10), Kd-network (depth 15), So-Net) is:

$$\mu = (87.4+89.2+90.0+90.7+90.6+91.8+90.9)/7=90.0857.$$

The mean and standard deviation of our method in 20 independent experiments are:

$$\bar{x} = 90.95924, \ s = 0.149275.$$

We choose the *t*-statistic as our statistical test method and give the step process of hypothesis testing as follows:

Step 1: set the null hypothesis $H_0 \leq \mu$, and the alternative hypothesis $H_A > \mu$

Step 2: calculate the $S_{\bar{x}}$ and t as:

$$S_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{0.149275}{\sqrt{20}} = 0.0333789$$

$$t = \frac{\bar{x} - \mu}{S_{\bar{x}}} = \frac{90.95924 - 90.0857}{0.0333789} = 26.16922$$

Look up the *t*-value table, we can get: $t_{0.001,19} = 3.883$

Therefore, we have $t = 26.16922 > t_{0.001,19} = 3.883$. Thus, we reject the null hypothesis if t < 26.16922 and accept the alternate hypothesis.

It can be seen that the difference between the proposed method and other methods is significant, and we have more than 99.9% confidence that our method is better than the average.

Second, looking into Table 2, we found that the overall classification accuracy of this method on the ModelNet40 dataset is slightly lower than the Kd-Net method. We think this is acceptable because the input of Kd-Net is $2^{15} \times 3$, which is much larger than that of this method.

Third, from all the baselines of recent years in the literature, references in CVPR, ICCV, ECCV etc., rarely use statistical hypothesis testing as the evaluation measurement. Researchers in the area of deep learning always use their best models to test on public datasets.