GaIA: Graphical Information Gain based Attention Network for Weakly Supervised Point Cloud Semantic Segmentation

Min Seok Lee* Seok Woo Yang* Sung Won Han†
School of Industrial and Management Engineering, Korea University
{karel, joshy, and swhan}@korea.ac.kr

Abstract

While point cloud semantic segmentation is a significant task in 3D scene understanding, this task demands a time-consuming process of fully annotating labels. To address this problem, recent studies adopt a weakly supervised learning approach under the sparse annotation. Different from the existing studies, this study aims to reduce the epistemic uncertainty measured by the entropy for a precise semantic segmentation. We propose the graphical information gain based attention network called GaIA, which alleviates the entropy of each point based on the reliable information. The graphical information gain discriminates the reliable point by employing relative entropy between target point and its neighborhoods. We further introduce anchor-based additive angular margin loss, ArcPoint. The ArcPoint optimizes the unlabeled points containing high entropy towards semantically similar classes of the labeled points on hypersphere space. Experimental results on S3DIS and ScanNet-v2 datasets demonstrate our framework outperforms the existing weakly supervised methods.

1. Introduction

Point cloud semantic segmentation is a fundamental task in the field of computer vision. With the success of deep neural networks, large-scale point cloud semantic segmentation on the 3D scene has drawn more attention due to its wide applications (e.g., augmented/virtual reality, autonomous driving, and robotics). However, a fully supervised method for point cloud semantic segmentation requires well-labeled point-wise annotations, and this entire process of data annotation is expensive [31, 32, 40, 25, 7, 43, 17, 41, 17, 20, 50, 30, 39]. To address this issue, recent studies have adopted a weakly supervised learning approach to train networks with partial annotations of point clouds. Previous studies [44, 5, 45, 48, 49, 16, 26, 47, 24] improved the semantic segmentation performance close to that of fully supervised one on small-scale datasets (e.g., ShapeNet [4] and PartNet [28]) as well as large-scale datasets (e.g., S3DIS [2] and ScanNet-v2 [8]).

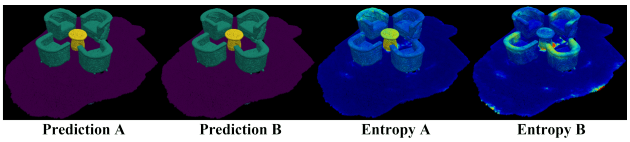

Figure 1. Comparison of performance recognition and information uncertainty. Prediction of the network A has higher uncertainty in the table compared with the network B.

In contrast to existing studies, this study focuses on alleviating epistemic uncertainty to obtain high-quality feature representations under sparse annotation. In Fig. 1, if two networks show a similar performance or visualization result, it is hard to determine which network is semantically well-embedded. To observe whether there is a difference in the estimation of the two networks, Shannon entropy [37] was employed for epistemic uncertainty quantification [21, 27]. In measuring the entropy of each point, it was observed that the reliability of the network prediction may differ even if the same result is obtained. Starting from this experimental result, the question was raised as to whether alleviating epistemic uncertainty improves segmentation performance along with satisfactory point cloud embedding. To address epistemic uncertainty reduction, we introduce two approaches: reducing the entropy of each point and effective optimization for points containing high entropy.

Reducing epistemic uncertainty is regarded as alleviating the entropy of each sample [15, 14, 38]. To reduce the entropy of each point, we treat points with low entropy as credible information to update the probability distribution of points containing high entropy. Reliable points near the ambiguous decision boundary of the network are identified by measuring relative entropy, because not all reliable points are important. As a relative entropy measure, this study introduces graphical information gain, which is de-

*Equal contribution.
†Corresponding author.

termined by the relative entropy between the entropy of the target point and that of its neighborhood. When a point has entropy lower than that of its neighborhood, it is more reliable. Based on the reliability, we enhance the point representation and update the point including the high entropy by propagating the credible information to the uncertain points.

Under the sparse annotation, effective optimization of the unlabeled points is important for achieving satisfactory semantic segmentation. Existing studies organize the relation network [45, 49] or class prototypical matrix [48] to optimize the unannotated points. For loss computation, the softmax function is widely employed to present class probability. However, the softmax has a limitation in terms of data optimization in that it can neither explicitly enhance the similarity of intra-class features nor discriminate inter-class features [9]. Moreover, previous studies equally focused on all unlabeled points during the optimization process. Although the points containing low entropy are semantically well-embedded in the optimization process, the network should focus more on optimizing the unannotated points with high entropy to improve segmentation performance. Therefore, it is necessary to overcome the drawbacks of softmax and address the optmization of highly uncertain points.

This study proposes a graphical information gain-based attention network (GaIA) for weakly supervised point cloud semantic segmentation. GaIA aims to reduce epistemic uncertainty using the graphical information gain and the anchor-based additive angular margin loss called ArcPoint. The graphical information gain measures the relative entropy between the entropy of the target point and that of its neighborhoods to discriminate reliable information. Based on relative entropy, GaIA updates the feature embedding of the unlabeled points containing high entropy toward semantically similar embedding of the labeled points. To address the limitation of softmax and focus on unlabeled point optimization, we introduce ArcPoint loss. By penalizing the unannotated points with high entropy using an additive angular margin in loss computation, ArcPoint optimizes the uncertain points embedded in the hypersphere toward a semantically similar embedding of the labeled points. The main contributions of this study are as follows:

- Epistemic uncertainty reduction is studied to improve weakly supervised point cloud semantic segmentation performance. To the best of our knowledge, this is the first approach to focus on epistemic uncertainty reduction for a performance gain in the weakly supervised point cloud semantic segmentation.

- For epistemic uncertainty reduction, we propose the graphical information gain to measure the relative entropy between the entropy of target point and that of its neighborhoods to identify reliable information.

- The proposed ArcPoint loss contributes to the epistemic uncertainty reduction by enabling the network to embed the unlabeled points with high entropy toward the reliable labeled points.

- GaIA improves mIoU by 2.2%p and 4.4%p compared with existing weakly supervised learning methods on two benchmark datasets (e.g., S3DIS and ScanNet-v2) under the 1 and 20 pts annotation.

2. Related work

2.1. Weakly supervised semantic segmentation on point cloud

Studies on 3D point cloud semantic segmentation have improved performance using fully annotated supervision learning [31, 32, 40, 25, 7, 43, 17, 41, 17, 20, 50, 30, 39]. Despite this achievement, annotating all point clouds remains a time-consuming task. To address this problem, recent studies have adopted a weakly supervised learning approach. Weakly supervised point cloud semantic segmentation performs segmentation with partial annotations for the point cloud. Existing studies generated semantically transformed types of point clouds, such as 2D segmentation maps [42], subcloud-level annotation [44], and superpoint [5]. With sparse annotation, previous approaches have employed pre-training method [16, 48], contrastive learning [16, 26, 24], and learning distribution consistency [45, 49, 24, 47] to learn spatial information of point clouds. For learning the topology of a point cloud, graph-structure was utilized to represent features of points [5, 26, 49]. Different from the previous approaches, we propose a novel weakly supervised framework that aims to reduce network uncertainty and effectively optimize unlabeled points.

2.2. Uncertainty quantification and reduction

Uncertainty quantification is important for precise decision-making in various domains [1], such as autonomous driving [11, 6] or medical image analysis [23, 36, 35, 29, 34]. Uncertainty in the predictive process is caused by three components: data uncertainty, epistemic uncertainty, and distributional uncertainty [33, 12, 21, 27]. Among these three types of uncertainty, this study focuses on epistemic uncertainty, which measures the information uncertainty in predicting network parameters given the data [12, 27]. Uncertainty can be reduced, such that the lower the uncertainty, the higher is the network performance is [15, 14, 27]. Based on this property, we introduce a network that focuses on epistemic uncertainty reduction to improve point cloud semantic segmentation performance. For the uncertainty quantification measure, Shannon entropy, which represents information uncertainty [37], is adopted, whereby the entropy of each point is estimated to identify

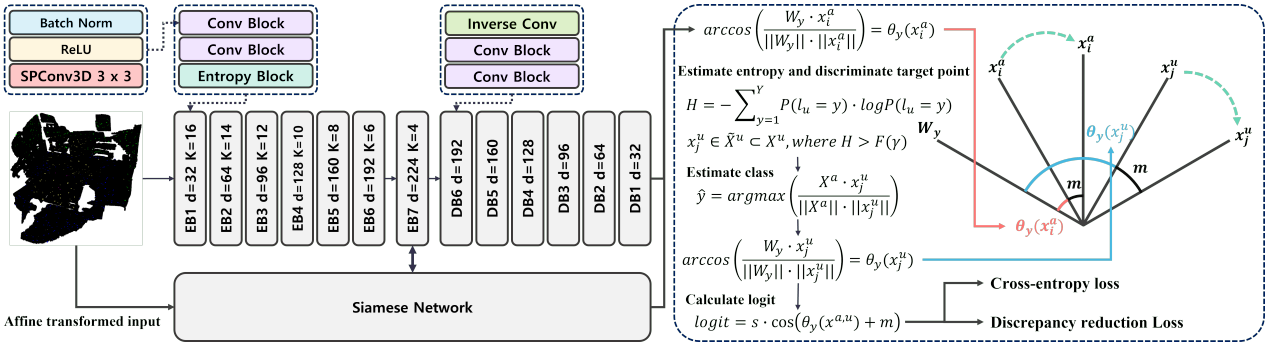

Figure 2. Overall architecture.

reliable information. Our approach updates uncertain points near the ambiguous decision boundary of the network by propagating credible features.

2.3. Sparse annotation embedding

Using point cloud data is a more attractive approach compared to a transformed representation (e.g., a voxel or mesh). However, it is difficult to employ raw point clouds due to their disordered and unstructured properties [31, 10]. Furthermore, it is challenging to generate a high-quality feature representations from partially annotated point clouds. Thus, existing studies focus on the feature representation of labeled points shared with unlabeled point clouds [45, 48, 49, 26, 24, 47]. To obtain the feature embedding, previous studies minimize the difference between the ground truth and projected label [42, 48]. In addition to the aforementioned studies, other approaches optimize the divergence between two probabilistic distributions [45, 49, 26, 24, 47]. In the training process, the above studies employed the softmax function. However, the softmax has a limitation when classifying an open-data set in that it is not in the training data [9]. Thus, the convergence of intra-class data and divergence of inter-class data should be enhanced to effectively embed unfamiliar data. Moreover, previous studies have optimized all unlabeled data equally. In contrast to these studies, we focus on the optimization of unlabeled points with high entropy for effective optimization. The uncertain points are closely embedded along with the semantically similar labeled points on the hypersphere by using the labeled points as anchor.

3. Method

3.1. GaIA overview

Architecture: GaIA is designed to alleviate epistemic uncertainty. To reduce the high entropy of the uncertain points, we organize the entropy block and ArcPoint loss. As depicted in Fig. 2, 3D U-Net is implemented as a backbone network with sub-manifold sparse convolution and sparse convolution as in [13, 20]. Input X is a point set of N points. Each point $x_i \in \mathbb{R}^6$ is represented by a concatenation of 3D coordinates and RGB colors, where $i \in \{1, ..., N\}$. Then, X is voxelized to a size of 0.02m. The semantic features are extracted by feeding X to a couple of convolution and entropy blocks. Each convolution block comprises a sequence of batch normalization-ReLU-sparse convolutional operations (SPConv3D). Subsequently, the entropy block computes the entropy variation between the target point entropy and entropy of its neighborhoods, referred to as graphical information gain in this study. As an attention weight, graphical information gain enhances reliable point representation and propagates the information to their neighborhoods. After extracting semantic features from the encoder blocks, X is reconstructed using a decoder. The entropy block at the decoder is excluded because applying the entropy block to each decoder block results in computational inefficiency. In fact, when a decoder with an entropy block is organized, the inefficiency increases with respect to the performance gain.

Learning strategy: To embed the unlabeled points, we adopt a Siamese network branch [3, 22] to GaIA. The Siamese branch maintains the consistency between the prediction of the original input X and that of the affine transformed input $aff(X)$. This learning strategy improves the embedding performance by imposing constraints on unannotated points [45]. For the affine transformation of a given input point X, we apply a random noise, flipped with the x and/or y axes, and rotated at random angles to the x axis. Subsequently, to achieve a more robust network against the sparse annotation, we impose more constraints on X by employing an elastic distortion. Initially, GaIA is trained excluding the Siamese branch for 100 epochs because the constraints result in unstable entropy for each point at earlier stages. After optimizing the original network, we adopt the Siamese branch to minimize the discrepancy between network predictions.

Optimization: Existing studies [45, 48, 49, 26, 24, 47] employed softmax cross-entropy loss and treat the points

Algorithm 1 Entropy block operation

1: **Input:** Point cloud representation $X \in \mathbb{R}^{N \times d}$.
2: Initialize: $\widetilde{X} = \mathcal{F}(X),\ where\ \widetilde{X} \in \mathbb{R}^{N \times Y}$ and graph $G(N, E) \leftarrow KNN(X_{loc},\ K)$.
3: Get H: $H_i = -\sum_{y=1}^{Y} P(x_i = y) \cdot logP(x_i = y)$, $where\ P(x_i) = softmax(x_i)$ and $x_i \in \widetilde{X}$.
4: Calibrate: $\widetilde{H}_i = \sum_{j \neq i}^{k} (D_{i,j})^{-2} \cdot H_j / \sum_{j \neq i}^{k} (D_{i,j})^{-2}$, $where\ x_j \in neighbor(x_i)$.
5: Get GI: $GI_i = |H_i - \widetilde{H}_i|$.
6: Neighbor aggregation: $x_i^n = (\sum_{j \neq i}^{k} x_j \otimes GI_j)/K$.
7: Update point embedding: $\widetilde{X} = \widetilde{X} + (\widetilde{X} \otimes GI) + \widetilde{X}^N$, $where\ x_i^n \in \widetilde{X}^N$.
8: **Output:** $O = \mathcal{F}(\widetilde{X}),\ where\ O \in \mathbb{R}^{N \times d}$.

equally for network optimization. Different from the previous approaches, this study focuses on optimizing the points including high entropy. Inspired by ArcFace [9], which addresses the limitation of conventional softmax cross-entropy loss, is adopted as a baseline form of the loss function. However, ArcFace loss cannot deal with a large number of unannotated points because it requires the ground truth in the training phase. Thus, we propose an anchor-based additive angular margin loss called ArcPoint. ArcPoint loss aims to embed the unlabeled points toward semantically similar points by employing labeled points regarded as anchors. In Fig. 2, ArcPoint at first optimizes the distance $\theta_y(x_i^a)$ between the class-prototypical weight matrix W_y and annotated point $x_i^a \in X^a$ on the normalized hypersphere. Afterward, unannotated points containing high entropy $x_j^u \in \widetilde{X}^u$ are identified and the angle between X^a and x_j^u is computed. Here, X^a functions as an anchor in leading the x_j^u towards the nearest class W_y. A more detailed computing process for ArcPoint loss is demonstrated in Section 3.3.

3.2. Graphical information gain

Graphical information gain (GI) measures the relative entropy between the entropy of the target point and that of its neighbors to identify reliable information. GI is theoretically based on information uncertainty [37]. The entropy H represents the information uncertainty using the probability of event i as follows: $H = -\sum_i P_i \cdot logP_i$.

That is, if the probability distribution of the classes is sparse, a network can make a reliable decision for class prediction. Focusing on this property, entropy is utilized through three phases to alleviate epistemic uncertainty: i) measure entropy of each point, ii) compute graphical information gain, and iii) update the point embeddings with reliable representations. As shown in Algorithm 1, the input point cloud $X \in \mathbb{R}^{N \times d}$ is projected onto $\widetilde{X} \in \mathbb{R}^{N \times Y}$ using the SPConv3D operation $\mathcal{F}(\cdot)$, where Y denotes the number of classes. In addition, based on the coordinates X_{loc}, the k-nearest neighbor algorithm is applied to the input X to identify the neighborhood. In line 3, entropy H_i for each point x_i is computed. To obtain GI, we aggregate the entropy of neighborhoods H_j that are inversely proportional to the Euclidean distance ($D_{i,j}$) between target point x_i and its neighborhood x_j in line 4. Inverse $D_{i,j}$ imposes more weights on the neighborhood entropy, which is geometrically close to the target x_i. In line 5, GI, which is regarded as relative entropy, is obtained by subtracting the calibrated entropy $\widetilde{H}$ from the original entropy H. When the target point contains lower entropy compared to that of its neighborhoods, the results are more reliable. As depicted in Fig. 3, it is recognized that the GI highlights reliable points with low entropy near the ambiguous decision boundary of the network. Subsequently, we enhance the reliable representations using $x_i \otimes GI_i$, and neighborhood information is aggregated along with normalization in lines 6 and 7. Based on both enhanced representations, the point embeddings are updated to reduce epistemic uncertainty. Finally, the entropy block reconstructs the updated representation $\widetilde{X} \in \mathbb{R}^{N \times C}$ to $O \in \mathbb{R}^{N \times d}$ by using the sparse convolutional operation $\mathcal{F}(\cdot)$. Further analysis of the GI is offered in the Supplementary-analysis on graphical information gain.

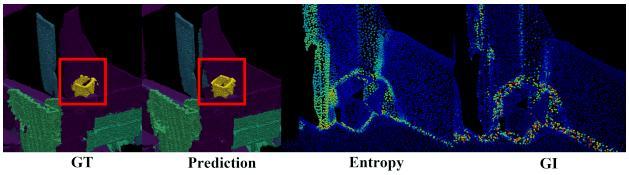

Figure 3. Visualization of a decision boundary and graphical information gain. Red points indicate high entropy and GI values.

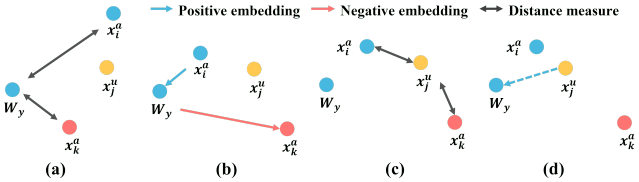

Figure 4. Embedding process of the ArcPoint loss.

3.3. Loss function design

Anchor based additive angular margin loss: ArcPoint is designed to effectively embed unannotated points by addressing the limitation of conventional softmax cross-entropy and ArcFace losses. Fig. 4-(a) and -(b) illustrate the original embedding process of ArcFace which embeds the intra-class anchor point similarly while discriminating the inter-class point. Following both (a) and (b), the angles between the unlabeled points and anchors are measured, in-

Algorithm 2 Anchor based additive angular margin loss

1: **Input:** labeled anchor $x_i^a \in X^a$, unannotated points $X^u \in \mathbb{R}^{N\times d}$, y^{th} class-prototypical weight matrix $W_y \in \mathbb{R}^d$, re-scaler s, and margin parameter m.
2: Get angle and add angular margin: $\theta_y(x_i^a) + m = arccos(\frac{W_y^\intercal \cdot x_i^a}{||W_y||\cdot||x_i^a||}) + m$
3: Calculate H: $H = -\sum_{y=1}^{Y} P(l_u = y)\cdot logP(l_u = y)$, $where\ \ P(l_u) = softmax(s\cdot(\frac{X^u\cdot W}{||X^u||\cdot||W||}))$, and $W \in \mathbb{R}^{d\times Y}$
4: Discriminate the points containing high entropy: $x_j^u \in \widetilde{X}^u \subset X^u,\ where\ H > F(\gamma)$
5: Estimate the nearest anchor: $\hat{y} = argmax(\frac{X^a\cdot x_j^u}{||X^a||\cdot||x_j^u||})$
6: Add angular margin for unannotated points: $\theta_y(x_j^u) + m = arccos(\frac{W_{\hat{y}}^\intercal \cdot x_j^u}{||W_{\hat{y}}||\cdot||x_j^u||}) + m$
7: Calculate final logit:

$$l = \begin{cases} s\cdot cos(\theta_y(x^i) + m), & \text{if } i \in \{a, u\},\ where \\ & x^a \in X^a\ \ and\ \ x^u \in \widetilde{X}^u \\ s\cdot cos(\theta_y(x^u)), & \text{otherwise} \end{cases}$$

8: **Output:** Final logit l

cluding different classes. Subsequently, in Fig. 4-(d), the nearest anchor of each unannotated point is determined by the smallest angle. Afterward, the unannotated point embedding is relocated toward the class which is the same as the nearest anchor. The detailed embedding process is presented in Algorithm 2.

In line 2, the i^{th} labeled anchor point $x_i^a \in \mathbb{R}^d$ belonging to the class y is embedded on the hypersphere by computing the angle $\theta_y(x_i^a)$ between x_i^a and W_y. Here, W_y denotes the y^{th} column of class-prototypical weight matrix $W \in \mathbb{R}^{d\times Y}$. The angle with an additive angular margin m is penalized to enhance intra-class intensity and inter-class distinction. To optimize the unannotated points along with epistemic uncertainty reduction, we focus on the points with high entropy. In lines 3 and 4, the entropy of unannotated points is calculated using a re-scaled logit l_u to discriminate the target points $\widetilde{X}^u$ that contain high entropy. Here, the function $F(\gamma)$ denotes the γ quantile of H, such that the higher area of γ in the distribution of H is adopted. Subsequently, to estimate the nearest anchor regarded as a class, we measure $cos\theta$ between the entire anchor X^a and target point x_j^u, which is the j^{th} instance of the target points $\widetilde{X}^u$. Following the estimation, the angular margin m is added to the angle $\theta_{\hat{y}}(x_j^u)$ measured by the estimated class weight $W_{\hat{y}}$ and x_j^u in line 6. For the final logit calculation, both $\theta_y(X^a)$ and $\theta_y(\widetilde{X}^u)$ are applied to the margin, except for the other cases. The logit passes the cross-entropy loss through a softmax function. In the inference phase, logit is computed without the additive angular margin as follows: $logit = s\cdot(\frac{X\cdot W}{||X||\cdot||W||})$. This optimization effect is discussed in Section 5.2 by visualizing similarities between X^a and $\widetilde{X}^u$ corresponding to each class.

Loss configuration: The loss function $\mathcal{L}$ comprises the ArcPoint based both softmax cross-entropy loss $\mathcal{L}_{ce}$ and distribution discrepancy reduction loss $\mathcal{L}_{sia}$, as follows: $\mathcal{L} = \mathcal{L}_{ce} + \mathcal{L}_{ce}^{aff} + \mathcal{L}_{sia}$. In Eq (1), A denotes the number of labeled points, and the annotated points are optimized by using the penalty term m. In Eq (2), when the Siamese branch is applied to GaIA, the segmentation loss $\mathcal{L}_{ce}^{aff}$ for the affine transformed input $aff(X)$ and distribution discrepancy reduction loss $\mathcal{L}_{sia}$, which are based on ArcPoint, are organized. The $\mathcal{L}_{sia}$ minimizes the L2 distance between all probabilistic predictions of the original network and those of the Siamese branch. In this process, the unannotated points are optimized. Here, the unlabeled points containing low entropy are not subject to the angular margin, but are involved in distance minimization.

$$\mathcal{L}_{ce} = -\frac{1}{A}\sum_{i=1}^{A} log\frac{e^{s\cdot cos(\theta_y(x_i^a)+m)}}{e^{s\cdot cos(\theta_y(x_i^a)+m)} + \sum_{j=1,j\neq y}^{Y} e^{s\cdot cos\theta_j}} \tag{1}$$

$$\mathcal{L}_{sia} = ||P(X) - P(aff(X))||_2, \quad where$$
$$P(X) = \frac{1}{N}\sum_{i=1}^{N} \frac{e^{s\cdot cos(\theta_y(x_i^{a,u})+m)}}{e^{s\cdot cos(\theta_y(x_i^{a,u})+m)} + \sum_{j=1,j\neq y}^{Y} e^{s\cdot cos\theta_j}} \tag{2}$$

4. Experiment

4.1. Experimental setup

Datasets: S3DIS [2] contains 271 scenes for six areas from three different buildings consisting of 3D RGB point clouds. Each point was annotated with one of 13 semantic categories. All the classes were used in the instance evaluation. GaIA was evaluated on two settings: i) Area 5 is used for testing and all others are utilized for training, ii) in the 6-fold cross validation each area is treated as the test set once. Experiments were also conducted on the ScanNet-v2 [8] which consisted of 1,613 scenes annotated with 20 classes. The dataset was split into 1,201 training, 312 validation, and 100 test scenes. To make it comparable to other approaches,the benchmark results are reported for the official test set.

Implementation details: For a sparse annotation setting, the points corresponding to the supervision ratio (1pt, 20 pts, and 1%) per class were labeled for each scene. GaIA was trained on a RTX A6000 GPU using the Adam optimizer with an initial learning rate of 0.01 and weight decay of 0.0001. The number of neighbors K was initially set to 16 and then reduced by 4 following the encoder block. The angular margin m was empirically determined to be

Table 1. Comparison with existing methods on S3DIS dataset.

Method	Supervision	Area 5	6-Fold
PointNet [31]	100%	41.1	47.6
PointNet++ [32]	100%	–	54.5
PointCNN [25]	100%	57.3	65.4
KPConv [41]	100%	67.1	70.6
MinkowskiNet [7]	100%	65.3	–
RandLA-Net [17]	100%	63.0	70.0
PointASNL [46]	100%	–	68.7
PointTransformer [50]	100%	70.4	**73.5**
Hou *et al*. [16]	100%	**72.2**	–
CBL [39]	100%	69.4	73.1
HybridCR [24]	100%	65.8	70.7
Zhang *et al*. [48]	1%	61.8	65.9
PSD [49]	1%	63.5	68.0
HybridCR [24]	1%	65.3	69.2
GaIA (Ours)	1%	**66.5**	**70.8**
Xu and Lee [45]	1pt (0.2%)	44.5	–
PSD [49]	1pt (0.03%)	48.2	–
HybridCR [24]	1pt (0.03%)	51.5	–
OTOC [26]	1pt (0.02%)	43.7	–
MIL Transformer [47]	1pt (0.02%)	51.4	–
GaIA (Ours)	1pt (0.02%)	**53.7**	–

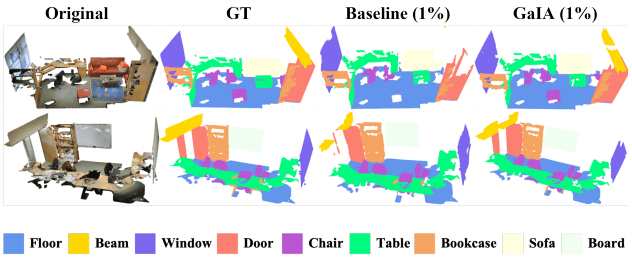

Figure 5. Comparison of qualitative results on S3DIS.

0.1 and re-scale factor s was set to 16. For S3DIS, we set the batch size to 150, and a batch size of 8 was employed for ScanNet-v2. GaIA was implemented using the PyTorch framework. As evaluation metric, the mean intersection over union (mIoU) was adopted.

4.2. Experimental results

S3DIS: GaIA was compared with existing fully supervised (100%) [31, 32, 25, 41, 17, 46, 7, 50, 16, 39] and weakly supervised (1pt and 1%) [45, 48, 49, 26, 47, 24] methods on S3DIS area 5 and 6-Fold, as listed in Tab. 1. Under the 1pt and 1% annotation on area 5, GaIA improved mIoU by 2.2%p and 1.2%p, respectively, compared to HybridCR [24]. In comparison of 6-Fold result on S3DIS, GaIA achieved the close performance on as that of the fully supervised state-of-the-art method [50] (-2.7%p) and surpassed the existing weakly supervised method [24] (+1.6%p). In Fig. 5, we visualized the qualitative results on S3DIS dataset. Compared to the baseline network excluded the Siamese branch, entropy blocks, and ArcPoint loss, GaIA precisely detected the classes, in particular, the beam and door. Additional visual comparison is reported in Supplementary.

Table 2. Comparison with existing methods on ScanNet-v2.

Method	Supervision	mIoU
PointNet++ [32]	100%	33.9
PointCNN [25]	100%	45.8
KPConv [41]	100%	68.4
RandLA-Net [17]	100%	64.5
PointASNL [46]	100%	66.6
MinkowskiNet [7]	100%	73.6
VMNet [19]	100%	74.6
BPNet [18]	100%	74.9
Mix3D [30]	100%	**78.1**
Zhang *et al*. [48]	1%	51.1
PSD [49]	1%	54.7
HybridCR [24]	1%	56.8
GaIA (Ours)	1%	**65.2**
Hou *et al*. [16]	20 pts / scene	55.5
OTOC [26]	20 pts / scene	59.4
MIL Transformer [47]	20 pts / scene	54.4
GaIA (Ours)	20 pts / scene	**63.8**
GaIA (Ours)	avg 7.8 pts / scene (1pt)	52.1

ScanNet-v2: The benchmark results of the ScanNet-v2 are listed in Tab. 2. Compared to the existing weakly supervised methods, HybridCR [24], GaIA improved mIoU by 8.4%p under the 1% annotation setting. Moreover, in limited annotations (LA) benchmark, GaIA outperformed Hou *et al*. [16] (+8.3%p), OTOC [26] (+4.4%p), and MIL Transformer [47] (+9.4%p). Remarkably, despite having more than 100× fewer annotations (1pt, 0.005%), GaIA exhibited a performance surpassing that of Zhang *et al*. (+1.0%p) and close to that of PSD (-2.6%p). As depicted in Fig. 6, GaIA was effective in epistemic uncertainty reduction compared to the baseline. When both networks exhibited similar segmentation results (cols 1 to 3), GaIA estimated the point cloud with higher reliability (rows 4 and 5). Although the segmentation results of both networks were unsatisfactory (cols 4 and 5), GaIA framework effectively alleviated epistemic uncertainty compared to the baseline.

5. Ablation study

5.1. Effectiveness of the proposed components

Ablation studies were conducted to analyze the contribution of each proposed component to performance gain. As listed in lines 1 and 2 of Tab. 3, the Siamese branch highly contributed to the performance gain compared with the baseline. This is because the Siamese branch is directly

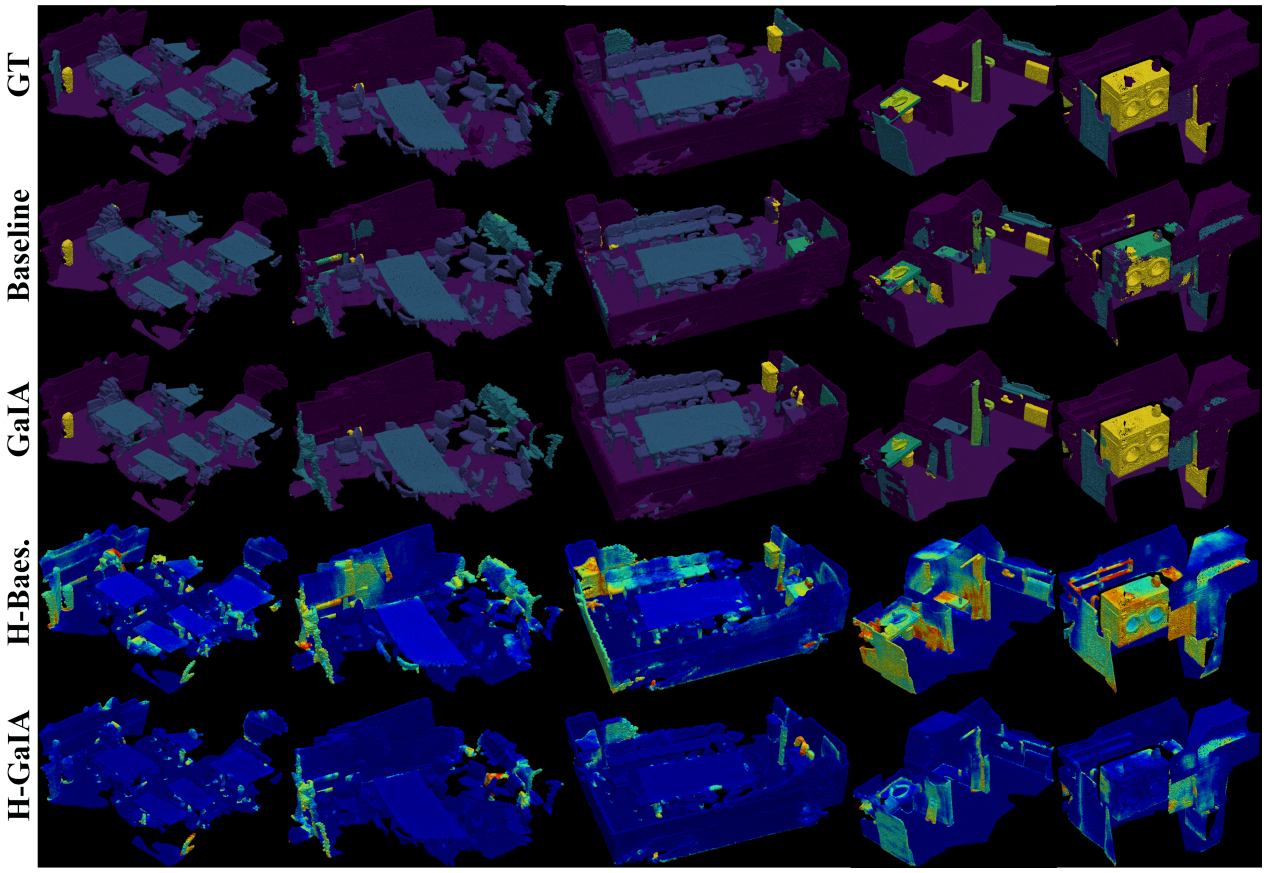

Figure 6. Comparison of qualitative results on ScanNet-v2 validation set. H denotes the entropy visualization.

Table 3. Comparison of the quantitative results on ScanNet-v2 validation set corresponding to the proposed components. (·) indicates officially measured test scores.

Base.	Sia.	EB.	AP.	AF.	1pt (0.005%)	1%
✓					33.2 (43.6)	42.7 (53.9)
✓	✓				37.4	49.5
✓	✓		✓		39.1	51.7
✓	✓	✓			40.8	52.4
✓	✓	✓		✓	41.1	52.8
✓	✓	✓	✓		41.9 (**52.1**)	54.9 (**65.2**)

involved in the optimization of unlabeled points, which occupy the largest proportion of the data. This tendency was also observed in a previous study [45]. When the proposed components were applied to the baseline with the Siamese branch, it was confirmed that the performance gain mainly originated from the entropy block (EB) compared with the ArcPoint loss (AP), as listed in lines 3 and 4. Under the 1pt annotation setting (0.005%), employing both components improved the performance by 8.7%p and 4.5%p compared to the baseline and Siamese network, respectively. We demonstrated more detailed analysis on the proposed components in Supplementary.

5.2. Effectiveness of ArcPoint loss

We conducted quantitative and qualitative experiments to validate the effectiveness of ArcPoint loss. In Tab. 3, ArcPoint loss was compared with the conventional softmax cross-entropy loss (line 4) and ArcFace (AF) loss (line 6). For a fair comparison, the Siamese branch with entropy blocks was employed. ArcFace outperformed the conventional softmax cross-entropy and L2 losses on 1pt and 1% annotations with 0.3%p and 0.4%p gains, respectively. Embedding on the hypersphere (e.g., ArcFace and ArcPoint) exhibited better performance compared with conventional losses. However, compared to ArcPoint loss, the gain of ArcFace was inevitably low because ArcFace could not be utilized in the optimization of the unlabeled points. In contrast, ArcPoint achieved improvement (i.e., 1.1%p and 2.5%p, respectively) by applying an additive angular margin to the unlabeled points containing high entropy. To verify the effectiveness of ArcPoint, we visualized the cosine similarities between the anchors and unannotated points. Excluding the entropy block from the GaIA, the net-

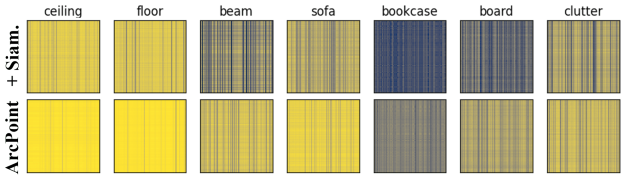

Figure 7. Comparison of cosine similarity. 50,000 anchors and unannotated points with high entropy were randomly sampled from the S3DIS dataset. In each heatmap, the rows and columns indicate anchors and unannotated points, respectively.

Table 4. Comparison of the performance on ScanNet-v2 validation set corresponding to $F(\gamma)$. $F(\gamma)$ and $(\cdot)$ denote the γ quantile of H and officially tested score, respectively.

Sup.	$0\uparrow$	$0.1\downarrow$	$0.3\downarrow$	$0.5\downarrow$	$0.5\uparrow$	$0.7\uparrow$	$F(0.9)\uparrow$
1%	52.1	49.3 (59.5)	49.5	49.8	53.0	53.8	**54.9 (65.2)**
1pt	39.2	37.1 (47.4)	37.7	38.1	40.6	41.1	**41.9 (52.1)**

work was compared with the baseline including the Siamese branch. In Fig. 7, dark-colored vertical lines indicate that the unlabeled points have low cosine similarity compared to other anchors. That is, the ArcPoint loss effectively optimizes the unlabeled points by employing both anchors and angular margin penalty.

5.3. Effectiveness of optimization with selective penalization

To validate the selective penalization for the optimization effect, which focuses on points with high entropy, $F(\gamma)$ values were experimented with in multiple ranges. In Tab. 4, it is observed that the more attention imposed on the points containing high entropy ($F(0.5$ to $0.9)\uparrow$), the higher the performance is compared with applying the penalty to the point with low entropy ($F(0.1$ to $0.3)\downarrow$). In particular, although all points including high entropy were equally treated in the penalization ($F(0)\uparrow$), the performance was reduced. This is because the points containing high entropy were considered under the same conditions, not selectively penalized. This tendency was consistently observed in both 1pt and 1% supervisions. In other words, it is effective to optimize the points, including high entropy, through selective penalization.

6. Discussion

This study aims to reduce epistemic uncertainty by using the entropy of each point for effective and precise point cloud semantic segmentation. This claim includes the premise that the lower entropy of the network, the more precise the semantic segmentation result is. However, 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. In Fig. 8, it is observed

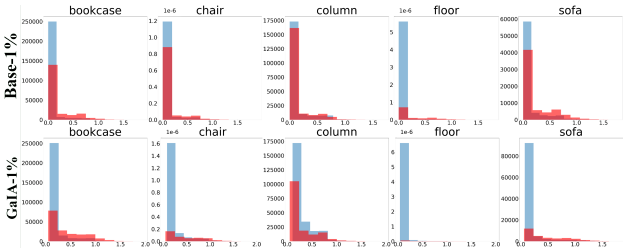

Figure 8. Comparison of entropy distribution with respect to prediction. The x and y axes indicate entropy and the number of samples, respectively. Distribution highlighted with red indicates distribution of false prediction.

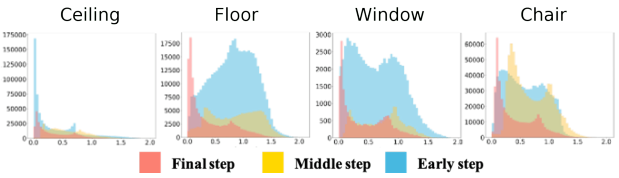

Figure 9. Comparison of point-wise entropy variation for false predictions during training steps. X-axis indicates the entropy.

that GaIA alleviated the number of false predictions with low entropy compared with the baseline which included the Siamese branch. In particular, GaIA reduced the false predictions for the classes floor and chair, by approximately 6 M and 0.6 M, respectively. Moreover, we observed the number of false predictions progressively decreased along with epistemic uncertainty reduction following the training steps, as depicted in Fig. 9. In other words, GaIA, which alleviates epistemic uncertainty, resulted in the reduction of false prediction with high reliability. Further analysis on the false prediction with high reliability is offered in Supplementary.

7. Conclusion

This study addressed epistemic uncertainty reduction in effective and precise point cloud semantic segmentation. Graphical information gain based attention network called GaIA was proposed. The graphical information gain and anchor-based additive angular margin loss called ArcPoint were main contributions of our approach. Specifically, the graphical information gain represents the reliable information by computing the relative entropy between the entropy of the target point and that of its neighborhoods. ArcPoint effectively optimizes unlabeled points containing high entropy. The experimental results of our method on two large-scale datasets demonstrated the improved performance of the proposed method compared with existing weakly supervised point cloud semantic segmentation methods.

References

[1] Dario Amodei, Chris Olah, Jacob Steinhardt, Paul Christiano, John Schulman, and Dan Mané. Concrete problems in ai safety. *arXiv preprint arXiv:1606.06565*, 2016.

[2] Iro Armeni, Ozan Sener, Amir R Zamir, Helen Jiang, Ioannis Brilakis, Martin Fischer, and Silvio Savarese. 3d semantic parsing of large-scale indoor spaces. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1534–1543, 2016.

[3] Jane Bromley, James W Bentz, Léon Bottou, Isabelle Guyon, Yann LeCun, Cliff Moore, Eduard Säckinger, and Roopak Shah. Signature verification using a "siamese" time delay neural network. *International Journal of Pattern Recognition and Artificial Intelligence*, 7(04):669–688, 1993.

[4] Angel X Chang, Thomas Funkhouser, Leonidas Guibas, Pat Hanrahan, Qixing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, et al. Shapenet: An information-rich 3d model repository. *arXiv preprint arXiv:1512.03012*, 2015.

[5] Mingmei Cheng, Le Hui, Jin Xie, and Jian Yang. Sspc-net: Semi-supervised semantic 3d point cloud segmentation network. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 1140–1147, 2021.

[6] Jiwoong Choi, Dayoung Chun, Hyun Kim, and Hyuk-Jae Lee. Gaussian yolov3: An accurate and fast object detector using localization uncertainty for autonomous driving. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 502–511, 2019.

[7] Christopher Choy, JunYoung Gwak, and Silvio Savarese. 4d spatio-temporal convnets: Minkowski convolutional neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3075–3084, 2019.

[8] Angela Dai, Angel X Chang, Manolis Savva, Maciej Halber, Thomas Funkhouser, and Matthias Nießner. Scannet: Richly-annotated 3d reconstructions of indoor scenes. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5828–5839, 2017.

[9] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4690–4699, 2019.

[10] Francis Engelmann, Theodora Kontogianni, Alexander Hermans, and Bastian Leibe. Exploring spatial context for 3d semantic segmentation of point clouds. In *Proceedings of the IEEE international conference on computer vision workshops*, pages 716–724, 2017.

[11] Di Feng, Lars Rosenbaum, and Klaus Dietmayer. Towards safe autonomous driving: Capture uncertainty in the deep neural network for lidar 3d vehicle detection. In *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, pages 3266–3273. IEEE, 2018.

[12] Yarin Gal et al. Uncertainty in deep learning. 2016.

[13] Benjamin Graham, Martin Engelcke, and Laurens Van Der Maaten. 3d semantic segmentation with submanifold sparse convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 9224–9232, 2018.

[14] George Francis Harpur. *Low entropy coding with unsupervised neural networks*. PhD thesis, Citeseer, 1997.

[15] George F Harpur and Richard W Prager. Development of low entropy coding in a recurrent network. *Network: computation in neural systems*, 7(2):277–284, 1996.

[16] Ji Hou, Benjamin Graham, Matthias Nießner, and Saining Xie. Exploring data-efficient 3d scene understanding with contrastive scene contexts. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15587–15597, 2021.

[17] Qingyong Hu, Bo Yang, Linhai Xie, Stefano Rosa, Yulan Guo, Zhihua Wang, Niki Trigoni, and Andrew Markham. Randla-net: Efficient semantic segmentation of large-scale point clouds. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11108–11117, 2020.

[18] Wenbo Hu, Hengshuang Zhao, Li Jiang, Jiaya Jia, and Tien-Tsin Wong. Bidirectional projection network for cross dimension scene understanding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14373–14382, 2021.

[19] Zeyu Hu, Xuyang Bai, Jiaxiang Shang, Runze Zhang, Jiayu Dong, Xin Wang, Guangyuan Sun, Hongbo Fu, and Chiew-Lan Tai. Vmnet: Voxel-mesh network for geodesic-aware 3d semantic segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 15488–15498, 2021.

[20] Li Jiang, Hengshuang Zhao, Shaoshuai Shi, Shu Liu, Chi-Wing Fu, and Jiaya Jia. Pointgroup: Dual-set point grouping for 3d instance segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4867–4876, 2020.

[21] Alex Kendall and Yarin Gal. What uncertainties do we need in bayesian deep learning for computer vision? *Advances in neural information processing systems*, 30, 2017.

[22] Gregory Koch, Richard Zemel, Ruslan Salakhutdinov, et al. Siamese neural networks for one-shot image recognition. In *ICML deep learning workshop*, volume 2. Lille, 2015.

[23] Tyler LaBonte, Carianne Martinez, and Scott A Roberts. We know where we don't know: 3d bayesian cnns for credible geometric uncertainty. *arXiv preprint arXiv:1910.10793*, 2019.

[24] Mengtian Li, Yuan Xie, Yunhang Shen, Bo Ke, Ruizhi Qiao, Bo Ren, Shaohui Lin, and Lizhuang Ma. Hybridcr: Weakly-supervised 3d point cloud semantic segmentation via hybrid contrastive regularization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 14930–14939, June 2022.

[25] Yangyan Li, Rui Bu, Mingchao Sun, Wei Wu, Xinhan Di, and Baoquan Chen. Pointcnn: Convolution on x-transformed points. In *NeurIPS*, 2018.

[26] Zhengzhe Liu, Xiaojuan Qi, and Chi-Wing Fu. One thing one click: A self-training approach for weakly supervised 3d semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1726–1736, 2021.

[27] Andrey Malinin and Mark Gales. Predictive uncertainty estimation via prior networks. *Advances in neural information processing systems*, 31, 2018.

[28] Kaichun Mo, Shilin Zhu, Angel X Chang, Li Yi, Subarna Tripathi, Leonidas J Guibas, and Hao Su. Partnet: A large-scale benchmark for fine-grained and hierarchical part-level 3d object understanding. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 909–918, 2019.

[29] Tanya Nair, Doina Precup, Douglas L Arnold, and Tal Arbel. Exploring uncertainty measures in deep networks for multiple sclerosis lesion detection and segmentation. *Medical image analysis*, 59:101557, 2020.

[30] Alexey Nekrasov, Jonas Schult, Or Litany, Bastian Leibe, and Francis Engelmann. Mix3d: Out-of-context data augmentation for 3d scenes. In *2021 International Conference on 3D Vision (3DV)*, pages 116–125. IEEE, 2021.

[31] Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 652–660, 2017.

[32] Charles R Qi, Li Yi, Hao Su, and Leonidas J Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *arXiv preprint arXiv:1706.02413*, 2017.

[33] Joaquin Quiñonero-Candela, Masashi Sugiyama, Anton Schwaighofer, and Neil D Lawrence. *Dataset shift in machine learning*. Mit Press, 2008.

[34] Jacob C Reinhold, Yufan He, Shizhong Han, Yunqiang Chen, Dashan Gao, Junghoon Lee, Jerry L Prince, and Aaron Carass. Validating uncertainty in medical image translation. In *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, pages 95–98. IEEE, 2020.

[35] Abhijit Guha Roy, Sailesh Conjeti, Nassir Navab, Christian Wachinger, Alzheimer's Disease Neuroimaging Initiative, et al. Bayesian quicknat: Model uncertainty in deep whole-brain segmentation for structure-wise quality control. *NeuroImage*, 195:11–22, 2019.

[36] Philipp Seeböck, José Ignacio Orlando, Thomas Schlegl, Sebastian M Waldstein, Hrvoje Bogunović, Sophie Klimscha, Georg Langs, and Ursula Schmidt-Erfurth. Exploiting epistemic uncertainty of anatomy segmentation for anomaly detection in retinal oct. *IEEE transactions on medical imaging*, 39(1):87–98, 2019.

[37] Claude Elwood Shannon. A mathematical theory of communication. *The Bell system technical journal*, 27(3):379–423, 1948.

[38] Lewis Smith and Yarin Gal. Understanding measures of uncertainty for adversarial example detection. *arXiv preprint arXiv:1803.08533*, 2018.

[39] Liyao Tang, Yibing Zhan, Zhe Chen, Baosheng Yu, and Dacheng Tao. Contrastive boundary learning for point cloud segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8489–8499, 2022.

[40] Lyne P. Tchapmi, Christopher Bongsoo Choy, Iro Armeni, JunYoung Gwak, and Silvio Savarese. Segcloud: Semantic segmentation of 3d point clouds. *2017 International Conference on 3D Vision (3DV)*, pages 537–547, 2017.

[41] Hugues Thomas, Charles R. Qi, Jean-Emmanuel Deschaud, Beatriz Marcotegui, François Goulette, and Leonidas J. Guibas. Kpconv: Flexible and deformable convolution for point clouds. *Proceedings of the IEEE International Conference on Computer Vision*, 2019.

[42] Haiyan Wang, Xuejian Rong, Liang Yang, Jinglun Feng, Jizhong Xiao, and Yingli Tian. Weakly supervised semantic segmentation in 3d graph-structured point clouds of wild scenes. *arXiv preprint arXiv:2004.12498*, 2020.

[43] Yue Wang, Yongbin Sun, Ziwei Liu, Sanjay E Sarma, Michael M Bronstein, and Justin M Solomon. Dynamic graph cnn for learning on point clouds. *Acm Transactions On Graphics (tog)*, 38(5):1–12, 2019.

[44] Jiacheng Wei, Guosheng Lin, Kim-Hui Yap, Tzu-Yi Hung, and Lihua Xie. Multi-path region mining for weakly supervised 3d semantic segmentation on point clouds. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4384–4393, 2020.

[45] Xun Xu and Gim Hee Lee. Weakly supervised semantic point cloud segmentation: Towards 10x fewer labels. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13706–13715, 2020.

[46] Xu Yan, Chaoda Zheng, Zhen Li, Sheng Wang, and Shuguang Cui. Pointasnl: Robust point clouds processing using nonlocal neural networks with adaptive sampling. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5589–5598, 2020.

[47] Cheng-Kun Yang, Ji-Jia Wu, Kai-Syun Chen, Yung-Yu Chuang, and Yen-Yu Lin. An mil-derived transformer for weakly supervised point cloud segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11830–11839, June 2022.

[48] Yachao Zhang, Zonghao Li, Yuan Xie, Yanyun Qu, Cuihua Li, and Tao Mei. Weakly supervised semantic segmentation for large-scale point cloud. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 3421–3429, 2021.

[49] Yachao Zhang, Yanyun Qu, Yuan Xie, Zonghao Li, Shanshan Zheng, and Cuihua Li. Perturbed self-distillation: Weakly supervised large-scale point cloud semantic segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 15520–15528, 2021.

[50] Hengshuang Zhao, Li Jiang, Jiaya Jia, Philip HS Torr, and Vladlen Koltun. Point transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 16259–16268, 2021.