

Article

Semi-Supervised Interior Decoration Style Classification with Contrastive Mutual Learning

Lichun Guo ^{1,*}, Hao Zeng ¹, Xun Shi ¹, Qing Xu ¹, Jinhui Shi ¹, Kui Bai ^{2,3}, Shuang Liang ⁴ and Wenlong Hang ^{2,3} ¹ College of Art and Design, Nanjing Audit University Jinshen College, Nanjing 210023, China² College of Computer and Information Engineering, Nanjing Tech University, Nanjing 211816, China; wlhang@njtech.edu.cn (W.H.)³ College of Artificial Intelligence, Nanjing Tech University, Nanjing 211816, China⁴ School of Internet of Things, Nanjing University of Posts and Telecommunications, Nanjing 210023, China

* Correspondence: guolichun@naujsc.edu.cn

Abstract: Precisely identifying interior decoration styles holds substantial significance in directing interior decoration practices. Nevertheless, constructing accurate models for the automatic classification of interior decoration styles remains challenging due to the scarcity of expert annotations. To address this problem, we propose a novel pseudo-label-guided contrastive mutual learning framework (PCML) for semi-supervised interior decoration style classification by harnessing large amounts of unlabeled data. Specifically, PCML introduces two distinct subnetworks and selectively utilizes the diversified pseudo-labels generated by each for mutual supervision, thereby mitigating the issue of confirmation bias. For labeled images, the inconsistent pseudo-labels generated by the two subnetworks are employed to identify images that are prone to misclassification. We then devise an inconsistency-aware relearning (ICR) regularization model to perform a review training process. For unlabeled images, we introduce a class-aware contrastive learning (CCL) regularization to learn their discriminative feature representations using the corresponding pseudo-labels. Since the use of distinct subnetworks reduces the risk of both models producing identical erroneous pseudo-labels, CCL can reduce the possibility of noise data sampling to enhance the effectiveness of contrastive learning. The performance of PCML is evaluated on five interior decoration style image datasets. For the average AUC, accuracy, sensitivity, specificity, precision, and F1 scores, PCML obtains improvements of 1.67%, 1.72%, 3.65%, 1.0%, 4.61%, and 4.66% in comparison with the state-of-the-art method, demonstrating the superiority of our method.

Keywords: semi-supervised learning; contrastive learning; interior decoration style**MSC:** 68T45; 68T07; 68T30

Citation: Guo, L.; Zeng, H.; Shi, X.; Xu, Q.; Shi, J.; Bai, K.; Liang, S.; Hang, W. Semi-Supervised Interior Decoration Style Classification with Contrastive Mutual Learning. *Mathematics* **2024**, *12*, 2980. <https://doi.org/10.3390/math12192980>

Academic Editor: Yiu Yin Raymond Lee

Received: 26 August 2024

Revised: 14 September 2024

Accepted: 23 September 2024

Published: 25 September 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

With the progressive improvement of living standards, there is an increasing focus on the aesthetic and functional aspects of interior decoration. Accordingly, the precise identification of interior decoration styles is of great significance for the guidance and instruction of interior design practice [1–3]. To date, many supervised learning methods based on deep neural networks have achieved great performances in interior design practice, such as interior decoration style (e.g., four images from different types of interior decoration styles in Figure 1) recognition [4,5], interior decoration style colorization [6], and interior decoration style design [7]. The success of deep learning models in this field is predominantly attributed to the availability of large amounts of labeled interior decoration images. Nevertheless, it is difficult to obtain a large number of labeled interior decoration images due to the requirements of professional knowledge and time consumption on labeling. Semi-supervised learning paradigm enables models leverage both limited labeled data and extensive unlabeled data, which significantly reduces the dependence

on annotations and progressively emerges as the mainstream paradigm for the interior decoration style recognition.



Figure 1. Examples of four types of interior decoration styles: (a) country style, (b) Chinese style, (c) European style, and (d) simple style.

The semi-supervised learning (SSL) paradigm aims to explore and leverage the internal knowledge underlying unlabeled data to enhance the performance of the model. To date, consistency learning [8] and pseudo-labeling [9] are two mainstream techniques that can be used for exploiting unlabeled data. The consistency-learning-based SSL methods enforce the model to generate consistent predictions for the same input under small perturbations [10]. The pseudo-labeling-based SSL methods generate pseudo-labels for unlabeled data, which are subsequently integrated with labeled data to retrain the model [11]. Furthermore, some other SSL methods attempt to integrate both consistency learning and pseudo-labeling techniques to boost the learning performance [12–14].

Although SSL methods based on above techniques have achieved promising results, several challenges significantly affect their robustness and may result in the degradation of model performance. First of all, most of consistency-learning-based SSL methods are based on the self-ensemble framework, which usually takes two subnetworks with identical architecture and enforces their predictions to be consistent. For instance, the mean teacher (MT) framework [15] is a typical consistency-learning-based SSL method which has a teacher subnetwork and a student subnetwork. Generally, the MT method and its extensions possess the following three characteristics: (1) the student and teacher networks share the same architecture; (2) the parameters of the teacher network are updated as the exponential moving average of those of the student network; and (3) consistency learning regularization is employed to ensure consistent predictions between the student and teacher networks. It is evident that the shared architecture leads to the homogenization of subnetworks. The model parameters of teacher network is a weighted mixture of the historical states of the student network; that means that the predictions of the teacher network are constrained by those of the student network. In addition, the consistency learning regularization enforces the consensus predictions of student and teacher networks, further limiting the diversity of the predictions of them. In summary, the coupling issue constrains MT-like SSL methods from generating diverse predictions, making the models prone to trapping confirmation bias and meaning that it is difficult for them to self-correct.

The second challenge arises from the unreliable pseudo-labels. In practice, unlabeled images usually cover different equipment and environments, thereby increasing the risk of the model making unreliable predictions. Potentially inaccurate pseudo-labels make the SSL's training and labeling loop collapse and degrade the performance of SSL methods. Furthermore, the inaccurate pseudo-labels will adversely affect the learning of internal

correlations among unlabeled images. The supervised contrastive learning strategy [16] has been demonstrated to have a superior performance in the discriminative feature learning of images. The core principle of supervised contrastive learning is that representations of similar samples should be closely aligned, while representations of different types of samples should be distinct. For unlabeled images, the pseudo-labels are usually employed to determine their categories. For instance, the authors of [17] leveraged the spatial consistency of weakly augmented images to generate similar samples, while dissimilar samples were constructed using a straightforward cross-image and pseudo-label weighting heuristic. However, utilizing pseudo-labels to generate samples may not align with actual semantic categories, potentially resulting in noisy sampling in contrastive learning. Due to the unreliability of predictions of unlabeled images, the use of self-generated pseudo-labels disrupts model training, leading to a progressive deterioration of model performance.

To tackle the aforementioned challenges, we propose a novel pseudo-label-guided contrastive mutual learning (PCML) framework for semi-supervised interior decoration style classification. Specifically, PCML employs two subnetworks with different architectures, thus directing the model generate diverse predictions. For labeled images, the inconsistent pseudo-labels generated by the two subnetworks are employed to identify images that are prone to misclassification. We then devise an inconsistency-aware relearning (ICR) regularization to perform a review training for these images. For unlabeled images, we introduce a class-aware contrastive learning (CCL) regularization to learn their discriminative feature representations using the corresponding pseudo-labels. Since the use of distinct subnetworks reduces the risk of both models producing identical erroneous pseudo-labels, CCL can reduce the possibility of noise data sampling to enhance the effectiveness of contrastive learning. We introduce a weighting module to CCL that emphasizes learning from highly probable samples within the same category while reducing the impact of unreliable noisy samples. The synergistic learning among the mutual learning framework, ICR regularization, and CCL regularization during the training process enables each subnetwork to selectively incorporate the reliable knowledge imparted by the other subnetwork, thereby mitigating the issue of confirmation bias. The primary contributions of this work can be summarized as follows:

- We propose a novel PCML framework to facilitate semi-supervised interior decoration style classification by exploiting the diversified pseudo-labels generated by distinct subnetworks.
- PCML integrates two novel modules: ICR regularization to direct the subnetworks review the labeled imaged with inconsistent predictions, and CCL regularization to learn discriminative feature representations of unlabeled images.
- The synergistic learning among the distinct subnetworks, ICR regularization, and CCL regularization helps the model overcome confirmation bias. Extensive experimental results demonstrate the superiority of PCML.

2. Related Works

2.1. Semi-Supervised Learning

By leveraging a large amount of unlabeled data, SSL methods can significantly improve the model performance. Current SSL approaches mainly focus on the consistency learning paradigm, the pseudo-labeling paradigm, or a combination of both to effectively exploit the extensive unlabeled data. Consistency-learning-based SSL methods [8,10,12–15,18–21] encourage the model to generate consistent predictions for the same input under small perturbations, such as input perturbations [8], feature perturbations [19], or network perturbations [14]. For instance, the π model presented in [16] directly uses the network's predictions of the same input under stochastic augmentation and dropout perturbations as the consistency targets. Based on [16], MT model introduces an additional teacher network with the same architecture. In [22], a sample relation consistency regularization is proposed to be integrated into the MT framework, which enables the model to capture the additional internal correlation information between unlabeled data. In addition, the local and global

structural consistencies [21] were developed to jointly learn spatial and geometric structural information, thereby enhancing the generalization capability of the MT model. In [14], three types of perturbations, i.e., input data, network, and feature perturbations, were employed to enhance model training and further improve the generalization of consistency learning. Despite the advancements, consistency-learning-based SSL methods typically employ two subnetworks with identical architectures and encourage consensus in their predictions, making the models prone to falling into confirmation bias and making it so that they find it difficult to self-correct.

Pseudo-labeling-based SSL methods follow the self-training [23] and pseudo-labeling [24] techniques to generate the pseudo-labels for unlabeled data by leveraging the model trained on labeled data. Subsequently, these pseudo-labels are incorporated into the labeled dataset to retrain the model [9,24–27]. To filter out the low quality pseudo-labels, most current SSL methods simply employ a predefined confidence threshold (e.g., 0.95) to discard potentially unreliable pseudo-labels with low confidence [24]. However, it is difficult to determine a reasonable confidence threshold for filtering out all the unreliable pseudo-labels. To address this, an entropy-based module [28] was designed to enable the model to generate low-entropy predictions for unlabeled data. In addition, a curriculum pseudo-labeling method was developed to adjust thresholds for different classes, thereby filtering out unreliable pseudo-labels [29]. Another SSL method introduced an uncertainty-aware pseudo-label selection strategy [30], which accounts for the effects of inadequate network calibration.

Some SSL methods integrated consistency learning and pseudo-labeling techniques. For example, [8] selected partially reliable pseudo-labels and guided the student network to learn from these reliable targets. The authors of [31] developed a cycled pseudo-label scheme to promote mutual consistency for challenging unlabeled data, thereby minimizing uncertain predictions. Despite the notable success of these SSL methods, most of them utilize subnetworks with identical architectures, thereby resulting in the homogenization problem. In view of this, some studies employed distinct subnetworks to enhance the diversity between them. For instance, the authors of [32] proposed the use of a convolutional neural network and a transformer as subnetworks. In addition, the authors of [33] developed a mutual correction framework utilizing two structurally distinct subnets with independent parameter updates for semi-supervised learning.

2.2. Contrastive Learning

The contrastive learning technique significantly advanced self-supervised representation learning [34–37]. The fundamental concept of contrastive learning is to draw together an anchor and a “positive” sample within the embedding space while simultaneously pushing the anchor away from the “negative” samples. In SSL, contrastive learning can fully leverage unlabeled data to learn discriminative visual representations. For example, a pseudo group contrast method [38] was developed to automatically rectify incorrect pseudo-labels. Based on [38], a reliability-aware contrastive self-ensemble framework [39] was proposed to select in-distribution unlabeled data to exploit the reliable internal correlation information, thereby enhancing the robustness of the SSL method. In addition, contrastive learning [40] was used to model pairwise similarities according to their pseudo-labels, which is beneficial for better prediction and avoids being trapped in local minimum. In [41], a graph-based contrastive learning scheme was developed to regularize the structure of the embeddings by using pseudo-labels of unlabeled data. Furthermore, contrastive learning regularization [42] was also used to improve the classification performance of the consistency regularization by well-clustered features of unlabeled data. In this paper, we extend the unsupervised contrastive learning technique to supervised scenarios [16] by using the diverse predictions generated from two distinct subnetworks, thereby facilitating the learning from samples within the same category while reducing the impact of unreliable noisy samples.

3. Pseudo-Label-Guided Contrastive Mutual Learning Framework

In this section, we first briefly introduce our pseudo-label-guided contrastive mutual learning framework. Subsequently, we elaborate on the techniques of inconsistency-aware re-learning and category-aware contrastive learning to enhance the semi-supervised interior decoration style classification performance, respectively.

3.1. Architecture Overview

We firstly introduce the basic formulations of SSL for interior decoration style classification. We have a set of training dataset D , which contains N fully labeled images $D^l = \left\{ \left(x_i^l, y_i^l \right) \right\}_{i=1}^N$ and M unlabeled images $D^u = \left\{ x_i^u \right\}_{i=N+1}^{N+M}$. In practice, we have $N \ll M$. $x_i^l \in \mathcal{X}$ denotes the i -th training image in D^l and $y_i^l \in \mathcal{Y} = \{1, 2, \dots, C\}$ denotes the corresponding one-hot ground-truth label of x_i^l .

In PCML, we employ two networks with different architectures, i.e., subnetwork A ($f_A(\cdot)$) and subnetwork B ($f_B(\cdot)$). We denote θ_A and θ_B as the model parameters of subnetworks A and B, respectively. The subnetwork f can be decomposed into a feature extractor $h(\cdot) : \mathcal{X} \rightarrow \mathcal{Z}$ and a classifier $g(\cdot) : \mathcal{Z} \rightarrow \mathcal{Y}$, parameterized by θ^h and θ^g , respectively. Here, $\mathcal{Z} \subset \mathbb{R}^Z$ represents the feature space of dimension Z . For subnetwork A, we have $f_A = g_A \circ h_A$ and $\theta_A = \left\{ \theta_A^g, \theta_A^h \right\}$. For subnetwork B, we have $f_B = g_B \circ h_B$ with $\theta_B = \left\{ \theta_B^g, \theta_B^h \right\}$. Our goal is to accurately predict the style of interior decoration images by using training dataset D .

The overall framework of the proposed PCML is shown in Figure 2. PCML employs subnetwork A and subnetwork B with different architectures to direct the model generate diverse predictions, thereby mitigating the homogenization problem. Utilizing the diverse predictions, PCML takes ICR regularization to conduct review training on labeled images that are prone to misprediction. For unlabeled images, PCML introduces CCL regularization to learn their discriminative feature representations using the corresponding pseudo-labels. Since the diverse predictions reduce the risk of both models producing identical erroneous pseudo-labels, CCL can reduce the possibility of noise data sampling to enhance the effectiveness of contrastive learning. Below, we introduce the main components of PCML, i.e., ICR regularization, and CCL regularization.

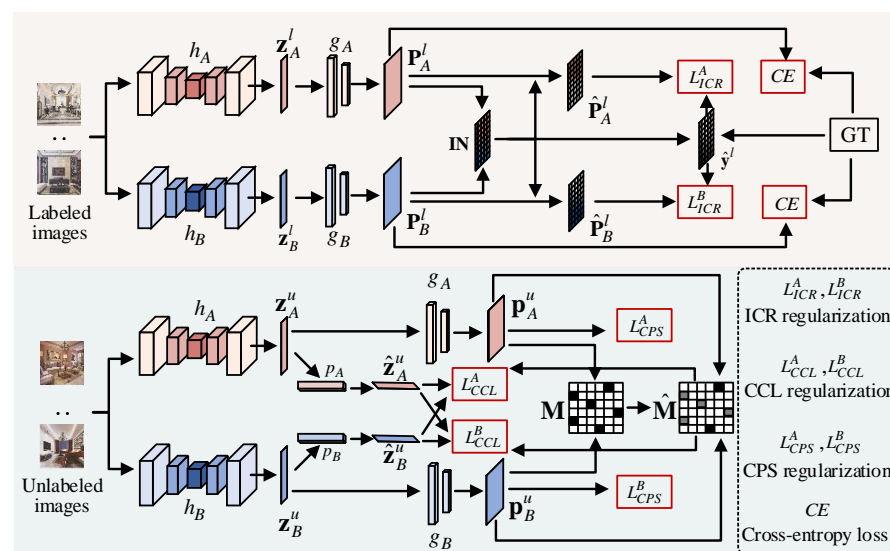


Figure 2. The architecture of PCML for medical image segmentation. A mini-batch input containing labeled data and unlabeled images are fed into subnetwork A and subnetwork B, respectively. The predictions P_A^l and P_B^l are used to calculate the ICR regularization and cross-entropy loss, respectively. The predictions P_A^u and P_B^u are used to calculate the CPS loss and the class-aware matrix, which is then used to guide the calculation of the CCL regularization.

3.2. Inconsistency-Aware Relearning

Assume that the input min-batch of data M contains N samples \mathbf{x}^M , comprising labeled image \mathbf{x}^l randomly sampled from D^l and unlabeled image \mathbf{x}^u randomly sampled from D^u . The input min-batch of images \mathbf{x}^M are fed into feature extractors $h_A(\cdot)$ and $h_B(\cdot)$, respectively. Then, we can obtain the feature representations as follows:

$$\mathbf{z}_A^l, \mathbf{z}_A^u = h_A(\mathbf{x}^M, \theta_A^h), \tag{1}$$

$$\mathbf{z}_B^l, \mathbf{z}_B^u = h_B(\mathbf{x}^M, \theta_B^h). \tag{2}$$

For labeled images, the probability outputs of subnetworks A and B can be obtained by using softmax function $\sigma(\cdot)$, which can be formulated as follows:

$$\mathbf{p}_A^l = \sigma(g_A(\mathbf{z}_A^l, \theta_A^g)), \tag{3}$$

$$\mathbf{p}_B^l = \sigma(g_B(\mathbf{z}_B^l, \theta_B^g)). \tag{4}$$

Since \mathbf{P}_A^l and \mathbf{P}_B^l originate from two different subnetworks, their consistency suggests that the corresponding pseudo-labels are highly accurate. If the two subnetworks yield different predictions for the same image, it indicates that at least one of the predictions is incorrect. As a result, we use the inconsistent predictions as indicators to identify the images that are prone to misclassification. Consequently, the subnetworks need to relearn these images to fully leverage the knowledge within the limited labeled data.

To this end, we propose an ICR regularization approach to selectively leverage the images with inconsistent predictions to retrain the two subnetworks. Specifically, the predicted one-hot label of annotated images obtained from subnetwork A and subnetwork B can be expressed as follows:

$$\mathbf{y}_A^l = \text{One-hot}(\mathbf{p}_A^l), \mathbf{y}_B^l = \text{One-hot}(\mathbf{p}_B^l), \tag{5}$$

where $\text{One-hot}(\cdot)$ transforms the predictions \mathbf{p}_A^l and \mathbf{p}_B^l to the responding hard labels \mathbf{y}_A^l and \mathbf{y}_B^l . The indexes of images with inconsistent predictions can be expressed as follows:

$$\mathbf{IN} = \mathbb{I}(\mathbf{y}_A^l \neq \mathbf{y}_B^l). \tag{6}$$

Here, $\mathbb{I}(\cdot)$ is the indicator function. Thus, the predictions corresponding to these images obtained from subnetworks A and B can be identified using operation $F(\cdot)$:

$$\hat{\mathbf{P}}_A^l = F(\mathbf{IN}, \mathbf{P}_A^l), \hat{\mathbf{P}}_B^l = F(\mathbf{IN}, \mathbf{P}_B^l). \tag{7}$$

Similarly, the truth labels corresponding to these images can be expressed as follows:

$$\hat{\mathbf{y}}^l = F(\mathbf{IN}, \mathbf{y}^l). \tag{8}$$

To facilitate the relearning of samples prone to misprediction, we introduce the ICR regularization, which directs the two subnetworks to focus more on these images. Thus, the ICR regularization employed for subnetworks A and B can be expressed as follows:

$$L_{ICR}^A = \text{MSE}(\hat{\mathbf{P}}_A^l, \hat{\mathbf{y}}^l), L_{ICR}^B = \text{MSE}(\hat{\mathbf{P}}_B^l, \hat{\mathbf{y}}^l). \tag{9}$$

where $\text{MSE}(\cdot)$ denotes the mean squared error loss.

3.3. Class-Aware Contrastive Learning

In the SSL paradigm, supervised contrastive learning [40,41] is often employed to guide the model to focus on the pairwise similarities of unlabeled data based on their pseudo-labels, thereby facilitating discriminative feature learning. Although this is effective, it is difficult to use it to guarantee the reliability of self-generated pseudo-labels, which can adversely affect the performance of semi-supervised classification. In view of this, we propose a class-aware contrastive learning (CCL) scheme to learn discriminative feature representations of unlabeled images by selectively leveraging reliable pseudo-labels. Different from [43], we employ two subnetworks with distinct architectures which can help the model reduce the risk of producing identical erroneous pseudo-labels and dispose of confirmation bias. Hence, CCL can reduce the possibility of noise data sampling to enhance the effectiveness of contrastive learning.

Specifically, the objective of classical supervised contrastive learning can be mathematically expressed as follows:

$$L_{SCL} = - \sum_i \frac{1}{J+1} \sum_{p \in P(i)} \log \frac{\exp(\hat{\mathbf{z}}_{A,i}^u \cdot \hat{\mathbf{z}}_{B,p}^u / \tau)}{\sum_{j=1}^N \mathbb{1}_{i \neq j} \exp(\hat{\mathbf{z}}_{A,i}^u \cdot \hat{\mathbf{z}}_{B,j}^u / \tau)}. \tag{10}$$

Here, $\hat{\mathbf{z}}_{A,i}^u$ and $\hat{\mathbf{z}}_{B,j}^u, \forall i, j = 1, 2, \dots, N$ represent the normalized embeddings output by the projection networks $p_A(\cdot)$ and $p_B(\cdot)$, respectively. Thus, we have $\hat{\mathbf{z}}_{A,i}^u = p_A(\mathbf{z}_{A,i}^u)$ and $\hat{\mathbf{z}}_{B,i}^u = p_B(\mathbf{z}_{B,i}^u)$. $P(i)$ is the set of indices of all positives according to their pseudo-labels. τ is a hyper-parameter for temperature scaling.

In practice, unlabeled images usually come different equipment and environments, thereby increasing the risk of the model making unreliable predictions. The positive samples identified by pseudo-labels in Equation (10) may be inaccurate due to their incorrect predictions. In view of this, we introduce CCL regularization for supervised contrastive learning among the reliable unlabeled images. Specifically, the probability outputs of unlabeled images from subnetwork A and subnetwork B can be expressed as follows:

$$\mathbf{p}_A^u = \sigma(g_A(\mathbf{z}_A^u, \theta_A^g)), \tag{11}$$

$$\mathbf{p}_B^u = \sigma(g_B(\mathbf{z}_B^u, \theta_B^g)). \tag{12}$$

We then use a threshold T to filter out some potentially unreliable unlabeled images. If $\mathbf{p}_{A,i}^u > T$, we assume the i -th unlabeled image has a high probability to be a reliable data and should be pulled closer with the same class. To achieve this, we define a class-aware matrix \mathbf{M} , each element of which can be formulated as follows:

$$m_{ij} = \begin{cases} 1, & \text{if } i = j \\ 1, & \text{if } y_{z_{A,i}^u} = y_{z_{B,j}^u} \text{ and } \mathbf{p}_{A,i}^u, \mathbf{p}_{B,j}^u > T \\ 0, & \text{otherwise} \end{cases} \tag{13}$$

where $y_{z_{A,i}^u}$ and $y_{z_{B,j}^u}$ denote the self-generated pseudo-labels. Although use class-aware matrix \mathbf{M} can filter out most of unreliable samples, the predictions of unlabeled images above threshold T still might be wrong. In view of this, we propose to reweigh these potentially unreliable unlabeled images to minimize the impact of them and maximize the utilization of their valuable information. Specifically, we direct the model to focus on high-confidence unlabeled images while diminishing the potential bias originating from unreliable ones. Thus, the class-aware matrix \mathbf{M} can be reformulated as $\hat{\mathbf{M}}$ through simply multiply their probability outputs. Each element \hat{m}_{ij} of $\hat{\mathbf{M}}$ can be defined as follows:

$$\hat{m}_{ij} = \begin{cases} \mathbf{P}_{A,i}^u \cdot \mathbf{P}_{B,j}^u \cdot m_{ij}, & \text{if } i \neq j \\ m_{ij}, & \text{otherwise} \end{cases} \tag{14}$$

Thus, the CCL regularization can be obtained by combining the class-aware matrix $\hat{\mathbf{M}}$ and supervised contrastive learning scheme, which can be formulated as follows:

$$L_{CCL}^A = - \sum_i \frac{1}{J+1} \sum_{p \in P(i)} \log \frac{\hat{m}_{ij} \cdot \exp(\hat{\mathbf{z}}_{A,i}^u \cdot \hat{\mathbf{z}}_{B,p}^u / \tau)}{\sum_{j=1}^N \mathbb{I}_{i \neq j} \exp(\hat{\mathbf{z}}_{A,i}^u \cdot \hat{\mathbf{z}}_{B,j}^u / \tau)}, \tag{15}$$

$$L_{CCL}^B = - \sum_i \frac{1}{J+1} \sum_{p \in P(i)} \log \frac{\hat{m}_{ij} \cdot \exp(\hat{\mathbf{z}}_{B,i}^u \cdot \hat{\mathbf{z}}_{A,p}^u / \tau)}{\sum_{j=1}^N \mathbb{I}_{i \neq j} \exp(\hat{\mathbf{z}}_{B,i}^u \cdot \hat{\mathbf{z}}_{A,j}^u / \tau)}. \tag{16}$$

3.4. Objective Function

Overall, the total objective function of subnetworks in PCML can be formulated as follows:

$$L_A = L_{sup}^A + L_{unsup}^A, \quad L_B = L_{sup}^B + L_{unsup}^B \tag{17}$$

where L_{sup}^A , L_{sup}^B , and L_{unsup}^A , L_{unsup}^B denote the supervised loss and the unsupervised loss for subnetwork A or subnetwork B, respectively. The supervised losses for subnetwork A and subnetwork B can be expressed as follows:

$$L_{sup}^A = CE(\mathbf{P}_A^l, \mathbf{y}) + L_{ICR}^A, \tag{18}$$

$$L_{sup}^B = CE(\mathbf{P}_B^l, \mathbf{y}) + L_{ICR}^B. \tag{19}$$

where $CE(\cdot)$ denotes the cross-entropy loss. The unsupervised losses for subnetwork A and subnetwork B can be formulated as follows:

$$L_{unsup}^A = \lambda_u \cdot L_{CPS}^A + \lambda_c \cdot L_{CCL}^A, \tag{20}$$

$$L_{unsup}^B = \lambda_u \cdot L_{CPS}^B + \lambda_c \cdot L_{CCL}^B. \tag{21}$$

Here, L_{CPS}^A and L_{CPS}^B denote the cross pseudo supervision (CPS) losses, e.g., $L_{CPS}^A = CE(\mathbf{P}_A^u, \mathbf{y}_B^u)$. λ_u and λ_c are hyper-parameters to balance the cross pseudo supervision loss and CCL regularization.

Compared with regular optimization of a single network, PCML needs to jointly optimize two different subnetworks. Therefore, compared with the regular training procedures of network backbone, PCML needs approximately $2 \times$ training time. Finally, we summarize the thorough optimization process of PCML framework in Algorithm 1.

Algorithm 1: Optimization of PCML Framework

Input: Labeled dataset D^l , unlabeled dataset D^u , trade-off coefficients λ_u and λ_c , max iteration IT , batch size.

Output: Parameters θ_A^T and θ_B^T of subnetworks A and B.

- 1 Initialize: Parameters θ_A^0 and θ_B^0
- 2 **for** $t = 0$ to $T - 1$ **do**
- 3 **for** each mini-batch M , **do**
- 4 Compute $\mathbf{z}_A^l, \mathbf{z}_A^u, \mathbf{z}_B^l$ and \mathbf{z}_B^u using Equations (1) and (2);
- 5 Compute \mathbf{P}_A^l and \mathbf{P}_B^l using Equations (3) and (4);
- 6 Compute $\mathbf{y}_A^l, \mathbf{y}_B^l$ and \mathbf{IN} using Equations (5) and (6);
- 7 Compute $\hat{\mathbf{P}}_A^l, \hat{\mathbf{P}}_B^l$ and $\hat{\mathbf{y}}^l$ using Equations (7) and (8);
- 8 Compute L_{ICR}^A and L_{ICR}^B using Equation (9);
- 9 Compute \mathbf{p}_A^u and \mathbf{p}_B^u using Equations (11) and (12);
- 10 Compute matrix $\hat{\mathbf{M}}$ using Equation (14);
- 11 Compute $\hat{\mathbf{z}}_{A,i}^u$ and $\hat{\mathbf{z}}_{B,i}^u, \forall i$;
- 12 Compute L_{CCL}^A and L_{CCL}^B using Equations (15) and (16);
- 13 // Optimize model parameter for subnetwork A
- 14 ;
- 15 Compute supervised loss L_{sup}^A with L_{ICR}^A and $CE(\mathbf{P}_A^l, \mathbf{Y}^l)$ using Equation (18);
- 16 Compute unsupervised loss L_{unsup}^A using Equation (20);
- 17 Optimize θ_A^t using Equation (17);
- 18 // Optimize model parameter for subnetwork B
- 19 Compute supervised loss L_{sup}^B with L_{ICR}^B and $CE(\mathbf{P}_B^l, \mathbf{Y}^l)$ using Equation (19);
- 20 Compute unsupervised loss L_{unsup}^B using Equation (21);
- 21 Optimize θ_B^t using Equation (17).
- 22 **end**
- 23 **end**

4. Experiments

In this section, we evaluate PCML using five real-world interior decoration style image datasets: (1) TV background wall, (2) chandelier, (3) living room, (4) dining room, and (5) bedroom (shown in Figure 3). First of all, we provide descriptions of the five datasets. Subsequently, we outline the comparison methods and offer the implementation details of the experiments. Finally, we present and analyze the experimental results.

4.1. Datasets and Pre-Processing

(1) *TV background wall dataset*: The available training dataset contains 1643 images of interior decoration style, labeled by four different types of styles (country style, Chinese style, European style, and simple style). Each type has 250, 270, 322, and 801 images, respectively, with each image having a resolution of 900×700 pixels. (2) *Chandelier dataset*: This dataset consists of 969 interior decoration style images. It contains 295, 169, 351, and 154 images for four types of styles, respectively. (3) *Living room dataset*: This dataset consists of 1489 interior decoration style images, involving 138 country style images, 248 Chinese style images, 523 European style images, and 580 simple style images, respectively. (4) *Dining room dataset*: This dataset consists of 520 images, involving 91 country style images, 98 Chinese style images, 178 European style images, and 153 simple style images, respectively. (5) *Bedroom dataset*: This dataset contains 643 interior decoration style images, involving 149 country style images, 119 Chinese style images, 191 European style images, and 184 simple style images, respectively.



Figure 3. Images from five datasets: (a) TV background wall, (b) chandelier, (c) living room, (d) dining room, and (e) bedroom. The first to the fourth rows denote the country style, Chinese style, European style, and simple style, respectively.

In the experiment, each dataset was randomly partitioned into three subsets: 70% for training, 10% for validation, and 20% for testing.

4.2. Experimental Setup

To validate the effectiveness of the proposed SSL framework for medical image segmentation, we conducted a comparison between our PCML and state-of-the-art SSL methods. The comparison methods include the following: (1) ResNet [44]; (2) DenseNet121 [45]; (3) MixMatch [26]; (4) ReMixMatch [46]; (5) CoMatch [41]; (6) FixMatch [27]; (7) MT model [15]; (8) SRC-MT [22]; (9) the proposed PCML.

In the experiment, we re-implemented all comparison methods using open-source code. We used DenseNet as the network backbone for the comparison methods MixMatch, ReMixMatch, CoMatch, FixMatch, MT, and SRC-MT, due to its superior performance. For the proposed PCML, we employed ResNet and DenseNet as subnetwork A and subnetwork B, respectively. For pseudo-labeling-based SSL comparison methods MixMatch, ReMixMatch, CoMatch, FixMatch, we applied a weak augmentation and a strong augmentation on the input data, respectively. For consistency-learning-based SSL methods MT and SRC-MT, the input image perturbation and random transformations, i.e., rotation, translation, horizontal flips, were applied to each image [22]. For pseudo-labeling-based SSL methods, we set the confidence threshold to 0.95 to keep consistency with other pseudo-labeling-based SSL comparison methods [27,29,30,32,38]. In addition, for consistency-learning-based SSL methods, the smoothing parameter exponential moving average was set to 0.99. All other settings of the comparison methods remained consistent with the original work. For the proposed PCML, the projection network contains two linear layers with 2048 and 64 hidden neurons. The trade-off coefficients λ_u and λ_c were set to 0.5 and 0.1 for all datasets. In addition, the threshold was set to 0.9 for living room dataset, and 0.8 for the rest datasets. For the used five datasets, subnetworks were trained by Adam optimizer for 10K iterations, with the fixed learning rate 0.0001. The batch size was set to 24, comprising 12 labeled image and 12 unlabeled images. The implementation of the proposed PCML framework was performed using PyTorch 2.4.1 on a two RTX 4090 GPUs.

We adopted AUC, accuracy (ACC), sensitivity (SEN), specificity (SPE), precision (PREC), and F1 score (F1) as the evaluation metrics to comprehensively evaluate the

classification performance of all comparison methods. The higher values of these metrics indicate better performance of the model.

4.3. Comparison with State-of-the-Art Methods

Tables 1–5 present the classification performance of all SSL comparison methods using 20% labeled images and 80% unlabeled images in training set on five image datasets. In our experiment, the classification results obtained by training the fully supervised models, i.e., ResNet and DenseNet, with 100% labeled data, are regarded as the upper-bound performance. Meanwhile, the classification results achieved by ResNet and DenseNet trained with only 20% labeled images serve as the baseline performance. The best classification results for each evaluation metric were highlighted in bold. Additionally, we listed the performance improvements achieved by all SSL methods relative to the better baseline method, DenseNet, with these enhancements presented in parentheses. Based on the experimental results provided in Tables 1–5, the following observations can be obtained.

Table 1. Performance comparison with different semi-supervised learning methods on TV background wall dataset.

Methods	Percentage		Metrics					
	Labeled	Unlabeled	AUC (%)↑	ACC (%)↑	SEN (%)↑	SPE (%)↑	PREC (%)↑	F1 (%)↑
DenseNet [45]	100%	0	97.44	94.04	86.12	95.09	86.18	86.08
ResNet [44]	100%	0	97.30	93.73	85.47	95.30	83.55	84.23
DenseNet [45]	20%	0	93.76	88.53	73.84	91.45	72.73	72.90
ResNet [44]	20%	0	93.31	89.30	76.39	91.22	75.50	75.58
MixMatch [26]	20%	80%	94.77 (1.01)	90.83 (2.29)	78.98 (5.14)	91.81 (0.36)	80.85 (8.12)	79.48 (6.58)
ReMixMatch [46]	20%	80%	95.48 (1.72)	90.52 (1.99)	76.11 (2.27)	92.10 (0.65)	80.42 (7.69)	77.55 (4.65)
CoMatch [41]	20%	80%	95.43 (1.67)	91.28 (2.75)	78.67 (4.48)	92.53 (1.08)	82.05 (9.32)	79.85 (6.95)
FixMatch [27]	20%	80%	95.29 (1.53)	89.76 (1.22)	77.26 (3.42)	91.81 (0.36)	77.19 (4.46)	76.90 (4.00)
MT [15]	20%	80%	95.55 (1.79)	90.06 (1.53)	74.03 (0.20)	91.51 (0.06)	81.07 (8.34)	75.98 (3.08)
SRC-MT [22]	20%	80%	95.84 (2.08)	91.36 (2.83)	80.11 (6.27)	93.08 (1.63)	81.33 (8.60)	79.76 (6.86)
PCML (Ours)	20%	80%	96.82 (3.06)	93.58 (5.05)	84.12 (10.3)	94.48 (3.03)	87.27 (14.5)	85.36 (12.5)

Table 2. Performance comparison with different semi-supervised learning methods on chandelier dataset.

Methods	Percentage		Metrics					
	Labeled	Unlabeled	AUC (%)↑	ACC (%)↑	SEN (%)↑	SPE (%)↑	PREC (%)↑	F1 (%)↑
DenseNet [45]	100%	0	99.61	97.28	95.58	97.91	93.85	94.61
ResNet [44]	100%	0	99.33	96.11	92.39	97.11	92.37	92.34
DenseNet [45]	20%	0	93.96	87.69	78.75	91.09	77.72	76.53
ResNet [44]	20%	0	93.71	86.92	75.01	90.37	75.97	73.51
MixMatch [26]	20%	80%	95.75 (1.79)	91.58 (3.89)	84.72 (5.97)	93.55 (2.47)	83.14 (5.42)	83.63 (7.10)
ReMixMatch [46]	20%	80%	96.51 (2.55)	92.10 (4.40)	86.13 (7.37)	93.86 (2.78)	85.22 (7.50)	85.03 (8.50)
CoMatch [41]	20%	80%	97.57 (3.61)	92.75 (5.05)	86.55 (7.80)	94.42 (3.33)	85.90 (8.18)	85.85 (9.32)
FixMatch [27]	20%	80%	97.66 (3.70)	93.39 (5.70)	86.86 (8.10)	95.07 (3.98)	85.54 (7.82)	86.11 (9.58)
MT [15]	20%	80%	96.38 (2.42)	92.75 (5.05)	87.45 (8.70)	94.55 (3.46)	86.38 (8.66)	86.70 (10.2)
SRC-MT [22]	20%	80%	97.35 (3.39)	92.49 (4.79)	88.66 (9.91)	94.42 (3.33)	84.50 (6.77)	84.07 (7.54)
PCML (Ours)	20%	80%	98.25 (4.29)	95.21 (7.51)	90.88 (12.1)	96.65 (5.56)	90.47 (12.8)	90.55 (14.0)

Table 3. Performance comparison with different semi-supervised learning methods on living room dataset.

Methods	Percentage		Metrics					
	Labeled	Unlabeled	AUC (%)↑	ACC (%)↑	SEN (%)↑	SPE (%)↑	PREC (%)↑	F1 (%)↑
DenseNet [45]	100%	0	94.31	92.42	83.30	94.26	83.91	83.52
ResNet [44]	100%	0	94.86	92.09	81.81	93.90	84.50	82.52
DenseNet [45]	20%	0	89.74	86.95	69.34	89.78	69.45	67.28
ResNet [44]	20%	0	91.06	87.88	70.41	90.89	68.94	67.81
MixMatch [26]	20%	80%	91.26 (1.53)	88.89 (1.94)	72.91 (3.57)	91.63 (1.85)	77.07 (7.62)	73.36 (6.07)
ReMixMatch [46]	20%	80%	92.12 (2.38)	89.39 (2.44)	72.91 (3.57)	92.19 (2.41)	74.09 (4.64)	75.81 (8.53)
CoMatch [41]	20%	80%	91.58 (1.84)	89.65 (2.69)	74.69 (5.36)	92.17 (2.38)	77.54 (8.08)	74.06 (6.77)
FixMatch [27]	20%	80%	92.53 (2.80)	88.72 (1.77)	70.55 (1.21)	91.63 (1.85)	76.60 (7.15)	69.57 (2.29)
MT [15]	20%	80%	91.91 (2.17)	87.96 (1.01)	71.77 (2.44)	91.16 (1.38)	73.74 (4.29)	69.70 (2.42)
SRC-MT [22]	20%	80%	92.41 (2.67)	88.22 (1.26)	71.98 (2.64)	91.07 (1.29)	74.23 (4.78)	70.43 (3.15)
PCML (Ours)	20%	80%	93.92 (4.19)	91.25 (4.29)	77.27 (7.93)	93.33 (3.55)	79.74 (10.3)	76.74 (9.46)

Table 4. Performance comparison with different semi-supervised learning methods on dining room dataset.

Methods	Percentage		Metrics					
	Labeled	Unlabeled	AUC (%)↑	ACC (%)↑	SEN (%)↑	SPE (%)↑	PREC (%)↑	F1 (%)↑
DenseNet [45]	100%	0	98.03	94.71	91.66	95.67	89.27	90.2
ResNet [44]	100%	0	97.87	93.75	89.48	95.12	86.94	88.06
DenseNet [45]	20%	0	91.89	87.26	76.37	89.71	74.64	75.24
ResNet [44]	20%	0	91.29	83.89	72.65	87.08	68.83	69.04
MixMatch [26]	20%	80%	95.20 (3.31)	89.18 (1.92)	79.68 (3.31)	90.46 (0.75)	80.20 (5.55)	78.60 (3.37)
ReMixMatch [46]	20%	80%	95.89 (4.00)	91.11 (3.85)	82.60 (6.23)	92.82 (3.11)	78.12 (3.48)	80.27 (5.03)
CoMatch [41]	20%	80%	96.08 (4.18)	89.66 (2.40)	82.20 (5.83)	91.93 (2.22)	75.89 (1.24)	76.62 (1.38)
FixMatch [27]	20%	80%	94.93 (3.04)	89.90 (2.64)	83.21 (6.85)	91.74 (2.03)	82.01 (7.36)	81.72 (6.48)
MT [15]	20%	80%	93.97 (2.08)	88.46 (1.20)	81.76 (5.39)	91.92 (2.21)	76.00 (1.36)	77.65 (2.41)
SRC-MT [22]	20%	80%	95.45 (3.56)	91.11 (3.85)	85.17 (8.80)	94.09 (4.39)	77.86 (3.22)	79.53 (4.29)
PCML (Ours)	20%	80%	97.23 (5.34)	93.51 (6.25)	88.96 (12.6)	94.75 (5.05)	86.35 (11.7)	87.59 (12.4)

Table 5. Performance comparison with different semi-supervised learning methods on bedroom dataset.

Methods	Percentage		Metrics					
	Labeled	Unlabeled	AUC (%)↑	ACC (%)↑	SEN (%)↑	SPE (%)↑	PREC (%)↑	F1 (%)↑
DenseNet [45]	100%	0	88.40	85.74	72.55	89.59	72.35	71.95
ResNet [44]	100%	0	88.57	85.16	72.58	89.03	71.12	70.68
DenseNet [45]	20%	0	81.35	78.71	57.78	84.75	59.35	56.82
ResNet [44]	20%	0	82.45	78.32	57.60	84.86	61.64	56.12
MixMatch [26]	20%	80%	84.04 (2.69)	81.84 (3.13)	65.42 (7.64)	86.67 (1.92)	63.59 (4.24)	64.01 (7.20)
ReMixMatch [46]	20%	80%	85.23 (3.88)	81.25 (2.54)	63.51 (5.74)	86.34 (1.58)	62.17 (2.82)	62.11 (5.30)
CoMatch [41]	20%	80%	85.20 (3.85)	80.86 (2.15)	65.62 (7.84)	86.32 (1.56)	62.25 (2.90)	62.25 (5.43)
FixMatch [27]	20%	80%	84.96 (3.61)	81.45 (2.73)	67.54 (9.77)	86.34 (1.58)	63.89 (4.54)	63.97 (7.16)
MT [15]	20%	80%	84.44 (3.09)	80.86 (2.15)	64.05 (6.27)	86.67 (1.92)	61.37 (2.02)	61.67 (4.86)
SRC-MT [22]	20%	80%	86.01 (4.66)	80.08 (1.37)	60.23 (2.46)	86.60 (1.84)	61.64 (2.29)	60.04 (3.23)
PCML (Ours)	20%	80%	86.71 (5.36)	82.42 (3.71)	69.26 (11.5)	86.80 (2.05)	65.10 (5.75)	66.52 (9.71)

In general, SSL methods consistently achieved improved classification performance compared to the baseline methods ResNet and DenseNet by leveraging unlabeled images. This verifies the efficacy of exploring the internal knowledge underlying unlabeled data. The proposed PCML framework exhibits superior classification results compared to other SSL methods across all evaluative metrics. These results confirm the effectiveness of integrating ICR regularization, which directs the two subnetworks to review labeled images with inconsistent predictions, and CCL regularization, which facilitates the learning of discriminative feature representations for unlabeled images in the training procedure.

The proposed method PCML exhibits promising classification results across five real-world interior decoration style image datasets. Compared to the baseline method DenseNet, it demonstrates average improvements of 4.45%, 5.36%, 10.9%, 3.85%, 11.0%, and 11.6% on AUC, ACC, SEN, SPE, PREC, and F1, respectively. Specifically, PCML can achieve consistent improvements over all the comparison methods across all metrics on the five image datasets. For the TV background wall dataset, PCML surpasses the classification performance of the most competitive method SRC-MT by 0.98%, 2.22%, 4.01%, 1.4%, 5.94%, and 5.6% on AUC, ACC, SEN, SPE, PREC, and F1, respectively. For the chandelier dataset, compared to the most competitive method FixMatch, PCML achieved the improvements of 0.59%, 1.82%, 4.02%, 1.58%, 4.93%, and 4.44%, respectively. For the living room dataset, PCML surpasses the classification performance of the most competitive method CoMatch by 2.34%, 1.6%, 2.58%, 1.16%, 2.2%, and 2.68% on AUC, ACC, SEN, SPE, PREC, and F1, respectively. For the dining room dataset, compared to the most competitive method SRC-MT, PCML achieved the improvements of 1.78%, 2.4%, 3.78%, 0.66%, 8.49%, and 8.06%, respectively. In addition, for the bedroom dataset, PCML surpasses the classification performance of the most competitive method MixMatch by 2.67%, 0.58%, 3.84%, 0.13%, 1.51%, and 2.51% on AUC, ACC, SEN, SPE, PREC, and F1, respectively. It is noteworthy to observe that the classification performance of our PCML surpasses the upper-bound performance on some evaluative metrics. For example, on TV background wall dataset, PCML outperforms the fully supervised DenseNet with 100% labeled data in terms of

PREC metric. In addition, the classification results of PCML closely approaches the upper-bound performance. The disparities between certain metrics are minimal, with less than 1% differences, such as the SPE and F1 classification metrics for the TV background wall dataset. This observation further validates the advantages gained by reviewing the labeled imaged with inconsistent predictions and learning discriminative feature representations of unlabeled images, which also explains our significant performance gains.

4.4. Analysis of the Proposed PCML Framework

4.4.1. Efficacy of Different Components

To obtain a better insight into the performance of the proposed PCML method, we conducted an ablation study to investigate the impact of different components in PCML. We listed the classification results of ablation study on the TV background wall dataset and the living room dataset, as shown in Tables 6 and 7. The ablation study includes the following models: (a) backbone network ResNet (Baseline 1); (b) backbone network DenseNet (Baseline 2); (c) subnetworks ResNet and DenseNet with CPS regularization (Scenario 1); (d) subnetworks ResNet and DenseNet with CPS and ICR regularizations (Scenario 2); (e) subnetworks ResNet and DenseNet with ICR and CCL regularizations (Scenario 3, PCML).

Table 6. Ablation study of different components on the segmentation performance of the TV background wall dataset.

ResNet	DenseNet	CPS	ICR	CCL	TV Background Wall Dataset					
					AUC (%)↑	ACC (%)↑	SEN (%)↑	SPE (%)↑	PREC (%)↑	F1 (%)↑
✓					93.31	89.30	76.39	91.22	75.50	75.58
	✓				93.76	88.53	73.84	91.45	72.73	72.90
✓	✓	✓			96.17	91.44	79.42	93.00	81.84	79.52
✓	✓	✓	✓		96.40	92.28	83.15	93.29	81.94	82.40
✓	✓	✓	✓	✓	96.82	93.58	84.12	94.48	87.27	85.36

Table 7. Ablation study of different components on the segmentation performance of the living room dataset.

ResNet	DenseNet	CPS	ICR	CCL	TV Background Wall Dataset					
					AUC (%)↑	ACC (%)↑	SEN (%)↑	SPE (%)↑	PREC (%)↑	F1 (%)↑
✓					91.06	87.88	70.41	90.89	68.94	67.81
	✓				89.74	86.95	69.34	89.78	69.45	67.28
✓	✓	✓			92.35	88.97	70.71	91.88	78.68	68.26
✓	✓	✓	✓		93.91	88.64	72.41	90.90	79.50	72.33
✓	✓	✓	✓	✓	93.92	91.25	77.27	93.33	79.74	76.74

As shown in Tables 6 and 7, Scenario 1 yields better classification performance than Baseline 1 and Baseline 2, which demonstrates the effectiveness of different subnetworks with CPS regularization. The architecture of different subnetworks can promote the diversified predictions, thereby preventing the two subnetworks from collapsing into each other. Scenario 2 can obtain better classification performance than Scenario 1, which verifies ICR regularization can conduct review training on the potential mis-predicted labeled images that enables the SSL model to leverage the limited reliable knowledge underlying labeled images. Moreover, compared to Scenario 2, Scenario 3 (PCML) demonstrates further enhancement in classification performance. These results underscore the importance of selectively utilizing more reliable unlabeled images for contrastive learning, which contributes to the overall robustness of the model.

In addition, we used Gradient Weighted Class Activation Mapping (Grad-CAM) [47] to visualize and localize the salient image regions that exert a substantial influence on the model's prediction score for a given class. To provide better interpretability of the proposed framework, we visualized Grad-CAMs of Scenario 1, Scenario 2, and PCML. As illustrated in Figures 4 and 5, the first row represents the Chinese style in the TV background wall

dataset and the living room dataset, respectively. The second row represents the European style in the TV background wall dataset and the living room dataset, respectively. We listed the original images in Figures 4a and 5a. In addition, Figures 4b and 5b display the Grad-CAM visualizations of Scenario 1. Figures 4c and 5c display the Grad-CAM visualizations of Scenario 2. Figures 4d and 5d display the Grad-CAM visualizations of PCML. The visualizations show that the proposed PCML method can learn the features of salient regions exhibiting style changes and focuses on the distinctive features associated with specific decoration style.

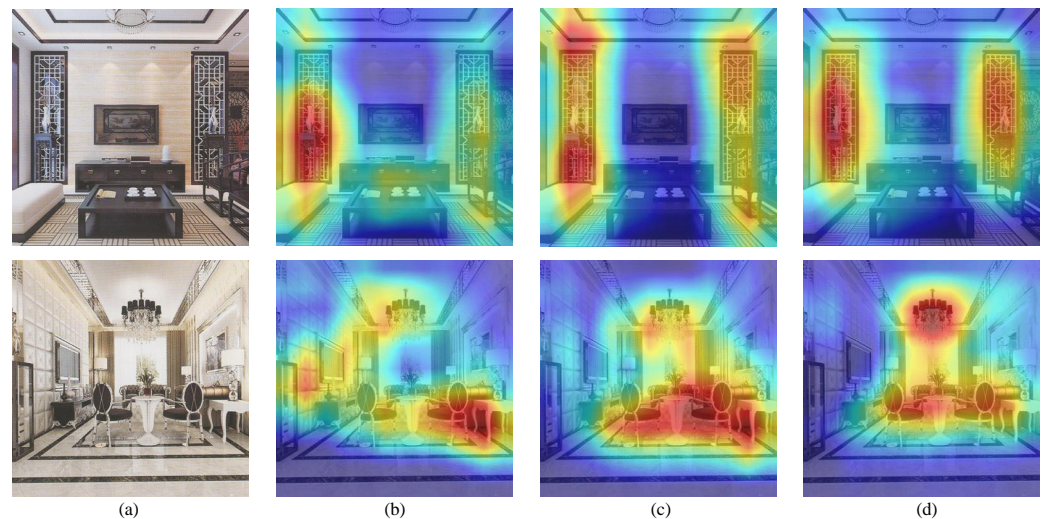


Figure 4. Grad-CAMs visualization of interior decoration style image attention regions from the TV background wall dataset: (a) original images; (b) Scenario 1; (c) Scenario 2; (d) Scenario 3. The first row denotes Chinese style and the second row denotes European style.

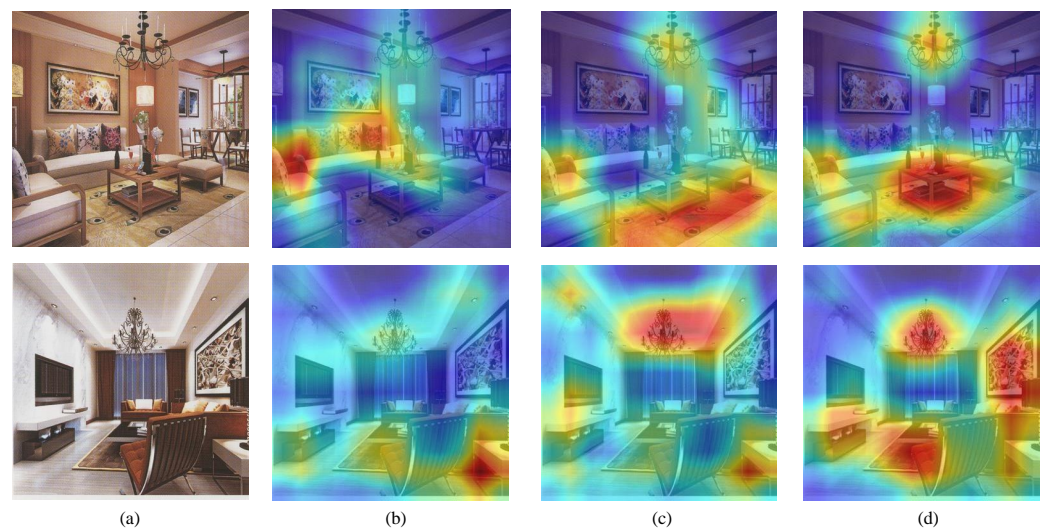


Figure 5. Grad-CAMs visualization of interior decoration style image attention regions from the Living room dataset: (a) Original images; (b) Scenario 1; (c) Scenario 2; (d) Scenario 3. The first row denotes Chinese style and the second row denotes European style.

4.4.2. Impact of Hyper-Parameters

We further investigated the influence of hyper-parameters on the classification performance of the proposed PCML. PCML contains three hyper-parameters— T , λ_u , and λ_c —where T is used to filter out the noise unlabeled images, λ_u is the weight for cross pseudo supervision loss, and λ_c is the weight for CCL regularization. In the experiment, we fixed two of the parameters and varied the other parameter to observe the classification

performance of PCML. The classification results were given in Table 8. The threshold T was used to filter out some potentially unreliable unlabeled images with low confidence. When $\lambda_u = 0.5$, $\lambda_c = 0.1$, with the increase in T , the classification performance is improved accordingly, demonstrating that threshold T can effectively filter out unreliable unlabeled images. As T keeps increasing, the classification performance degrades, which means that some reliable unlabeled images were excluded. In addition, for parameters λ_u and λ_c , we can observe the similar phenomena, as shown in Table 8. In total, PCML exhibits stable predictions despite variations in both parameters λ_u and λ_c , which demonstrates the effectiveness of the proposed pseudo-label-guided contrastive mutual learning framework.

Table 8. Sensitivity of trade-off coefficients on the classification performance (%) of the TV background wall dataset.

$\lambda_u = 0.5$						Metrics
$\lambda_c = 0.1$	AUC	ACC	SEN	SPE	PREC	F1
$T = 0.5$	96.24	91.28	78.46	92.58	82.79	79.37
$T = 0.6$	96.12	91.97	82.01	92.58	82.55	81.71
$T = 0.7$	96.50	91.97	80.21	92.89	84.75	81.86
$T = 0.8$	96.82	93.58	84.12	94.48	87.27	85.36
$T = 0.9$	96.51	91.67	81.17	93.00	82.19	80.74
$T = 0.8$						Metrics
$\lambda_c = 0.1$	AUC	ACC	SEN	SPE	PREC	F1
$\lambda_u = 0.05$	95.31	90.37	78.91	91.62	78.53	78.23
$\lambda_u = 0.1$	95.59	91.67	81.53	92.18	82.20	81.46
$\lambda_u = 0.3$	96.50	92.81	82.24	93.88	84.80	82.79
$\lambda_u = 0.5$	96.82	93.58	84.12	94.48	87.27	85.36
$\lambda_u = 0.7$	95.94	92.97	82.06	94.44	83.84	82.34
$T = 0.8$						Metrics
$\lambda_u = 0.5$	AUC	ACC	SEN	SPE	PREC	F1
$\lambda_c = 0.05$	96.30	92.28	80.61	93.18	85.32	82.11
$\lambda_c = 0.1$	96.82	93.58	84.12	94.48	87.27	85.36
$\lambda_c = 0.3$	96.51	92.58	82.24	93.87	82.91	82.26
$\lambda_c = 0.5$	95.23	91.82	81.19	93.70	81.03	80.61
$\lambda_c = 0.7$	94.95	90.67	79.40	92.97	78.49	78.05

4.4.3. Impact of Input Noise

To evaluate the influence of noise on the classification performance of the proposed PCML framework, we utilize different kinds of input image perturbations as input noise to evaluate the performance of our PCML. Specifically, random transformations including rotation, translation, and horizontal flips were applied to each input image. The rotation angle was randomly set in the range of -10 to 10 degrees. Horizontal and vertical translations for pixels were applied within a range of -2% to 2% of the image width. Additionally, the input image was randomly flipped horizontally and vertically with a probability of 50% . We use the above input noise into one subnetwork and two subnetworks to evaluate the performance of the proposed method. It includes the following scenarios: (a) without input noise (setting 1); (b) with input noise to one subnetwork (setting 2); (c) with input noise to two subnetworks (setting 3). As illustrated in Table 9, the experimental results across different settings show minimal variation, which indicates that the proposed PCML is robust to the input perturbation noise.

Table 9. Impact of input perturbation noise on the classification results (%) obtained by PCML.

Settings		TV Background Wall					
		AUC	ACC	SEN	SPE	PREC	F1
1		96.82	93.58	84.12	94.48	87.27	85.36
2		96.24	92.13	80.82	93.47	82.63	81.02
3		96.09	91.67	81.93	93.26	80.11	80.64
Settings		Chandelier					
		AUC	ACC	SEN	SPE	PREC	F1
1		98.25	95.21	90.88	96.65	90.47	90.55
2		97.96	94.43	89.38	96.29	89.63	89.11
3		98.90	95.47	91.73	96.71	90.16	90.78
Settings		Living Room					
		AUC	ACC	SEN	SPE	PREC	F1
1		93.92	91.25	77.27	93.33	79.74	76.74
2		93.34	90.91	78.47	93.03	79.02	77.58
3		93.16	91.25	79.19	93.12	79.07	78.26
Settings		Dining Room					
		AUC	ACC	SEN	SPE	PREC	F1
1		97.23	93.51	88.96	94.75	86.35	87.59
2		97.84	94.71	88.08	95.46	89.71	88.68
3		98.15	93.99	85.16	94.30	87.59	85.86
Settings		Bedroom					
		AUC	ACC	SEN	SPE	PREC	F1
1		86.71	82.42	69.26	86.80	65.10	66.52
2		87.68	82.03	67.81	87.7	70.43	63.14
3		88.32	82.23	68.81	86.98	70.58	64.37

5. Conclusions

In this paper, we propose a pseudo-label-guided contrastive mutual learning framework, named PCML, to facilitate semi-supervised interior decoration style classification by harnessing reliable knowledge within limited labeled data and selectively utilizing the reliable unlabeled data for learning discriminative feature representations. The framework employs two subnetworks with different architectures, and integrates two novel modules: inconsistency-aware relearning regularization, and class-aware contrastive learning regularization. The different subnetworks can direct the model generate diverse predictions. Thus, inconsistency-aware relearning regularization to can perform a review training for these images with different predictions. In addition, class-aware contrastive learning regularization can learn the discriminative feature representations of unlabeled images using the corresponding reliable pseudo-labels. More importantly, the synergistic learning among the mutual learning framework, inconsistency-aware relearning regularization, and class-aware contrastive learning regularization during the training process enables each subnetwork to selectively incorporate the reliable knowledge imparted by the other subnetwork, thereby mitigating the issue of confirmation bias. The comprehensive performance evaluations on multiple interior decoration style image datasets demonstrate the superiority of the proposed PCML over existing SSL methods. In future work, we plan to further comprehensively evaluate our PCML on more complex interior decoration tasks.

Author Contributions: Conceptualization, W.H.; methodology, K.B. and H.Z.; software, K.B.; validation, L.G., S.L. and J.S.; data curation, Q.X. and X.S.; writing—original draft preparation, L.G. and K.B.; writing—review and editing, W.H.; supervision, L.G. and W.H.; funding acquisition, W.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Natural Science Foundation of the Higher Education Institutions of Jiangsu Province under Grant (23KJB520012).

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

PCML	pseudo-label-guided contrastive mutual learning
ICR	inconsistency-aware relearning
CCL	class-aware contrastive learning
SSL	semi-supervised learning
MT	mean teacher
Grad-CAM	Gradient Weighted Class Activation Mapping

References

- Liu, S.; Bo, Y.; Huang, L. Application of image style transfer technology in interior decoration design based on ecological environment. *J. Sens.* **2021**, *2021*, 9699110. [\[CrossRef\]](#)
- Xu, J.; Li, M.; Huang, D.; Wei, Y.; Zhong, S. A comparative study on the influence of different decoration styles on subjective evaluation of hotel indoor environment. *Buildings* **2022**, *12*, 1777. [\[CrossRef\]](#)
- Weiss, T.; Yildiz, I.; Agarwal, N.; Ataer-Cansizoglu, E.; Choi, J.W. Image-Driven Furniture Style for Interactive 3D Scene Modeling. In Proceedings of the Computer Graphics Forum, Geneva, Switzerland, 20–23 October 2020; Wiley Online Library: Hoboken, NJ, USA, 2020; Volume 39, pp. 57–68.
- Kim, J.; Lee, J.K. Stochastic detection of interior design styles using a deep-learning model for reference images. *Appl. Sci.* **2020**, *10*, 7299. [\[CrossRef\]](#)
- Tian, J.; Zakaria, S.A. Application of Image Classification Algorithm Based on Deep Learning in Residential Interior Design Style Recognition. *Rev. Ibér. Sist. Tecnol. Inf.* **2023**, *E63*, 340–352.
- Tong, H.; Wan, Q.; Kaszowska, A.; Panetta, K.; Taylor, H.A.; Agaian, S. ARFurniture: Augmented reality interior decoration style colorization. *Electron. Imaging* **2019**, *31*, 1–9. [\[CrossRef\]](#)
- Wu, Z.; Jia, X.; Jiang, R.; Ye, Y.; Qi, H.; Xu, C. CSID-GAN: A Customized Style Interior Floor Plan Design Framework Based on Generative Adversarial Network. *IEEE Trans. Consum. Electron.* **2024**. [\[CrossRef\]](#)
- Yu, L.; Wang, S.; Li, X.; Fu, C.W.; Heng, P.A. Uncertainty-aware self-ensembling model for semi-supervised 3D left atrium segmentation. In Proceedings of the Medical Image Computing and Computer Assisted Intervention—MICCAI 2019: 22nd International Conference, Shenzhen, China, 13–17 October 2019; Proceedings, Part II 22; Springer: Berlin/Heidelberg, Germany, 2019; pp. 605–613.
- Su, J.; Luo, Z.; Lian, S.; Lin, D.; Li, S. Mutual learning with reliable pseudo label for semi-supervised medical image segmentation. *Med. Image Anal.* **2024**, *94*, 103111. [\[CrossRef\]](#)
- Sajjadi, M.; Javanmardi, M.; Tasdizen, T. Regularization with stochastic transformations and perturbations for deep semi-supervised learning. *Adv. Neural. Inf. Process. Syst.* **2016**, *29*, 1–9.
- Iscen, A.; Tolia, G.; Avrithis, Y.; Chum, O. Label propagation for deep semi-supervised learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 5070–5079.
- Chen, X.; Yuan, Y.; Zeng, G.; Wang, J. Semi-supervised semantic segmentation with cross pseudo supervision. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 2613–2622.
- Ouali, Y.; Hudelot, C.; Tami, M. Semi-supervised semantic segmentation with cross-consistency training. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 12674–12684.
- Liu, Y.; Tian, Y.; Chen, Y.; Liu, F.; Belagiannis, V.; Carneiro, G. Perturbed and strict mean teachers for semi-supervised semantic segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 4258–4267.
- Tarvainen, A.; Valpola, H. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 1–10.
- Khosla, P.; Teterwak, P.; Wang, C.; Sarna, A.; Tian, Y.; Isola, P.; Maschinot, A.; Liu, C.; Krishnan, D. Supervised contrastive learning. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 18661–18673.

17. Zhong, Y.; Yuan, B.; Wu, H.; Yuan, Z.; Peng, J.; Wang, Y.X. Pixel contrastive-consistent semi-supervised semantic segmentation. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 7273–7282.
18. Rasmus, A.; Berglund, M.; Honkala, M.; Valpola, H.; Raiko, T. Semi-supervised learning with ladder networks. *Adv. Neural Inf. Process. Syst.* **2015**, *28*, 1–9.
19. Miyato, T.; Maeda, S.I.; Koyama, M.; Ishii, S. Virtual adversarial training: A regularization method for supervised and semi-supervised learning. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *41*, 1979–1993. [[CrossRef](#)] [[PubMed](#)]
20. Luo, Y.; Zhu, J.; Li, M.; Ren, Y.; Zhang, B. Smooth neighbors on teacher graphs for semi-supervised learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 8896–8905.
21. Hang, W.; Feng, W.; Liang, S.; Yu, L.; Wang, Q.; Choi, K.S.; Qin, J. Local and global structure-aware entropy regularized mean teacher model for 3d left atrium segmentation. In Proceedings of the Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, 4–8 October 2020; Proceedings, Part I 23; Springer: Berlin/Heidelberg, Germany, 2020; pp. 562–571.
22. Liu, Q.; Yu, L.; Luo, L.; Dou, Q.; Heng, P.A. Semi-supervised medical image classification with relation-driven self-ensembling model. *IEEE Trans. Med Imaging* **2020**, *39*, 3429–3440. [[CrossRef](#)] [[PubMed](#)]
23. Grandvalet, Y.; Bengio, Y. Semi-supervised learning by entropy minimization. *Adv. Neural Inf. Process. Syst.* **2004**, *17*, 529–536.
24. Lee, D.H. Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In Proceedings of the Workshop on Challenges in Representation Learning, ICML, Atlanta, GA, USA, 16–21 June 2013; Volume 3, p. 896.
25. Li, Y.; Chen, J.; Xie, X.; Ma, K.; Zheng, Y. Self-loop uncertainty: A novel pseudo-label for semi-supervised medical image segmentation. In Proceedings of the Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, 4–8 October 2020; Proceedings, Part I 23; Springer: Berlin/Heidelberg, Germany, 2020; pp. 614–623.
26. Berthelot, D.; Carlini, N.; Goodfellow, I.; Papernot, N.; Oliver, A.; Raffel, C.A. Mixmatch: A holistic approach to semi-supervised learning. *Adv. Neural Inf. Process. Syst.* **2019**, *32*, 1–11.
27. 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. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 596–608.
28. Kalluri, T.; Varma, G.; Chandraker, M.; Jawahar, C. Universal semi-supervised semantic segmentation. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 5259–5270.
29. Zhang, B.; Wang, Y.; Hou, W.; Wu, H.; Wang, J.; Okumura, M.; Shinozaki, T. Flexmatch: Boosting semi-supervised learning with curriculum pseudo labeling. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 18408–18419.
30. Rizve, M.N.; Duarte, K.; Rawat, Y.S.; Shah, M. In defense of pseudo-labeling: An uncertainty-aware pseudo-label selection framework for semi-supervised learning. *arXiv* **2021**, arXiv:2101.06329.
31. Wu, Y.; Xu, M.; Ge, Z.; Cai, J.; Zhang, L. Semi-supervised left atrium segmentation with mutual consistency training. In Proceedings of the Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, 27 September–1 October 2021; Proceedings, Part II 24; Springer: Berlin/Heidelberg, Germany, 2021; pp. 297–306.
32. Li, Y.; Wang, X.; Yang, L.; Feng, L.; Zhang, W.; Gao, Y. Diverse cotraining makes strong semi-supervised segmentor. *arXiv* **2023**, arXiv:2308.09281.
33. Wang, Y.; Xiao, B.; Bi, X.; Li, W.; Gao, X. MCF: Mutual correction framework for semi-supervised medical image segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 15651–15660.
34. Chen, T.; Kornblith, S.; Norouzi, M.; Hinton, G. A simple framework for contrastive learning of visual representations. In Proceedings of the International Conference on Machine Learning, PMLR, Virtual, 13–18 July 2020; pp. 1597–1607.
35. He, K.; Fan, H.; Wu, Y.; Xie, S.; Girshick, R. Momentum contrast for unsupervised visual representation learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 9729–9738.
36. Henaff, O. Data-efficient image recognition with contrastive predictive coding. In Proceedings of the International Conference on Machine Learning, PMLR, Virtual, 13–18 July 2020; pp. 4182–4192.
37. Wu, Z.; Xiong, Y.; Yu, S.X.; Lin, D. Unsupervised feature learning via non-parametric instance discrimination. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 3733–3742.
38. Wang, X.; Gao, J.; Long, M.; Wang, J. Self-tuning for data-efficient deep learning. In Proceedings of the International Conference on Machine Learning, PMLR, Virtual, 18–24 July 2021; pp. 10738–10748.
39. Hang, W.; Huang, Y.; Liang, S.; Lei, B.; Choi, K.S.; Qin, J. Reliability-aware contrastive self-ensembling for semi-supervised medical image classification. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Singapore, 18–22 September 2022; Springer: Berlin/Heidelberg, Germany, 2022; pp. 754–763.
40. Zhang, Y.; Zhang, X.; Li, J.; Qiu, R.C.; Xu, H.; Tian, Q. Semi-supervised contrastive learning with similarity co-calibration. *IEEE Trans. Multimed.* **2022**, *25*, 1749–1759. [[CrossRef](#)]
41. Li, J.; Xiong, C.; Hoi, S.C. Comatch: Semi-supervised learning with contrastive graph regularization. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Virtual, 11–17 October 2021; pp. 9475–9484.

42. Lee, D.; Kim, S.; Kim, I.; Cheon, Y.; Cho, M.; Han, W.S. Contrastive regularization for semi-supervised learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 3911–3920.
43. Yang, F.; Wu, K.; Zhang, S.; Jiang, G.; Liu, Y.; Zheng, F.; Zhang, W.; Wang, C.; Zeng, L. Class-aware contrastive semi-supervised learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 14421–14430.
44. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
45. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.
46. Berthelot, D.; Carlini, N.; Cubuk, E.D.; Kurakin, A.; Sohn, K.; Zhang, H.; Raffel, C. Remixmatch: Semi-supervised learning with distribution alignment and augmentation anchoring. *arXiv* **2019**, arXiv:1911.09785.
47. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-cam: Visual explanations from deep networks via gradient-based localization. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 618–626.

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