

# A Novel Point Cloud Registration Method Based on ROPNet

Yuan Li <sup>†</sup>, Fang Yang <sup>†</sup> and Wanning Zheng <sup>\*†</sup> 

The Engineering Research Center of Metallurgical Automation and Measurement Technology, Wuhan University of Science and Technology, Wuhan 430081, China

\* Correspondence: zhengwanning@wust.edu.cn

<sup>†</sup> Current address: School of Information Science and Engineering, Wuhan University of Science and Technology, Wuhan 430081, China.

**Abstract:** Point cloud registration is a crucial preprocessing step for point cloud data analysis and applications. Nowadays, many deep-learning-based methods have been proposed to improve the registration quality. These methods always use the sum of two cross-entropy as a loss function to train the model, which may lead to mismatching in overlapping regions. In this paper, we designed a new loss function based on the cross-entropy and applied it to the ROPNet point cloud registration model. Meanwhile, we improved the ROPNet by adding the channel attention mechanism to make the network focus on both global and local important information, thus improving the registration performance and reducing the point cloud registration error. We tested our method on ModelNet40 dataset, and the experimental results demonstrate the effectiveness of our proposed method.

**Keywords:** deep learning; point cloud registration; cross-entropy loss; channel attention mechanism

## 1. Introduction

Point cloud registration is the basis of 3D reconstruction [1,2], 3D localization [3–6], pose estimation [7–9], and other fields. With the development of deep learning, point cloud registration develops from traditional iterative closest point-based (ICP) [10] methods to approaches based on deep learning, such as PCRNNet [11], D3feat [12], iterative distance-aware similarity matrix convolution network (IDAM) [13], point cloud registration with deep attention to the overlap region (PREDATOR) [14], robust point matching (RPMNet) [15], etc. In recent years, these methods have achieved good registration results under certain conditions. However, most of the current models do not perform well in overlapping regions. In this paper, to solve the above issue, we will take measures from two aspects, i.e., the loss function and the backbone network. To demonstrate the effectiveness, we use the representative overlapping points network (ROPNet) [16] as an example in the following context.

Loss functions measure the difference between the real value and the predicted value in deep learning and play important roles in back-propagation. Depending on different tasks, loss functions can be roughly divided into two categories. The first one is used for regression tasks, such as mean absolute (MAE), mean square error (MSE), Huber loss, etc. The other one is used for classification tasks, such as focal loss [16], center loss [17], circle loss [18], dice loss [19], cross-entropy loss, etc. The point cloud registration task involves the classification problem, and often uses the cross-entropy as the loss function. However, researchers [19,20] found that, in some cases cross-entropy loss can not achieve good results. For example, when the data between positive and negative samples are seriously unbalanced, negative samples will dominate the training of the model. In this case, by using the original cross-entropy loss, the model cannot predict accurately. To solve this problem, many researchers have made a lot of improvements based on the original cross-entropy, such as weighted cross-entropy (WCE) [21], balanced cross-entropy (BCE) [22], dice loss [19], focal loss [20], etc. Most of the above approaches improve the cross-entropy



**Citation:** Li, Y.; Yang, F.; Zheng, W. A Novel Point Cloud Registration Method based on ROPNet. *Sensors* **2023**, *23*, 993. <https://doi.org/10.3390/s23020993>

Academic Editor: Jiayi Ma

Received: 7 December 2022

Revised: 3 January 2023

Accepted: 11 January 2023

Published: 15 January 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

loss by introducing weights; however, these metrics often neglect the relationship between the corresponding points in overlapping regions.

In this paper, to improve the registration quality on overlapping areas, we introduce a new way of calculating loss function based on the cross-entropy. Specifically, we use the product of two cross-entropy loss instead of the sum. We prove the validity of our proposed loss function in theory in Section 3.1. Moreover, to testify the effectiveness of the proposed loss, we apply it to a recently proposed network, i.e., the ROPNet, which originally uses the sum of two cross-entropy to supervise the overlapping area. In addition, the ROPNet considers more global information but ignores local information, which leads to mis-registration at places where there are details. Therefore, we propose to add a channel attention mechanism to focus on important local information, and use it to complement the self-attention mechanism in the original model. A large number of experiments show that, compared with the original ROPNet scheme, the proposed method has a certain improvement in overlapping point cloud registration. Our main contributions are as follows:

- A new loss function based on cross-entropy is designed, which helps improve the point clouds registration accuracy, especially in overlapping regions.
- The performance of the ROPNet is also improved by adding a channel attention mechanism.
- A large number of experiments on commonly used datasets were conducted and verified the effectiveness of our proposed method.

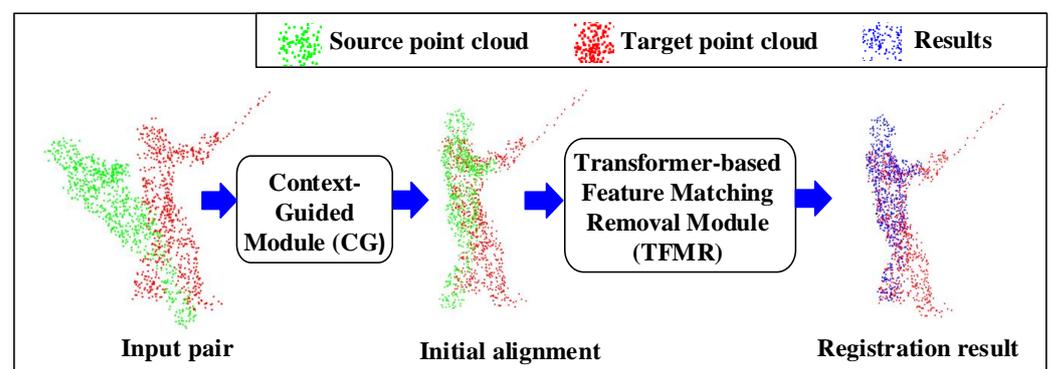
The rest of the paper is organized as follows. Section 2 introduces the related work. In Section 3, a new loss function based on the cross-entropy loss is proposed. Experiments are presented in Section 4 and conclusions are drawn in Section 5.

## 2. Related Work

In this section, we introduce some related work and preliminary knowledge, including the ROPNet model, cross-entropy loss, and attention mechanism.

### 2.1. ROPNet

ROPNet is a point cloud registration model that typically uses representative points in overlapping regions for registration. As shown in Figure 1, the ROPNet consists of a context-guided (CG) module and a transformer-based feature matching removal (TFMR) module.



**Figure 1.** The original point cloud registration model of ROPNet.

Generally speaking, ROPNet can be divided into two stages: coarse alignment and fine alignment. In the first stage, the source point cloud and the target point cloud are input. The CG module is used to extract global features from the input and compute the overlap score between two point clouds. The extracted global features are then used for coarse alignment, and the non-representative points are removed based on the overlap score. In

the second stage, point features are generated based on the offset attention [23,24]. The representative points are used to generate accurate correspondences. Finally, the precise registration of point clouds is carried out.

## 2.2. Cross-Entropy Loss

The cross-entropy loss is a widely used loss function to classification problems. In the point cloud registration model [16,25],  $p$  is defined as the true distribution. It can be written as

$$p(x_i) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

where  $x_i$  refers to any point in the source point cloud. If the point in the source point cloud matches the point in the target point cloud, which means  $i = j$ , then  $p(x_i) = 1$ . Otherwise,  $p(x_i) = 0$ .

At the same time,  $q$  is defined as the predicted probability distribution, which represents the probability that the source point cloud matches the target point cloud. It can be written as:

$$q(x_i) = \frac{e^{x_i}}{\sum_{j=1}^n e^{x_j}}, \quad (2)$$

where  $n$  represents the number of points in the point cloud.

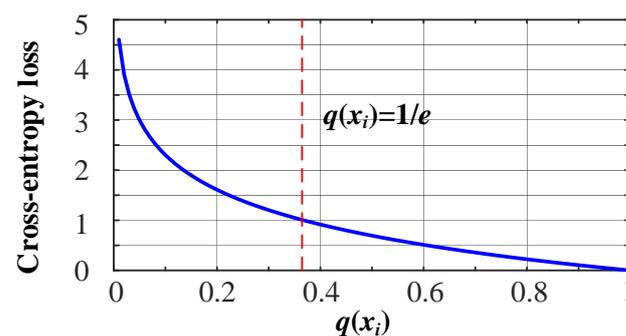
Then the cross-entropy  $H(p, q)$  is defined as:

$$H(p, q) = - \sum_{i=1}^n p(x_i) \log(q(x_i)), \quad (3)$$

Substituting (1) and (2) to (3), the final cross-entropy loss for  $p(x_i) = 1$  can be simplified as

$$H(p, q) = - \log(q(x_i)). \quad (4)$$

Figure 2 shows the curve of simplified cross-entropy loss. We can see that  $q < \frac{1}{e}$  when  $H(p, q) > 1$ . Moreover, the larger the  $q$ , the smaller the cross-entropy  $H(p, q)$  is. It indicates that the difference between the true value and the predicted value will be smaller with more accurate classification.



**Figure 2.** Simplified cross-entropy loss.

## 2.3. Attention Mechanism

In recent years, a series of attention mechanisms have been proposed to improve the performance of the deep learning models, such as channel attention mechanism (CAM) [26], spatial attention mechanism (SAM) [3], mixed attention [27,28], etc. The attention mechanism focuses on the important part of the obtained information. For example, the channel attention mechanism focuses on the correlation between different channels. It captures and reinforces the important features on each channel. The spatial attention mechanism focuses on the areas associated with the task. More succinctly, channel attention mechanisms focus on what is important and spatial attention mechanisms focus on where is important. It is

found in [29] that channel-related information is more important than spatial information when embedding the attention module of point cloud features. Therefore, in this paper, we mainly improve the model performance by adding the channel attention mechanism—see Figure 3.

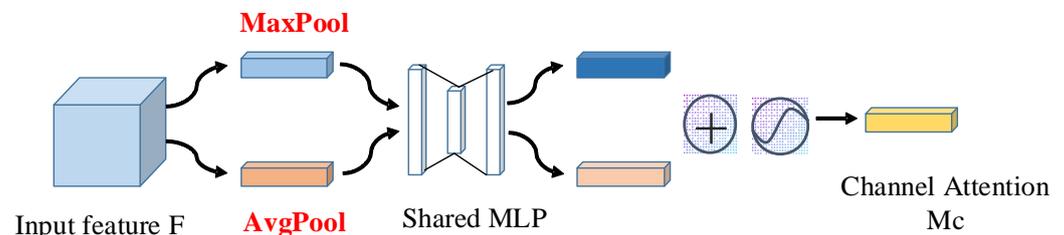


Figure 3. Channel attention module.

### 3. Proposed Method

In this section, we improve the ROPNet by introducing the revised cross-entropy and the channel attention mechanism.

#### 3.1. Modification of Loss Function

In most overlap-based point cloud registration models including ROPNet, the overlap loss is defined as:

$$\begin{aligned} loss_0 &= H_1(p_1, q_1) + H_2(p_2, q_2) \\ &= \frac{1}{2N} \sum_i \sum_j p_1 \cdot \log(q_1) + \frac{1}{2M} \sum_i \sum_j p_2 \cdot \log(q_2), \end{aligned} \quad (5)$$

where  $H_1(p_1, q_1)$ ,  $H_2(p_2, q_2)$  are the overlap supervisory losses of the two point clouds.  $N$  and  $M$  are the number of points in the source and target point clouds.  $p_1, p_2$  are the true distribution of the source and target point clouds, and  $q_1, q_2$  are the predicted probability distribution of the source and target point clouds, respectively.

Next, we modify the loss function to reduce the cross-entropy. Since the overlap between the source and target point clouds is high in most point cloud registration tasks, we mainly consider the case when  $\frac{1}{e} \leq q \leq 1$ . We can see from Figure 2 that  $0 \leq H(p, q) \leq 1$  when  $\frac{1}{e} \leq q \leq 1$ . According to the square sum inequality, that is, for any  $a, b \in R$ ,  $(a + b)^2 \geq 2ab$ , we propose a new loss function shown by Definition 1.

**Definition 1.** An improved loss function with cross-entropy is defined as:

$$\begin{aligned} loss_1 &= H_1(p_1, q_1) \times H_2(p_2, q_2) \\ &= \frac{1}{2N} \sum_i \sum_j p_1 \cdot \log(q_1) \times \frac{1}{2M} \sum_i \sum_j p_2 \cdot \log(q_2), \end{aligned} \quad (6)$$

We prove that the proposed loss function (6) performs better than the original one (5) by comparing the absolute value of the derivative of  $loss_0$  and  $loss_1$  as follows.

Denote  $H_k(p_k, q_k)$  as any one of the cross-entropy loss, where  $k = 1, 2$ . According to (4), the cross-entropy loss can be expressed as

$$H_k(p_k, q_k) = -\log(q_k(x_i)). \quad (7)$$

Take the derivative of  $H_k(p_k, q_k)$ . It can be written as

$$\begin{aligned} \frac{\partial H_k(p_k, q_k)}{\partial x_j} &= \frac{\partial H_k(p_k, q_k)}{\partial q_k(x_i)} \times \frac{\partial q_k(x_i)}{\partial x_i} = \frac{\partial -\log(q_k(x_i))}{\partial q_k(x_i)} \times \frac{\partial \frac{e^{x_i}}{\sum_{j=1}^n e^{x_j}}}{\partial x_i} \\ &= -\frac{1}{q_k(x_i)} \times \frac{\frac{\partial e^{x_i}}{\partial x_j} \times \sum_{j=1}^n e^{x_j} - e^{x_i} \times e^{x_j}}{(\sum_{j=1}^n e^{x_j})^2} \\ &= -\frac{\sum_{j=1}^n e^{x_j}}{e^{x_i}} \times \frac{\frac{\partial e^{x_i}}{\partial x_j} \times \sum_{j=1}^n e^{x_j} - e^{x_i} \times e^{x_j}}{(\sum_{j=1}^n e^{x_j})^2}. \end{aligned} \quad (8)$$

When  $i = j$ ,  $\partial e^{x_i} / \partial x_j = \partial e^{x_j} / \partial x_j$ . (7) can be written as

$$\begin{aligned} \frac{\partial H_k(p_k, q_k)}{\partial x_j} &= -\left(1 - \frac{e^{x_j}}{\sum_{j=1}^n e^{x_j}}\right) \\ &= -(1 - q_k(x_j)) \\ &= -1 + q_k(x_j). \end{aligned} \quad (9)$$

When  $i \neq j$ ,  $\partial e^{x_i} / \partial x_j = 0$ . The derivation of  $H_k(p_k, q_k)$  with  $i \neq j$  can be written as

$$\frac{\partial H_k(p_k, q_k)}{\partial x_j} = -\frac{e^{x_j}}{\sum_{j=1}^n e^{x_j}} = q_k(x_j). \quad (10)$$

Combining (7) and (9), the derivation of the loss function in different conditions can be expressed as

$$\frac{\partial H_k(p_k, q_k)}{\partial x_j} = \begin{cases} q_k(x_j) & i \neq j \\ -1 + q_k(x_j) & \text{if } i=j \end{cases} \quad (11)$$

Then, to compare the values of the original and proposed loss functions, we compute the derivation of  $loss_0$  and  $loss_1$ . We denote  $y_0$  and  $y_1$  as the derivation of  $loss_0$  and  $loss_1$ , respectively. Substituting (1) and (11),  $y_0$  and  $y_1$  can be calculated as

$$y_0 = -1 + q_1(x_j) - 1 + q_2(x_j) \quad (12)$$

$$y_1 = (-1 + q_1(x_j))H_2(p_2, q_2) + H_1(p_1, q_1)(-1 + q_2(x_j)) \quad (13)$$

According to Figure 2, when  $\frac{1}{e} \leq q \leq 1$ , it has  $0 \leq H(p, q) = -\log(q(x_i)) \leq 1$ . Since  $0 \leq q_{1j}, q_{2j} \leq 1$ , we have  $y_0 \leq y_1 \leq 0$ . Therefore,  $|y_1| \leq |y_0|$ .

The above analysis shows that the absolute value of the derivative of our proposed loss function is smaller than that of the original loss function. It means that the proposed loss function has a better performance. Moreover, since  $q \geq \frac{1}{e}$  in most point cloud alignment tasks, the proposed loss function can be generalized in more tasks based on overlapping point regions.

### 3.2. Channel Attention-Based ROPNet

Figure 4a shows the original structure of TMFR module, which contains offset-attention blocks. The offset attention block is used as the self-attention to generate global features. The TMFR module fuses the multi-scale features obtained from multiple attention modules. However, the original self-attention module only extracts global features, but ignores the local features. Hence, to make the features more complete, we add a channel attention module after each of the original attention module. As a matter of fact, there are a lot of ways to add the channel attention model. Empirically, we use a hybrid attention mechanism as shown in Figure 4b. The improvement of the proposed model with the attention mechanism can be demonstrated by experiments presented in Section 4.

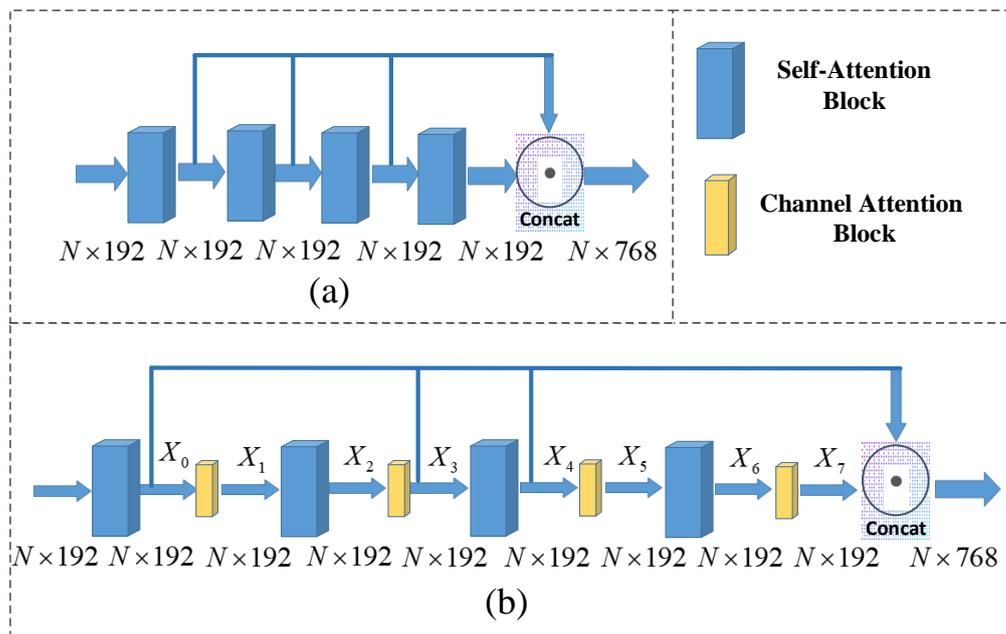


Figure 4. (a) Initial TMFR structure; (b) improved structure.

3.3. Overall Improved Architecture

An overall architecture of the improved ROPNet is presented in Figure 5. Source' represents the point cloud after the initial transformation of the source point cloud.  $T_0$  is the initial transformation matrix,  $T_1$  is the final transformation matrix. The improvement has two aspects. One is that we added a channel attention mechanism in the feature extraction part to focus more on the local features. The other is that an improved loss function presented in Section 3.2 is used after obtaining the true distribution  $p_1$  and  $p_2$ .

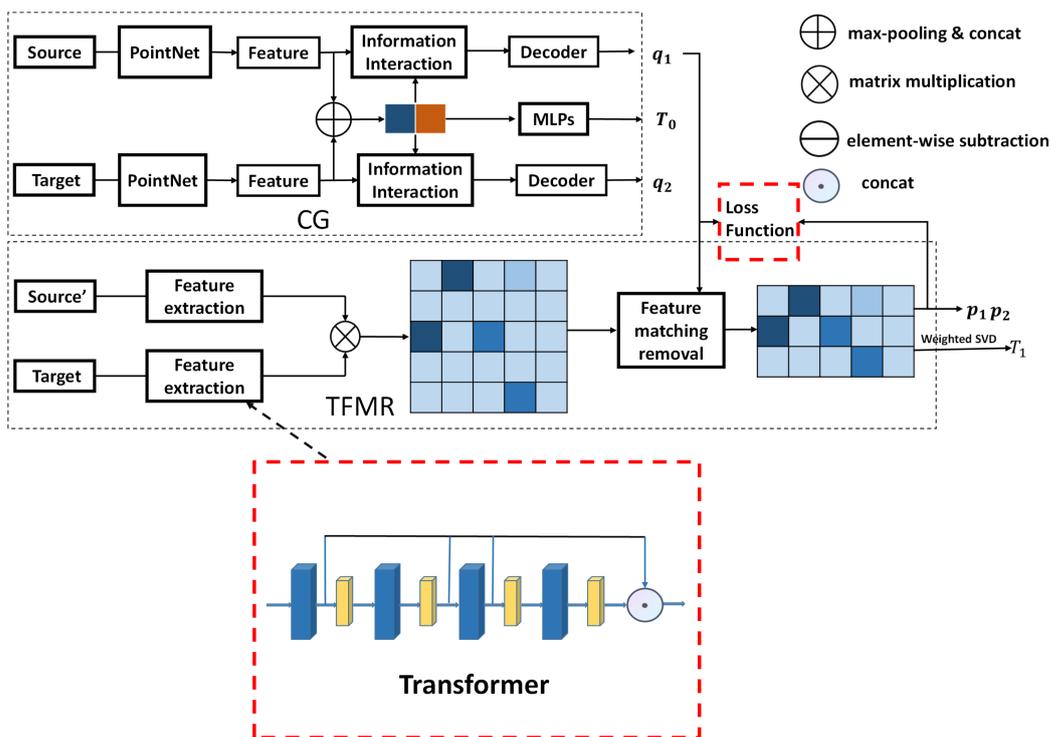


Figure 5. An overall architecture of the improved ROPNet.

## 4. Experiments

We conducted a series of experiments to demonstrate the effectiveness and advantage of our proposed method. Firstly, we introduce the dataset and the evaluation metrics. Secondly, we compared our proposed loss function with the original one by experiments. Thirdly, we conducted the ablation experiment to show the effectiveness of our proposed model with attention mechanism. Finally, we use both the improved model and the improved loss function to show the advantage of our proposed method. We keep the same experimental settings, parameters, and evaluation indicators with [15].

### 4.1. Dataset and Evaluation Metrics

As with [15], this paper used dataset ModelNet40 [30], including 40 categories and 12311 CAD models. We trained the model in the first 20 categories and tested in the remaining 20. Since there are some symmetrical categories in ModelNet40, we divided the data into two categories. One is the total objects (TO), and the other is the asymmetric objects (AO). No symmetrical objects, such as tent or vase, exist in AO. According to RPMNet [15], we generated the source and target point clouds by sampling twice to take 1024 points. We randomly generated three Euler angles within  $[0^\circ, 45^\circ]$  and translations within  $[-0.5, 0.5]$  on each axis. Then, we projected all points to a random direction and deleted thirty percent of them to generate a partial point cloud. In addition, the model is trained on noisy data. The noise is sampled from  $N(0, 0.01^2)$  and clipped to  $[-0.05, 0.05]$ .

In the following experiments, we set the average isotropic rotation and translation error as the evaluation metrics, which are

$$Error(R) = \arccos \frac{\text{tr}(\hat{R}^{-1}R)}{2} \quad (14)$$

$$Error(t) = \left\| \hat{R}^{-1}t - \hat{t} \right\|_1. \quad (15)$$

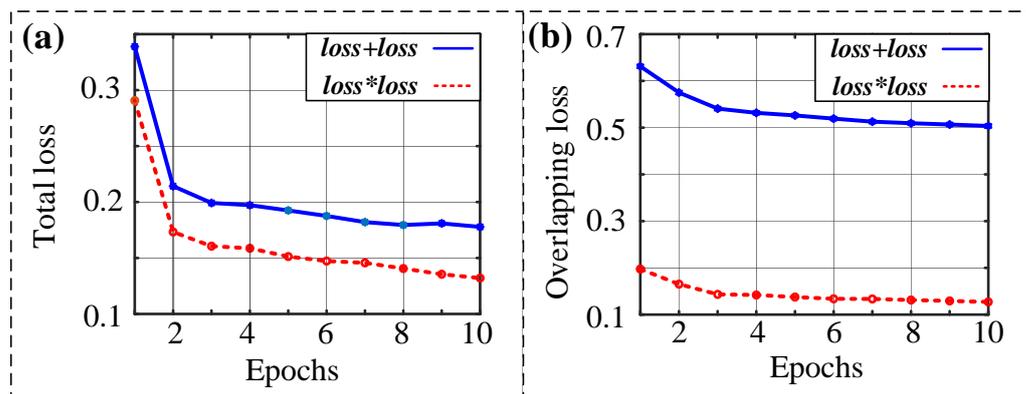
$\hat{R}$  and  $\hat{t}$  are the ground truth transformation, while  $R$  and  $t$  are the predicted transformation. The mean absolute error  $MAE(R)$  is used for the rotation matrix and  $MAE(t)$  is used for the translation vector.

### 4.2. Improvement by New Loss Function

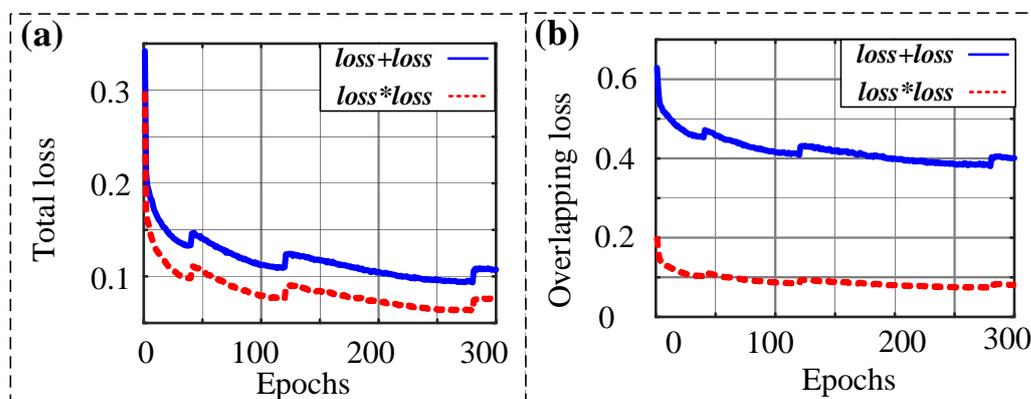
In this subsection, we compare the results obtained from different loss functions. In the following tables and legends of the figures, denote  $loss^*$  as the proposed loss function, and  $loss +$  as the original loss function. Denote  $0^*$  as the loss functions without considering the overlapping region.

First, let us observe the cross-entropy loss convergence of different loss functions under 10 epochs. Figure 6a depicts the total loss function of the two methods. We can see that the total loss of the proposed loss function is smaller than the total loss of the original one. In addition, the starting point of the decline of the proposed loss function is lower than the original starting point. Figure 6b shows the overlapping loss calculated by different loss functions. It can be seen that both curves decrease with the increase in epochs. The trend of the two curves is consistent. More importantly, the product of cross-entropy loss, i.e.,  $loss_1$ , is smaller than the sum of cross-entropy loss functions, i.e.,  $loss_0$ . The above results illustrate that the proposed loss function has a better performance.

In order to obtain more reliable results, we calculate different loss functions after training for 300 epochs. The convergence of the total loss and the overlapping loss are shown in Figure 7a,b. The results show that the two loss functions have the same trend and the proposed loss function is smoother and smaller than the original one.



**Figure 6.** Comparison of loss function decline in the case of 10 epochs: (a) total loss; (b) overlapping loss.



**Figure 7.** Comparison of loss function decline in the case of 300 epochs: (a) total loss; (b) overlapping loss.

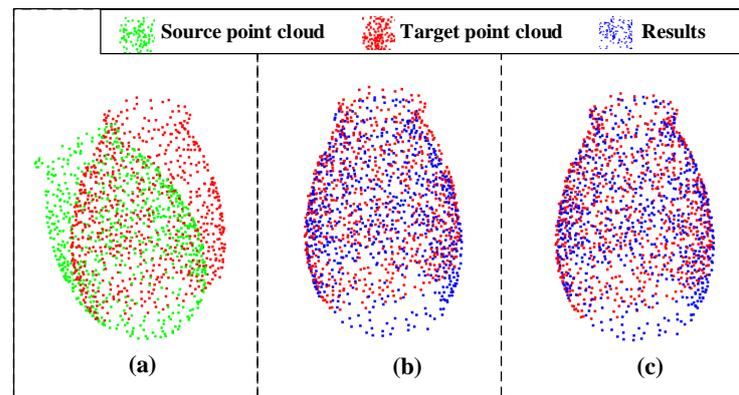
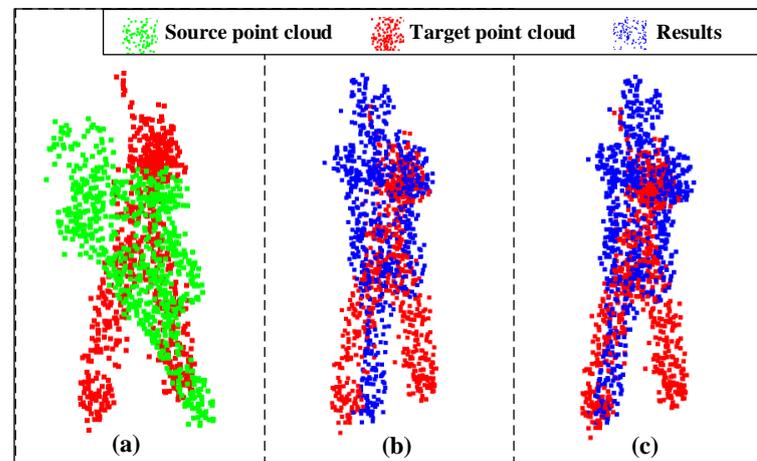
Next, we conducted two experiments to show the evaluation metrics of different loss functions. Table 1 shows the results of the evaluation metrics in the original and proposed loss functions after training for 300 epochs, and Table 2 shows the results of the evaluation metrics after training for 600 epochs. It is important to note that our table is divided into two parts: the top half being the results obtained in the noiseless condition and the bottom half in the noisy condition. We can see that in both tables, most of the metrics of the proposed loss function are smaller than the metrics of the original loss function, which indicates that our proposed loss function performs better than the original one. Furthermore, we found that in most cases, the larger the original value of the metrics, the better error reduction is. It is also worth noticing that comparing with the loss function without considering the overlapping areas, the proposed loss function improves the accuracy of the registration significantly. Taking the results training for 300 epochs as an example, the analysis results can be tested via alignment visualization results for some data, as shown in Figures 8 and 9. From these figures, we can see that our proposed method improves the alignment effects.

**Table 1.** Results of metrics with different loss functions training for 300 epochs.

	AO				TO			
	Error(R)	Error(t)	MAE(R)	MAE(t)	Error(R)	Error(t)	MAE(R)	MAE(t)
noiseless								
loss*loss	<b>1.2994</b>	<b>0.0138</b>	<b>0.6186</b>	<b>0.0066</b>	<b>1.6687</b>	<b>0.0146</b>	<b>0.7634</b>	<b>0.0069</b>
loss+loss	1.3926	0.0141	0.6287	0.0067	1.7064	0.0151	0.7951	0.0072
noise								
loss*loss	<b>1.47</b>	<b>0.016</b>	<b>0.7378</b>	<b>0.0076</b>	<b>1.8856</b>	<b>0.0178</b>	<b>0.9303</b>	<b>0.0085</b>
loss+loss	1.5239	0.0163	0.7516	0.0078	1.9385	0.0179	0.9606	0.0086

**Table 2.** Results of metrics with different loss functions training for 600 epochs.

	AO				TO			
	Error(R)	Error(t)	MAE(R)	MAE(t)	Error(R)	Error(t)	MAE(R)	MAE(t)
noiseless								
loss+loss	1.191	0.0113	0.5089	0.0053	1.4868	0.0127	0.6514	0.006
0*loss	2.8917	0.0311	1.5149	0.0148	3.2228	0.0327	1.6547	0.0156
loss*loss	<b>1.1819</b>	<b>0.011</b>	<b>0.5001</b>	<b>0.0052</b>	<b>1.4255</b>	<b>0.0122</b>	<b>0.6241</b>	<b>0.0058</b>
noise								
loss+loss	1.3935	<b>0.0135</b>	0.6502	<b>0.0065</b>	1.7197	0.0154	0.8458	0.0074
0*loss	3.0078	0.0327	1.6225	0.0156	3.5492	0.0362	1.8584	0.0173
loss*loss	<b>1.3308</b>	0.0136	<b>0.6305</b>	0.0066	<b>1.6164</b>	<b>0.0145</b>	<b>0.7754</b>	<b>0.0070</b>

**Figure 8.** Registration visualization under 300 epochs: (a) input; (b) ROPNet; (c) ours.**Figure 9.** Registration visualization under 300 epochs with Gaussian noise: (a) input; (b) ROPNet; (c) ours.

### 4.3. Ablation Experiment

In this subsection, we compare the results of the original ROPNet model and the ROPNet model with channel attention mechanism. Then, we analyze the results obtained from our proposed solution.

Table 3 shows the results of the evaluation metrics in the original and proposed ROPNet model. Compared with the original ROPNet, the results show that most of the metrics, especially the metrics in Schemes 1 and 3 are smaller than the original one. The larger the original value of metrics, the better the performance of the proposed model.

**Table 3.** Results of metrics with different models.

	AO				TO			
	Error(R)	Error(t)	MAE(R)	MAE(t)	Error(R)	Error(t)	MAE(R)	MAE(t)
noiseless								
Original result	1.191	0.0113	0.5089	0.0053	1.4868	0.0127	0.6514	0.006
Scheme 1	<b>1.1754</b>	<b>0.0101</b>	<b>0.4759</b>	<b>0.0047</b>	1.4637	0.0121	0.6508	0.0057
Scheme 2	1.233	0.0113	0.5304	0.0054	1.4461	0.0123	0.6536	0.0058
Scheme 3	1.2215	0.0114	0.5057	0.0053	<b>1.3976</b>	<b>0.0117</b>	<b>0.6215</b>	<b>0.0056</b>
noise								
Original result	1.3935	0.0135	0.6502	0.0065	1.7197	0.0154	0.8458	0.0074
Scheme 1	<b>1.2797</b>	<b>0.0128</b>	<b>0.6055</b>	<b>0.0062</b>	1.7157	0.0149	0.8235	0.0072
Scheme 2	1.3741	0.0133	0.6389	0.0065	1.6765	0.0152	0.8302	0.0073
Scheme 3	1.3357	0.0136	0.6162	0.0066	<b>1.5287</b>	<b>0.0141</b>	<b>0.7399</b>	<b>0.0068</b>

Next, we discuss the results when selecting different positions to fuse the features. As mentioned in Section 3.2, there are a lot of ways to add the channel attention model. Figure 10 illustrates three fusion schemes with different positions. In specific, Figure 10a shows the stepwise fusion method, Figure 10b shows the feature extracted through the channel attention module, and Figure 10c shows a mixed method. Table 3 depicts the results of feature fusion with different positions. It can be seen that Scheme 1 and Scheme 3 has better performance than Scheme 2, because Scheme 2 ignores the global information. Moreover, Scheme 1 fuses multiple layers of feature and leads to the feature redundancy. So, we choose Scheme 3 as the feature fusion method.

### 4.4. Combination of Two Improvements

In this subsection, we combine the proposed loss function with the improved model to show the effectiveness of our proposed method on ModelNet40.

Table 4 shows the results of evaluation metrics in different methods. We can see that the results with the improved loss function and improved model has the best performance comparing with others. When comparing the metric results in AO utilizing both the improved loss function and model with the original metric results, we can see from Table 4 that Error(R) and Error(t) with no noise are reduced by 4.83% and 14.16%. Error(R) and Error(t) with noise are reduced by 9.60% and 8.89%. When comparing the metric results utilizing both the improved loss function and model with the results only utilizing the improved loss function, Error(R) and Error(t) with no noise are reduced by 4.10% and 11.82%. The Error(R) and Error(t) with noise are reduced by 5.34% and 9.56%. When comparing the metric results utilizing both the improved loss function and model with the results only utilizing the improved model, Error(R) and Error(t) with no noise are reduced by 7.20% and 20.18%. The Error(R) and Error(t) with noise are reduced by 5.69% and 9.56%. The results show that the proposed method with the improved loss function and the improved model is effective in improving the performance of the model.

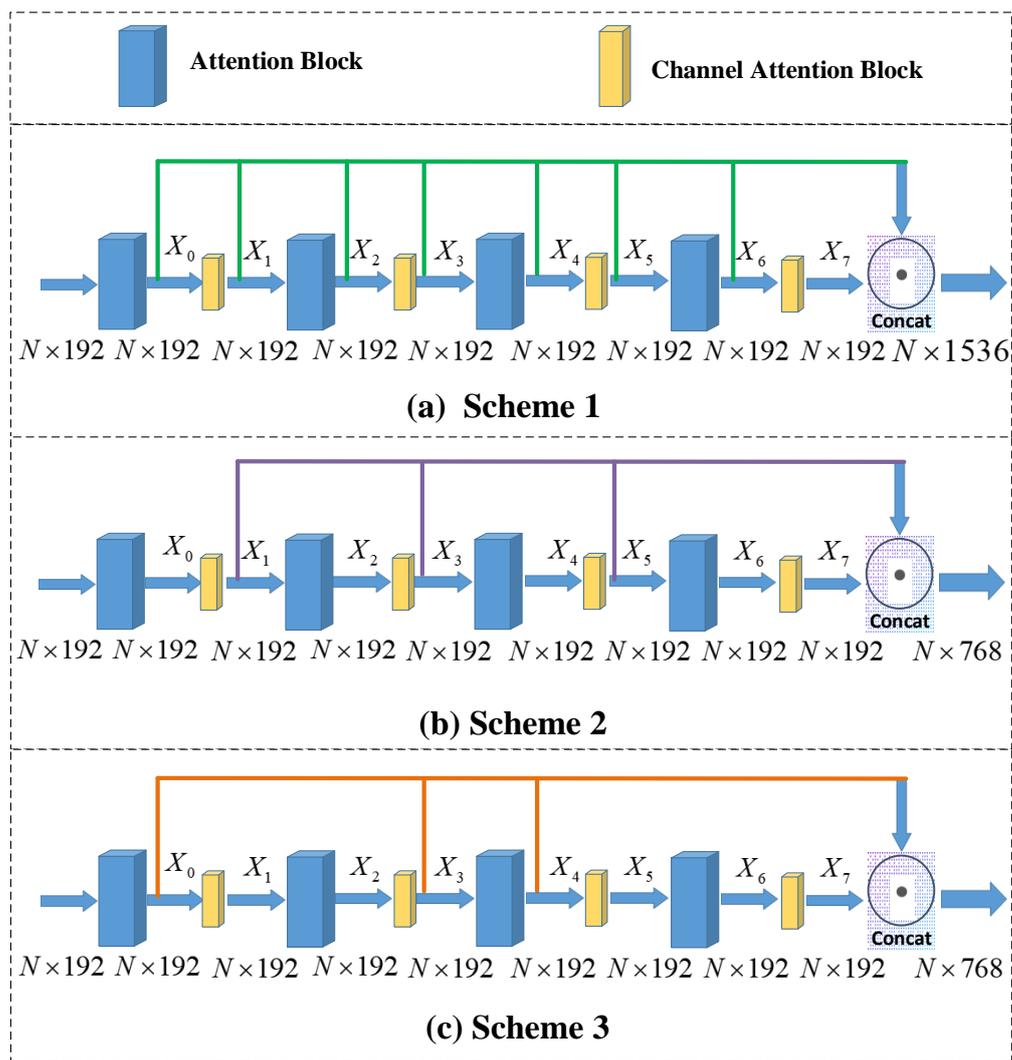


Figure 10. Comparison of different schemes: (a) Scheme 1 concat[x0, x1, x2, x3, x4, x5, x6, x7]; (b) Scheme 2 concat[x1, x3, x5, x7]; (c) Scheme 3 concat[x0, x3, x4, x7].

Table 4. Overall improved results on ModelNet40.

	AO				TO			
	Error(R)	Error(t)	MAE(R)	MAE(t)	Error(R)	Error(t)	MAE(R)	MAE(t)
noiseless								
Original result	1.191	0.0113	0.5089	0.0053	1.4868	0.0127	0.6514	0.006
loss*loss	1.1819	0.011	0.5001	0.0052	1.4255	0.0122	0.6241	0.0058
Scheme 3	1.2215	0.0114	0.5057	0.0053	1.3976	0.0117	0.6215	0.0056
Final result	<b>1.1335</b>	<b>0.0097</b>	<b>0.4717</b>	<b>0.0046</b>	<b>1.3855</b>	<b>0.011</b>	<b>0.5749</b>	<b>0.0052</b>
noise								
Original result	1.3935	0.0135	0.6502	0.0065	1.7197	0.0154	0.8458	0.0074
loss*loss	1.3308	0.0136	0.6305	0.0066	1.6164	0.0145	0.7754	0.0070
Scheme 3	1.3357	0.0136	0.6162	0.0066	<b>1.5287</b>	0.0141	<b>0.7399</b>	0.0068
Final result	<b>1.2597</b>	<b>0.0123</b>	<b>0.5953</b>	<b>0.0059</b>	1.5916	<b>0.0137</b>	0.7467	<b>0.0066</b>

Figures 11 and 12 show the comparison results of ICP, RPMNet, ROPNet, and the proposed model in the form of bar charts. We can see that our method is superior to both RPMNet and the original ROPNet. When comparing the metric results with RPMNet, Error(R) and Error(t) with no noise are reduced by 17.44% and 34.91%. The Error(R) and

Error(t) with noise are reduced by 9.41% and 22.60%, respectively. Comparing with ICP, our method performs better, the Error(R) and Error(t) is reduced by a factor of more than ten.

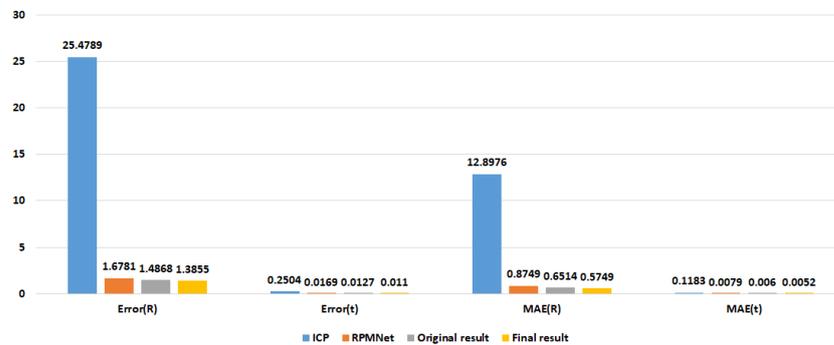


Figure 11. Registration results on ModelNet40.

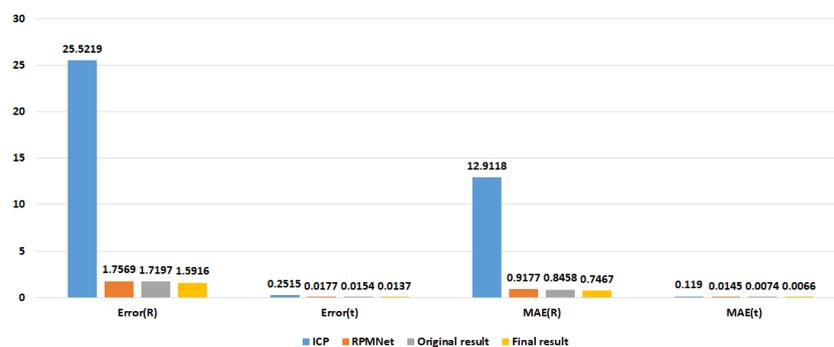


Figure 12. Registration results on ModelNet40 with Gaussian noise.

We also tested our method on the Stanford Bunny dataset. The Stanford Bunny dataset is composed of 35,947 points. We rotated the Bunny model by 30 degrees around the y-axis. Then, 1500 points are selected from the original point cloud and the point cloud after rotation as the source and target point clouds. The final result is shown in Figure 13.

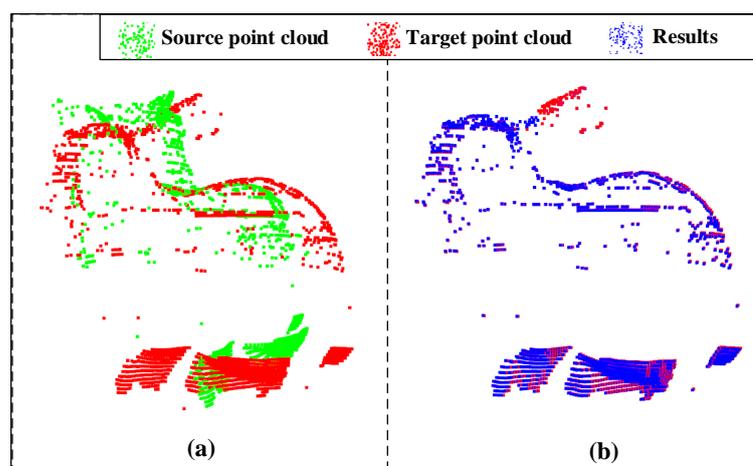


Figure 13. Registration results of Stanford bunny model: (a) input; (b) result error (R): 0.3403; Error(t): 0.0219.

### 5. Conclusions

In this paper, we designed a new way of using the cross-entropy loss, which integrates two cross-entropy losses by the product instead of sum. To verify the effectiveness of this

loss, we applied it to the point cloud registration network ROPNet. In addition, we improve the original ROPNet model by introducing the channel attention mechanism, which focus on the local features. A large number of experiments show that the proposed method improves the accuracy and efficiency of the point cloud registration task. Compared to the original ROPNet, the Error(R) and Error(t) with noise are reduced by 4.83% and 14.16%. Compared to the RPMNet, the Error(R) and Error(t) with noise are reduced by 9.41% and 22.60%. Compared to other traditional methods such as ICP the Error(R) and Error(t) are even reduced by more than ten times. In this way, the proposed method is expected to contribute to more point cloud registration models.

**Author Contributions:** Conceptualization, F.Y. and W.Z.; methodology, Y.L., F.Y. and W.Z.; software, validation, investigation, data curation, writing—original draft preparation, and visualization, Y.L.; writing—review and editing, F.Y. and W.Z.; supervision, W.Z.; funding acquisition, F.Y., and W.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Natural Science Foundation of China under Grant 62101392, and in part by the Stability Module trial production Project of the State Grid Hubei Power Research Institute under Grant 2022420612000585.

**Institutional Review Board Statement:** Not applicable

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Izadi, S.; Kim, D.; Hilliges, O.; Molyneaux, D.; Newcombe, R.; Kohli, P.; Shotton, J.; Hodges, S.; Freeman, D.; Davison, A.; et al. KinectFusion: Real-time 3D reconstruction and interaction using a moving depth camera. In Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology, Santa Barbara, CA, USA, 16–19 October 2011; pp. 559–568.
2. Huang, X.; Mei, G.; Zhang, J.; Abbas, R. A comprehensive survey on point cloud registration. *arXiv* **2021**, arXiv:2103.02690.
3. Elhousni, M.; Huang, X. Review on 3d lidar localization for autonomous driving cars. *arXiv* **2020**, arXiv:2006.00648.
4. Nagy, B.; Benedek, C. Real-time point cloud alignment for vehicle localization in a high resolution 3D map. In Proceedings of the 15th European Conference on Computer Vision (ECCV) Workshops, Munich, Germany, 8–14 September 2018.
5. Zhang, A.; Min, Z.; Zhang, Z.; Meng, M.Q.-H. Generalized point set registration with fuzzy correspondences based on variational Bayesian Inference. *IEEE Trans. Fuzzy Syst.* **2022**, *30*, 1529–1540. [[CrossRef](#)]
6. Fu, Y.; Brown, N.M.; Saeed, S.U.; Casamitjana, A.; Baum, Z.M.C.; Delaunay, R.; Yang, Q.; Grimwood, A.; Min, Z.; Blumberg, S.B.; et al. DeepReg: A deep learning toolkit for medical image registration. *arXiv* **2020**, arXiv:2011.02580.
7. Li, Y.; Harada, T. Leopard: Learning partial point cloud matching in rigid and deformable scenes. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 20–25 June 2022; pp. 5554–5564.
8. Min, Z.; Wang, J.; Pan, J.; Meng, M.-Q.H. Generalized 3-D point set registration with hybrid mixture models for computer-assisted orthopedic surgery: From isotropic to anisotropic positional error. *IEEE Trans. Autom. Sci. Eng.* **2020**, *18*, 1679–1691. [[CrossRef](#)]
9. Min, Z.; Liu, J.; Liu, L.; Meng, M.-Q.H. Generalized coherent point drift with multi-variate gaussian distribution and watson distribution. *IEEE Robot. Autom. Lett.* **2021**, *6*, 6749–6756. [[CrossRef](#)]
10. Besl, P.J.; McKay, N.D. Method for registration of 3-D shapes. Sensor fusion IV: Control paradigms and data structures. *Spie* **1992**, *1611*, 586–606.
11. Sarode, V.; Li, X.; Goforth, H.; Aoki, Y.; Srivatsan, R.A.; Lucey, S.; Choset, H. Pcnnet: Point cloud registration network using pointnet encoding. *arXiv* **2019**, arXiv:1908.07906.
12. Bai, X.; Luo, Z.; Zhou, L.; Fu, H.; Quan, L.; Tai, C.-L. D3feat: Joint learning of dense detection and description of 3d local features. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 6359–6367.
13. Li, J.; Zhang, C.; Xu, Z.; Zhou, H.; Zhang, C. Iterative distance-aware similarity matrix convolution with mutual-supervised point elimination for efficient point cloud registration. In Proceedings of the 16th European Conference on Computer Vision, Glasgow, UK, 23–28 August 2020; Springer: Cham, Switzerland, 2020; pp. 378–394.
14. Huang, S.; Gojcic, Z.; Usvyatsov, M.; Wieser, A.; Schindler, K. Predator: Registration of 3d point clouds with low overlap. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 4267–4276.
15. Yew, Z.J.; Lee, G.H. Rpm-net: Robust point matching using learned features. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 11824–11833.

16. Zhu, L.; Liu, D.; Lin, C.; Yan, R.; Gómez-Fernández, F.; Yang, N.; Feng, Z. Point cloud registration using representative overlapping points. *arXiv* **2021**, arXiv:2107.02583.
17. Wen, Y.; Zhang, K.; Li, Z.; Qiao, Y. A discriminative feature learning approach for deep face recognition. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; Springer: Cham, Switzerland, 2016; pp. 499–515.
18. Sun, Y.; Cheng, C.; Zhang, Y.; Zhang, C.; Zheng, L.; Wang, Z.; Wei, Y. Circle loss: A unified perspective of pair similarity optimization. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 6398–6407.
19. Li, X.; Sun, X.; Meng, Y.; Liang, J.; Wu, F.; Li, J. Dice loss for data-imbalanced NLP tasks. *arXiv* **2019**, arXiv:1911.02855.
20. Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal loss for dense object detection. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2980–2988.
21. Pihur, V.; Datta, S.; Datta, S. Weighted rank aggregation of cluster validation measures: A monte carlo cross-entropy approach. *Bioinformatics* **2007**, *23*, 1607–1615. [[CrossRef](#)] [[PubMed](#)]
22. Jadon, S. A survey of loss functions for semantic segmentation. In Proceedings of the 2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), Viña del Mar, Chile, 27–29 October 2020; pp. 1–7.
23. Guo, M.H.; Cai, J.X.; Liu, Z.N.; Mu, T.J.; Martin, R.R.; Hu, S.M. Pct: Point cloud transformer. *IEEE Comput. Vis. Media* **2021**, *7*, 187–199. [[CrossRef](#)]
24. Zhao, H.; Jiang, L.; Jia, J.; Torr, P.; Koltun, V. Point transformer. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 10–17 October 2021; pp. 16259–16268.
25. Xu, H.; Liu, S.; Wang, G.; Liu, G.; Zeng, B. Omnet: Learning overlapping mask for partial-to-partial point cloud registration. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 10–17 October 2021; pp. 3132–3141.
26. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141.
27. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. Cbam: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, March 2018; pp. 3–19.
28. Fu, J.; Liu, J.; Tian, H.; Li, Y.; Bao, Y.; Fang, Z.; Lu, H. Dual attention network for scene segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 3146–3154.
29. Qiu, S.; Wu, Y.; Anwar, S.; Li, C. Investigating attention mechanism in 3d point cloud object detection. In Proceedings of the 2021 International Conference on 3D Vision (3DV), London, UK, 1–3 December 2021; pp. 403–412.
30. Wu, Z.; Song, S.; Khosla, A.; Yu, F.; Zhang, L.; Tang, X.; Xiao, J. 3d shapenets: A deep representation for volumetric shapes. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1912–1920.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.