

Improving 3D Semi-supervised Learning by Effectively Utilizing All Unlabelled Data

Sneha Paul, Zachary Patterson, and Nizar Bouguila

Concordia Institute for Information Systems Engineering,
Concordia University, Montreal, Canada
sneha.paul@mail.concordia.ca
{zachary.patterson,nizar.bouguila}@concordia.ca

Abstract. Semi-supervised learning (SSL) has shown its effectiveness in learning effective 3D representation from a small amount of labelled data while utilizing large unlabelled data. Traditional semi-supervised approaches rely on the fundamental concept of predicting pseudo-labels for unlabelled data and incorporating them into the learning process. However, we identify that the existing methods do not fully utilize all the unlabelled samples and consequently limit their potential performance. To address this issue, we propose AllMatch, a novel SSL-based 3D classification framework that effectively utilizes all the unlabelled samples. AllMatch comprises three modules: (1) an adaptive hard augmentation module that applies relatively hard augmentations to the high-confident unlabelled samples with lower loss values, thereby enhancing the contribution of such samples, (2) an inverse learning module that further improves the utilization of unlabelled data by learning what not to learn, and (3) a contrastive learning module that ensures learning from all the samples in both supervised and unsupervised settings. Comprehensive experiments on two popular 3D datasets demonstrate a performance improvement of up to 11.2% with 1% labelled data, surpassing the SOTA by a significant margin. Furthermore, AllMatch exhibits its efficiency in effectively leveraging all the unlabelled data, demonstrated by the fact that only 10% of labelled data reaches nearly the same performance as fully-supervised learning with all labelled data. The code of our work is available at: github.com/snehaputul/AllMatch.

Keywords: Semi-supervised Learning · Point Cloud · 3D Vision

1 Introduction

In the realm of 3D computer vision, Point Cloud applications are on the rise, finding extensive use in diverse domains. This increasing popularity of Point Cloud applications across various 3D computer vision domains has heightened the demand for annotated data, a resource-intensive and expensive task [16]. To tackle this

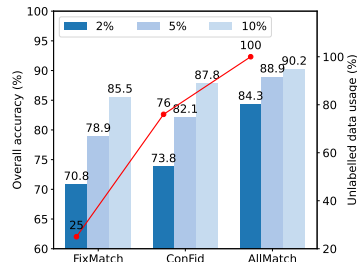


Fig. 1: Performance in comparison to unlabelled data usage.

issue, 3D semi-supervised learning (SSL) emerges as a promising avenue. By capitalizing on scarce labelled data in conjunction with abundant unlabelled data, this approach has demonstrated its efficacy in different computer vision applications, including 3D object recognition. Most popular forms of SSL (e.g. [21], [28]) involve predicting pseudo-labels for the unlabelled data and using *only* the high-confident predictions for learning. However, due to the hand-designed nature of the high-confidence-based filtering, most of the unlabelled samples never contribute to the learning (Figure 1).

Some of the more recent works propose different techniques to increase the number of unlabelled samples the model learns from. For example, FlexMatch [28] introduced a dynamic threshold strategy, where the threshold value is adjusted according to the model’s learning status using curriculum learning. SoftMatch [3] proposed a trade-off between quality and quantity to increase unlabelled data usage. While there has been some progress in this direction, these methods are specially developed for the image domain and have not been explored in the context of 3D SSL. Furthermore, as shown in Figure

2 (left) for 3D SSL, FixMatch utilizes about 25% of the unlabelled data, while the current state-of-the-art (SOTA) method, ConFid [8], utilizes about 76% of the unlabelled data. Moreover, we find that the correctly predicted pseudo-labels do not contribute much towards the learning, as they have higher confidence in prediction and lower loss (Figure 2 (right)).

In this work, we propose a novel semi-supervised framework named AllMatch that handles the above-mentioned problems and ensures learning from *all* the unlabelled data. AllMatch consists of three modules, all of which improve different aspects of increasing unlabelled data usage. While the concepts of these modules are not entirely new, they have never been explored in 3D SSL, especially with the focus on increasing the utilization of unlabelled data. The modules in AllMatch are: (a) An adaptive hard augmentation(AHA) module that increases the augmentation strength of the high-confident unlabelled samples with lower loss. This module ensures better utilization of high-confident samples that the model has already learned, thus leaving no further contribution to the model’s learning process. (b) An inverse learning module that further ensures better utilization of unlabelled samples by assigning inverse pseudo labels to the low-confident predictions and learning *what not to learn*. (c) A contrastive learning (CL) module that learns from all the unlabelled data in unsupervised learning settings. This module ensures that the samples excluded from pseudo-label-based learning also

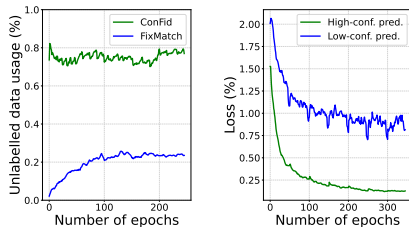


Fig. 2: (left) Pseudo label utilization over the training epochs by FixMatch and ConFid. While ConFid does better utilization than FixMatch, none properly utilizes all unlabelled data. (right) Loss over training for high and low-confident pseudo-labels. The high-confident samples are selected for unsupervised learning, but low loss results in marginal contribution to learning.

contribute to the learning process by participating in CL. Additionally, we utilize a modified supervised contrastive loss [12] for learning from the labelled data in a similar learning setting as the unlabelled data.

We evaluate AllMatch on two popular point cloud datasets, ModelNet40 [25] and ScanObjectNN [23], with different amounts of labelled sets. We compare AllMatch with the existing SSL methods and find that AllMatch outperforms the existing SOTA by up to 11.2% and 10.45% on ScanObjectNN and ModelNet40 datasets, respectively, with a minimal amount of labelled data. AllMatch also shows superior performance with a lower amount of labelled set, indicating its effectiveness in utilizing all unlabelled samples. With our proposed method, only 10% of the labelled data performs almost as well as in a fully supervised setting with all the labelled data. Furthermore, our method is more computationally efficient than the previous SOTA, ConFid [8], which requires 500 epochs of training compared to 250 in ours. We also conduct an in-depth ablation study and sensitivity analysis demonstrating the importance of all three components for AllMatch’s superior performance. Here, the AHA module shows the highest impact on the model’s performance, followed by inverse learning and contrastive learning modules. Our main contributions to this work can be summarized as follows:

- We propose a novel semi-supervised 3D classification framework named AllMatch that effectively utilizes the whole unlabelled set to improve the performance.
- We introduce an adaptive hard augmentation module that efficiently utilizes the higher-confident samples with lower loss and an inverse learning module to further boost the utilization of unlabelled samples by learning from low-confident predictions.
- We introduce a contrastive learning module to ensure learning from the remainder of the unlabelled data that are not utilized by pseudo-label-based learning.
- We improve the SOTA for 3D semi-supervised learning by 11.2%, and 10.45% on the ScanobjectNN and ModelNet40 datasets, respectively. We also show that AllMatch achieves almost similar results in a fully-supervised setting with only 10% of labelled data.

2 Related Work

2.1 Semi-supervised Learning

There has been significant progress in semi-supervised learning in recent years, mostly for image representation learning. Most dominant approaches in semi-supervised learning can be divided into two main categories: consistency regularization [22, 26] and entropy minimization [1, 2, 21]. The consistency regularization learns by enforcing consistency in the prediction of the model under different perturbations. On the other hand, entropy minimization learns by reducing the entropy in the prediction of the model on the unlabelled data. One of

the most prominent methods in this category is FixMatch [21]. The basic concept of FixMatch is to predict the pseudo-label for the unlabelled data from a weakly augmented of the input and then utilize this as the supervision for a strongly augmented sample if the confidence of the pseudo-label is high. One main issue with this approach is its reliance on the quality of the pseudo-labels. Also, the fixed thresholding nature results in very low utilization of the unlabelled data as the low confidence prediction does not contribute towards the learning process. Recently, several methods have been proposed to increase the utilization of the unlabelled data by proposing variants of the thresholding concept, including Dash [27], FlexMatch [28], CoMatch [15], SimMatch [29], ConMatch [13].

Dash [27] selects unlabelled samples based on a criterion where their cross-entropy loss is smaller than an adjustable threshold value. FlexMatch [28] is another popular work that utilizes curriculum learning to adapt the threshold value, taking into account the class-wise learning status. This dynamic threshold is then applied to select pseudo-labels with high confidence. CoMatch [15] introduces a co-training framework that involves the interaction between class probabilities and embeddings. The embeddings impose a smoothness constraint on class probabilities to enhance the quality of pseudo-labels. These refined pseudo-labels are used as targets for training both the classification head with cross-entropy loss and the projection head with a graph-based contrastive loss. SimMatch [29] adeptly aligns similarity relationships at dual levels, utilizing a memory buffer for annotated examples to amplify the synergy between semantic and instance pseudo-labels. ConMatch [13] elevates model efficacy through the integration of confidence-guided consistency regularization. This approach tackles the hurdle of learning from a restricted pool of labelled data by promoting consistent predictions on unlabelled data, all the while taking into account the confidence levels associated with those predictions. SequenceMatch [17] enhances semi-supervised learning by minimizing prediction distribution divergence between weakly and strongly augmented examples. It underscores the significance of diverse augmentations and consistency constraints in the learning process. CHMatch [24] integrates instance-level prediction matching and contrastive graph-level matching through the use of a memory-bank-based adaptive threshold strategy. It employs hierarchical label-guided graph matching to enhance contrastive feature learning, providing a more robust approach to feature representation. While semi-supervised methods exhibit promising performance in the image domain, a notable gap exists in the SSL literature for point clouds, particularly in the context of classification tasks.

2.2 Semi-supervised Learning in Point Cloud

Semi-supervised learning has recently been explored by a few prior works on the 3D point cloud classification. For example, [4] introduces a semi-supervised learning framework, leveraging consistency constraints to ensure uniform predictions on both original and perturbed point clouds. This approach mitigates overfitting the limited labelled data available. Moreover, it incorporates pseudo-label generation to provide high-quality labels for unlabelled point clouds, thereby enhanc-

ing the supervision of the discriminative model. On the other hand, [7] presents a multi-modal semi-supervised learning framework featuring an instance-level consistency constraint and a multi-modal contrastive prototype (M2CP) loss. M2CP works towards minimizing intra-class feature variations by optimizing the distance to class prototypes for each object. ConFid [8] is another prominent work that specifically addresses class-imbalanced data scenarios, which is a prevalent challenge in the context of point clouds. ConFid takes into account class-level confidence through a resampling strategy designed to prevent bias towards high-confident classes. Nevertheless, all existing methods fall short of fully harnessing the potential of unlabelled data. This suggests a potential research gap in the realm of SSL-based 3D classification that maximizes the utilization of all available unlabelled data.

3 Method

3.1 Preliminaries on SSL

Semi-supervised methods learn from a small amount of labelled data and a relatively large amount of unlabelled data in supervised and unsupervised settings, respectively. Let, $X_{lb} = (x_i, y_i)_{i=1}^n$ and $X_{ulb} = (x_i)_{i=1}^N$ represent the labelled and unlabelled sets with $n \ll N$. Here, $x_i \in \mathcal{R}^{m \times 3}$ represents a 3D point cloud with m points, and y_i is the corresponding class label. Consequently, the semi-supervised learning can be formulated as:

$$\min_{\theta} \left[\sum_{(x_i, y_i) \in X_{lb}} \mathcal{L}_{sup}(x_i, y_i, \theta) + \omega \sum_{x_i \in X_{ulb}} \mathcal{L}_{unsup}(x_i, \theta) \right], \quad (1)$$

Here, θ represents the learnable parameters, \mathcal{L}_{sup} and \mathcal{L}_{unsup} are supervised loss (cross-entropy loss, $H(x_i, y_i)$) and unsupervised loss. ω is the weight that balances the importance of the unsupervised loss.

Unsupervised learning is the key component of semi-supervised learning, exhibiting variations across different SSL methods. Consistency regularization is one of the common unsupervised losses that is learned by forcing the predictions of the same sample to be consistent under different perturbations (augmentations) [1, 2, 20]. Another popular form of unsupervised learning is entropy minimization [21, 28], which learns from the unlabelled data by predicting a pseudo-label for each sample. Our method is built over the entropy minimization concept from FixMatch [21]. More specifically, FixMatch first generates a class prediction for an unlabelled sample from a weakly augmentation of the sample and considers it as a pseudo-label if the confidence of the prediction is above a pre-defined threshold. Finally, these high-confident pseudo-labels are used as supervision for a strong augmentation of the sample.

Let, $x_w^{(i)} = \phi(x_{ulb}^i)$ and $x_s = \omega(x_{ulb}^i)$ be the weakly augmented and strongly augmented samples, and $p_m(p_z(x))$ be the output of the model, where, p_z is the encoder and p_m is the classifier. The unsupervised loss of FixMatch can be defined as follows:

$$\mathcal{L}_{unsup} = \frac{1}{\mu B} \sum_{b=1}^{\mu B} \mathbb{1}(\max(p_m(p_z(x_w^{(i)}))) \geq \tau) \cdot H(\operatorname{argmax}(p_m(p_z(x_w^{(i)}))), p_m(p_z(x_s^{(i)}))), \quad (2)$$

where τ is the fixed confidence threshold, μ is the ratio of unlabelled to labelled batch size, B is the batch size and H represents the cross-entropy loss function.

Nevertheless, a fixed (high) threshold results in under-utilization of the unlabelled data as the model does not utilize the low-confident samples. To address this, FlexMatch [28] and ConFid [8] propose dynamic pseudo labelling to adjust the threshold value considering the class-wise learning status by curriculum learning. In this context, the learning status refers to the ratio of predicted labels to the target label of a class. Specifically, the threshold value is low for high learning status classes to encourage the learning process, and vice-versa. Yet, the model does not utilize all the available unlabelled samples. Following, we describe our proposed solution, AllMatch, which effectively utilizes all the unlabelled samples for learning.

3.2 AllMatch

AllMatch consists of three components, all aimed towards improving the utilization of the unlabelled data: adaptive hard augmentation, inverse learning and contrastive learning. Following, we discuss each of the components in detail.

Adaptive hard augmentation (AHA). One limitation of the high confidence-based thresholding method is that the selected (highly confident) samples are already well learned by the model [9]. As a result, the loss for these samples is very low and, therefore, does not contribute towards the learning of the model. Following [9], we refer to these high-confident samples as *easy samples* and propose an adaptive hard augmentation (AHA) module to further utilize the easy samples. The basic idea is to apply relatively hard augmentations on the easy samples to introduce further regularization while applying the usual strong augmentations for the remaining samples. Note that the alternation between extra hard augmentation and strong augmentation plays a crucial role, as excess augmentation can potentially have a detrimental effect on the model. While the role of hard augmentation is extensively studied in the image semi-supervised learning literature [9], it’s not well explored in 3D point clouds.

We define the easy samples based on a historical loss computed until the current epoch t . More specifically, we calculate the historical loss for the $i - th$ sample as the exponential moving average (EMA):

$$\mathcal{H}_i^t = (1 - \kappa) \cdot \mathcal{H}_i^{t-1} + \kappa \cdot l_i^t, \quad (3)$$

where, l_i^t is computed as:

$$l_i^t = H(\operatorname{argmax}(p_m(p_z(x_w^{(i)}))), p_m(p_z(x_s^{(i)}))), \quad (4)$$

The lower historical loss for sample i indicates an easy sample on which we apply additional hard augmentations. To divide the samples into easy and hard samples based on the historic loss, we utilize the OTSU [18] function that returns a binary indicator \mathcal{T}_i for each sample. Here, $\mathcal{T}_i = 0$ indicates easy sample, and $\mathcal{T}_i = 1$ indicates hard sample. Finally, for an easy sample, we apply a series of transformations to act as hard augmentation, including random scaling, rotation, translation and jitters. The AHA module can be represented as:

$$Augmented\ sample(x_i) = \begin{cases} \mathcal{K}(x_i), \mathcal{T}_i = 1 \\ \mathcal{K}'(x_i), \mathcal{T}_i = 0 \end{cases}, \quad (5)$$

Here, $\mathcal{K}'(x_i)$ represents the extra hard augmentation, while $\mathcal{K}(x_i)$ represents hard augmentation. In the implementation, we start applying the AHA module after certain epochs of training, defined by ‘warm-up epochs.’ Since the loss is relatively high for all samples at the start of training, adding the AHA from the start of training does not add any benefit.

Inverse learning. Another challenge in pseudo-labeling-based semi-supervised learning comes from the ambiguity in the predictions of the models. More specifically, the samples with low confidence (high ambiguity) are discarded by the pseudo-labeling concept. These low-confident samples do not contribute to the model’s learning process, as pseudo-labels cannot be assigned to them. Consequently, a significant number of unlabelled samples go unused. However, these neglected unlabelled samples can be re-purposed to optimize the model inversely by learning *what not to learn*.

As demonstrated in [6] for SSL in the image domain, after a certain number of training iterations, the correct classes are situated in one of the top- k predictions (i.e., the k classes with the highest predicted probabilities) of the model. In essence, the model is extremely confident in its predictions for containing the right class in one of these top- k classes. Therefore, it is advantageous to treat the remaining classes (those outside the top- k predictions) as *wrong* class for the input. Motivated by [6], we incorporate inverse learning, which learns to avoid predicting such classes.

For Inverse learning, following the convention of FixMatch, we first predict a pseudo-label from a weakly augmented sample and find the value of k for which the model achieves 100% accuracy on the strongly augmented samples at the current epoch. Then, we sort the classes in order of highest to lowest confidence using a ranking function, *Rank*. Finally, for the classes with $Rank > k$ we perform inverse learning on the strongly augmented samples as:

$$\mathcal{L}_{inv} = -\frac{1}{B} \sum_{i=1}^B \sum_{c=1}^C \mathbb{1}[Rank(p_m(p_z(x_w^{(i)}))) > k] \cdot \log(1 - p_m(p_z(x_s^{(i)}))), \quad (6)$$

Contrastive learning (CL). While the AHA and inverse learning module increase the utilization of the unlabelled data, the samples that are neither very

high-confident nor very low-confident are not utilized by any of the above modules. To ensure learning from the remainder of the sample, we incorporate a contrastive loss on the unlabelled set that learns from all the samples in an unsupervised setting. Contrastive methods learn from positive and negative samples, with the aim of pushing the positive pairs closer in the embedding space and moving them away from the negative sample. In this context, positive samples are the perturbation of the same sample, while all other samples are considered negative. Contrastive learning has shown remarkable success in learning representation in different computer vision domains [5, 11, 20]. Since contrastive learning does not require labels, it ensures total usage of unlabelled data.

For each unlabelled sample, $x_{ulb}^{(i)} \in X_{ulb}$, a weak perturbation ϕ and a strong perturbation ω is applied to generate a positive pair. In our case, we apply rotation and random scale as weak augmentation, and for strong augmentation, a couple of different augmentations (e.g. jitter, translation, rotation, random scale) are randomly applied. The augmented samples are then fed to the encoder and projection head that generates the embedding $z_i = p_z(y|x_{ulb_i})$, where $[z_{w_i}, z_{s_i}] \in z_i$. Finally, the unsupervised contrastive loss [5] function can be described as:

$$\mathcal{L}_{con} = -\frac{1}{2B} \sum_{i=1}^{2B} \log \frac{\exp(z_i, z_{\kappa(i)}/\tau)}{\sum_{k=1}^{2B} \mathbb{1}_{[k \neq i]} \exp(z_i, z_k/\tau)}, \quad (7)$$

where, $\kappa(i)$ is the index of the second augmented sample, $\mathbb{1}_{[k \neq i]}$ is an indicator function which returns 1 when k is not equal to i , and 0 otherwise. τ is a temperature parameter. Note that the pair samples are the same as the ones used for the unsupervised loss mentioned above. So, the contrastive loss does not increase the computational complexity.

Additionally, we utilize a modified supervised version of contrastive loss (supervised contrastive loss [12]) on the labelled set. Like the unsupervised contrastive loss, the supervised contrastive loss learns from the augmented positive and negative samples. However, the additional label information benefits the model by pushing together samples that belong to the same class. The supervised contrastive loss can be formulated as follows:

$$\mathcal{L}_{supcon} = \sum_{i=1}^{2B} -\frac{1}{2B_{y_i} - 1} \sum_{j=1}^{2B} \mathbb{1}_{[i \neq j]} \cdot \mathbb{1}_{[y_i = y_j]} \cdot \log \frac{\exp(z_i, z_j/\tau)}{\sum_{k=1}^{2B} \mathbb{1}_{[k \neq i]} \exp(z_i, z_k/\tau)}, \quad (8)$$

where, y_i represents class label of sample i . $\mathbb{1}_{[y_i = y_j]}$ is an indicator function that returns 1 when the samples i and j belongs to same class label. $2B_{y_i} - 1$ represents number of positive samples from class y_i .

3.3 Total Loss

Finally, the total loss of AllMatch is the sum of all the described above:

$$\mathcal{L} = \mathcal{L}_{sup} + \mathcal{L}_{unsup} + \alpha \cdot \mathcal{L}_{supcon} + \beta \cdot \mathcal{L}_{con} + \gamma \cdot \mathcal{L}_{inv}, \quad (9)$$

Algorithm 1: AllMatch Algorithm

Input: Labelled batch $\mathcal{B}_l = \{(x_i, y_i)\}^M$, unlabelled batch $\mathcal{B}_u = \{u_i\}^N$

- 1 Apply AHA to \mathcal{B}_l as Eq. 5
- 2 $l_{sup} = \mathcal{L}_{sup}(\mathcal{B}_l)$
- 3 $l_{unsup} = \mathcal{L}_{unsup}(\mathcal{B}_u)$ // via Eq. 2
- 4 $l_{con} = \mathcal{L}_{con}(\mathcal{B}_u)$ // via Eq. 7
- 5 $l_{supcon} = \mathcal{L}_{supcon}(\mathcal{B}_l)$ // via Eq. 8
- 6 **for** $i \leftarrow 1$ **to** N **do**
- 7 **Compute** the rank using, $Rank(p_m(p_z(x_w^{(i)})))$ **for** $c \leftarrow 1$ **to** C **do**
- 8 **if** $Rank(p_m(p_z(x_w^{(i)}))) > k$ **then**
- 9 $l_{inv} += \log(1 - p_m(p_z(x_s^{(i)})))$
- 10 **end**
- 11 **end**
- 12 **Compute** and update historical loss, \mathcal{H}_i // via Eq. 3
- 13 **Update** mark $\tau_i = \mathbb{1}(\mathcal{H}_i \leq \text{OTSU}(\mathcal{H}_i))$
- 14 **end**
- 15 $l_{total} = l_{sup} + l_{unsup} + \alpha \cdot l_{supcon} + \beta \cdot l_{con} + \gamma \cdot l_{inv}$

Here, α, β and γ are loss factors balancing the importance of each loss. We find optimal values of each factor from the empirical study described in Section 4.5.

4 Experiments

In this section, we discuss the experimental results of our study on the ModelNet40 and ScanobjectNN datasets. First, we discuss the main result, followed by an in-depth ablation study of our proposed modules and sensitivity analysis of different module-specific parameters.

4.1 Implementation Details

We evaluate AllMatch on two popular point cloud datasets, ModelNet40 [25] and ScanObjectNN [23]. For the implementation details and hyper-parameters, we nearly follow the existing SOTA, ConFid [8], for a fair comparison. We use PointTransformer [10] as the encoder. We train the model with SGD optimizer with a learning rate of 0.00005 for 350 epochs. The unlabelled to labelled batch size ratio is 4, and the labelled batch size is 24. We utilize rotation and random scale as weak augmentation and random scaling, rotation, translation and jitter as strong augmentation. Note that, as a geometric entity, the underline representation of the point clouds must remain invariant under specific transformations [19]. Thus, we only chose the augmentation methods that do not violate this special property of the point clouds. For the AHA module, various combinations of strong augmentation are applied as an additional hard augmentation to the easy samples. A detailed sensitivity analysis on the augmentation strength is discussed in Section 4.5.

Table 1: Comparison of AllMatch with the SOTA methods in 3D object classification task on ModelNet40 dataset with different amounts of labelled sets. Results of the prior works are adopted from [8]. Here, ‘Overall Acc’ and ‘Mean Acc’ are the overall accuracy and the average of per-class accuracy.

Method	2%		5%		10%	
	Overall Acc	Mean Acc	Overall Acc	Mean Acc	Overall Acc	Mean Acc
PCT [10]	71.1	61.0	77.1	69.2	84.6	77.2
PL [14]	69.7	59.6	78.3	69.0	85.1	77.7
Flex-PL [28]	66.7	54.9	74.2	62.3	83.2	70.3
ConFid-PL [8]	74.4	61.9	80.6	73.5	86.5	80.4
FixMatch [21]	70.8	62.7	78.9	71.1	85.5	79.4
Dash [27]	71.5	63.0	79.7	71.8	85.9	80.1
FlexMatch [28]	70.1	61.2	80.5	70.4	86.2	78.7
ConFid-Match [8]	73.8	64.1	82.1	74.3	87.8	82.5
AllMatch	84.3	82.6	89.0	87.9	90.2	89.8

4.2 Main Result

The overall comparison to the previous works is presented in Tables 1 and 2 for ModelNet-40 and ScanObjectNN datasets, respectively. Here, PL [14], FixMatch [21], Dash [27], and FlexMatch [28] were initially proposed for the image domain and later adopted to the PointCloud, whereas ConFid [8] is specifically designed for the point cloud. The results for other prior works are directly adopted from the reported results in ConFid [8]. Following the experimental setups of previous literature on this task, we report the results for different amounts of labelled samples. More specifically, we evaluate the performance on three different sizes of labelled sets: 2%, 5%, and 10% for the ModelNet40 dataset and 1%, 2%, and 5% for the ScanObjectNN dataset. We report the overall accuracy and the mean of class-wise accuracy (mean accuracy) on the test set.

As we observe from these tables, AllMatch outperforms SOTA by a significant margin across all sizes of unlabelled sets, demonstrating substantial improvement. For example, with only 2% labelled data from ModelNet40 (Table 1), AllMatch outperforms ConFid by a notable 10.48% and 18.52% in overall and mean accuracy. While we also see considerable improvements for 5% and 10% data, the biggest improvement is observed when the least amount of labelled data (2%) is utilized. This indicates the effectiveness of AllMatch in learning from a very small amount of labelled data.

Table 2 illustrates a similar on the ScanObjectNN dataset. With only 1% of the labelled set, AllMatch outperforms the SOTA by a superior 11.20% on overall accuracy and 14.67% on mean accuracy. The maximum performance gain is again observed for the least amount of labelled samples.

Table 2: Comparison of AllMatch with the SOTA methods in 3D object classification task on ScanObjectNN dataset with different amounts of labelled sets. Results of the prior works are adopted from [8]. Here, ‘Overall Acc’ and ‘Mean Acc’ are the overall accuracy and the average of per-class accuracy.

Method	1%		2%		5%	
	Overall Acc	Mean Acc	Overall Acc	Mean Acc	Overall Acc	Mean Acc
PCT [10]	32.1	26.1	44.7	36.5	56.6	50.0
PL [14]	31.2	25.8	47.5	38.6	58.1	51.5
Flex-PL [28]	29.2	24.2	47.2	37.6	60.1	51.9
Confid-PL [8]	32.6	27.1	48.8	41.5	63.1	55.2
FixMatch [21]	33.5	27.6	47.4	39.9	59.4	52.4
Dash [27]	35.1	29.3	50.3	44.1	62.8	60.3
FlexMatch [28]	34.2	26.2	48.5	39.7	63.4	57.2
ConFid-Match [8]	38.2	32.7	57.0	48.6	69.4	65.5
AllMatch	49.4	47.4	60.8	58.9	77.5	76.0

4.3 Ablation Study

In this section, we perform a detailed ablation study to understand the importance of different components in our proposed method. We start with the main ablation on the 3 main components of AllMatch, then perform another ablation on the contrastive learning approach.

Main ablation. Table 3 presents the results for the main ablation study. Here, we ablate all the combinations of the three main components of

AllMatch: adaptive hard augmentation (AHA), inverse learning and contrastive learning (CL). First, we drop these modules individually and then in groups to investigate their significance on overall performance. We perform this study with 2% of the labelled data and report the overall accuracy (OA).

As we observe from this table, when we drop the AHA module individually, the method’s performance declines by 4.29%. This is considerably higher than the drops observed by dropping inverse learning or contrastive loss, making AHA the most important module in AllMatch. We find the inverse learning to be the second most important component, removing which results in a 3.77% drop in accuracy. Finally, we find that contrastive learning also plays a critical role. Removing the contrastive learning also results in a 1.34% drop in accuracy. We find similar trends when two modules are removed at a time. We observe a performance drop of 6.37% when we remove both the AHA and inverse learning modules. Similarly, removing the AHA and contrastive learning module results in a 4.9% decline in AllMatch’s performance. The least accuracy we find by utilizing a single component is 77.9% with contrastive learning. Finally, we find

Table 3: Ablation study of different proposed components of AllMatch on ModelNet40.

Inv. lear.	AHA	CL	OA
✓	✓	✓	84.28
✓	✓		82.94
✓		✓	79.98
	✓	✓	80.50
		✓	77.90
✓			79.37
	✓		80.10
			74.80

that removing all components drastically drops the performance by 74.8%. This further shows the importance of each of the components in AllMatch.

Ablation on contrastive learning. Next, we perform another study on our contrastive learning module, which consists of supervised and self-supervised contrastive loss. Here, we perform another ablation study by removing the contrastive learning components in different combinations. Similar to the previous study, we do this study on 2% of labelled data from ModelNet40. The results of this study are presented in Table 4. As we find from this table, removing both of the components results in

a considerable drop in performance, with unsupervised contrastive loss having a greater impact on the final performance. More specifically, when both the supervised and unsupervised contrastive losses are employed, AllMatch achieves its highest performance, 84.28%. By dropping the unsupervised contrastive loss, the method’s overall accuracy drops sharply by 1.21%. However, when the supervised loss is removed, the overall accuracy declines by 0.8%. This proves AllMatch is gaining better learning signals from the unsupervised contrastive loss component. Finally, when both contrastive losses are removed, AllMatch returns with AHA and inverse learning setting with an overall loss of 82.94%.

Table 4: Ablation study of contrastive loss components of AllMatch on ModelNet40 dataset.

Sup.	Unsup.	OA
✓	✓	84.28
	✓	83.47
✓		83.06
		82.94

4.4 Performance on Increased Labelled Set Size

In this section, we study the performance of different amounts of labelled data. Specifically, we investigate how an increase in the amount of labelled data impacts performance. To this end, we increase the labelled set size to all labelled data with a 10% interval (Figure 3). It is evident that when only a small amount of labelled data is available (2% - 5%), the overall accuracy of AllMatch is relatively low and gradually increases until it reaches 10% of the labelled set. With 10% labelled data, the overall accuracy reaches 90.19%, compared to 90.75% with the whole labelled set. In other words, the performance with 10% of labelled samples is almost close to the performance with 100% of the labelled data. This again proves the fact that AllMatch properly utilizes all the unlabelled data, making it possible for 10% labelled data to reach such performance.

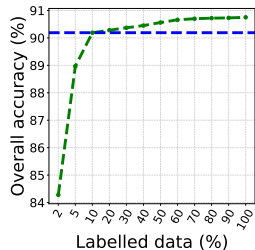


Fig. 3: Impact of labelled set size.

4.5 Sensitivity Study

In this section, we perform detailed sensitivity studies on different parameters of the AllMatch. Following, We discuss the details of these experiments.

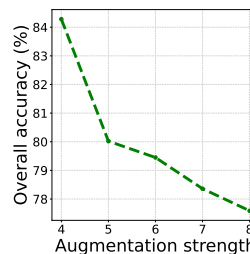
Table 5: Impact of different values of loss factors.

α	Overall accuracy	β	Overall accuracy	γ	Overall accuracy
0.2	84.28	0.2	84.28	0.2	83.22
0.5	83.38	0.5	83.46	0.5	83.38
1.0	81.96	1.0	81.24	1.0	84.28

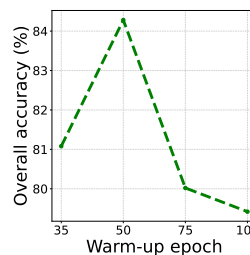
(a) α vs acc. (b) β vs acc. (c) γ vs acc.

Loss factor. The final loss function for training AllMatch consists of multiple losses and includes weight factors to balance the impact of individual loss components. Table 5 illustrates the influence of different loss factors on AllMatch’s overall performance. The experiments are conducted on the ModelNet40 dataset with 2% of labelled data. It is observed that the overall accuracy of AllMatch is sensitive to the values of these factors. Specifically, lower loss factor values lead to higher performances for both supervised and unsupervised contrastive losses, while higher values for the inverse loss factor prove beneficial. This indicates that the inverse loss plays a more significant role in contributing to AllMatch’s performance compared to the contrastive loss.

Augmentation strength. Augmentation strength is an essential hyperparameter for the Adaptive Hard Augmentation (AHA) module of AllMatch. It indicates the combination of different types of augmentation randomly applied as an additional hard augmentation to the easy samples. In Figure 4, we evaluate the overall performance of AllMatch for different augmentation strengths ranging from 4 to 8. Here, the numbers represent the total number of transformations applied to a sample. As we find from Figure 4, the best result is observed for using 4 transformations. While the augmentations are essential for learning from easy samples, applying a very high number of augmentations makes the optimization of the model difficult.

**Fig. 4:** Sensitivity study on the augmentation strengths.

Warm-up epoch for AHA. As discussed in the method section, the AHA module comes into effect after a certain epoch of training. Given that AHA introduces an additional hard augmentation to easy samples, the downstream task becomes challenging for the model. If AHA is applied right at the beginning of the training process, it can negatively impact the model’s performance. To this end, warm-up epochs enable the model to learn representations from all samples efficiently during the initial training. As depicted in Figure 5, AllMatch shows superior performance when

**Fig. 5:** Sensitivity study on the impact of different warm-up epochs.

warmed up for 50 epochs. However, prolonged warm-up periods diminish performance, as they reduce the effective action period of AHA.

Epoch and learning rate. Finally, we study two hyper-parameters that are important for the optimization of the model, epochs and learning rate. As depicted in Figure 6a, we observe that AllMatch’s performance improves by increasing the number of epochs, reaching its peak at 350 epochs with a best performance of 84.28%. However, extending the number

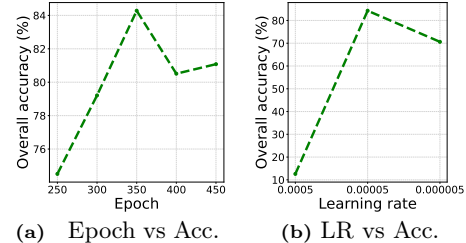


Fig. 6: Sensitivity study of epoch and learning rate on ModelNet40.

of epochs beyond this point results in a decline in performance. This is considerably lower than the previous SOTA, which requires 500 epochs of training. This further shows the benefit of utilizing the unlabelled data properly. In Figure 6b, we analyze the impact of different learning rates on AllMatch’s performance. Notably, lower learning rates contribute to improved performance, whereas higher values lead to a collapse in the model’s performance. For instance, at a learning rate of 0.0005, AllMatch achieves only 12% overall accuracy, emphasizing the sensitivity of the model’s performance to the choice of learning rate.

5 Conclusion

In this work, we identify that existing 3D semi-supervised techniques fail to capitalize the whole unlabelled data. To solve this issue, we introduce AllMatch, a novel semi-supervised method for point cloud classification that effectively utilizes the entire unlabelled set to improve the learnt representation. Comprised of adaptive hard augmentation, inverse learning, and contrastive learning components, AllMatch outperforms existing SOTA methods by a significant margin, showcasing superior performance even with minimal (1%) labelled sets. Remarkably, AllMatch, with just 10% of the labelled set, attains a performance close to the fully supervised learning with all samples, underscoring its efficiency in minimizing reliance on labelled data.

Limitations. While our proposed method has made significant improvement in efficiently leveraging the entire unlabelled set, there are still opportunities for further exploration in this promising direction. In the AHA module, we successfully incorporated selected transformation-invariant augmentation techniques, showcasing a positive step forward. Expanding our exploration to encompass a more diverse range of augmentation methods could prove to be a fruitful avenue for additional advancements. Furthermore, we only explored contrastive learning as the self-supervised module to ensure the utilization of all unlabelled samples, leaving room for exploring different versions and alternatives of contrastive loss.

References

1. Berthelot, D., Carlini, N., Cubuk, E.D., Kurakin, A., Sohn, K., Zhang, H., Raffel, C.: Remixmatch: Semi-supervised learning with distribution matching and augmentation anchoring. In: ICLR (2019)
2. Berthelot, D., Carlini, N., Goodfellow, I., Papernot, N., Oliver, A., Raffel, C.A.: Mixmatch: A holistic approach to semi-supervised learning. NeurIPS (2019)
3. Chen, H., Tao, R., Fan, Y., Wang, Y., Wang, J., Schiele, B., Xie, X., Raj, B., Savvides, M.: Softmatch: Addressing the quantity-quality tradeoff in semi-supervised learning. In: ICLR (2022)
4. Chen, L., Zhang, Y., Lin, Y., Jiang, M., Huang, Y., Lei, Y.: Consistency-based semi-supervised learning for point cloud classification. In: PRAI. pp. 440–445 (2021)
5. Chen, T., Kornblith, S., Norouzi, M., Hinton, G.: A simple framework for contrastive learning of visual representations. In: ICML. pp. 1597–1607 (2020)
6. Chen, Y., Tan, X., Zhao, B., Chen, Z., Song, R., Liang, J., Lu, X.: Boosting semi-supervised learning by exploiting all unlabeled data. In: CVPR. pp. 7548–7557 (2023)
7. Chen, Z., Jing, L.: Multimodal semi-supervised learning for 3d objects. In: BMVC (2021)
8. Chen, Z., Jing, L., Yang, L., Li, Y., Li, B.: Class-level confidence based 3d semi-supervised learning. In: WACV. pp. 633–642 (2023)
9. Gui, G., Zhao, Z., Qi, L., Zhou, L., Wang, L., Shi, Y.: Enhancing sample utilization through sample adaptive augmentation in semi-supervised learning. In: CVPR. pp. 15880–15889 (2023)
10. Guo, M.H., Cai, J.X., Liu, Z.N., Mu, T.J., Martin, R.R., Hu, S.M.: Pct: Point cloud transformer. Computational Visual Media **7**, 187–199 (2021)
11. Jiang, L., Shi, S., Tian, Z., Lai, X., Liu, S., Fu, C.W., Jia, J.: Guided point contrastive learning for semi-supervised point cloud semantic segmentation. In: ICCV. pp. 6423–6432 (2021)
12. Khosla, P., Teterwak, P., Wang, C., Sarna, A., Tian, Y., Isola, P., Maschinot, A., Liu, C., Krishnan, D.: Supervised contrastive learning. NeurIPS pp. 18661–18673 (2020)
13. Kim, J., Min, Y., Kim, D., Lee, G., Seo, J., Ryoo, K., Kim, S.: Conmatch: Semi-supervised learning with confidence-guided consistency regularization. In: ECCV. pp. 674–690 (2022)
14. Lee, D.H., et al.: Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In: Workshop on challenges in representation learning, ICML. p. 896 (2013)
15. Li, J., Xiong, C., Hoi, S.C.: Comatch: Semi-supervised learning with contrastive graph regularization. In: ICCV. pp. 9475–9484 (2021)
16. Liu, F., Lin, G., Foo, C.S., Joshi, C.K., Lin, J.: Point discriminative learning for data-efficient 3d point cloud analysis. In: IEEE Int. Conf. on 3D Vision. pp. 42–51 (2022)
17. Nguyen, K.B.: Sequencematch: Revisiting the design of weak-strong augmentations for semi-supervised learning. In: WACV. pp. 96–106 (2024)
18. Otsu, N.: A threshold selection method from gray-level histograms. TSMC **9**, 62–66 (1979)
19. Qi, C.R., Su, H., Mo, K., Guibas, L.J.: Pointnet: Deep learning on point sets for 3d classification and segmentation. In: CVPR. pp. 652–660 (2017)

20. Roy, S., Etemad, A.: Scaling up semi-supervised learning with unconstrained unlabelled data. In: AAAI. pp. 14847–14856 (2024)
21. Sohn, K., Berthelot, D., Carlini, N., Zhang, Z., Zhang, H., Raffel, C.A., Cubuk, E.D., Kurakin, A., Li, C.L.: Fixmatch: Simplifying semi-supervised learning with consistency and confidence. NeurIPS pp. 596–608 (2020)
22. Tarvainen, A., Valpola, H.: Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. In: NeurIPS. pp. 1195–1204 (2017)
23. Uy, M.A., Pham, Q.H., Hua, B.S., Nguyen, D.T., Yeung, S.K.: Revisiting point cloud classification: A new benchmark dataset and classification model on real-world data. In: ICCV. pp. 1588–1597 (2019)
24. Wu, J., Yang, H., Gan, T., Ding, N., Jiang, F., Nie, L.: Chmatch: Contrastive hierarchical matching and robust adaptive threshold boosted semi-supervised learning. In: CVPR. pp. 15762–15772 (2023)
25. Wu, Z., Song, S., Khosla, A., Yu, F., Zhang, L., Tang, X., Xiao, J.: 3d shapenets: A deep representation for volumetric shapes. In: CVPR. pp. 1912–1920 (2015)
26. Xie, Q., Dai, Z., Hovy, E., Luong, T., Le, Q.: Unsupervised data augmentation for consistency training. In: NeurIPS. pp. 6256–6268 (2020)
27. Xu, Y., Shang, L., Ye, J., Qian, Q., Li, Y.F., Sun, B., Li, H., Jin, R.: Dash: Semi-supervised learning with dynamic thresholding. In: ICML. pp. 11525–11536 (2021)
28. Zhang, B., Wang, Y., Hou, W., Wu, H., Wang, J., Okumura, M., Shinozaki, T.: Flexmatch: Boosting semi-supervised learning with curriculum pseudo labeling. NeurIPS pp. 18408–18419 (2021)
29. Zheng, M., You, S., Huang, L., Wang, F., Qian, C., Xu, C.: Simmatch: Semi-supervised learning with similarity matching. In: CVPR. pp. 14471–14481 (2022)