RGM Supplementary Material

1. More Implementation Details of RGM

1.1. Local Feature Extractor

Our local feature extractor contains five MLPs of size (64, 64, 128, 256, 512), and uses ReLU as the activation function. The output features of the first four layers are concatenated together and inputted into the fifth layer. Finally, we concatenate the features of the first four layers with those of the fifth layer to output the final high dimensional local features V.

1.2. Edge Generator Based on Transformer

The edge generator consists of transformer, concatenation and softmax. The transformer consists of several stacked encoder-decoder layers. The encoder takes node features $\mathcal{F}_{\mathbf{X}}, \mathcal{F}_{\mathbf{Y}}$ as input and encodes them into a embedding space by using multi-head attention. The number of heads in multi-head attention is 4, and the feature embedding dimension is set to 1024. The decoder associates features based on the co-attention mechanism, and ouputs the embedding feature $\mathcal{T}_{\mathbf{X}}, \mathcal{T}_{\mathbf{Y}}$. The architecture of the transformer is shown in Figure 1, where (a) and (b) represent how to obtain the embedded features $\mathcal{T}_{\mathbf{X}}$ and $\mathcal{T}_{\mathbf{Y}}$, respectively. Their inputs are different, but the weights are the same. For the softmax, we adopt column normalization.

1.3. Graph Feature Extractor and AIS Module

In graph feature extractor, we set the dimension of node self-correlation feature $\mathcal{F}_{x_i}^{corr}$ to 512, and the dimension of new node feature $\mathcal{F}'_{x_i}$ is also set to 512. In AIS module, the calculation formula of instance normalization is as follows:

$$A'_{ij} = \frac{A_{ij} - \mu}{\sqrt{\sigma^2 + \epsilon}} * \gamma + \beta, \tag{1}$$

$$\mu = \frac{1}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} A_{ij}, \tag{2}$$

$$\sigma^2 = \frac{1}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} (A_{ij} - \mu)^2, \tag{3}$$

where A' is the result obtained after instance normalization, ϵ represents a small positive value to avoid denominator 0, γ and β are learnable affine parameters. If we use softmax instead of instance normalization, there will be the phenomenon as shown in Figure 2, and we can see that the second element of the first column has a smaller value than the last element of the second column, but it becomes larger after using softmax on the two columns separately.

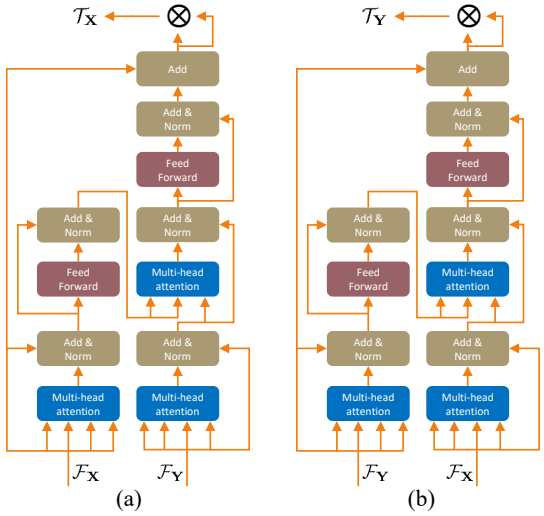


Figure 1: The architecture of the transformer, $\otimes$ represents matrix multiplication.

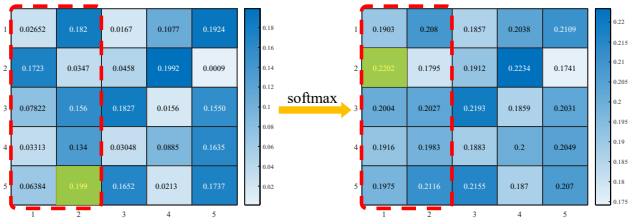


Figure 2: Columnwise softmax of the affinity matrix. Green represents the element with the largest value in the red box.

The fixed number of iterations is 20 for Sinkhorn. Because the feature dimension of graph feature extractor is 512, the size of learnable weights in affinity layer is 512x512.

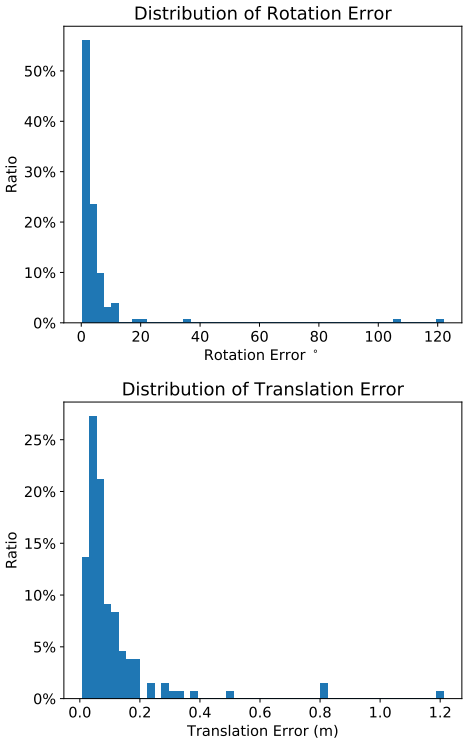


Figure 3: Experimental results on 3DMatch dataset.

Setting	method	MIE(R)	MIE(t)	MAE(R)	MAE(t)	CCD	Recall
Clean	ICP	2.866	0.02265	5.7820	0.04968	0.01986	85.56%
	FGR	0.001	0.00001	**0.0010**	0.00002	0.00005	**100.00%**
	RPM-Net	0.139	0.00093	0.2847	0.00198	0.00186	96.56%
	IDAM	7.756	0.05623	13.8014	0.11481	0.07644	51.53%
	DeepGMR	0.011	0.00007	0.0358	0.00015	0.00015	99.79%
	RGM	**<0.001**	**<0.00001**	0.0094	**<0.00001**	**<0.00001**	**100.00%**
Gaussian Noise	ICP	2.893	0.02319	5.8934	0.05097	0.05173	83.89%
	FGR	7.985	0.05075	13.6898	0.10500	0.08048	29.23%
	RPM-Net	0.429	0.00363	0.7936	0.00775	0.04111	95.41%
	IDAM	1.818	0.01565	3.4658	0.03177	0.05380	55.81%
	DeepGMR	1.065	0.00653	2.0160	0.01380	0.04885	64.37%
	RGM	**0.143**	**0.00122**	**0.2810**	**0.00257**	**0.03995**	**98.19%**
Partial to Partial	ICP	12.207	0.13438	23.8867	0.29026	0.11853	9.12%
	FGR	26.873	0.18319	47.4291	0.37820	0.13043	5.08%
	RPM-Net	2.917	0.03698	5.6309	0.08199	0.09150	49.37%
	IDAM	10.090	0.10870	19.2204	0.22631	0.13025	1.81%
	DeepGMR	46.901	0.25167	74.9521	0.51363	0.15012	0.28%
	RGM	**0.887**	**0.00791**	**1.6236**	**0.01696**	**0.08201**	**87.47%**

Table 1: Experimental results on ShapeNet dataset.

2. Experiments on ShapeNet and 3DMatch

In order to evaluation on more datasets, we did experiments on ShapeNet[1] and 3DMatch[3]. We compare the proposed RGM with ICP, FGR, RPM-Net, IDAM, and DeepGMR on ShapeNet using the same experimental protocol in Section 5.3 - 5.5. As shown in Table 1, the results are consistent with the original the paper. Second, we register point cloud pairs with more than 50% overlapping ratio in 3DMatch, in which data preprocessing is the same as in [2]. The results are shown in Fig. 3. Overall, the proposed method works well on ShapeNet and 3DMatch.

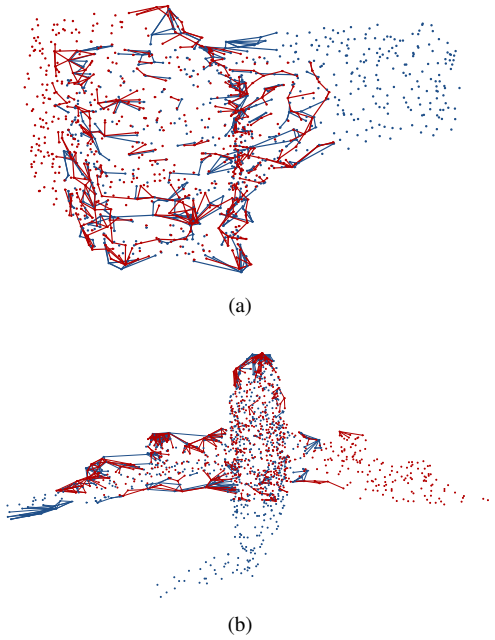


Figure 4: Learned graphs. Down-sampled points and edges with large weights are illustrated. Please note that most edges are in the overlapping region and the edges of the two graphs roughly correspond to each other. These two properties are expected for better registration.

method	ICP	FGR	RPM-Net	IDAM	DeepGMR	RGM
time(ms)	24.6	41.0	78.6	1.3	50.5	180.9

Table 2: Computational efficiency

3. Computational Efficiency

We compare the inference time of our method and the comparison methods on the partial-to-partial dataset. Please note that ICP and FGR are executed on CPU, and the other methods are executed on GPU. We performed these experiments on a 4.2GHz Intel i7-7700K and a Nvidia GTX1080. The results are shown in Table 2.

4. Visualizing the learned graph

In order to prove the importance of our edge generator, we show the learned graph in Fig. 4.

5. Accuracy of Predicted Correspondences

In order to directly show that our network can learn very accurate correspondences, we statistically analyze the distances between the predicted corresponding points and the ground-truth corresponding points for each category.

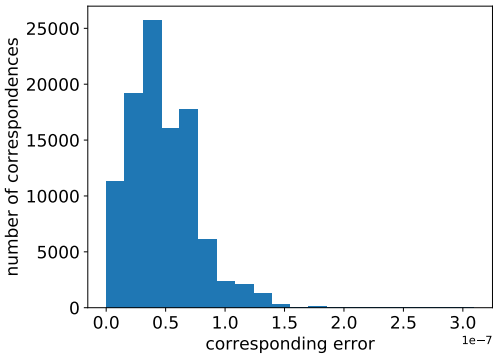


Figure 5: The histogram of corresponding errors in the clean experiment for category ”car”.

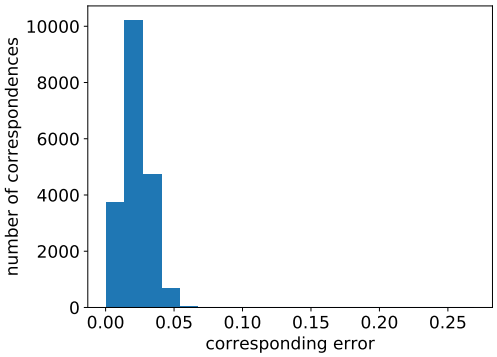


Figure 6: The histogram of corresponding errors in the noise experiment for category ”bowl”.

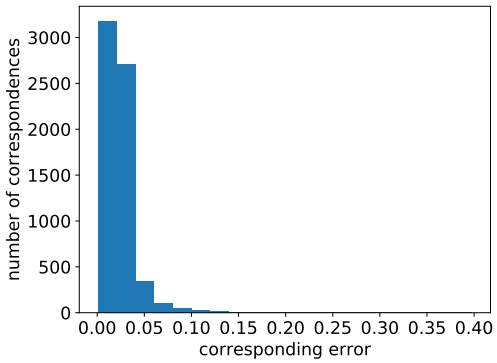


Figure 7: The histogram of corresponding errors in the partial-to-partial experiment for category ”lamp”.

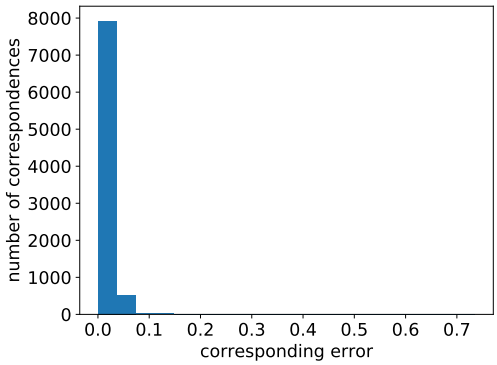


Figure 8: The histogram of corresponding errors in the unseen categories experiment for category ”tent”.

First, we use the ground-truth transformation to align the source point cloud with the target point cloud, and for each point in the source point cloud, the transformed poistion is its ground-truth corresponding point. Then for each corresponding point pair found by our method, we calculate the distance between the source point's ground-truth corresponding point and the corresponding target point in the pair, and this distance is named as corresponding error. We calculate the corresponding error for all the predicted correspondences in registering all point clouds in each category, and draw the distance distribution histogram for each category. Figure 5, Figure 6, Figure 7 and Figure 8 are the corresponding error histograms for one example category in the clean, noise, partial-to-partial and unseen categories experiments, respectively.

Figure 5 is the corresponding error histogram for category ”car” in the clean experiment. The horizontal axis represents the corresponding error, and the vertical axis represents the number of correspondences in a specific error bin. Please note that the scale of the horizontal axis is 1e-7, because in the clean experiment, all the predicted correspondences are completely correct in registration point clouds in this categroy and the corresponding error is 0 in theory. The very small corresponding error comes from the error in representing floating point number. The histogram of corresponding errors for all categories in clean experiment are shown in Figure 9 and Figure 10.

Figure 6 is the corresponding error histogram for category ”bowl” in the noise experiment. The corresponding error histograms for all categories in noise experiment are shown in Figure 11 and Figure 12. We can find that almost all the corresponding errors are less than 0.05, which is consistent with the clipped value of Gaussian noise in our experiment setting.

Figure 7 is the corresponding error histogram for category ”lamp” in the partial-to-partial experiment. Partial-to-partial is the most challenging case for point cloud registration. The corresponding error histograms for all categories in the partial-to-partial experiment are Figure 13 and Figure 14. Although this situation is more dificult, the majority of the corresponding errors are still less than 0.05.

Figure 8 is the corresponding error histogram for category ”tent” in the unseen categories experiment. Performance on unseen categroy demonstrate the generalization

capability of a learned model. The corresponding error histograms for all 20 unseen categories in this experiment are shown Figure 15. The majority of the corresponding errors are still very small.

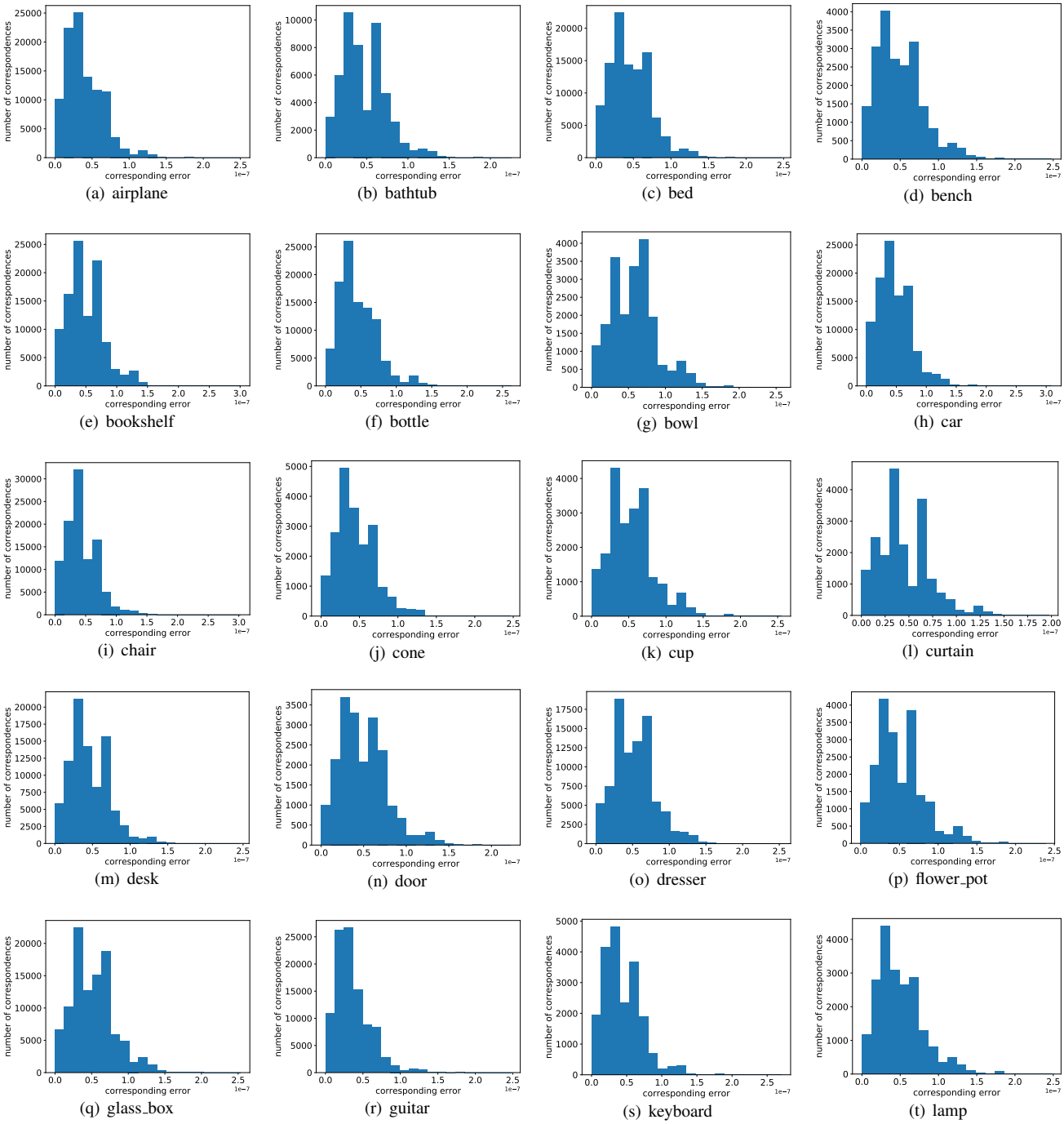
These histograms of corresponding error show that the majority of the correspondences found by using our method are very accurate. That is why high registration accuracy is achieved in the registration experiments.

References

[1] Angel X. Chang, Thomas Funkhouser, Leonidas Guibas, Pat Hanrahan, Qixing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, and Hao and Su. Shapenet: An information-rich 3d model repository. *Computer ence*, 2015. 2

[2] Z. Gojcic, C. Zhou, J. D. Wegner, L. J. Guibas, and T. Birdal. Learning multiview 3d point cloud registration. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020. 2

[3] Andy Zeng, Shuran Song, Matthias Nießner, Matthew Fisher, Jianxiong Xiao, and Thomas Funkhouser. 3dmatch: Learning local geometric descriptors from rgb-d reconstructions. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017. 2



(a) airplane (b) bathtub (c) bed (d) bench

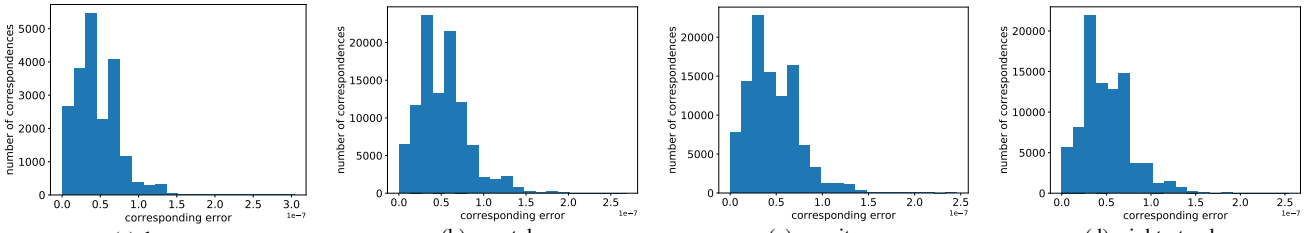
(e) bookshelf (f) bottle (g) bowl (h) car

(i) chair (j) cone (k) cup (l) curtain

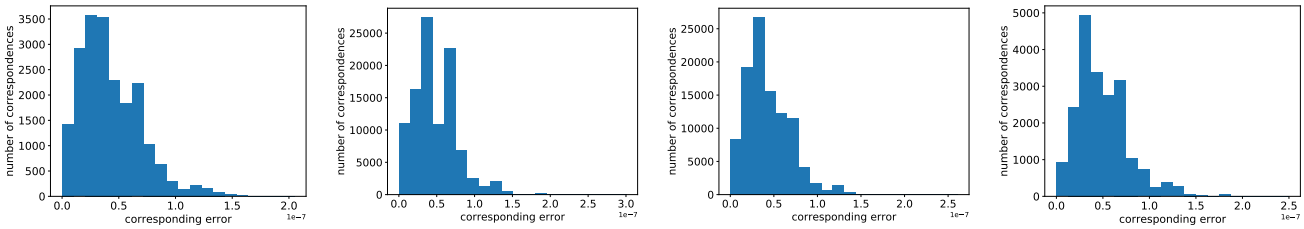
(m) desk (n) door (o) dresser (p) flower_pot

(q) glass_box (r) guitar (s) keyboard (t) lamp

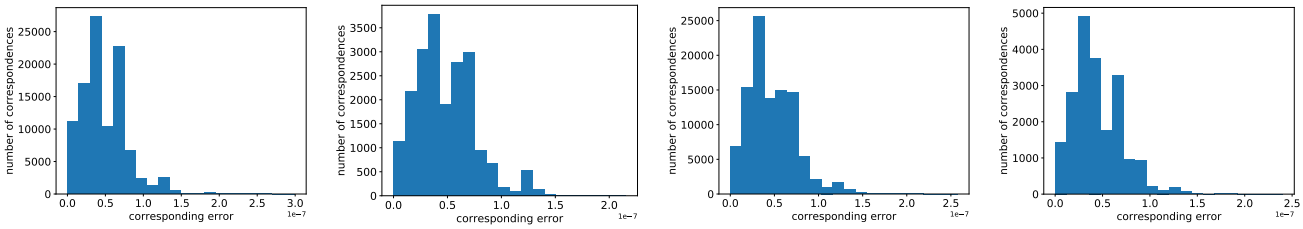
Figure 9: The histogram of the corresponding errors in the clean experiment of the first 20 categories.



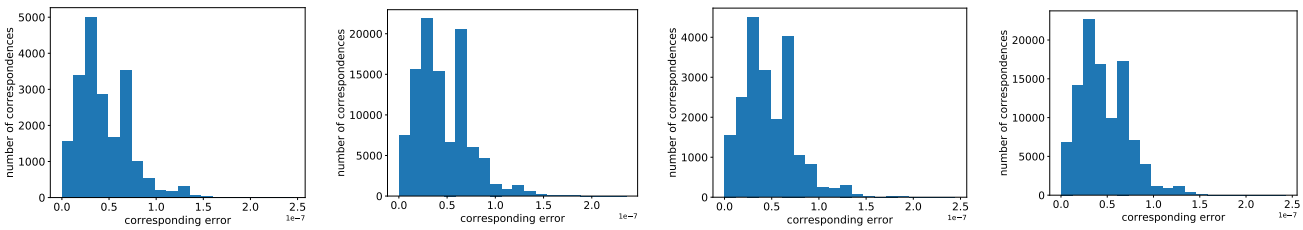
(a) laptop (b) mantel (c) monitor (d) night_stand



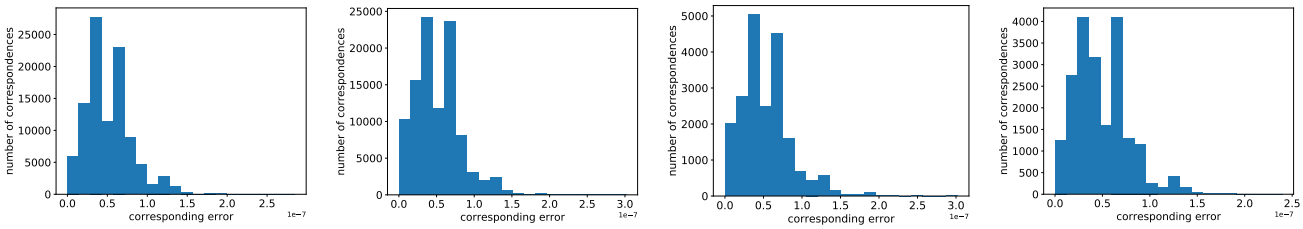
(e) person (f) piano (g) plant (h) radio



(i) range_hood (j) sink (k) sofa (l) stairs

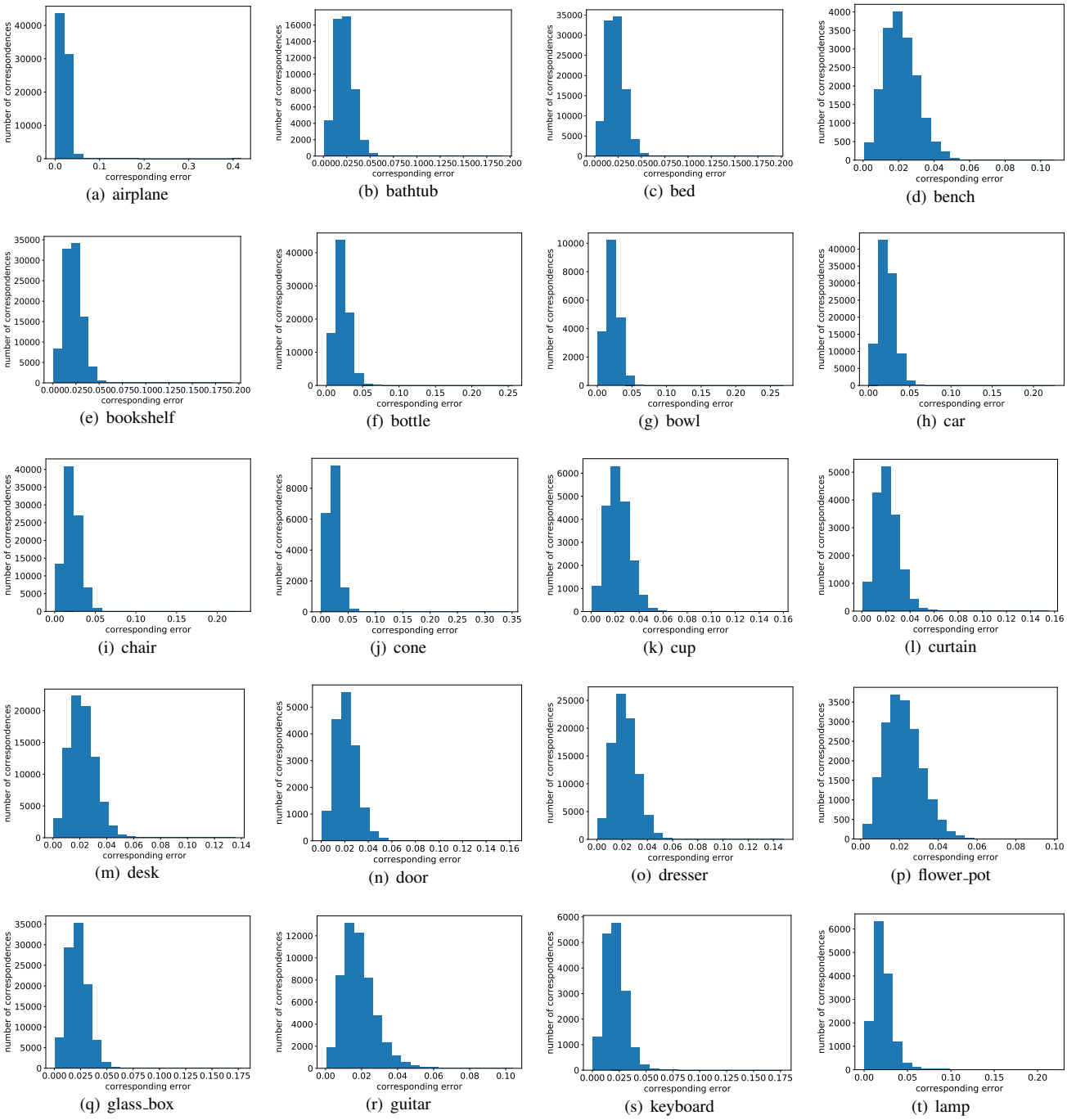


(m) stool (n) table (o) tent (p) toilet



(q) tv_stand (r) vase (s) wardrobe (t) xbox

Figure 10: The histogram of the corresponding errors in the clean experiment of the last 20 categories.



(a) airplane (b) bathtub (c) bed (d) bench

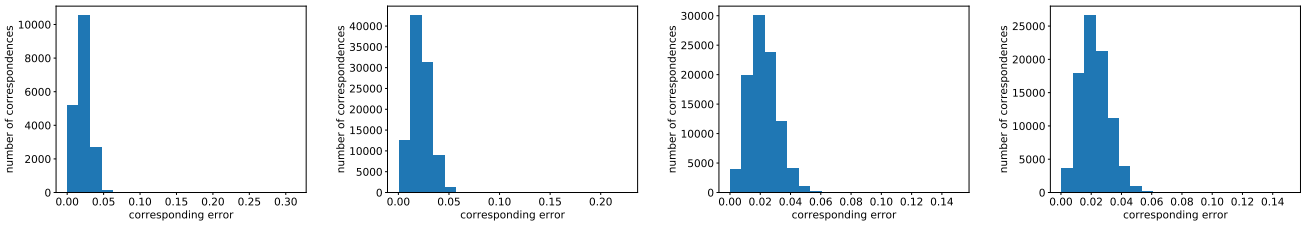
(e) bookshelf (f) bottle (g) bowl (h) car

(i) chair (j) cone (k) cup (l) curtain

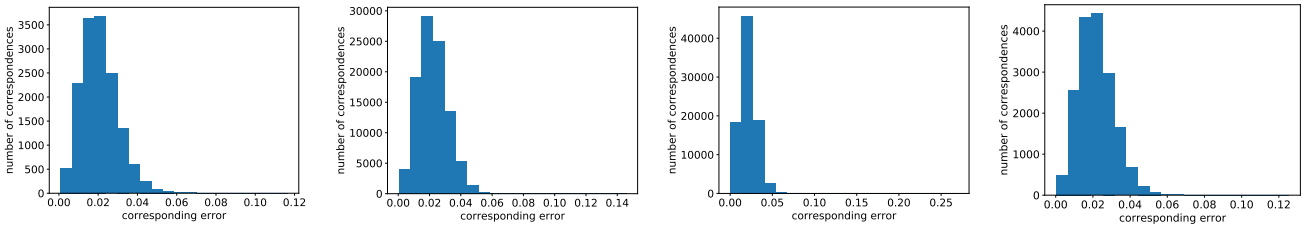
(m) desk (n) door (o) dresser (p) flower_pot

(q) glass_box (r) guitar (s) keyboard (t) lamp

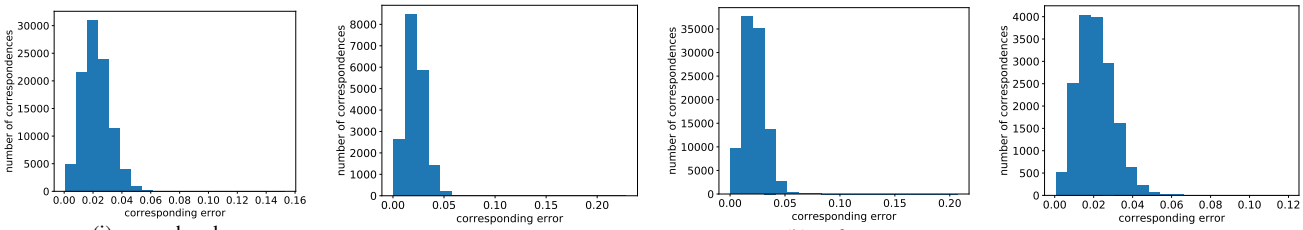
Figure 11: The histogram of the corresponding errors in the noise experiment of the first 20 categories.



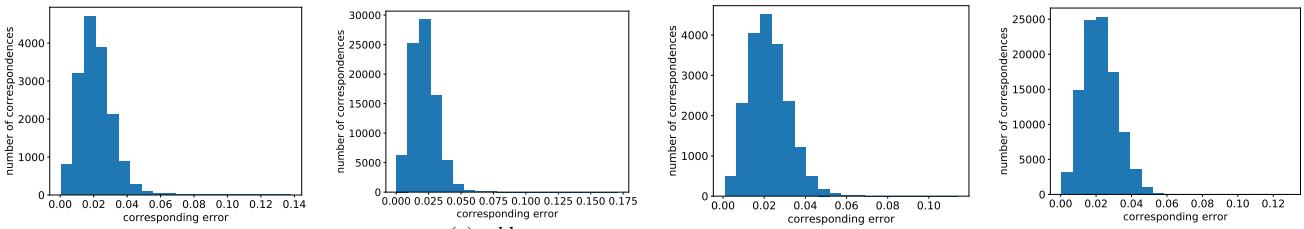
(a) laptop (b) mantel (c) monitor (d) night_stand



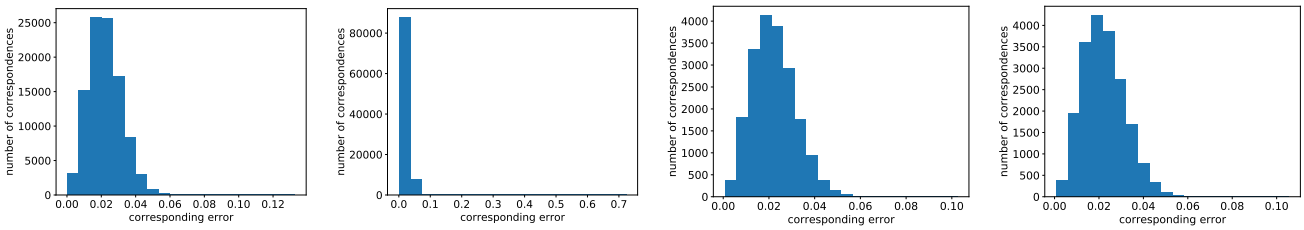
(e) person (f) piano (g) plant (h) radio



(i) range_hood (j) sink (k) sofa (l) stairs

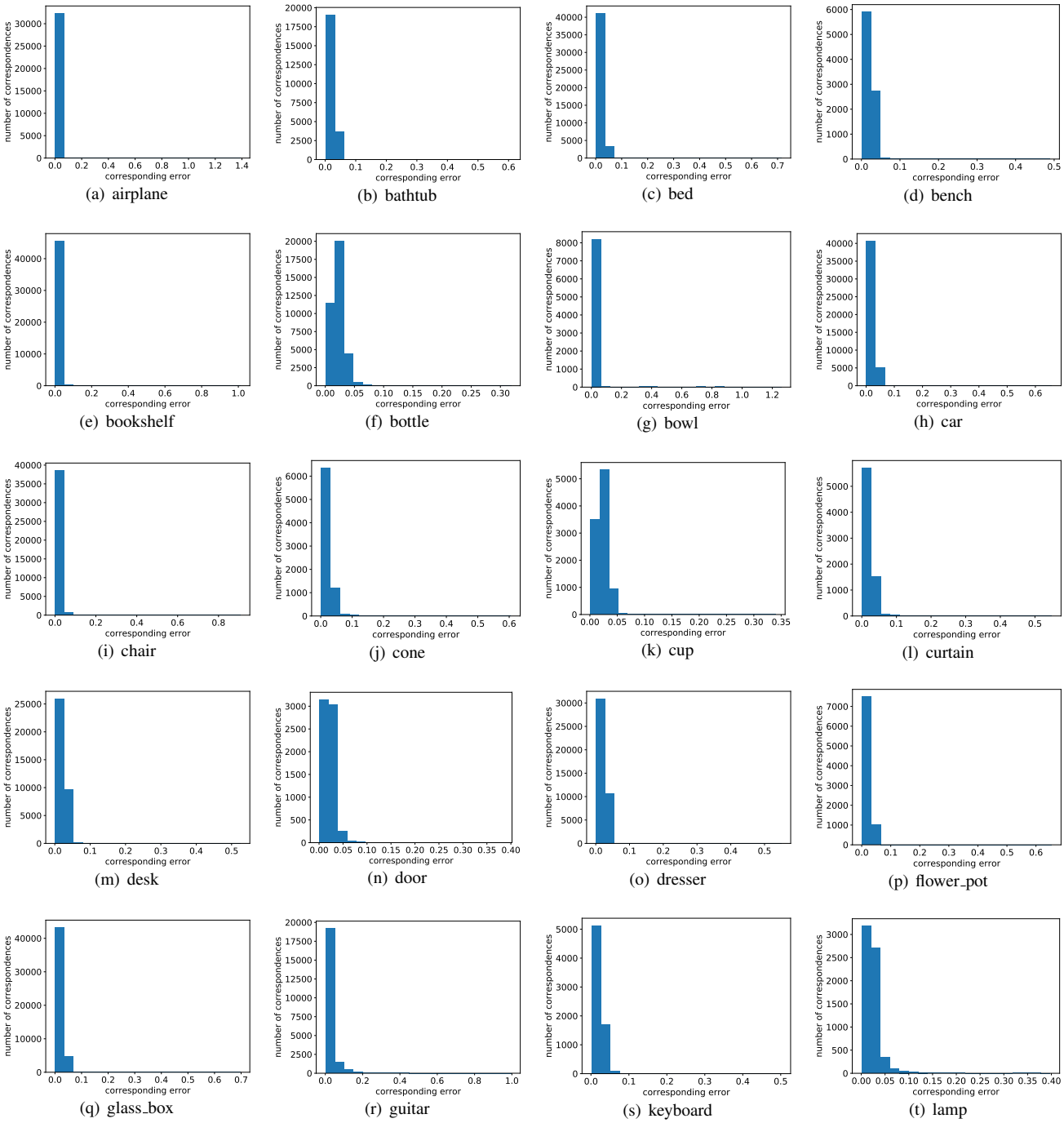


(m) stool (n) table (o) tent (p) toilet



(q) tv_stand (r) vase (s) wardrobe (t) xbox

Figure 12: The histogram of the corresponding errors in the noise experiment of the last 20 categories.



(a) airplane (b) bathtub (c) bed (d) bench

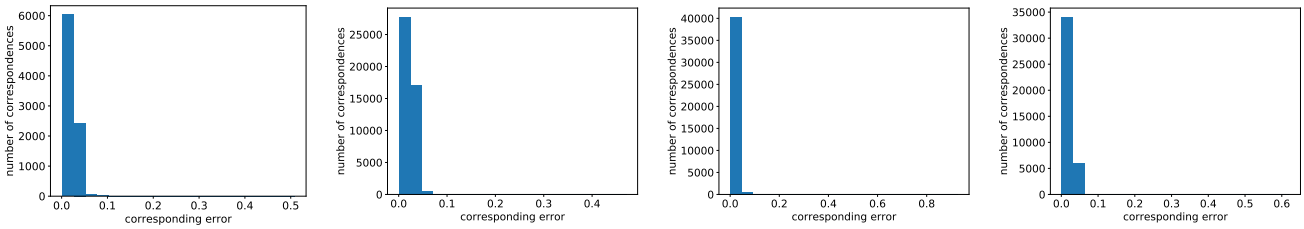
(e) bookshelf (f) bottle (g) bowl (h) car

(i) chair (j) cone (k) cup (l) curtain

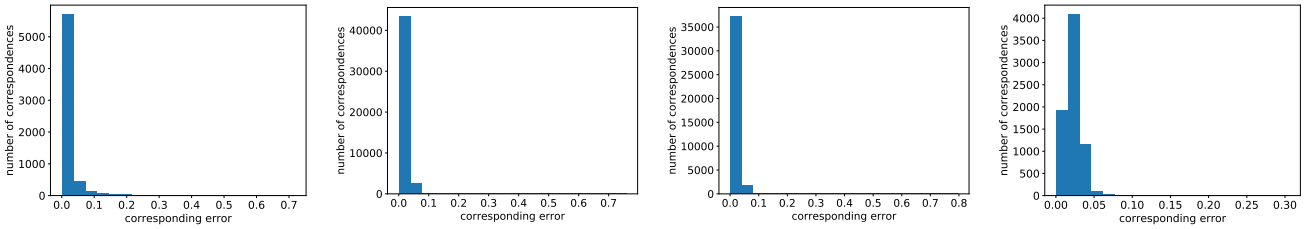
(m) desk (n) door (o) dresser (p) flower_pot

(q) glass_box (r) guitar (s) keyboard (t) lamp

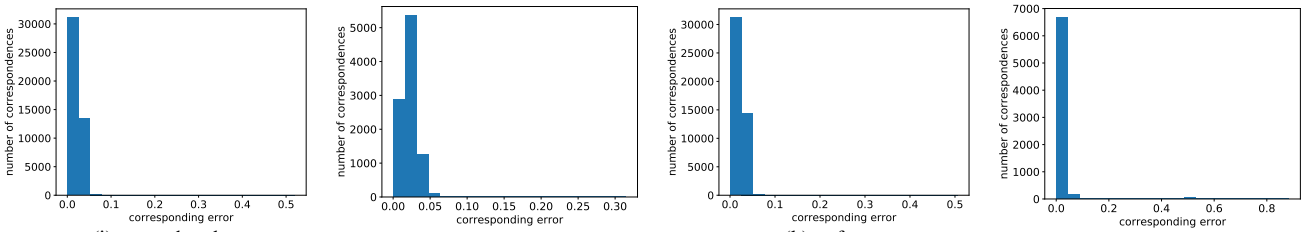
Figure 13: The histogram of the corresponding errors in the partial-to-partial experiment of the first 20 categories.



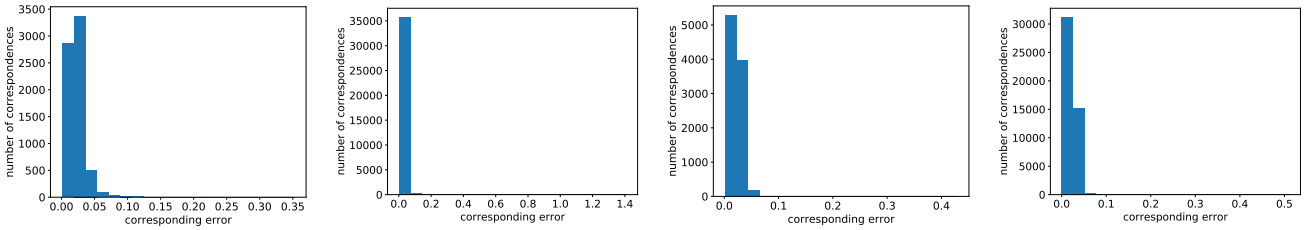
(a) laptop (b) mantel (c) monitor (d) night_stand



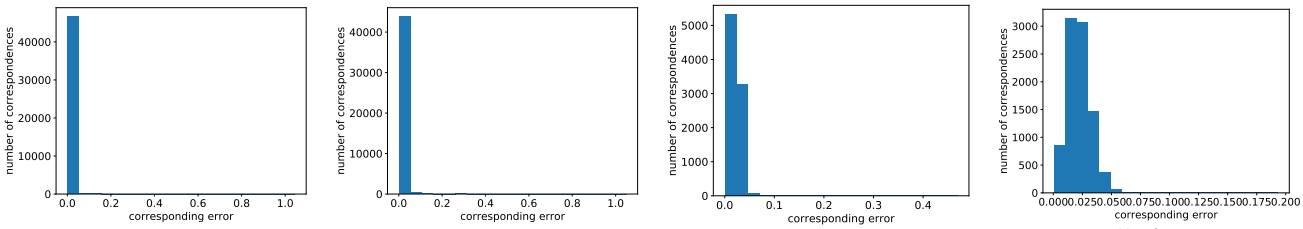
(e) person (f) piano (g) plant (h) radio



(i) range_hood (j) sink (k) sofa (l) stairs

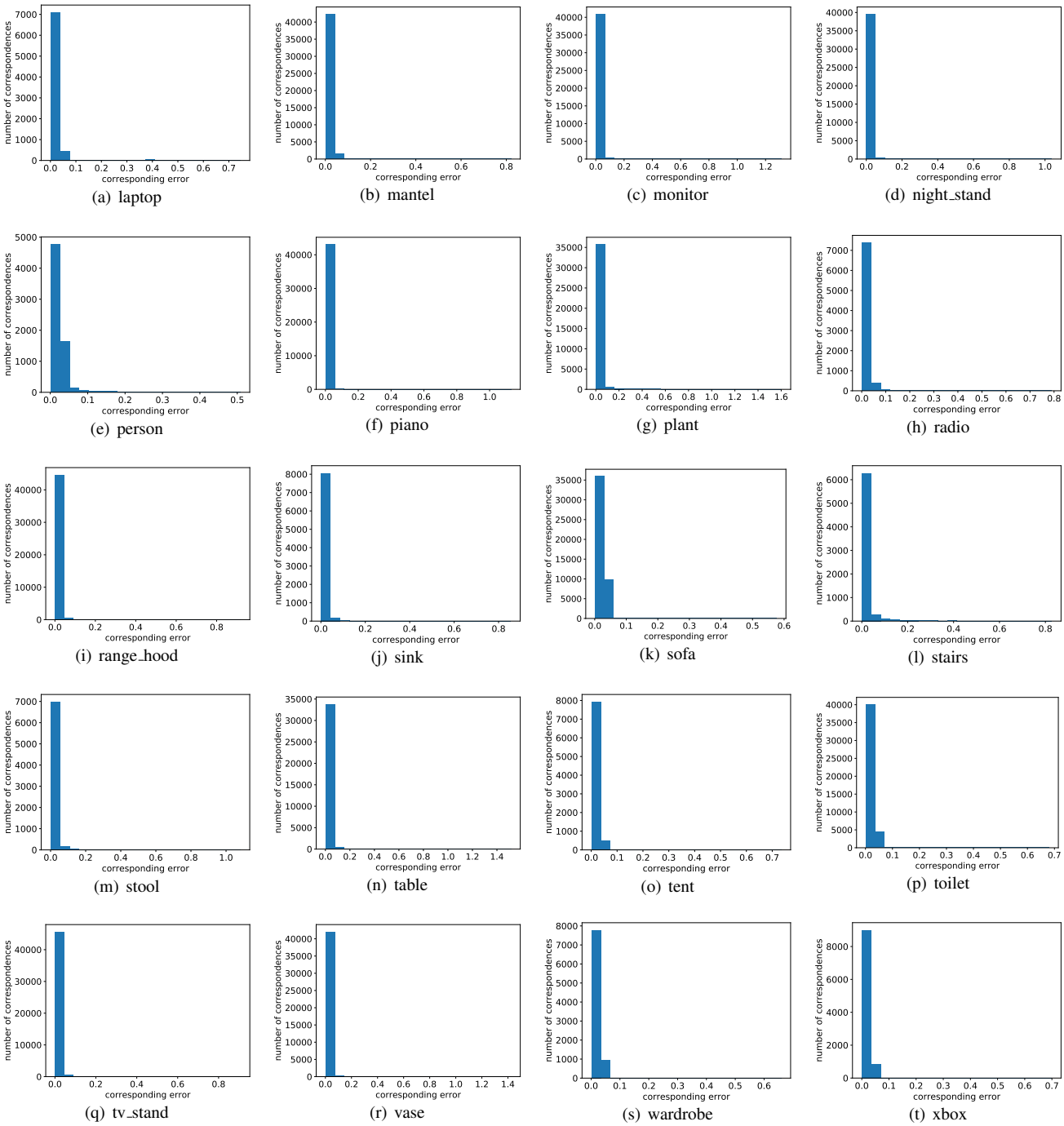


(m) stool (n) table (o) tent (p) toilet



(q) tv_stand (r) vase (s) wardrobe (t) xbox

Figure 14: The histogram of the corresponding errors in the partial-to-partial experiment of the last 20 categories.



(a) laptop (b) mantel (c) monitor (d) night_stand

(e) person (f) piano (g) plant (h) radio

(i) range_hood (j) sink (k) sofa (l) stairs

(m) stool (n) table (o) tent (p) toilet

(q) tv_stand (r) vase (s) wardrobe (t) xbox

Figure 15: The histogram of the corresponding errors in the unseen categories experiment of the last 20 categories.