Delve into Base-Novel Confusion: Redundancy Exploration for Few-Shot Class-Incremental Learning

Haichen Zhou[1] , **Yixiong Zou**[1*] , **Ruixuan Li**[1] , **Yuhua Li**[1] and **Kui Xiao**[2]

[1]Huazhong University of Science and Technology
[2]Hubei University

{m202273832, yixiongz, rxli, idcliyuhua}@hust.edu.cn, xiaokui@hubu.edu.cn

Abstract

Few-shot class-incremental learning (FSCIL) aims to acquire knowledge from novel classes with limited samples while retaining information about base classes. Existing methods address catastrophic forgetting and overfitting by freezing the feature extractor during novel-class learning. However, these methods usually tend to cause the confusion between base and novel classes, i.e., classifying novel-class samples into base classes. In this paper, we delve into this phenomenon to study its cause and solution. We first interpret the confusion as the collision between the novel-class and the base-class region in the feature space. Then, we find the collision is caused by the label-irrelevant redundancies within the base-class feature and pixel space. Through qualitative and quantitative experiments, we identify this redundancy as the short-cut in the base-class training, which can be decoupled to alleviate the collision. Based on this analysis, to alleviate the collision between base and novel classes, we propose a method for FS-CIL named Redundancy Decoupling and Integration (RDI). RDI first decouples redundancies from base-class space to shrink the intra-base-class feature space. Then, it integrates the redundancies as a dummy class to enlarge the inter-base-class feature space. This process effectively compresses the base-class feature space, creating buffer space for novel classes and alleviating the model's confusion between the base and novel classes. Extensive experiments across benchmark datasets, including CIFAR-100, *mini*ImageNet, and CUB-200-2011 demonstrate that our method achieves state-of-the-art performance.

1 Introduction

Deep Neural Networks have achieved remarkable success in many domains [Liu *et al.*, 2020; Li *et al.*, 2022; Daihong *et al.*, 2022]. However, real-life scenarios often involve streaming data, posing challenges to obtaining sufficient data for

---

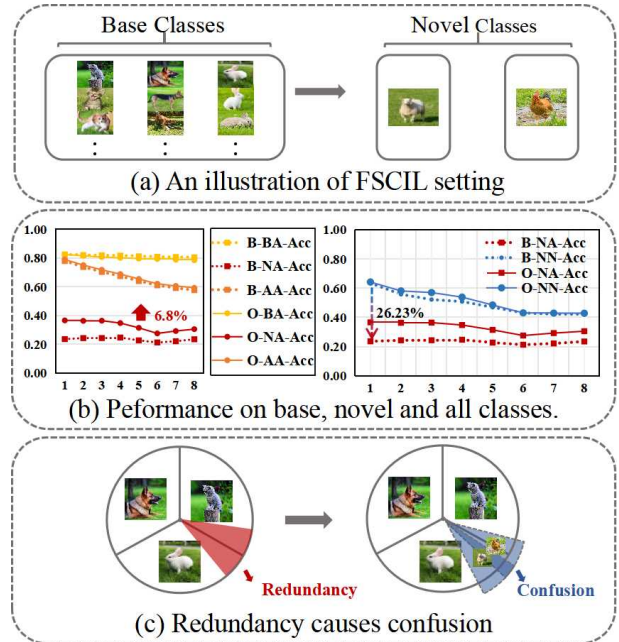*denotes the corresponding author.

Figure 1: (a) FSCIL entails mastering base classes through ample examples and incrementally addressing novel classes with few samples. (b) Left: Performance on base classes, novel classes, and all classes. BA-Acc, NA-Acc, and AA-Acc denote the average accuracy of classifying Base-class, Novel-class, and All-class samples against All encountered classes, respectively. B- represents using the Baseline fixed-backbone method, while O- represents using Ours. Right: Performance on novel-class samples. NN-Acc denotes the average accuracy of classifying Novel-class samples against only encountered Novel classes. We can see the model tends to classify novel-class samples into base classes, which is a major cause of the low performance in FSCIL. We term this problem as the confusion between novel and base classes, which we aim to interpret and alleviate in this paper. (c) Intuitive interpretation of the confusion in the feature space and the pixel space.

continuous learning of new knowledge, especially with limited samples. To address this, Few-Shot Class-Incremental Learning (FSCIL) [Tao *et al.*, 2020] was introduced. FSCIL entails learning base classes with an adequate number of samples in the base session, followed by learning novel classes using only a few samples in incremental sessions (Fig.1a).

The primary challenges in Few-Shot Class-Incremental Learning (FSCIL) center around the catastrophic forgetting of learned knowledge and the overfitting [Hu *et al.*, 2019] to novel classes due to limited data. To address these challenges, current approaches typically initiate training on base classes and then freeze the backbone while learning novel classes [Snell *et al.*, 2017]. However, although the fixed-backbone method maintains the model's performance on base classes, it impedes effective generalization to novel classes.

To investigate this issue, we briefly plot the performance of classifying base-class, novel-class, and all-class samples into all encountered classes, denoted as BA-Acc, NA-Acc, and AA-Acc, separately in Fig. 1b(left). We can see the yellow dashed line notably surpasses the red dashed line, signifying a considerable decrease in performance on novel-class samples compared to base-class ones. Then, we divide the novel-class performance into classifying novel-class samples against only novel classes and against all (i.e., base + novel) classes, denoted as NN-Acc and NA-Acc, respectively in Fig. 1b(right). We can observe the prediction accuracy significantly drops by 26.23% when base classes are introduced, which is a major cause of the low performance on novel-class samples. This phenomenon indicates the model tends to misclassify novel-class samples into base classes, and we call it the confusion between base and novel classes. Although this phenomenon widely exists in current FSCIL methods, only a few works [Wang *et al.*, 2023] attempted to explore its cause.

In this study, we delve into the cause of the confusion between base and novel classes, focusing on the redundancy in the feature space and the pixel space. We interpret the confusion as the collision between the novel-class region and the base-class region in the feature space, which is caused by the redundant area of base-class regions (Fig. 1c). Through qualitative and quantitative experiments, we find such redundancy reflects shortcuts of the model training, which can be decoupled to alleviate the collision between base and novel classes.

Based on the above analysis, to alleviate the confusion between base and novel classes, we propose an approach named Redundancy Decoupling and Integration (RDI) to address the redundancy within the base-class feature space. We first decouple the redundancy in the feature space by finding the label-irrelevant regions, and then map these regions to the pixel space for image masking, which would shrink the intra-base-class space. Then, we integrate these label-irrelevant regions by classifying masked images into a dummy class, which further compresses the base-class feature space and reserves a buffer space for novel classes. Consequently, the reserved space would reduce the collision between base and novel-class regions in the feature space, therefore alleviating the confusion between base and novel classes.

The contributions of this work are summarized as follows:

- Our experimental analysis reveals that the model's generalization on novel classes is majorly limited by the confusion between base and novel classes, predominantly attributed to label-irrelevant redundancies in the base-class feature space.

- Based on the analysis, we propose an innovative approach to mitigate the confusion between novel and base classes, which firstly decouples the label-irrelevant redundancies in both the feature space and the pixel space, and then integrates these redundancies to reserve a butter space for novel classes.

- Experimental results not only showcase the efficacy of our approach in enhancing generalization capacity on novel classes but also demonstrate its achievement of state-of-the-art performance on benchmark datasets, including CIFAR-100, CUB200, and *mini*ImageNet.

2 Interpretation of Base-Novel Confusion

In this section, we initially outline the Few-Shot Class-Incremental Learning (FSCIL) task and the baseline method. Then, we conduct experiments to interpret the cause of the confusion between base and novel classes.

2.1 Problem Set-up

Few-shot Class Incremental Learning (FSCIL) is a learning paradigm that involves mastering base classes with ample examples before tackling novel classes with limited samples. Typically, FSCIL employs a sequence of T training sessions denoted as $\{D_{\text{train}}^0, \ldots, D_{\text{train}}^{T-1}\}$, each associated with a distinct label space $\{C^0, \ldots, C^{T-1}\}$, where there is no overlapping between them. During each session, only the session-specific training set D_{train}^t is available, comprising examples $\{(x_i, y_i)\}_{i=1}^{n^t}$ with $y_i \in C^t$, and n^t denotes the number of examples. The testing set D_{test}^t, however, includes samples from all previously encountered and current classes, denoted as $D_{\text{test}}^t = \{(x_i, y_i)\}_{i=1}^{m^t}$, where $y_i \in \bigcup_{i=0}^t C^i$. The initial session $(t = 0)$ involves a training set with ample examples per class. The model is built as $f(x) = W^T \phi(x)$, where $\phi(x)$ represents feature extraction and W denotes the classifier for base class categorization. Subsequent incremental sessions $(0 < t < T)$ consisting of limited samples are organized as N-way K-shot tasks. In these sessions, the model learns novel classes while maintaining performance on base classes.

2.2 Baseline Description

The baseline model follows a structured approach for incremental learning. In the base session, the model undergoes training on D_{train}^0 utilizing the cross-entropy loss:

$$L = \frac{1}{N} \sum_{i=1}^{N} -log \frac{e^{\tau cos(x_i, w_c)}}{e^{\tau cos(x_i, w_c)} + \sum_{m \neq c} e^{\tau cos(x_i, w_m)}} \quad (1)$$

Moving to the incremental session, each base class's classifier is replaced with the class's average embedding, known as the prototype. The prototype calculation is denoted as

$$p_c = \frac{1}{n_c^0} \sum_{j=1}^{D_{train}^0} \mathbb{I}(y_j = c)\phi(x_j) \quad (2)$$

where $\mathbb{I}$ represents the indicator function, $\phi(\cdot)$ is the feature extractor, and n_c^0 denotes the count of examples in class c. During the incremental session, the classifier for base classes remains fixed, while for novel classes, the classifier can be

refined through fine-tuning or by employing the average embedding. During evaluation in session T, an input x is predicted using its output feature $\phi(x)$ and the classifier W, by:

$$\hat{y} = argmax_k \quad cos(\phi(x), w_k), \forall 1 \le k \le \sum_{i=0}^{T} N^i \quad (3)$$

2.3 Interpretation of Confusion Between Novel and Base Classes

In Fig. 1b, we conclude that the major limitation of the novel-class generalization: the model tends to classify novel-class test samples into base classes, termed "Base-Novel Confusion". In this section, we delve into this confusion from the redundancy in the feature space.
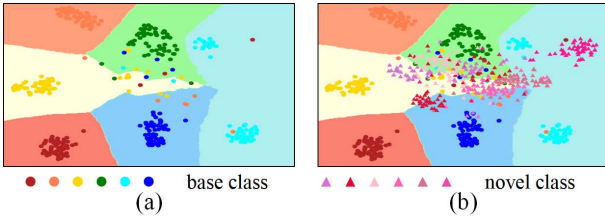

Figure 2: (a) Feature distribution and decision boundary of six base classes. (b) Feature distribution of six introduced novel classes, based on the six existing base classes. The introduced novel classes demonstrate a potential for confusion with the base classes.

Confusion represented as collisions in the feature space
To explore the confusion between base and novel classes, we first plot the feature distribution for both base and novel classes in Fig. 2. Initially, we randomly select sixty samples from each of six randomly sampled base classes in CIFAR-100 and visualize the distribution of feature points (circles) and decision boundaries using t-SNE [Van der Maaten and Hinton, 2008], as shown in Fig. 2a. To simulate the baseline model's learning process of novel classes, we then visualize the feature distribution of six newly introduced (randomly sampled) novel classes together with base classes, as depicted in Fig. 2b. Notably, these novel classes, represented as triangles, fall within the space allocated for base classes, with some even closely approaching the feature points (circles) representing base classes. Although we can see clear clusters even for novel classes, novel samples are located within the decision boundary for each base class, leading to collisions in feature spaces and confusing the model to classify novel-class samples into base classes. In essence, the collision exists because the model learns to occupy the entire feature space during the base-class training. Therefore, a natural question arises: *do base classes really need to occupy so much space to make the model discriminative on them?*

Redundancies caused by the ordinary classification loss
To answer the above question, we look back into the classification loss utilized to train the baseline model. The ordinary cross-entropy loss in Eq.1 can be rewritten as illustrated in Fig.3. To minimize this loss, the model needs to increase $e^{\tau cos(x_i,w_c)}$ and decrease $\sum_{m \ne c} e^{\tau cos(x_i,w_m)}$. In other
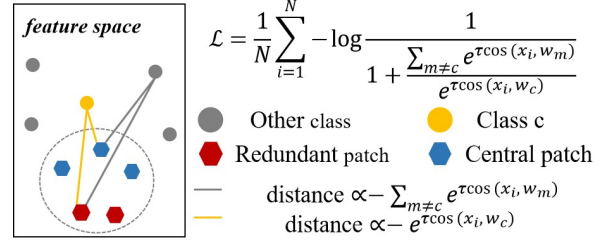

Figure 3: Because of the misalignment between decreasing $\sum_{m \ne c} e^{\tau cos(x_i,w_m)}$ and increasing $e^{\tau cos(x_i,w_c)}$, the ordinary cross-entropy loss could push the model to learn redundant features that are both distant from other classes and the class m.

words, the model learns to (1) push features in the class c close to class c's centroid while (2) pulling away these features from other classes. However, these two targets may not be aligned, because **features away from other classes may not be close to the class** c, which could lead to redundancies in the feature space. For example, in Fig.3, although red hexagons (could be features of an instance or an image patch) are pulled away from other classes (the gray circle) by decreasing $\sum_{m \ne c} e^{\tau cos(x_i,w_m)}$, it may not be close to class c's centroid (the yellow circle). This means the features represented by red hexagons, although may help the classification of the class c, may not really capture the central characteristics of the class c, which could be understood as a kind of shortcut that overfits the class c.

Dataset	$e^{\tau cos(x,w_c)}$		$\sum_{m \ne c} e^{\tau cos(x,w_m)}$	
	Redundant Patches	Central Patches	Redundant Patches	Central Patches
CIFAR-100	4.50	69.69	1.70	3.10
*mini*ImageNet	4.21	36.66	0.87	1.60
CUB200	7.28	20.72	1.31	5.67

Table 1: Similarities between patches and the ground truth class as well as other classes. Redundant patches demonstrate smaller similarities from both kinds of classes, consistent with Fig.3.

To validate our hypothesis, we compute the similarity of each patch $f(a,b) \in \mathbb{R}^d$ of the feature map $F \in \mathbb{R}^{h \times w \times d}$ using a pre-trained model from its ground truth class, and categorize the patches into central and redundant patches by thresholding the similarity. The central patch is close to its ground truth class, while the redundant patch is far from that class. Subsequently, we utilize $e^{\tau cos(x_i,w_c)}$ and $\sum_{m \ne c} e^{\tau cos(x_i,y_m)}$ to compute the average similarity between patches and classes, as depicted in Tab.1. Redundant patches are not only distant from both the ground truth class but also far from other classes. This quantitatively indicates that the model tends to learn label-irrelevant but discriminatory redundant patches to minimize $\sum_{m \ne c} e^{\tau cos(x_i,y_m)}$, thereby reducing the overall loss.

Redundancies as label-irrelevant regions in the heatmap
To ascertain whether redundancies are genuinely label-irrelevant, we examine them in the pixel space. Initially, we scrutinize the feature map of base classes under the pre-trained baseline model, visualizing the output from the last

convolutional layer as a heat map using ProtoNet [Snell *et al.*, 2017] (Fig.4, 'Feature Map'). Although the pre-trained model can identify semantic objects with higher activations, not all activated regions necessarily align with the label. For instance, in the visualization (third row of Fig.4), activated regions corresponding to both a fish and a part of stones were identified, despite the image being labeled as a fish. The parts of the stones are considered irrelevant to the label, while the fish is relevant. We subsequently isolate central and redundant patches from the original feature map by their similarities with the ground truth class following Tab.1. We can see the redundant patches correspond to label-irrelevant but discriminatory regions. Based on the above experiments, we conclude that the ordinary cross-entropy loss leads the model to focus not only on label-relevant features (central patches) but also on label-irrelevant redundancies for discrimination.

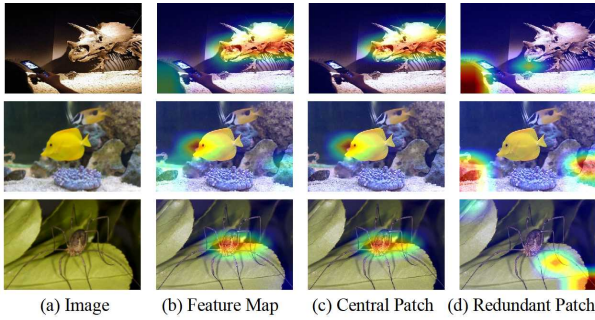(a) Image   (b) Feature Map   (c) Central Patch   (d) Redundant Patch

Figure 4: Decoupling of feature map into central and redundant patches. The redundant patches correspond to label-irrelevant objects in the base-class pixel space.

Since redundant features also occupy spaces within the decision boundary of each base class, it could lead to the collision between base and novel classes in the feature space. This inspires us to revisit the feature space to explore the relationship between redundancies and collision.

Redundancies lead to collisions in the feature space
We first decouple the Activated Label-Irrelevant (ALI) features and Activated Label-Relevant (ALR) features from the original feature map, where ALI corresponds to the redundant patches and ALR corresponds to the central patches. Then, we visualize the feature distribution and decision boundaries of ALI features and ALR features from the six previously selected classes, as shown in Fig.5. We can see that ALI features (gray fork) are distributed around ALR features, and ALR features exhibit a more compact distribution compared to the original feature distribution. In other words, these ALI features, acting as redundancies in feature space, expand the base-class space. Without these redundancies, the base-class space would be more compact.

Therefore, we hypothesize that these redundancies in the base-class feature space are the primary cause of the model's confusion between base and novel classes. To substantiate this, we integrate all ALI features as a dummy class and visualize the decision boundaries of base classes (using ALR features) and the dummy class (grey region). Then, we plot the feature distribution of novel classes (Fig.5 right). We observe
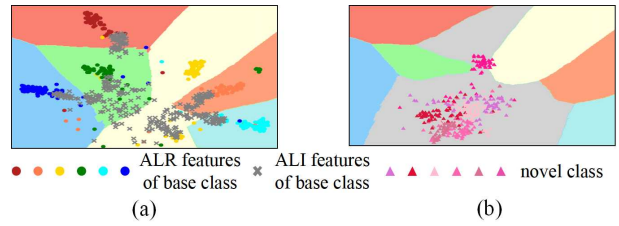

Figure 5: (a) Visualization of the distribution of ALR and ALI features from base classes. Pixel-space Redundancies (ALI features) imply redundancy in feature space and enlarge the base-class space. (b) Visualization of the distribution of newly introduced novel classes, based on the existing ALR and ALI features. The gray areas indicate regions of base-class redundancies. We can see novel classes are located in this gray region, indicating that redundancy in the base-class feature space leads to confusion.

that novel classes are concentrated in the grey region where redundant ALI features are distributed, which confirms our speculation that the redundancies lead to the collision.

Conclusion and Discussion
We conclude that the confusion between novel base classes can be interpreted as the collisions in the feature space. Based on the analysis of the regular classification loss, we find that it will encourage the model to learn label-irrelevant but discriminatory patches, which could be redundancies. To investigate the relationship between redundancies and collision in feature space, we visualize the redundancies in feature space and find that it is the redundancies in base-class space that mainly cause the collision between base and novel classes.

3 Proposed Method

While the baseline approach addresses catastrophic forgetting to some extent by employing a fixed feature extractor, it compromises the model's capacity to generalize over novel classes. This raises a natural question:
    How can we improve novel-classes generalization while maintaining base-classes performance?

Based on the above analysis, we pinpoint the key factors leading to inadequate generalization of existing methods in novel classes—the confusion between novel and base classes. This confusion predominantly stems from redundancy within the base-class feature space. In response, we introduce our method, Redundancy Decoupling and Integration (RDI), for Few-Shot Class Incremental Learning (FS-CIL). RDI comprises two modules. The first module decouples redundancy—Activated Label-Irrelevant (ALI) features—from the base-class feature space, utilizing Activated Label-Relevant (ALR) features for base-class classification. In the second module, we integrate redundancies as a dummy class to contribute to ALR-based recognition. An overview of our method is illustrated in Fig.6.

3.1 Shrink Intra-Base-Class Space by Decoupling Redundancy from Base-Class Space

For a more precise decoupling of the base-class space, we propose a two-step approach. Firstly, we pre-train the model using the cross-entropy loss, obtaining a classifier
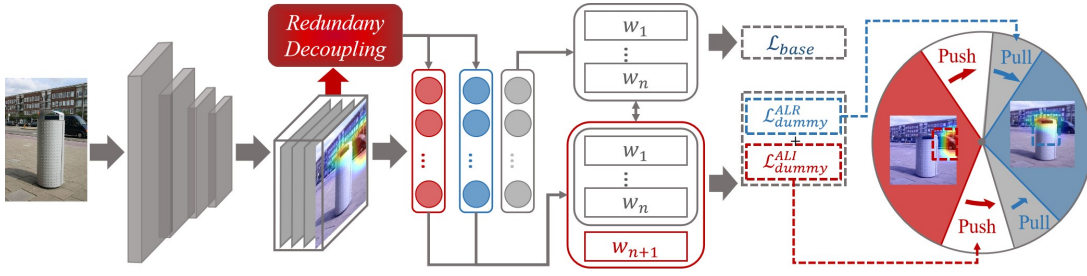
Figure 6: Our FSCIL framework: (1) Decoupling redundant ALI features from the base-class space and utilizing ALR features for base-class classification, minimizing intra-base-class space (loss term $\mathcal{L}_{dummy}^{ALR}$). (2) Integrating redundant ALI features to generate samples of a dummy class for ALR-based recognition, reserving space for novel classes and reducing inter-base-class space (loss term $\mathcal{L}_{dummy}^{ALI}$).

$W_{\text{base}} = [w_1, \ldots, w_n] \in \mathbb{R}^{d \times n}$ and a feature extractor $\Phi(\cdot)$. Here, n is the number of base classes. Next, we initiate the decoupling process by segmenting the feature map output from the pre-trained model into ALR (*Activated Label-Relevant*) and ALI (*Activated Label-Irrelevant*) regions.

Based on the analysis in Section 2, we aim to decouple redundancies by computing the distance between each patch $f(a, b) \in \mathbb{R}^d$ of the feature map $F \in \mathbb{R}^{h \times w \times d}$ and its ground truth class and generating an ALR mask $M^{ALR} \in \mathbb{R}^{h \times w}$:

$$M^{ALR}(a, b) = \sigma(\sum_{k=1}^{d} f_k(a, b) w_{y_p}^k) \qquad (4)$$

where $\sigma(\cdot)$ is an activation function with a threshold ϱ as a hyperparameter, y_p is the predicted label, and $M^{ALR} \in \mathbb{R}^{h \times w}$ is a 0-1 mask, Considering the possibility of misclassified samples and incorrectly activated feature maps after the pre-training stage, we suggest using the predicted label instead of the real label for all samples to ensure the validity of the feature map. The predicted label y_p can be obtained by:

$$y_p = argmax \quad (\frac{W_{base}}{\|W_{base}\|_2})^T (\frac{\Phi(x)}{\|\Phi(x)\|_2}) \qquad (5)$$

Subsequently, the ALR feature f^{ALR} is obtained as:

$$f^{ALR} = Pooling(M^{ALR}F) \qquad (6)$$

We label the ALR feature with its real label, i.e., $y^{ALR} = y$. Consequently, the cross-entropy loss of using the ALR feature for classification can be represented as:

$$\mathcal{L}_{base}^{ALR} = CE\left((\frac{W_{base}}{\|W_{base}\|_2})^T(\frac{f^{ALR}}{\|f^{ALR}\|_2}), y\right), \qquad (7)$$

The classification loss using the original feature is denoted as $\mathcal{L}_{base}$. The total loss is then expressed as a weighted sum of the original loss and the ALR loss: $\mathcal{L} = \mathcal{L}_{base} + \lambda \mathcal{L}_{base}^{ALR}$. Through $\mathcal{L}_{base}^{ALR}$, the model guides central patches to approach the ground truth class and distance themselves from other classes. This strategy mitigates shortcuts in the base-class training, effectively decoupling redundancies from the base-class space.

3.2 Reduce Inter-Base-Class Space by Integrating Redundancies as a Dummy Class

In contrast to previous methods, our innovative approach integrates redundancies as a dummy class to contribute to ALR-based recognition. This process compresses intra-base-class space and reserves a space for novel classes, effectively mitigating confusion between novel and base classes.

Building on the insights from Section 2, we categorize patches in the feature map into two groups: central patches (ALR) and redundant patches (ALI). Subsequently, we obtain the ALI mask from the ALR mask, denoted as $M^{ALI}(a, b) = 1 - M^{ALR}(a, b)$. Utilizing this ALI mask, we derive the ALI feature $f^{ALI} \in \mathbb{R}^d$.

$$f^{ALI} = Pooling(M^{ALI}F) \qquad (8)$$

Afterward, the ALI feature is labeled as $n + 1$, indicated as $y^{ALI} = n + 1$. To set aside space for novel classes, we adjust the classifier by incorporating the dummy class, yielding $W_{dummy} = [W_{base}, w_{n+1}] \in \mathbb{R}^{d \times (n+1)}$. Subsequent to this modification, the loss for the ALR feature can be updated:

$$\mathcal{L}_{dummy}^{ALR} = CE\left((\frac{W_{dummy}}{\|W_{dummy}\|_2})^T(\frac{f^{ALR}}{\|f^{ALR}\|_2}), y\right) \qquad (9)$$

Then, we can calculate the loss for the ALI feature:

$$\mathcal{L}_{dummy}^{ALI} = CE\left((\frac{W_{dummy}}{\|W_{dummy}\|_2})^T(\frac{f^{ALI}}{\|f^{ALI}\|_2}), n+1\right) \qquad (10)$$

The total is expressed as:

$$\mathcal{L} = \mathcal{L}_{base} + \lambda \mathcal{L}_{dummy}^{ALR} + \beta \mathcal{L}_{dummy}^{ALI} \qquad (11)$$

In this equation, the first term ensures base-class performance. The second term facilitates the separation of redundancies from the base-class space, leading to a reduction in intra-base-class space. The third term introduces a dummy class to compress inter-base-class space, thereby creating room for novel classes. The combination of the last two terms alleviates the model's confusion between novel and base classes. λ and β are hyper-parameres.

Before the incremental session, we extract base-class average embedding (prototypes) as classifiers for base classes using Eq.2. During the incremental session, we maintain the feature extractor $\Phi(\cdot)$ fixed and employ it to extract features from novel-class examples for inference.

4 Experiment

In this section, we introduce benchmark datasets and implementation details, validate our method's effectiveness through qualitative and quantitative analyses, perform ablation experiments to analyze component contributions, and compare our method's performance with state-of-the-art approaches on benchmark datasets.

4.1 Implementation Details

Datasets & Split: Following [Yang *et al.*, 2023], Our method is evaluated on CIFAR-100 [Krizhevsky *et al.*, 2009], *mini*ImageNet [Russakovsky *et al.*, 2015], and CUB200 [Wah *et al.*, 2011]. CIFAR-100 has 60,000 images (64×64) across 100 classes, *mini*ImageNet has 60,000 images (84×84) representing 100 classes from ImageNet [Deng *et al.*, 2009], and CUB200 comprises 11,788 fine-grained images (224×224). Both *mini*ImageNet and CIFAR-100 have 60 base classes initially and 40 novel classes distributed across 8 incremental sessions as 5-way 5-shot tasks. For CUB200, the initial session has 100 base classes, and 100 novel classes are distributed over 10 incremental sessions as 10-way 5-shot tasks. **Training Details:** Our training protocol follows [Yang *et al.*, 2023], utilizing the ResNet-12 for CIFAR-100 and *mini*ImageNet, and pretrained ResNet-18 for CUB200. Batch sizes of 128 and 100 are employed for the base session and incremental sessions, respectively.

4.2 Qualitative and Quantitative Analysis

This section aims to substantiate key claims regarding our method's efficacy through a series of experiments:

- **Decoupling Redundancy**
- **Compressing Base-Class Space**
- **Alleviating the Collision in Feature Space**
- **Mitigating Base-Novel Confusion**

Decoupling Redundancy from Base-class Space

We randomly choose images from the *mini*ImageNet dataset to visualize the feature map. The heat maps in Fig.7 show changes in the feature map. The top row represents the baseline, and the bottom row showcases our method. Significantly, our method directs the model's attention to the label-relevant region, effectively decoupling label-irrelevant redundancies from the base-class space.
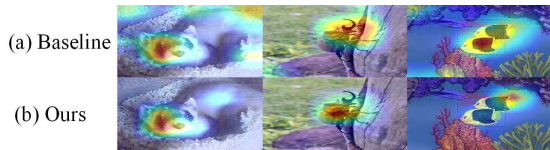

(a) Baseline

(b) Ours

Figure 7: Compared to the baseline method, ours can decouple redundancies and make the model focus on label-relevant regions.

Compressing Base-class Space

To quantitatively showcase our method's efficacy in reducing base-class space, we measure intra-base-class and inter-base-class cosine distances on CIFAR-100 (Fig.8). Our approach (solid red line) adeptly diminishes both intra-base-class and inter-base-class space, consistently placing cumulative distribution functions to the left of the baseline (gray dashed line).

Alleviating Collisions in Feature Space

To visually highlight our method's efficacy, we randomly selected twelve base and ten novel classes from CIFAR-100, each with twenty samples. Visualizing the feature distribution

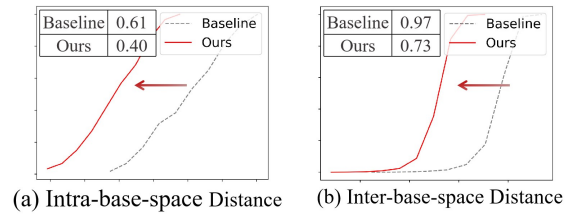(a) Intra-base-space Distance   (b) Inter-base-space Distance

Figure 8: The frequencies cumulative distribution diagram depicts inter-base-class and intra-base-class distance, where the abscissa represents cosine distance, and the ordinate shows cumulative frequency. The average cosine distance is presented in the upper left corner. Our method (red solid line) consistently positions to the left of the baseline (gray dotted line), indicating reductions in both inter-base-class and intra-base-class distances.

after inserting novel classes using both baseline and our methods (Fig.9), the left side shows novel classes (green squares) are mixed with base classes (orange dots) in the baseline. In contrast, on the right side, with our method, novel classes are located within the dummy class (* marked in gray) region and are separated with base classes, showcasing our method's ability to alleviate feature space collisions.

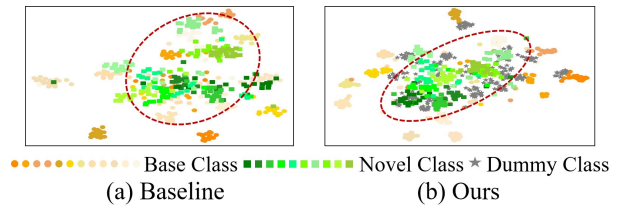Base Class ▪▪▪▪▪▪ Novel Class ★ Dummy Class
(a) Baseline   (b) Ours

Figure 9: The red dotted line outlines the extent of novel classes, showing collisions between novel and base classes in the baseline approach. In contrast, our method introduces a dummy class to reserve space for novel classes, effectively alleviating the collisions.

Mitigating the Model's Confusion

We quantitatively demonstrate our method's effectiveness in reducing the model's confusion between novel and base classes on *mini*ImageNet. Using NN-acc and NA-Acc metrics in Section 2, the larger difference between them indicates increased confusion between base and novel classes. This evaluation, depicted in Fig. 10, spans both the baseline and our model. Results unequivocally emphasize our approach's success in mitigating the model's confusion.

4.3 Ablation Study

We conducted ablation studies on three benchmark datasets, summarizing key findings in Table 3. Results highlight our approach's pivotal role in enhancing final session accuracy and improving novel-class generalization. Consistently across benchmark datasets, introducing both $\mathcal{L}_{dummy}^{ALR}$ and $\mathcal{L}_{dummy}^{ALI}$ enhances the model's ability to learn novel classes. The combination of two loss terms proves highly effective in reducing confusion, leading to substantial improvements across all classes. These ablation experiments provide insights into contributions of each component, emphasizing the overall effectiveness of our proposed method.

Method	Accuracy in each session(%)									Average
	0	1	2	3	4	5	6	7	8	Acc.
CEC [Zhang *et al.*, 2021]	73.07	68.88	65.26	61.19	58.09	55.57	53.22	51.34	49.14	59.53
LIMIT [Zhou *et al.*, 2022b]	73.81	72.09	67.87	63.89	60.70	57.77	55.67	53.52	51.23	61.84
Meta FSCIL [Chi *et al.*, 2022]	74.50	70.10	66.84	62.77	59.48	56.52	54.36	52.56	49.97	60.79
Self-promoted [Zhu *et al.*, 2021]	64.10	65.86	61.36	57.45	53.69	50.75	48.58	45.66	43.25	54.52
FACT [Zhou *et al.*, 2022a]	74.60	72.09	67.56	63.52	61.38	58.36	56.28	54.24	52.10	62.24
Data-free Replay [Liu *et al.*, 2022]	74.40	70.20	66.54	62.51	59.71	56.58	54.52	52.39	50.14	60.78
ALICE [Peng *et al.*, 2022]	79.00	70.50	67.10	63.40	61.20	59.20	58.10	56.30	54.10	63.21
Ours	**81.45**	**77.02**	**72.73**	**68.95**	**65.75**	**63.02**	**61.07**	**59.01**	**56.72**	**67.30**
NC-FSCIL(w/ mixup) [Yang *et al.*, 2023]	82.52	76.82	73.34	69.68	66.19	62.85	60.96	59.02	56.11	67.50
Ours(w/ mixup)	**82.52**	**78.00**	**73.69**	**69.73**	**66.27**	**63.48**	**61.46**	**59.66**	**57.48**	**68.03**

Table 2: Top-1 average accuracy on all seen classes during each incremental session on CIFAR-100. Results for the compared methods are sourced from [Yang *et al.*, 2023].
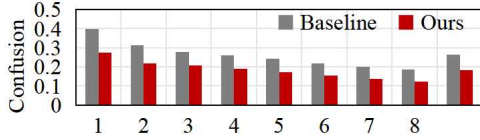

Figure 10: Comparing the confusion levels between novel and base classes on the baseline model (gray column) and ours(red column), the results demonstrate that ours effectively alleviates confusion.

Figure 11: Detailed average accuracy during each incremental session on *mini*ImageNet and CUB200.

4.4 Comparison with the State-of-the-Art Methods

We comprehensively compare our proposed method with existing approaches across three FSCIL benchmarks. Performance curves for *mini*ImageNet and CUB200 are shown in Figure 11, and detailed performance values for CIFAR100 are provided in Table 2. Our method consistently outperforms previous approaches in average accuracy across the benchmarks, including a notable improvement over NC-FSCIL [Yang *et al.*, 2023], a recent top-performing method. Specifically, our method achieves marked improvements of 1.37%, 0.94%, and 0.76% in the last session on CIFAR-100, *mini*ImageNet, and CUB200, respectively. Moreover, we substantiate our superiority by attaining an average accuracy enhancement of 1.74% on *mini*ImageNet.

Method	CIFAR-100		*mini*ImageNet		CUB200	
	novel	average	novel	average	novel	average
$\mathcal{L}_{base}$	19.62	54.53	23.50	57.64	40.78	58.63
$\mathcal{L}_{base} + \mathcal{L}_{dummy}^{ALR}$	22.77	54.87	26.82	58.58	42.05	58.92
$\mathcal{L}_{base} + \mathcal{L}_{dummy}^{ALI}$	24.40	56.28	23.75	56.94	42.59	59.61
$\mathcal{L}_{base} + \mathcal{L}_{dummy}^{ALR} + \mathcal{L}_{dummy}^{ALI}$	26.42	56.72	30.52	59.25	44.13	60.20

Table 3: Ablation study on three benchmark datasets.

5 Related Work

Few-Shot Class-Incremental Learning (FSCIL) [2019; 2017; 2021] seamlessly combines Class Incremental Learning (CIL) [2018; 2019; 2022] and Few-Shot Learning (FSL) [2019; 2020; 2018]. FSCIL prioritizes mastering base classes with abundant examples initially and then adeptly addressing novel classes with few instances in each incremental session.

FACT [Zhou *et al.*, 2022a] addresses feature distribution discrepancies between novel and base classes by introducing forward compatibility. It allocates space for novel classes
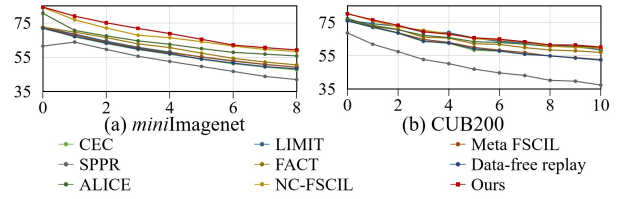
through virtual instances from instance fusion. However, concerns may arise regarding the accuracy of synthesized samples in representing real-world instances.

N-FSCIL [Yang *et al.*, 2023], the leading approach, pre-assigns classifier prototypes for the entire label spectrum to rectify feature-classifier misalignment for old classes. It reserves space for novel classes by referencing class prototypes to accommodate both base and novel categories. However, this method requires prior knowledge of novel class counts, limiting its practical applicability.

TEEN [Wang *et al.*, 2023], closely related to our work, addresses the lower performance of novel classes in FSCIL, highlighting the model's tendency to misclassify them into base classes. It introduces a training-free prototype calibration method based on semantic similarity between base and novel classes. In contrast, our study thoroughly analyzes the root cause of low novel class performance from a feature distribution perspective, proposing a more effective and widely applicable solution to this issue.

6 Conclusion

In this paper, we address the challenge of confusion between novel and base classes within the Few-Shot Class-Incremental Learning framework, particularly when using a fixed backbone trained on base classes. Through a thorough analysis of the root cause behind the model's confusion in existing methods – redundancy within the feature space of the base class leading to collisions between novel and base classes in feature space – we introduce the Redundancy Decoupling and Integration (RDI) method. Extensive evaluations on benchmark datasets affirm RDI's effectiveness, significantly enhancing novel-class generalization and achieving state-of-the-art performance across all classes.

Acknowledgments

References

[Aljundi *et al.*, 2018] Rahaf Aljundi, Francesca Babiloni, Mohamed Elhoseiny, Marcus Rohrbach, and Tinne Tuytelaars. Memory aware synapses: Learning what (not) to forget. In *Proceedings of the European conference on computer vision (ECCV)*, pages 139–154, 2018.

[Chen *et al.*, 2019] Wei-Yu Chen, Yen-Cheng Liu, Zsolt Kira, Yu-Chiang Frank Wang, and Jia-Bin Huang. A closer look at few-shot classification. *arXiv preprint arXiv:1904.04232*, 2019.

[Chi *et al.*, 2022] Zhixiang Chi, Li Gu, Huan Liu, Yang Wang, Yuanhao Yu, and Jin Tang. Metafscil: a meta-learning approach for few-shot class incremental learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14166–14175, 2022.

[Daihong *et al.*, 2022] Jiang Daihong, Zhang Sai, Dai Lei, and Dai Yueming. Multi-scale generative adversarial network for image super-resolution. *Soft Computing*, 26(8):3631–3641, 2022.

[Deng *et al.*, 2009] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

[Dong *et al.*, 2021] Songlin Dong, Xiaopeng Hong, Xiaoyu Tao, Xinyuan Chang, Xing Wei, and Yihong Gong. Few-shot class-incremental learning via relation knowledge distillation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 1255–1263, 2021.

[Hou *et al.*, 2019] Saihui Hou, Xinyu Pan, Chen Change Loy, Zilei Wang, and Dahua Lin. Learning a unified classifier incrementally via rebalancing. In *Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition*, pages 831–839, 2019.

[Hu *et al.*, 2019] Wenpeng Hu, Zhou Lin, Bing Liu, Chongyang Tao, Zhengwei Tao Tao, Dongyan Zhao, Jinwen Ma, and Rui Yan. Overcoming catastrophic forgetting for continual learning via model adaptation. In *International conference on learning representations*, 2019.

[Krizhevsky *et al.*, 2009] Alex Krizhevsky, Geoffrey Hinton, et al. Learning multiple layers of features from tiny images. 2009.

[Li *et al.*, 2022] Ruibo Li, Chi Zhang, Guosheng Lin, Zhe Wang, and Chunhua Shen. Rigidflow: Self-supervised scene flow learning on point clouds by local rigidity prior. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16959–16968, 2022.

[Liu *et al.*, 2020] Weide Liu, Chi Zhang, Guosheng Lin, Tzu-Yi Hung, and Chunyan Miao. Weakly supervised segmentation with maximum bipartite graph matching. In *Proceedings of the 28th ACM International Conference on Multimedia*, pages 2085–2094, 2020.

[Liu *et al.*, 2022] Huan Liu, Li Gu, Zhixiang Chi, Yang Wang, Yuanhao Yu, Jun Chen, and Jin Tang. Few-shot class-incremental learning via entropy-regularized data-free replay. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXIV*, pages 146–162. Springer, 2022.

[Peng *et al.*, 2022] Can Peng, Kun Zhao, Tianren Wang, Meng Li, and Brian C Lovell. Few-shot class-incremental learning from an open-set perspective. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXV*, pages 382–397. Springer, 2022.

[Rebuffi *et al.*, 2017] Sylvestre-Alvise Rebuffi, Alexander Kolesnikov, Georg Sperl, and Christoph H Lampert. icarl: Incremental classifier and representation learning. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 2001–2010, 2017.

[Russakovsky *et al.*, 2015] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115:211–252, 2015.

[Snell *et al.*, 2017] Jake Snell, Kevin Swersky, and Richard Zemel. Prototypical networks for few-shot learning. *Advances in neural information processing systems*, 30, 2017.

[Sung *et al.*, 2018] Flood Sung, Yongxin Yang, Li Zhang, Tao Xiang, Philip HS Torr, and Timothy M Hospedales. Learning to compare: Relation network for few-shot learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1199–1208, 2018.

[Tao *et al.*, 2020] Xiaoyu Tao, Xiaopeng Hong, Xinyuan Chang, Songlin Dong, Xing Wei, and Yihong Gong. Few-shot class-incremental learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12183–12192, 2020.

[Van der Maaten and Hinton, 2008] Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(11), 2008.

[Wah *et al.*, 2011] Catherine Wah, Steve Branson, Peter Welinder, Pietro Perona, and Serge Belongie. The caltech-ucsd birds-200-2011 dataset. 2011.

[Wang *et al.*, 2020] Haofan Wang, Zifan Wang, Mengnan Du, Fan Yang, Zijian Zhang, Sirui Ding, Piotr Mardziel, and Xia Hu. Score-cam: Score-weighted visual explanations for convolutional neural networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pages 24–25, 2020.

[Wang *et al.*, 2022] Zifeng Wang, Zizhao Zhang, Chen-Yu Lee, Han Zhang, Ruoxi Sun, Xiaoqi Ren, Guolong Su, Vincent Perot, Jennifer Dy, and Tomas Pfister. Learning to prompt for continual learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 139–149, 2022.

[Wang *et al.*, 2023] Qi-Wei Wang, Da-Wei Zhou, Yi-Kai Zhang, De-Chuan Zhan, and Han-Jia Ye. Few-shot class-incremental learning via training-free prototype calibration. *arXiv preprint arXiv:2312.05229*, 2023.

[Wu *et al.*, 2019] Yue Wu, Yinpeng Chen, Lijuan Wang, Yuancheng Ye, Zicheng Liu, Yandong Guo, and Yun Fu. Large scale incremental learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 374–382, 2019.

[Yang *et al.*, 2023] Yibo Yang, Haobo Yuan, Xiangtai Li, Zhouchen Lin, Philip Torr, and Dacheng Tao. Neural collapse inspired feature-classifier alignment for few-shot class incremental learning. *arXiv preprint arXiv:2302.03004*, 2023.

[Zhang *et al.*, 2021] Chi Zhang, Nan Song, Guosheng Lin, Yun Zheng, Pan Pan, and Yinghui Xu. Few-shot incremental learning with continually evolved classifiers. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12455–12464, 2021.

[Zhou *et al.*, 2022a] Da-Wei Zhou, Fu-Yun Wang, Han-Jia Ye, Liang Ma, Shiliang Pu, and De-Chuan Zhan. Forward compatible few-shot class-incremental learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9046–9056, 2022.

[Zhou *et al.*, 2022b] Da-Wei Zhou, Han-Jia Ye, Liang Ma, Di Xie, Shiliang Pu, and De-Chuan Zhan. Few-shot class-incremental learning by sampling multi-phase tasks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.

[Zhu *et al.*, 2021] Kai Zhu, Yang Cao, Wei Zhai, Jie Cheng, and Zheng-Jun Zha. Self-promoted prototype refinement for few-shot class-incremental learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6801–6810, 2021.