Supplementary

The remainder of the supplementary materials are organized as follows. Section 1 explains ablation studies related to graphical information gain. Section 2 describes further analysis of the ArcPoint loss function. Section 3 presents a trade-off between epistemic uncertainty reduction and false prediction with high reliability. Finally, Section 4 compares the S3DIS visualization results of Baseline and GaIA.

Table 1. Comparison of components applied in entropy block on ScanNet-v2. $(\cdot)$ and $*$ denote the officially measured test scores and excluded neighborhood normalization, respectively.

H_k	$(D_k^{-2} \cdot H_k)$	GI	X^N	1%
✓		✓	✓	52.7
✓	✓	✓	✓	54.9 (65.2)
✓	✓		✓	52.1*
✓	✓	✓		52.8
✓	✓		✓	53.6
✓	✓	✓	✓	54.9 (65.2)

1. Analysis on graphical information gain

1.1. Neighborhood aggregation based on Euclidean distance

We conducted ablation studies to investigate the effect of Euclidean distance-based neighbor aggregation in the entropy block. Experiments were performed on a network with a Siamese branch and ArcPoint. As listed in lines 1 and 2 of Tab. 1, we compared the original entropy aggregation $\widetilde{H} = \sum_k H_k$ with the aggregation, which is inversely proportional to the Euclidean distance of the neighborhood, $\widetilde{H} = \sum_k (D_k)^{-2} \cdot H_k / \sum_j (D_k)^{-2}$. When we calibrated H_k based on the Euclidean distance, the performance surpassed that of only employing the original entropy by 2.2%p.

1.2. Effectiveness of neighbor aggregation

To investigate which information is more important, either reliable point-wise attention or reliable neighbor information for updating the point representation, ablation studies were conducted. Lines 4 and 5 of Tab. 1 show that the neighborhood aggregation $X = X + (X \otimes GI) + X^N$ outperformed the reliable point-wise attention $X = X + (X \otimes GI)$. Here, aggregation without normalization showed unsatisfactory performance compared with the point-wise attention $X = X + (X \otimes GI)$ because of the overfitting problem. In other words, normalized neighbor aggregation was effective in updating the uncertain point representation compared with only enhancing the reliable points.

Table 2. Comparison of applying entropy block to GaIA framework. The results of ScanNet-v2 are measured by validation set.

Dataset	Enc.	Dec.	Sec/iter	1%
S3DIS	✓		2.4 sec	66.5
	✓	✓	3.9 sec	66.8
ScanNet-v2	✓		4.3 sec	54.9
	✓	✓	6.2 sec	55.3

1.3. Entropy block organization

Applying the entropy block to each decoder block results in computational inefficiency. Therefore, we exclude the entropy block from the decoder in our network. When we organized both the encoder and decoder with the entropy block, the inefficiency increased compared with the performance gain. In line 2 of Tab. 2, the training time (seconds per iteration) increased by 1.5 seconds while the performance improved 0.1%p at 1% on S3DIS dataset. With a similar tendency, for the ScanNet-v2 dataset, the seconds per iteration increased by 1.9 seconds in comparison with the performance gain of 0.2%p. Based on the experimental results, we excluded the entropy block from the decoder. This means that obtaining high-quality feature representations contributes significantly to performance gain.

1.4. Comparison of entropy reduction

GaIA was compared with the baseline network by visualizing entropy to investigate the entropy reduction effect. As depicted in Fig. 1, we observed that GaIA more broadly alleviated the entropy overall training steps compared to the baseline. This is because GI contributed to updating the uncertain points near the ambiguous decision boundary of the network toward the semantically similar points. As mentioned in Section 3.2 from the original study, GI discriminates the relative entropy between the entropy of the target point and that of its neighbors to identify reliable information.

2. Analysis on ArcPoint loss

We addressed that ArcPoint loss optimizes unlabeled points toward semantically similar points by penalizing the unannotated points using an additive angular margin. In addition to the analysis in Sections 5.2 and 6 from the original study, we compared ArcPoint loss with the conventional softmax loss function by using the distribution of angle. The angle was measured using the class-wise prototypical weight matrix W_y and the unlabeled point as follows: $\theta_y(x_j^u) = arccos(\frac{W_y^\intercal \cdot x_j^u}{||W_y|| \cdot ||x_j^u||})$. As depicted in Fig. 8 from the original study, for the baseline network employing softmax loss, the ratio of false predictions in the column, chair, bookcase, and board classes were higher compared with ArcPoint. Moreover, as shown in Fig. 2, the angle distri-

bution of the baseline network commonly exhibited a long tail distribution for the mentioned classes. This is because the baseline could not enhance both intra-class density and inter-class distinction. In contrast, ArcPoint showed a short-tail distribution, which indicates that ArcPoint effectively optimized the unlabeled points for all classes compared to the existing softmax loss. To observe angle variation corresponding to training steps, the distributions of angle were compared. As depicted in Fig. 3, the average level of distance, which was measured by $\theta_y(x_j^u)$, gradually decreased during the training. To verify the effectiveness of ArcPoint, we visualized the cosine similarities between the anchors and unannotated points in Fig. 4.

3. Trade-off between epistemic uncertainty reduction and false prediction with high reliability

3.1. Reduction of false predictions

As mentioned in Section 6 of the original study, although the network contains low epistemic uncertainty, it can still estimate incorrectly. Hence, the entropy distribution was examined for each class and the distributions of true and false predictions were compared. As depicted in Fig. 5, it is observed that GaIA alleviated the number of false predictions with low entropy compared with the baseline which included the Siamese branch. False predictions with high reliability were reduced for all classes.

3.2. Point-wise entropy variation

Based on the experimental results in Fig. 2 and 3, a few unannotated points (in the column, sofa, bookcase, and board classes) could not be optimized well despite further training steps. Indeed, for these classes, the more training the network, the closer the distribution to bimodal is. To examine this optimization discrepancy, we visualized the point-wise entropy variation corresponding to both network prediction and training steps. In Fig. 6, we observed the number of false predictions progressively decreased along with epistemic uncertainty reduction following the training steps. Although the uncertainty reduction significantly improved the semantic segmentation performance compared with the baseline, it was confirmed that the uncertainty reduction resulted in a trade-off that false predictions with high reliability increased under the 1% annotation. To investigate whether this trade-off is consistently observed under the sparser annotation, we visualized the entropy distribution under the 1pt annotation. Similar to the 1% annotation, Fig. 7 exhibited a ratio of false prediction and the average level of entropy increased.

3.3. Consistency for false prediction with high reliability

Based on these observations, we assumed that the points, which were inaccurately predicted with high reliability in the early training step, consistently had low entropy. To validate this assumption, we examined the consistency of false predictions with high reliability. As listed in Tab. 3, the number of false predictions for each class was measured corresponding to the training steps. Consistency was observed in the column, sofa, bookcase, and board classes, which exhibited a trade-off. The ratio of false predictions with high reliability for the mentioned classes commonly occupied higher levels over the training steps compared with other classes.

4. Comparison of qualitative results on S3DIS

As depicted in Fig. 8, GaIA performed better, which is closer to Ground Truth, than the baseline.

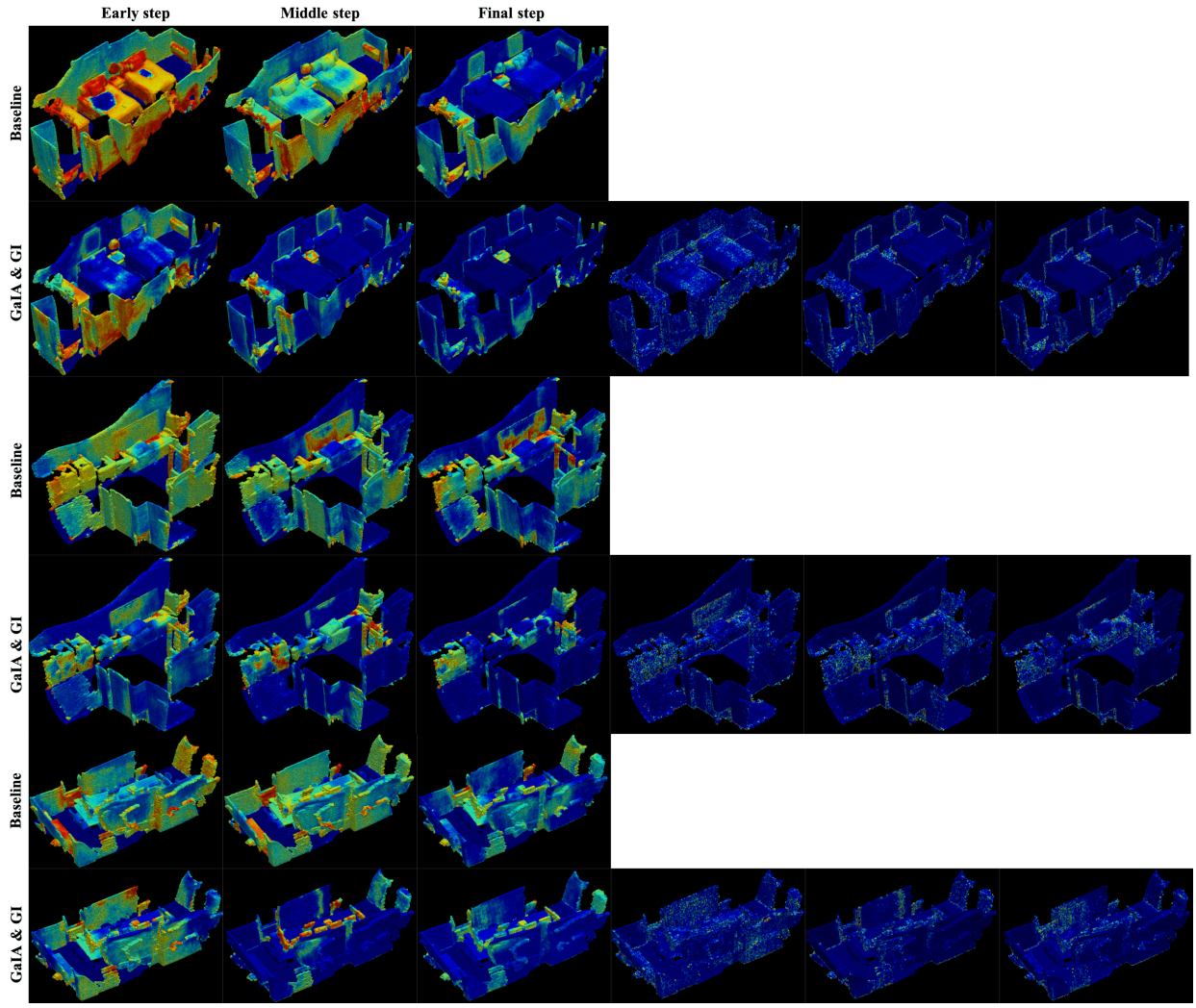

Figure 1. Comparison of entropy variation.

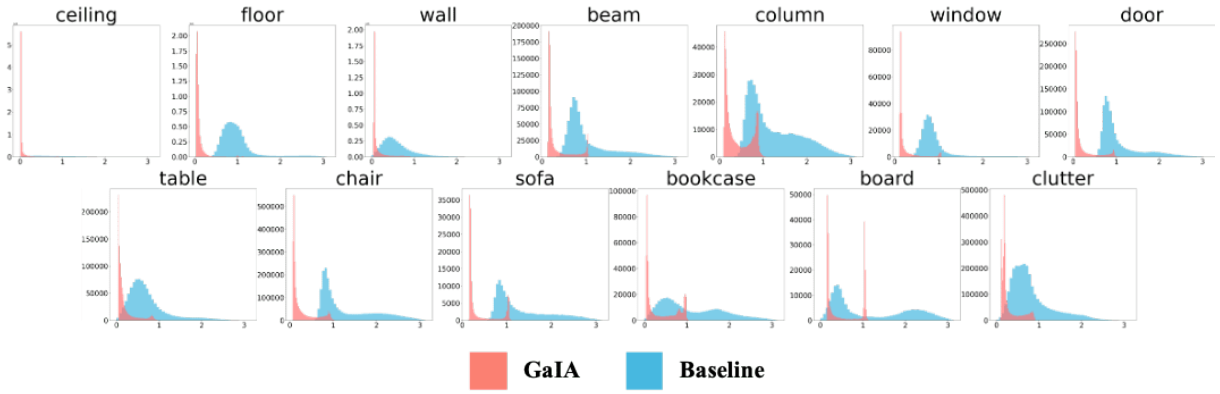

Figure 2. Comparison with ArcPoint and conventional softmax loss under the 1% annotation. The x-axis indicates the angle between the class-wise prototypical weight matrix W_y and the unannotated point.

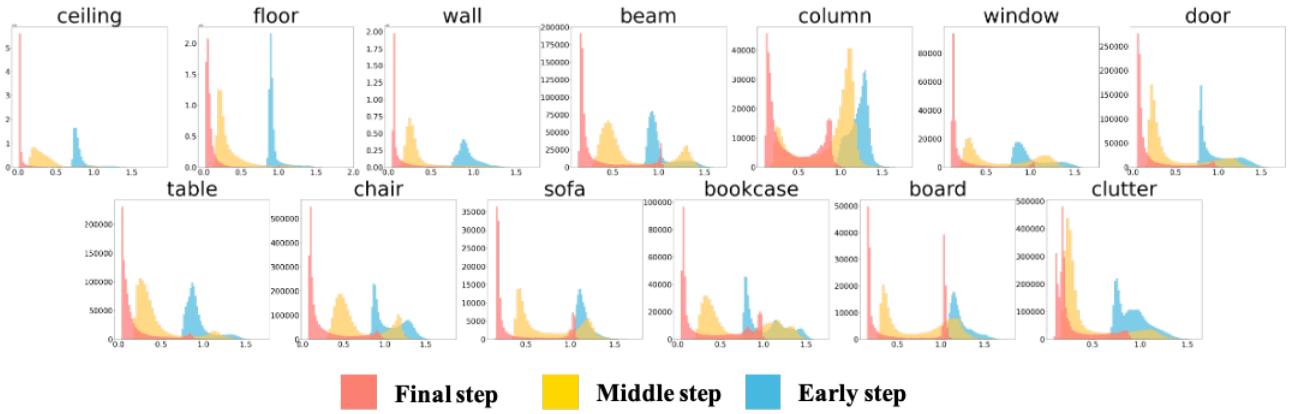

Figure 3. Comparison of angle variation corresponding to training steps under the 1% annotation. The x-axis indicates the angle between the class-wise prototypical weight matrix W_y and the unannotated point.

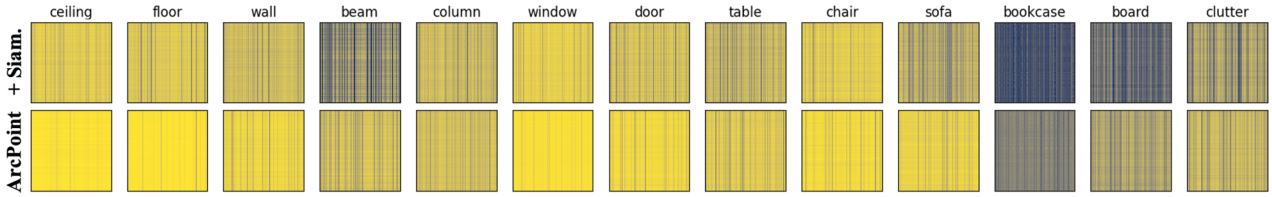

Figure 4. Comparison of cosine similarity. 50,000 anchors and unannotated points with high entropy were randomly sampled from the S3DIS dataset. In each heatmap, the rows and columns indicate anchors and unannotated points, respectively.

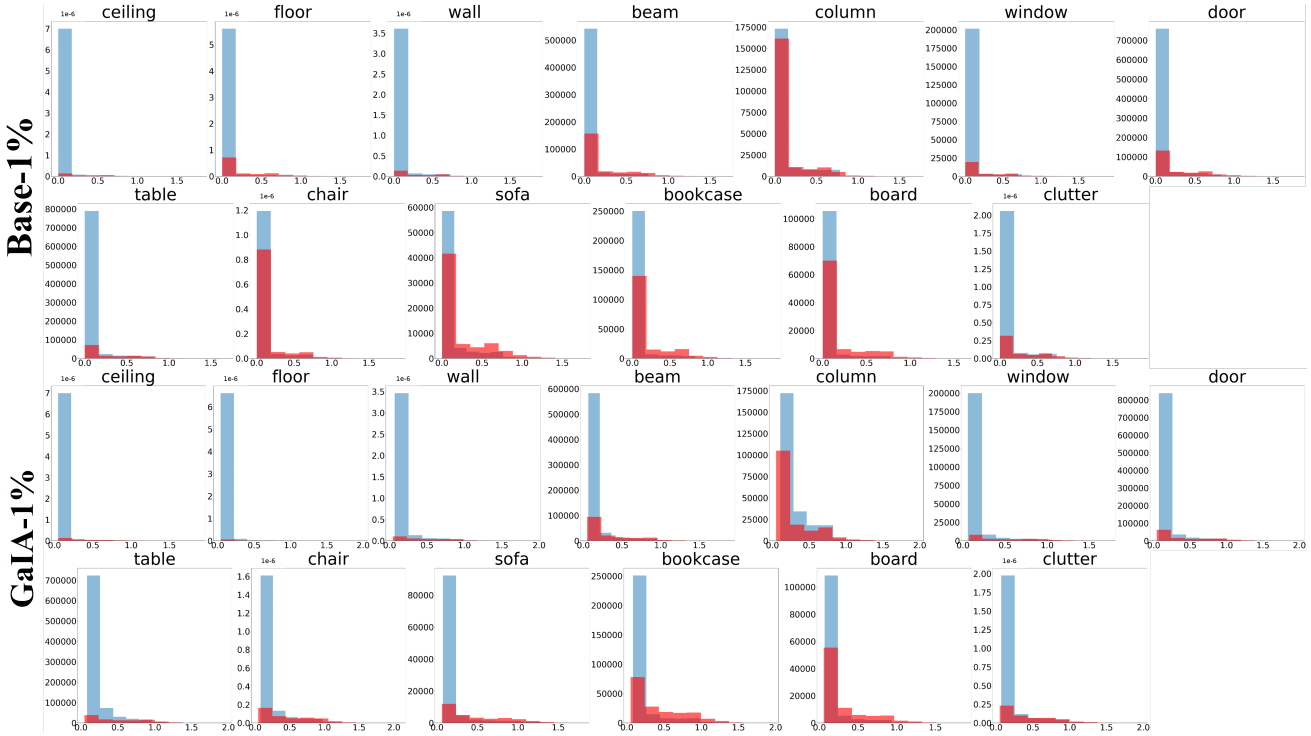

Figure 5. Comparison of entropy distribution with respect to prediction. The x and y axes indicate entropy and the number of samples, respectively. Distribution highlighted with red indicates distribution of false prediction.

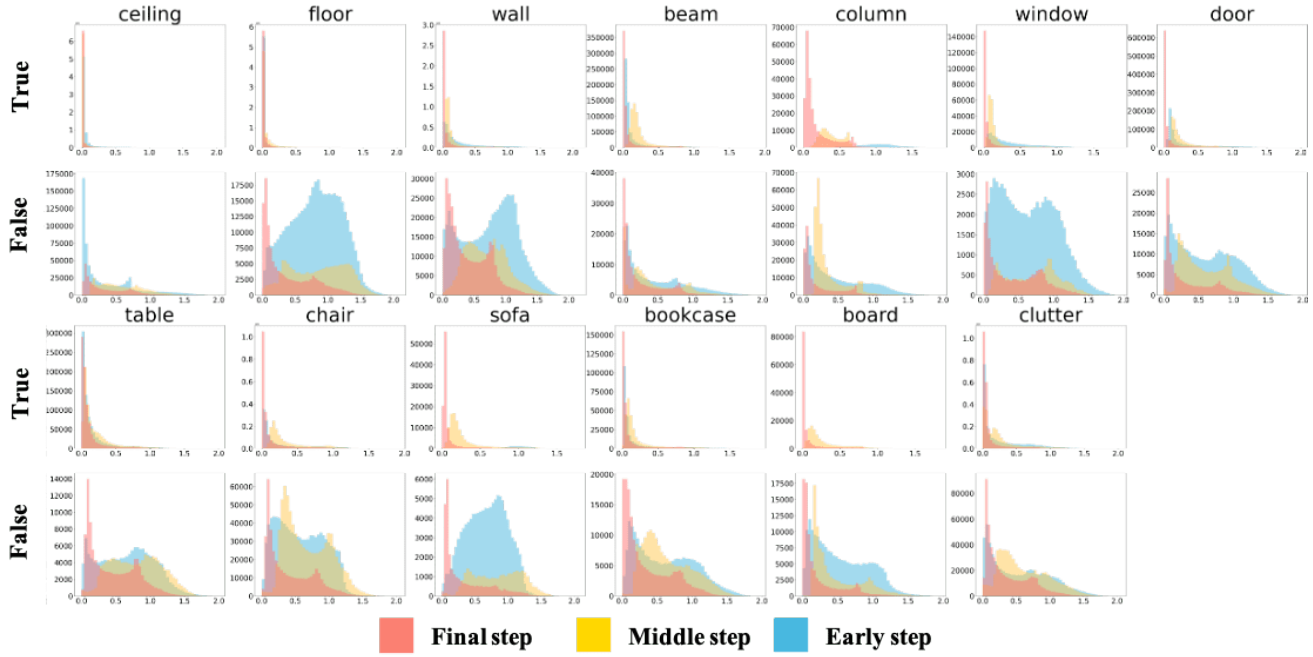

Figure 6. Comparison of point-wise entropy variation corresponding to prediction and training steps under the 1% annotation. The x-axis indicates the entropy.

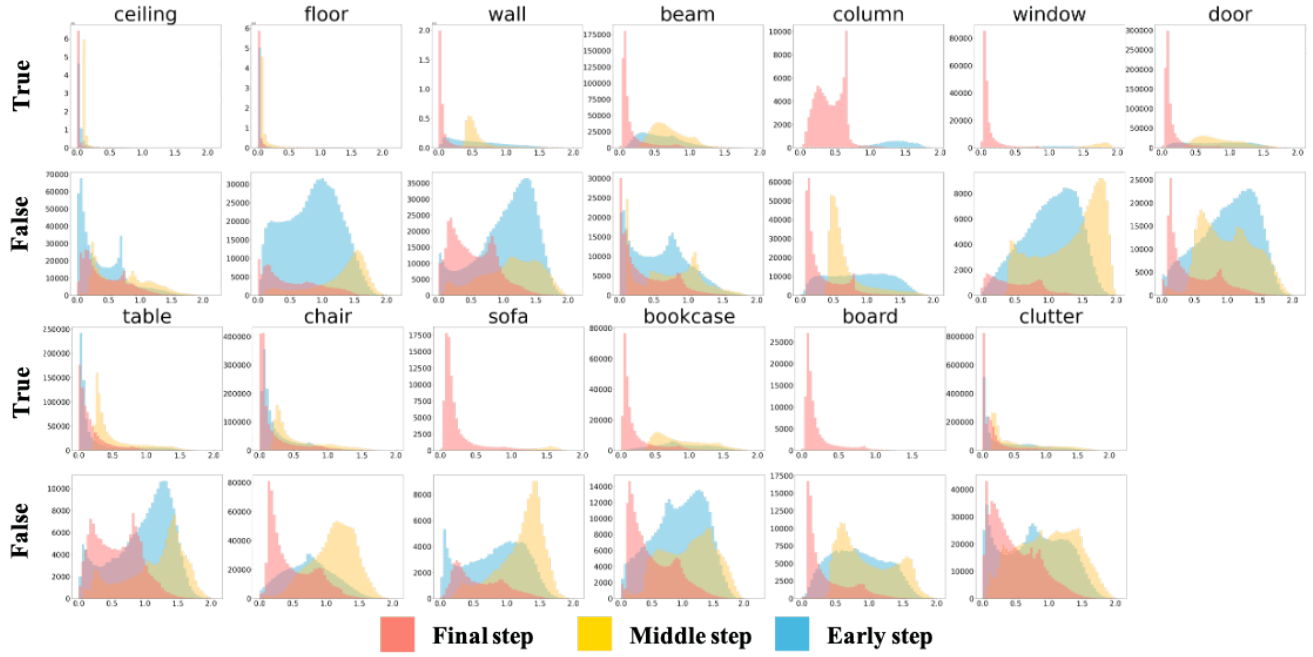

Figure 7. Comparison of point-wise entropy variation corresponding to prediction and training steps under the 1pt annotation. The x-axis indicates the entropy.

Table 3. Comparison of the number of false predictions corresponding to training steps.

Class	Early	Mid	Final	Early & Mid	Mid & Final	Early & Mid & Final
Ceiling	662,251	380,898	281,615	344,494	197,691	193,557
Floor	456,441	413,410	122,183	295,146	98,293	93,063
Wall	637,130	418,767	286,331	274,121	138,686	108,794
Beam	193,450	252,106	151,278	164,453	121,664	100,832
Column	365,368	315,687	**154,541**	303,338	143,604	**139,579**
Window	75,755	107,105	19,849	60,126	18,173	15,699
Door	370,788	291,378	121,271	211,820	75,458	59,609
Table	154,154	159,808	107,214	113,659	78,010	71,358
Chair	1,031,409	752,985	413,881	629,131	279,760	244,589
Sofa	115,168	60,429	**30,085**	58,586	25,965	**25,787**
Bookcase	222,362	210,304	**168,781**	186,488	149,283	**144,135**
Board	205,920	110,817	**83,586**	110,592	74,966	**74,833**
Clutter	680,719	625,378	533,408	403,932	321,078	250,165

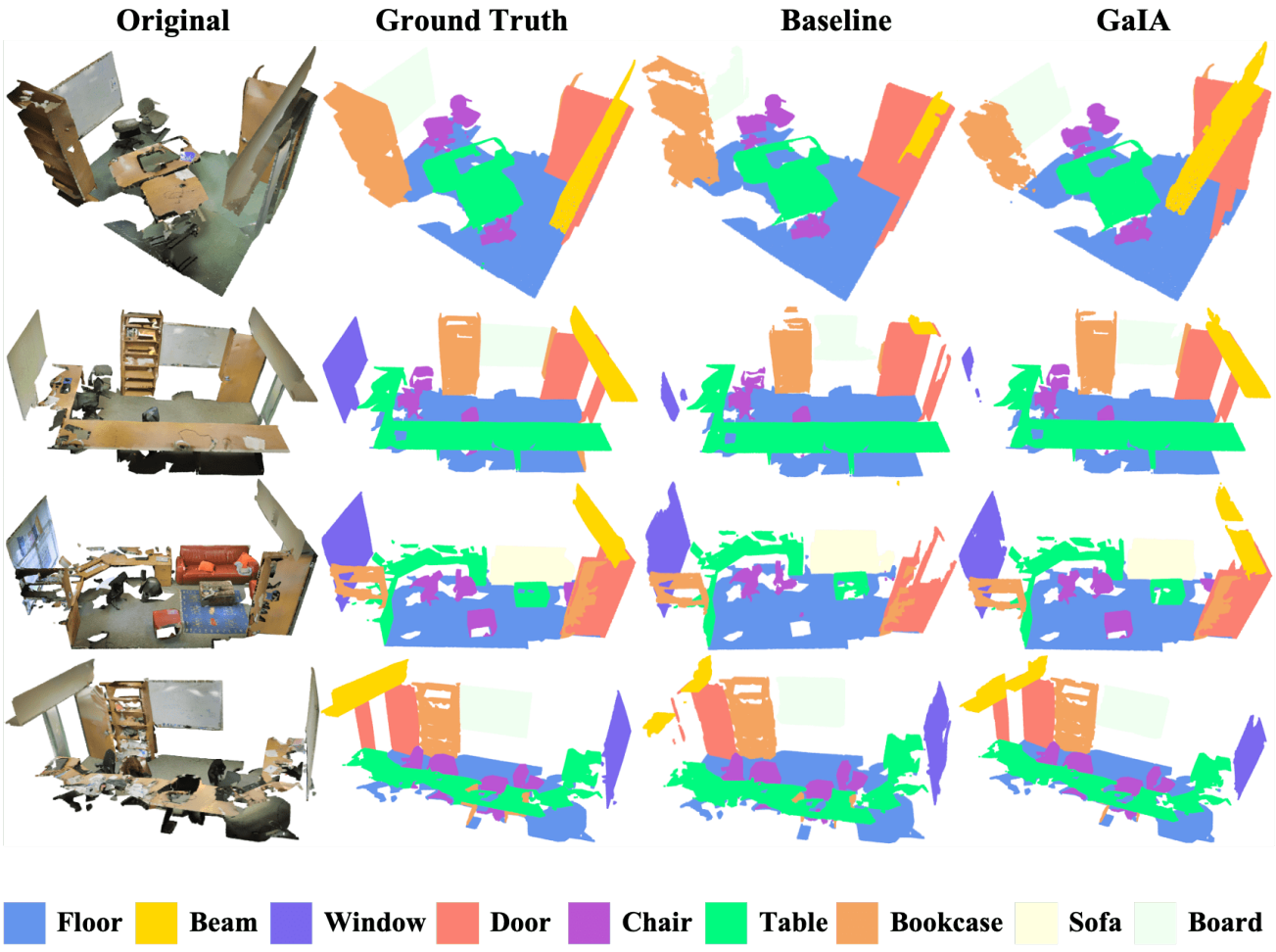

Figure 8. Comparison of qualitative results on S3DIS.