基于不确定性感知深度聚类的图像特征自监督学习

摘要

在当前图像处理的机器学习模型中，标注精准的图像数量受限是一个关键瓶颈，这凸显了自监督学习的重要性。自监督学习涉及为大量未标注图像赋予伪标签。通过用这些伪标签训练卷积神经网络，所学习的特征表示或卷积模式，可以作为预训练模型，转移到下游的图像处理任务中。现有的自监督视觉学习方法通常人为构造伪标签，例如，通过翻转或颜色修改。然而，我们认为，类别信息是数据集中自然存在的、内在的信息，作为伪标签，在各种情境中具有特别有效的表现。这凸显了基于聚类的方法与其他复杂技术之间的显著性能差异。

因此，我们提出了一种新的基于聚类的自监督学习方法。该方法迭代地融合高层和低层特征表示进行聚类，随后为样本赋予伪标签及其不确定性，这些伪标签随后用于端到端的视觉表示学习。通过在诸如 ImageNet 该种大型数据集上进行评估，我们证明了我们的方法超越了现有大多数常见自监督方法。

关键词：自监督学习；聚类；数据扩充；表征学习

1 引言

深度学习网络的效能往往得益于基于预先存在的模型进行训练。这些预训练模型提供关键的特征表示，有助于提高模型在后续任务中的泛化能力，并在标注数据不足的情况下加速训练过程 [9, 23]。通常，通过利用大型标注数据集（如 ImageNet）进行监督学习来实现此类预训练。然而，标注预训练数据集的规模限制了预训练特征表示的进一步提升，从而影响了许多新型深度学习架构的性能潜力 [3, 10]。然而对此类数据集的规模进行进一步提升需要大量数据清理和细致的手动标注。鉴于此，利用大量未标注的网络数据变得尤为重要，它允许我们绕过手动标注数据所伴随的大量劳动和时间限制，同时仍能提取有价值的见解和模式。然而，仅仅用原始数据替代标签会导致不理想的结果。因此，我们可以为大量的不带标签的数据设计一种伪标签，例如将图像的颜色变为灰色，然后将灰色的图像作为一种伪标签。伪标签的设计正是自监督学习的核心问题。

自监督学习是无监督学习的变种。在自监督学习中包含了两个关键，分别是前置任务和下游任务。在前置任务当中，我们需要根据设计的伪标签来进行卷积神经网络的训练。通过训练，卷积神经网络的卷积池化层代表的是对图像数据的一种特征提取模式。伪标签的设计是根据原始数据本身。例如，在 [16] 中，Larsson 等人提出了一种基于上色的伪标签方法用于进行自监督学习，正如刚刚所提到的，他们将彩色图像转化为了灰色的黑白图像，然后将图像的上色作为了一种前置任务。在 [4] 中，Chen 等人对图像进行了旋转，旋转的角度包括了 90 度，180 度等，随后将旋转之后的图像作为网络的输入，旋转的角度作为训练的监督信号。

以上方法都是对原始的数据（图像）进行一些人为的变换来得到伪标签进而进行自监督学习。但是与上面两种方法不同的是，在 [2] 中，Caron 等人提出了一种基于聚类的自监督方法。在该方法中，作者对图像的高级特征进行聚类，并且将聚类的结果作为了图像的类标，随后基于生成的伪类标来进行卷积神经网络的训练。以上方法都体现了在自监督学习中，我们的目的是通过设计伪类标，来让网络学习到图像等数据的特征提取模式，用于作为下游任务网络的预训练模型。常见的下游任务包括了图像分类、图像分割等。在自监督学习领域，预设任务至下游任务的流畅过渡是至关重要的。这种无缝的过渡不仅构成了自监督学习的核心架构，而且在实现理论与实践的融合方面发挥了关键作用。它有效地弥合了从未标注数据中提取知识到将这些知识应用于实际世界情景的差距。这一过程不仅提高了学习模型对未标注数据的理解和利用效率，而且增强了模型在处理真实世界问题时的适应性和准确性。通过这种方式，自监督学习展现出其在数据驱动的应用中的巨大潜力，为解决现实世界复杂问题提供了一种创新且高效的方法。

与一些基于原始图像本身构建的伪标签相比，通过聚类的方法生成伪类标的形式更符合数据集本身的特性。现有的大多数数据集是基于对象的类别信息来进行收集的，例如常用的 CIFAR-10 数据集包含了 10 类物体的图像（飞机、汽车、鸟、猫、鹿、狗、青蛙、马、船和卡车）。可见，类别信息是这些图像数据集中天然的内在标签，如果通过自监督学习得到的图像特征也能够反应数据的类别差异的话，能够帮助我们更好地解决下游任务。这就解释了在许多关于自监督学习的研究中，基于聚类的自监督方法都表现出了更加优异的性能 [8,13,14,17,18]。聚类能力在直观地解析和利用这些潜在的分类方面，使其成为自监督学习背景下一种特别有效的策略。基于这一视角，我们提出了一种不确定性感知深度聚类（Uncertainty-Awared Deepcluster，简称 UAD）的新型基于聚类的自监督学习方法。该方法采用循环过程，最初融合高层和低层特征。然后，它将这些特征进行聚类以生成伪类别，并为每个伪标签分配一个计算出的不确定性。随后，网络在这些伪标签及其相应不确定性的指导下进行训练。

在提出的研究方法中，通过整合低级和高级特征来增强聚类效果，进而产生更加准确可信的伪标签。处理过程中，实例首先通过卷积网络，其中高级特征从深层网络中提取，而低级特征则源自最初的卷积层。针对早期卷积层通常存在的高通道数问题，采用了一种随机选择的策略进行维度降低。接下来，将低级和高级特征的通道融合在一起。之后，这些特征会从矩阵格式转换为向量格式，为每个实例构建一个全面的向量表示。这一过程实现了低级和高级特征的有效融合。随后，根据 Caron 等人在 2018 年的研究 [2]，采用了 k-means 或成对实例聚类（PIC）等聚类算法对这些特征进行分类。最后，每个簇中的实例都被赋予一个统一的伪标签。

在提出的不确定性感知深度聚类（UAD）框架中，特别设计了一种针对由聚类结果派生出的伪标签的不确定性估计机制。在传统做法中，一批实例输入用于分类时，伪标签会被直接应用于交叉熵损失函数，并通过反向传播过程调整模型权重。这种处理通常基于一个假设：所有伪标签的可靠性都是均等的。然而，由于伪标签置信度的显著波动——如特征的模糊性或类内变异性所引起的 [1, 32]，这一假设可能存在缺陷。将同等的权重赋予不太可靠的伪标签和更可靠的伪标签，可能会误导模型，进而影响其准确性。为此，通过借鉴模糊聚类的概念 [19]，提出了一种新颖的不确定性估计方法。在提出的方法中，聚类完成后，采用最小-最大方法来规范化每个样本与其所属簇中心的距离，使其落在 [0,1] 的范围内。但注意到，如果直接将这种归一化值作为样本在分类交叉熵损失中的不确定性使用，可能会导致模型欠拟合。

为了克服这一挑战，我们对规范化过程进行了改进，并引入了一个超参数。这个调整后的值作为伪标签的不确定性。在基于伪标签进行分类和误差反向传播的过程中，这种不确定性被考虑在相应样本的损失项中。这种方法使得模型能够有效地从高置信度的伪标签中学习，同时最小化低置信度标签对学习不利视觉特征的影响。

该方法的创新体现在两个主要方面：首先，它在传统基于聚类的自监督学习方法之上进行了进一步的发展，通过整合低层次和高层次的特征，实现了更加可靠伪标签的生成。其次，该方法提倡在自监督学习过程中，不应将所有伪标签视为等同。本研究还提出了一种创新的估计基于聚类的自监督学习中伪标签不确定性的方法，这一新颖做法激发了自监督学习其他领域对于人工设计的伪标签中不确定性评估的重新思考，及其在引领更优秀特征表示方面的潜在作用。

文章的结构安排如下：第 2 节对自监督学习、深度聚类以及不确定性的基本概念进行了探讨。第 3 节则深入介绍了本研究提出的不确定性感知深度聚类（Uncertainty-Awared Deepcluster，简称 UAD）方法。第 4 节对复现细节，代码贡献和模型的使用细节进行了介绍，第 5 节展示了对该模型的分析，包括与其他方法的对比研究。最终，第 6 节对本研究的主要发现和贡献进行了总结。

2 相关工作

本节对自监督学习和不确定性的概念以及相关的研究进行了介绍。

2.1 自监督学习

作为无监督学习的一种领先变体，自监督学习在表示学习领域，尤其是在图像处理的视觉特征提取方面，已经引起了广泛关注 [5]。在 [20] 中，Noroozi 和 Favaro 提出了一个创新的方法，即训练卷积神经网络（CNN）通过解决 Puzzle 的预设任务来学习物体部分的特征映射及其空间排列。此外，Noroozi 等人在 [21] 中进一步提出了利用图像变换（主要是缩放和平铺）的自监督学习方法，其中经过修改的图像作为预设任务的伪标签。而在 [30] 中，Zhang 等人介绍了一种不同数据子集相互预测的方法，其中部分通道被用作该自监督任务的伪标签。

在本研究中，我们用 $f_\theta(x)$ 来代表图像的视觉表示，其中 x 指原始图像，而 θ 代表模型的参数。通常，$f_\theta(x)$ 被视为卷积网络中间层的输出结果。针对一系列特定的图像，本研究致力于有效学习一个理想的 $f_\theta(x)$，这一过程对于增强计算机视觉领域中多种下游任务的表现具有重要价值。

自监督学习的主要目标是从原始数据中派生出有效的伪标签。以一个包含 N 张图像的数据集为例，可以表示为 $\{x_1, x_2, x_3, ..., x_N\}$，针对每张图像，我们将生成对应的伪标签 y_n。自监督学习的精髓在于利用这些伪标签作为指导，通过反向传播学习图像的视觉表征 $f_\theta(x_n)$。这一过程可以用以下公式表达：

$$\min_{\theta,V} \sum_{n=1}^{N} \ell(y_n, V f_\theta(x_n)), \tag{1}$$

其中，V 代表紧接在表征提取之后的分类器，而 ℓ 为模型中采用的损失函数。通过这种方式学习的视觉模式 f_θ 可作为下游任务的预训练框架，其有效性在很大程度上取决于伪标签的设计和结构。

设计自监督学习中的伪标签，可以从三个方面入手。首先，通过对原始数据进行变换，如裁剪图像，可以创造伪标签。其次，构建与回归任务相关联的连续变量，例如预测图像颜色或邻近区块，也是一种可行的方法。此外，利用数据中的内在类别信息，可以通过聚类为图像生成伪类别，随后将这些伪类别作为伪标签使用。这种运用聚类技术的方法被称为深度聚类。

基于聚类的自监督学习方法的核心目标是通过对卷积网络输出的视觉特征 $f_\theta(X)$ 进行聚类，从而形成伪标签。在这一框架下，每张图像 x_n 被赋予一个在集合 Z 中的潜在伪标签 z_n。利用分类器 W，模型在这些伪标签的引导下进行反向传播。相较于其他伪标签类型，聚类标签更能有效捕捉数据集内固有的类别信息。通过视觉特征的持续聚类和反向传播，学习到的视觉模式 f_θ 被应用于多种下游任务。参数 θ 和 W 的优化可表达为：

$$\min_{\theta,W} \sum_{n=1}^{N} \ell(z_n, W f_\theta(x_n)), \tag{2}$$

此公式与方程式 (1) 的形式保持一致。遵循 [2,28] 的方法，我们选择 k-均值作为聚类手段。具体而言，目标 z_n 通过解决如下优化问题而确定：

$$\min_{C\in\mathbb{R}^{d\times k}} \sum_{n=1}^{N} \left[\min_{z_n\in\{0,1\}^k \text{s.t.} z_n^\top \mathbf{1}=1} \|C z_n - f_\theta(x_n)\|_2^2 \right] \tag{3}$$

其中，C 代表每列与一个簇心对齐的矩阵，而 k 是簇心的数量。一般而言，k 的值被视为先验知识，并在下游任务中进行验证。聚类是通过获取整个数据集的 f_θ 实现的，随后我们计划利用随机梯度下降方法进行 T 次反向传播。这种交替的优化方案可能导致平凡解的出现，因此，确保两个优化目标间有效互动至关重要。为避免平凡的参数化问题，实施了重新分配空聚类和基于聚类分配均匀分布的批量采样等策略 [2]。

2.2 不确定性

在决策理论领域，" 不确定性" 这一术语用于描绘那些由于信息缺乏而无法精确刻画当前状况或有效预测未来结果的情形 [15]。机器学习的实践涉及基于不完整样本的智能预测，这些样本可能无法全面代表整个数据集，从而在学习过程中引入不确定性，这一点在各种算法或模型中均是如此。另外，在大数据环境下，数据的多样化表现形式、庞大的特征维度和类别数量进一步放大了这种不确定性。面对诸如数据缺失、庞杂的解决方案空间、数据噪声、样本间依赖性违反、不平衡数据导致的长尾分布以及大量超参数等挑战，机器学习建模过程的复杂性显著提高，从而影响了传统算法的有效性 [26]。

根据 [12] 的研究，存在两种主要的不确定性类型：偶然不确定性（数据不确定性）和认知不确定性（模型不确定性）。数据不确定性指的是数据生成过程中固有的随机性，例如数据的潜在错误标注。而模型不确定性则反映了系统对正确模型的认识不足，导致在做出准确预测时存在不确定性。在 [25] 中，作者特别指出，在分类任务中由于预测目标的不连续性，分类器输出空间产生了另一种特定形式的不确定性。这种特殊的不确定性，即从分类器输出向量到最终类别决定的过渡，被定义为输出不确定性。

数据不确定性 在机器学习领域，数据通常以“特征-标签”对 $\{x, y\}$ 的形式表示。数据收集过程中，由于环境因素和标注者的影响，收集得到的数据往往与真实的特征-标签对存在偏差，这种偏差即为“数据不确定性”。例如，假设数据中的标签 y 存在噪声，那么数据不确定性

可以用以下公式表示：

$$d(x) := f(x) - y, \tag{4}$$

这里 $f(x)$ 是估计的观测值，而 y 是对应的真实值。我们假定 $d(x)$ 遵循正态分布 $d(x) \sim N(0, \sigma^2)$，因此，估计 σ^2 成为评估数据不确定性的关键。

模型不确定性　模型固有的不确定性反映了系统的认知条件，这受模型架构的选择、优化方法的设计以及超参数配置等多种因素的影响。与一般的随机现象不同，模型不确定性实际上映射了我们对模型预测准确性的信心程度。例如，在回归任务中，可以通过预测值方差的量化来评估这种不确定性，具体公式为：

$$m(x) := \sigma^2 + \frac{1}{T}\sum_{t=1}^{T}\left(f^{\widehat{\omega_t}}(x) - E(y)\right)^2, \tag{5}$$

其中，$E(y) = \frac{1}{T}\sum_{t=1}^{T} f^{\widehat{\omega_t}}(x)$ 是进行 T 次测试的预期输出，$f^{\widehat{\omega_t}}(x)$ 是利用模型参数 $\widehat{\omega}_t$ 的估计输出，而 $m(x)$ 则是用于衡量模型不确定性的指标。

输出不确定性　接下来探讨了 [25] 中提出的输出不确定性概念。考虑一个分类任务的数据集 $D_{\text{train}} = \{(X, Y)\} \subset \mathbb{R}^n \times \{0,1\}^C$，其中 C 表示类别的数量。当一个样本输入模型时，它会产生一个维度为 C 的向量。该向量中的每个元素，尤其是第 i 个值，代表该样本属于第 i 类的概率。输出不确定性的目的在于量化模型输出向量与真实标签向量之间的差距，可以通过以下公式来表达：

$$o(x) := -\sum_{c=1}^{C} p_c \log p_c \tag{6}$$

其中，p_c 是样本属于第 c 类的概率。如图1所示，不同类型的不确定性在机器学习的不同阶段发挥作用 [31]。

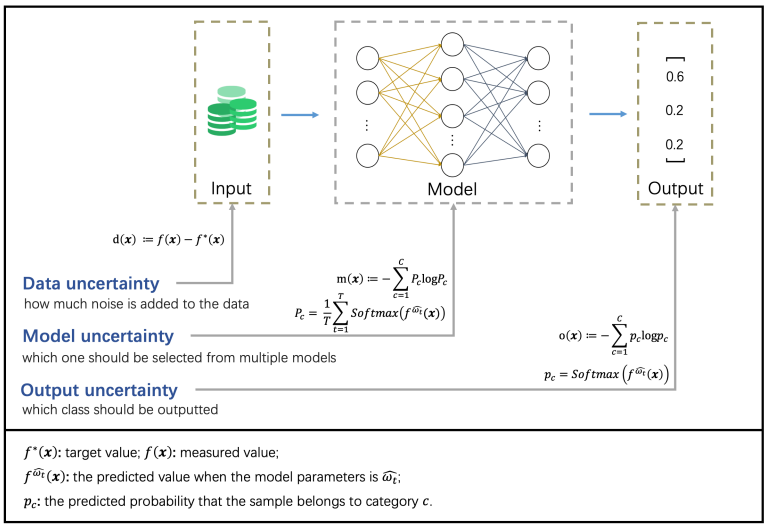

图 1. 不同种类的不确定性 [31].

3　本文方法

本节对本文所提出的不确定性感知深度聚类（Uncertainty-Awared Deepcluster，简称 UAD）方法进行介绍。

3.1 本文方法概述

首先我们对提出方法的整体框架进行介绍，关于 UAD 的整体框架如图2所示。假设当前我们现在有一批无标签图像数据 $\{x_i\}_{i=1,2,N}$，其中 $x_i \in \mathbb{R}^{H\times W\times C}$，$N$ 表示的是图像数据的数据量。H，W 和 C 分别表示了一张图像的高度、宽度以及通道数。此外我们有一个卷积神经网络，卷积神经网络的特征抽取层表示为：

$$f_\theta(x_i) = f^L(f^{L-1}\dots f^3(f^2(f^1(x_i)))), i = 1, 2, ...N \tag{7}$$

其中 L 为特征抽取的层数，一般情况下，f_l 由一层卷积层和一层池化层组成，共同作用于提取图像的不同层次特征。通过层层叠加的特征提取过程，我们能够从原始图像中捕捉到丰富的视觉信息，为后续的聚类分析和伪标签生成奠定基础。通过将所有图像输入到特征提取层，可以得到每个图像的高级特征表示，用公式表达为：

$$\{f_\theta(x_i)\}, i = 1, 2, ...N \tag{8}$$

现有的基于聚类的自监督方法采用基于每一个样本的高级特征使用聚类算法进行聚类，根据[2], 假设这里使用了 K-mean 方法来进行聚类。聚类过程的损失函数如下所示：

$$\min_{C\in\mathbb{R}^{d\times k}} \sum_{n=1}^{N} \left[\min_{z_i\in\{0,1\}^k \text{s.t.} z_n^\top \mathbf{1}=1} \|Cz_i - f_\theta(x_i)\|_2^2 \right], i = 1, 2, ..., N \tag{9}$$

其中，k 是簇心的数量，z_i 是第 i 个样本的伪标签向量。在 z_i 向量中第 q 值为 1，其余为 0，表示 x_i 这个样本属于第 q 个簇。通过应用聚类算法，我们为每幅图像分配了一个伪类别标签。与直接基于原始图像数据生成的伪标签相比，这种基于聚类的伪标签方法能够更加有效地捕捉图像数据内在的类别差异。这种差异的捕捉为我们后续的步骤提供了支持，即利用误差反向传播算法来提取出能够显著体现类别差异的高级图像特征。现在的无标签数据可以用“特征 = 伪类标”的形式进行表示：

$$\{x_i, z_i\}, i = 1, 2, ...N \tag{10}$$

这里我们的研究认为，如果在对样本进行聚类时只使用高级特征进行聚类的话，会存在以下两点局限性：**（1）训练冷启动** 在卷积神经网络的训练初期，网络参数通常通过随机方式初始化。这种初始化方式可能导致网络在早期阶段难以有效捕捉到图像的关键特征，从而影响到聚类的效果。这种现象被称为“训练冷启动”，它可能导致网络在初始几轮训练时的表现不稳定或者效果较差。因此，仅依赖于这些早期阶段提取的高级特征进行聚类可能会限制模型对于数据内在结构的理解。；**（2）伪类标与数据关联性差** 如果聚类仅基于高级特征进行，那么生成的伪类标可能与原始数据的实际内容和结构存在一定的脱节。高级特征虽然能够捕捉到数据的抽象信息，但可能忽略了一些对于区分不同类别至关重要的细微差别。这种情况下，伪类标可能不能准确反映图像的真实类别，导致与数据的关联性降低，进而影响后续基于这些伪类标的训练过程和模型的准确性。

基于以上问题，我们的方法提出了一种将低级特征与高级特征融合的机制，通过这种方式来弥补仅使用高级特征所带来的限制。具体来说，我们在聚类过程中同时考虑了卷积神经网络中较浅层的低级特征和较深层的高级特征。低级特征通常包含了更多关于图像原始细节

和局部纹理的信息，而高级特征则包含了更抽象和全局的图像内容。我们的融合机制首先涉及到特征提取的改进。在提取特征时，我们不仅从网络的深层获取高级特征，也从初级层次提取低级特征。然后，通过设计一个特征融合层，将这两类特征有效地结合起来。该过程可以表示为：

$$f^* = F\left(f^L, f^1\right), \tag{11}$$

其中，F 为提出的特征融合机制。在我们的研究中，我们对低级和高级特征进行了创新性的转换和融合。对于低级特征，我们首先实施了随机通道筛选策略，以降低通道数并减少维度。然后，我们将降维后的特征矩阵进行拼接，转换成一维向量。对于高级特征，鉴于其较少的通道数，我们直接将特征矩阵拼接后转化为一个向量。这样，我们分别获得了代表高级特征和低级特征的两个向量。

在探索如何有效融合这两种特征的过程中，我们尝试了多种方法，包括向量相加、相乘、进行矩阵乘法，甚至引入高斯噪声以提高模型的鲁棒性。然而，经过一系列实验比较后，我们发现简单地将两个向量直接拼接在一起不仅操作简便，而且在效果上也最为显著。因此，我们选择了这种直接拼接的方式来作为我们的特征融合策略。这种方法的优势在于它既保留了每种特征的独特信息，又通过融合增强了特征表征的能力，从而为后续的分析和处理提供了一个更加全面和丰富的特征表示。接下来我们根据融合后的特征来进行聚类：

$$\min_{C\in\mathbb{R}^{d\times k}} \sum_{n=1}^{N} \left[\min_{z_i\in\{0,1\}^k \text{s.t.} z_n^\top \mathbf{1}=1} \left\| Cz_i - f^*\left(x_i\right)\right\|_2^2 \right], i = 1, 2, ..., N \tag{12}$$

经过聚类过程后，我们得到了样本的一个伪标签，接下来我们带有伪标签的数据划分为 Batch 的形式，不断将一个 Batch 的数据的特征 $\{x_i\}, i = 1, 2, ...B$ 输入网络中，首先经过特征抽取 f_θ，随后通过一个分类器 f_W 来对数据对应的伪标签进行预测，用公式表达为：

$$\hat{y_i} = f_w\left(f^*\left(x_i\right)\right), i = 1, 2, ..., B \tag{13}$$

其中 $W \in \mathbb{R}^{h\times k}$ 为分类器的参数矩阵，h 为特征抽取过程完成后样本最高级特征向量的维度。我们采用了机器学习分类任务中广泛应用的交叉熵损失函数来计算模型的损失。交叉熵损失是评估分类模型性能的一种有效方法，它量化了模型预测的概率分布和实际标签的概率分布之间的差异，用公式可以表达为：

$$\text{Cross} - \text{EntropyLoss} = -\sum_{i=1}^{B}\sum_{c=1}^{k} y_{ic}\log(\hat{y_{ic}}) \tag{14}$$

在现有的自监督学习方法中，一个常常被忽视的问题是伪标签的不确定性。这些伪标签通常是主观设定的，而没有考虑到它们固有的不确定性。我们认为，在自监督学习中，对伪标签的不确定性进行量化，并在训练过程中为不确定性较低的样本赋予更大的权重，可以显著提高模型学习到的特征提取模式的可靠性。因此，在自监督学习的框架下，同时定义伪标签及其不确定性显得尤为重要。基于这一认识，我们提出了一种新颖的方法，即基于聚类的伪标签不确定性度量准则。具体来说，在聚类过程结束后，样本在其融合特征空间中被分为 k 个簇。借鉴模糊聚类的概念，我们认为，簇内距离中心点更近的样本具有更高的置信度，即它们作为伪标签的不确定性较低。反之，距离簇心较远的样本的伪标签不确定性就较高。为了量化这一概念，我们设计了一个基于样本与其所属簇心距离的不确定性度量函数 $U(x)$。关于

$U(x)$ 的设计细节我们将在后面进行介绍。接下来，我们可以将不确定性函数引入交叉熵损失函数中，如下所示：

$$\text{Cross} - \text{EntropyLoss}^{*} = -\sum_{i=1}^{B} U(x_i) \sum_{c=1}^{k} y_{ic} \log(\hat{y_{ic}}) \tag{15}$$

通过引入不确定性，我们的方法为伪标签的使用提供了一个更加科学和可靠的量化标准，从而使整个学习过程更加精准和有效。接下来我们对提出的特征融合模块 F 和不确定性度量模块 U 的设计细节以及动机进行解释。

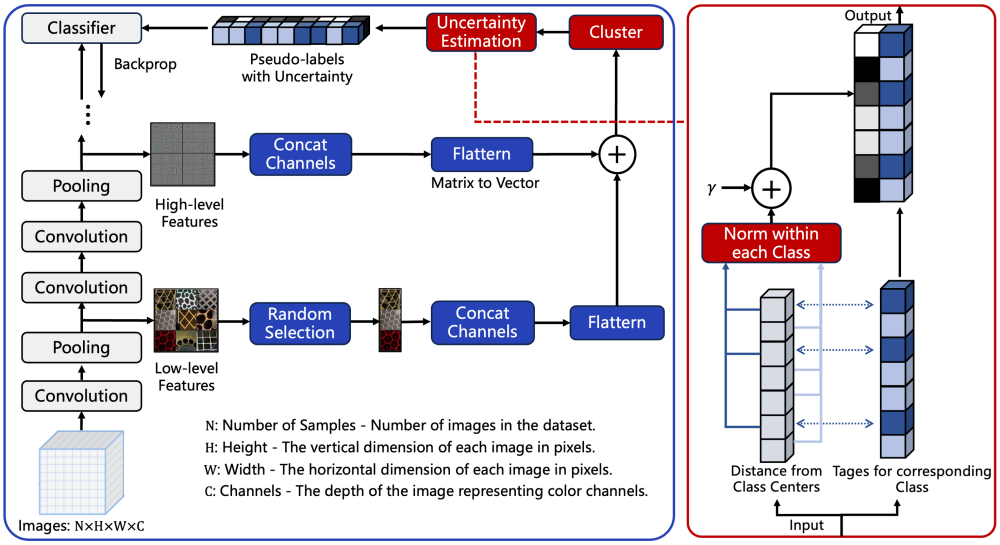

图 2. 本研究提出的方法框架如下：首先，我们分别从卷积神经网络（ConvNet）的第一个和最后一个池化层提取低层特征和高层特征。接着，这些特征与用于聚类的随机矩阵进行融合。聚类过程中生成的伪标签带有不确定性信息，这些伪标签被用来指导分类器和 ConvNet 的反向传播过程。通过这种方式，我们有效地将低层和高层特征结合，以增强模型对数据的理解和处理能力。

3.2 特征融合

在目前流行的基于聚类的自监督学习方法中，主流的趋势倾向于设计复杂的对抗样本策略和对比学习模式，以改善传统方法。然而，正如我们的研究所展示的，简洁而直接的方法同样可以产生显著的效果。我们提出的特征融合机制正是一个典型例证。这种机制简单地将低级特征与高级特征进行拼接，用于后续的聚类过程。令人惊喜的是，这样一个直观的方法能够有效地提高了自监督学习的性能。

在我们的研究中，我们首先假设低级特征为 $f^1 \in \mathbb{R}^{H^1 \times W^1 \times C^1}$。在观察现有的众多网络结构，例如 AlexNet，我们注意到在卷积神经网络的初级卷积池化层中获取的每个特征内二维矩阵的维度往往非常大。当这些低级特征与高级特征融合时，会引发两个主要问题：

1. 高级特征的效果减弱：如果低级特征的维度过高且未经过处理，那么在与高级特征融合后，所得到的特征向量将主要由低级特征决定。这将导致高级特征的作用被大幅度降低，甚至可能被忽略。

2. 融合特征的维度灾难：过高的低级特征维度会导致最终的高级特征维度过大。特别是在处理如 ImageNet 这样的大规模数据集时，这种高维度特征会对计算机硬件，尤其是显卡内存提出极高的要求。

为了解决这些问题，我们提出了对低级特征通道数进行了控制。具体来说，我们采用了一种随机筛选的方法，从 C^1 个通道中随机抽取一部分通道。例如，在我们的实验中，低级特征的通道数 C^1 为 96，而经过随机筛选后的通道数 C^{1*} 降至 16。这个过程可以用以下公式表示：

$$f^{1*} = R\left(f^1\right), \tag{16}$$

其中 $f^{1*} \in \mathbb{R}^{H^1 \times W^1 \times C^{1*}}$。通过这种方式，我们既能够保持低级特征的关键信息，又能有效地防止维度灾难和确保高级特征在融合过程中的作用不被削弱。接下来我们将所有的通道矩阵拼接在一起，组成一个低级特征二维矩阵，如下所示：

$$\left[(f_1^{1*})^T, (f_2^{1*})^T, ..., (f_{C^{1*}}^{1*})^T\right] \tag{17}$$

最后我们将矩阵转化为一个向量的形式, 最后最终低级特征的表示，如式18所示：

$$f_{low} = flattern(\left[(f_1^{1*})^T, (f_2^{1*})^T, ..., (f_{C^{1*}}^{1*})^T\right]) \tag{18}$$

接下来对于高级特征的处理与低级特征大致相同，除了去掉了通道特征筛选这一步（因为高级特征的维度一般情况下比较低，例如在 Alexnet 中最终的高级特征维度为 $13 \times 13 \times 256$。处理后高级特征的表示如式19所示。

$$f_{high} = flattern(\left[(f_1^{L})^T, (f_2^{L})^T, ..., (f_{C^{L}}^{L})^T\right]) \tag{19}$$

最终我们将低级特征向量和低级特征向量拼接在一起就得到了最终的融合特征。如式20所示。

$$F\left(x\right) = \left[f_{low}, f_{high}\right] \tag{20}$$

提出的特征融合方法非常简单容易理解, 通过分析我们的方法有以下几个特性：

增强的特征表示能力：我们的方法通过融合低级和高级特征，有效地结合了低级特征中丰富的细节信息和高级特征中的抽象信息。低级特征捕捉了图像的局部纹理、边缘和形状等细节，而高级特征则提供了全面的语义洞察。这种综合性的融合显著增强了模型对图像内容及其细节的理解，从而提升了整体特征的表达能力。通过巧妙地平衡细节与抽象信息，我们确保了高级特征不被低级特征所掩盖，同时赋予低级特征以充分的重要性。

灵活性和适应性：我们的方法设计考虑到了不同应用场景和数据集的需求，因此具有高度的灵活性。例如，根据特定任务的要求，我们可以调整低级特征的筛选比例，或者针对不同的网络架构调整融合策略。这种方法不仅适用于多种卷积神经网络，而且提升了其通用性。此外，该方法的设计为未来的扩展和改进提供了灵活的基础，从而为进一步的研究开辟了新的可能性。

3.3 不确定性评估

伪标签由人为设计而成，因此受到主观因素的影响，其不确定性具有一定的变动性。因此，在进行自监督学习时，不仅需要明确定义伪标签的形式，还应重视对伪标签不确定性的量

化评估。这一评估的重要性在于，通过将伪标签的不确定性纳入特征表示学习过程，我们能够训练出更加可靠和有效的特征提取模式，进而为下游任务提供更强的支持。简而言之，对伪标签不确定性的细致分析和利用，将为自监督学习提供更深层次的理解。

因此，在特征融合机制的基础上，我们借鉴了模糊聚类的概念，提出了一个基于聚类的伪标签不确定性模块。众所周知，聚类过程本质上是基于样本特征将其划分到不同的群簇中。借鉴现有的基于聚类的自监督方法，我们同样采用了 k-means 算法进行聚类。以 k-means 为例，我们在聚类过程中随机选择 k 个簇心，然后根据每个样本相对于这些簇心的聚类情况，将其归入相应的簇中。随后，我们迭代更新簇心，并重新计算每个簇内的元素。最终，每个样本都被唯一地划归到一个簇中，并将该簇所代表的类别定义为该样本的类别。如式14所示，在传统的基于聚类的自监督方法中，我们通常直接利用伪类标和分类器，通过交叉熵损失进行误差反向传播，以此来训练网络参数。

在提出的 UAD 中，我们从聚类机制的角度出发，定义了一个伪类标不确定性度量函数 $U(x)$。这一函数在后续的交叉熵损失计算中为更可信的类标赋予更高的权重。具体而言，我们认识到在聚类过程后，每个样本都被分配到某个簇中。假设簇的总数为 k，样本 x_i 被分配到的簇标记为 z_i，且它与簇心的距离为 d_i。令 S_j 表示属于第 j 个簇的样本集合，可以定义为：

$$S_j = \{x_i \mid z_i = j\}, j = 1, 2, ..., k \tag{21}$$

随后，我们在每个簇内部实施了 min-max 归一化处理，对样本与其簇心的距离 d 进行标准化。设定 d^j_{min} 和 d^j_{max} 分别为第 j 个簇内距离的最小值和最大值。对于属于第 j 个簇的每个样本 x_i，其与簇心的距离 d_i 经过归一化处理可以表示为：

$$\widetilde{d}_i = \frac{d_i - d^j_{min}}{d^j_{max} - d^j_{min}}, \tag{22}$$

其中 $\widetilde{d}_i$ 是归一化后的距离值。这个归一化过程确保了每个簇内样本距离的相对大小被保留，同时消除了不同簇之间距离尺度的差异，从而为后续的处理提供了一个统一的距离度量标准。伪类标不确定性 $U(x_i)$ 通过样本 x_i 与其簇心的归一化距离 $\widetilde{d}_i$ 进行量化，定义如下：

$$U(x_i) = 1 - \widetilde{d}_i + \gamma, \tag{23}$$

其中，γ 是一个超参数。归一化后的距离值被映射至 0 到 1 的范围内，通过从 1 减去这个距离值，我们得到了一个基础的不确定性度量。但是，若直接将此度量应用于后续的分类任务，会导致所有样本在梯度下降过程中的损失都乘以一个介于 0 到 1 之间的数值，类似于降低学习率的效果，这可能最终导致过拟合现象。因此，我们引入了超参数 γ 来调节这一效应。最终，这些不确定性度量可以被应用于交叉熵损失计算中，以赋予那些具有较低确定性的样本更大的权重，正如式15所示。这种方法旨在优化模型的整体性能，确保低确定性样本在学习过程中发挥更大的影响。

Algorithm 1 不确定性感知深度聚类（UAD）方法

Input 无标签图像数据集 $\{x_i\}_{i=1}^{N}$
Output 训练好的模型参数 θ 和 W
初始化卷积神经网络参数 θ
for 每次迭代 **do**
  for 每个图像 x_i **do**
    使用网络 f_θ 提取特征
  end for
  初始化融合特征集合 $F^* = \{\}$
  for 每个图像 x_i **do**
    提取低级特征 $f_{low}(x_i)$
    提取高级特征 $f_{high}(x_i)$
    融合特征 $f^*(x_i) = \text{concatenate}(f_{low}(x_i), f_{high}(x_i))$
    将 $f^*(x_i)$ 添加到 F^*
  end for
  使用 K-means 聚类算法对 F^* 进行聚类
  for 每个簇 j **do**
    计算簇 j 中每个样本到簇心的距离 d_i
    对簇内样本距离执行 min-max 归一化 $\widetilde{d}_i$
  end for
  for 每个样本 x_i **do**
    计算伪类标不确定性 $U(x_i) = 1 - \widetilde{d}_i + \gamma$
  end for
  使用梯度下降法更新网络参数 θ 和 W，考虑不确定性 $U(x_i)$
end for
return θ, W

4 复现细节

4.1 与已有开源代码对比

复现过程中引用参考了 [2] 发布的基于聚类的自监督方法代码，该方法是基础版本的深度聚类模型，（没有我们改进的特征融合以及不确定性度量模块），关于引用代码详细描述使用情况如图3所示。在 Baseline 中，关于 Backbone 的搭建，以及通过聚类进行高级特征的提取，聚类，并最终使用高级特征的误差反传都进行了定义。为了实现特征融合，我们首先对主函数中的特征抽取 $compute_features(dataloader, model, N)$ 函数进行了重新定义，具体，如图4所示。可以看到，与原代码相比，整个特征抽取模块都进行了改写，改写之后的函数能够对网络王的低级特征和高级特征都进行捕获，除此之外，我们还将低级特征融合与低级特征进行和融合拼接。原始代码使用的是高级特征进行聚类，但是在我们提出的新方法中，我们使用了融合特征进行聚类。 此外，我们借助了模糊聚类的思想定义了伪类标的不确定性，并将不

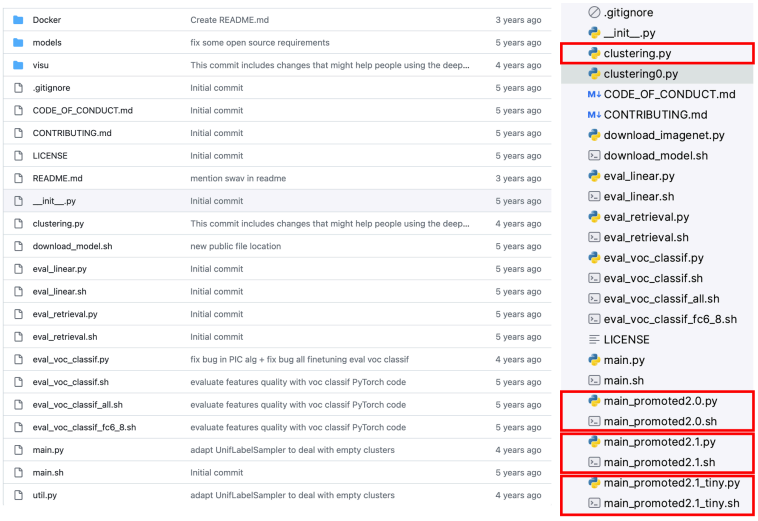

图 3. 代码复现以及改进 (主要的改进代码用红框进行了标注).

```
low_features = sob.data.cpu().numpy()
selection = random.sample(range(0, int(low_features.shape[1])), int(8))
low_features = low_features[:, selection, :, :]
amount_samples = low_features.shape[0]
low_features = low_features.reshape(amount_samples, -1)

if i == 0:
    h_features = np.zeros((N, aux.shape[1]), dtype='float32')
    l_features = np.zeros((N, low_features.shape[1]), dtype='float32')

aux = aux.astype('float32')
low_features = low_features.astype('float32')
if i < len(dataloader) - 1:
    h_features[i * args.batch: (i + 1) * args.batch] = aux
    l_features[i * args.batch: (i + 1) * args.batch] = low_features
else:
    # special treatment for final batch
    h_features[i * args.batch:] = aux
    l_features[i * args.batch:] = low_features
```

图 4. 特征融合模块（部分展示），我们对高级特征低级特征进行了抽取和变换.

确定性引入到模型的分类训练中。为了实现这一功能，我们对整个 *clustering.py* 文件都进行了重写，重写后的文件命名为 *clustering.py*，而原来的文件被保留为 *clustering*0*.py*。这样做的目的是为了尽可能简化代码量，避免一些不必要的代码工作量。我们对原始 *clustering.py* 的函数都进行了详细分析和学习，发现了一个非常重要的细节。特别是在 *faiss* 提供的聚类库函数中，聚类的结果不仅包括了样本对应的簇标签，还包括了样本与簇心的距离。这一发现对我们的研究至关重要，因为它意味着我们无需重新计算样本与簇心的距离，从而大大简化了计算过程。

我们进一步利用了这个距离信息，将其直接用于伪类标不确定性的计算中。具体而言，我们采用了基于距离的不确定性评估方法。这种方法的优点在于它直接利用了 *faiss* 聚类结果中已有的距离数据，从而避免了额外的计算负担。如图 5 所示，我们展示了 *clustering.py* 中如何处理这些距离数据，并将其融入我们的不确定性评估框架。

为了更好地融合传统聚类方法和不确定性评估，我们在 *clustering.py* 中实现了一个新的功能，它能够在聚类过程中同时计算每个样本的簇标签和与簇心的距离，并据此计算不确定

性。这种方法的关键在于能够为每个样本赋予一个不确定性权重，这个权重随后用于模型的分类训练过程。我们认为，通过这种方式，模型能够更有效地从数据中学习，特别是在处理具有高度不确定性的样本时。

我们的方法的核心思想是，通过对不确定性的量化评估和利用，我们能够更全面地利用无标签数据集的信息，从而提升自监督学习方法的性能。这种方法的应用不仅限于图像聚类任务，还可以扩展到其他类型的无监督或自监督学习任务中，为深度学习模型的训练提供了一个新的视角。

```python
def __getitem__(self, index):
    """
    Args:
        index (int): index of data
    Returns:
        tuple: (image, pseudolabel) where pseudolabel is the cluster of index datapoint
    """
    path, pseudolabel = self.imgs[index]
    uncertainty = self.distance[index]
    img = pil_loader(path)
    if self.transform is not None:
        img = self.transform(img)
    return img, pseudolabel,uncertainty
```

图 5. 保留样本距离簇心的距离.

```python
for i in range(len(data)):
    self.images_lists[I[i]].append(i)
    self.distance_lists[I[i]].append(self.D[i])

self.distance_minmax_lists = [[] for i in range(self.k)]
for i in range(len(self.distance_lists)):
    self.distance_minmax_lists[i].append(min(self.distance_lists[i]))
    self.distance_minmax_lists[i].append(max(self.distance_lists[i]))
for i in range(len(self.distance_lists)):
    # print(len(self.distance_lists[i]))
    if(len(self.distance_lists[i])>1 and self.distance_minmax_lists[i][1] > self.distance_minmax_list
        for j in range(len(self.distance_lists[i])):
            # print(self.distance_lists[i][j], self.distance_minmax_lists[i][0], self.distance_minma
            self.distance_lists[i][j] = (((self.distance_lists[i][j] -  self.distance_minmax_lists[i]
```

图 6. 将距离转化为不确定性（部分展示，完整代码请参考原代码）.

```python
softmax_func = nn.Softmax(dim=1)
soft_output = softmax_func(output)
log_output = torch.log(soft_output) #这里可以再加速
for ii in range(len(log_output)):
    log_output[ii][target_var[ii]] = log_output[ii][target_var[ii]] * uncertainty[ii]
loss = crit(log_output, target_var)

# record loss
losses.update(loss.item(), input_tensor.size(0))
```

图 7. 损失函数的重新定义（部分展示，完整代码请参考原代码）.

4.2 实验环境

实验环境搭建以及模型参数设置参考了 [2]。

- Python 安装版本 2.7
- SciPy 和 scikit-learn 包

- PyTorch 安装版本 0.1.8 (pytorch.org)
- CUDA 8.0
- Faiss 安装
- ImageNet 数据集

4.3 使用说明

首先我们要进行自监督学习的前置任务训练，即根据大量不带标签的数据进行低级特征以及高级特征的抽取、聚类、不确定性度量和误差反传，这里使用代码的话就直接对 $main_promoted2.1.sh$ 进行修改，把数据集组织成需要的形式，然后运行脚本就可以。

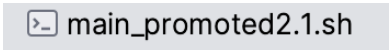

图 8. 自监督学习的前置任务训练.

```
DIR="/media/data/wq/imagenet/ILSVRC2012_img_train"
ARCH="alexnet"
LR=0.05
WD=-5
K=10000
WORKERS=12
EXP="/media/data/wq/deepcluster_exp/exp_pro2.1"
PYTHON="/home/wqq/anaconda3/envs/pytorch/bin/python"

mkdir -p ${EXP}

CUDA_VISIBLE_DEVICES=1,2,3 ${PYTHON} main_promoted2.1.py ${DIR} --exp ${EXP} --arch ${ARCH} \
  --lr ${LR} --wd ${WD} --k ${K} --sobel --verbose --workers ${WORKERS}
```

图 9. 模型自监督前置任务训练脚本.

训练好的模型被保存到参数指定的文件夹中，如图10所示。

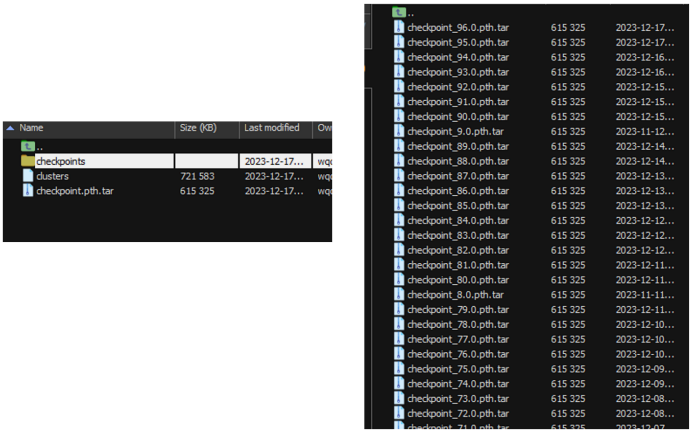

图 10. 自监督前置任务训练得到的模型文件

然后我们可以将训练好的模型用到下游任务里面，因为时间以及数据集规模的原因，这里还使用了 ImageNet 图像分类作为下游任务来进行下一步训练（ImageNet 图像分类作为下游任务是已有工作验证自监督效果常用的方式之一）。具体来说，我们将前置任务训练的模型

的特征提取部分作为的下游任务的预训练模型。这里使用的话直接修改 $eval_linear.sh$ 的参数就可以。

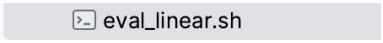

图 11. 下游任务训练.

```
DATA="/media/data/wq/tiny-imagenet-200/"
MODELROOT="/home/wq/projects/deepcluster"
MODEL="/media/data/wq/deepcluster_exp/exp_pro2.1_tiny/checkpoints/checkpoint_15.0.pth.tar"
EXP="/media/data/wq/deepcluster_exp/linear_tiny"

PYTHON="/home/wqq/anaconda3/envs/pytorch/bin/python"

mkdir -p ${EXP}

CUDA_VISIBLE_DEVICES=2,3 ${PYTHON} eval_linear.py --model ${MODEL} --data ${DATA} --conv 4 --lr 0.01 \
  --wd -7 --tencrops --verbose --exp ${EXP} --workers 12
```

图 12. 模型自监督下游任务训练脚本.

4.4 创新点

本文方法的创新点如下：

- 结合低级和高级特征的融合机制：UAD 方法通过融合来自卷积神经网络不同层次的低级和高级特征，有效地结合了低级特征中的细节信息和高级特征中的抽象信息。这种融合策略不仅增强了模型对图像内容和细节的理解，还提高了整体特征表达的能力
- 利用模糊聚类的思想定义伪类标不确定性：UAD 方法借鉴了模糊聚类的概念，创新性地定义了基于样本与其簇心距离的伪类标不确定性。这种方法允许模型在聚类过程中考虑每个样本的不确定性，从而更有效地处理数据。
- 将不确定性引入模型的分类训练：通过将伪类标不确定性引入到分类模型的训练过程中，UAD 方法能够为不同样本赋予不同的权重。这样做可以提高模型在处理高度不确定性样本时的性能和准确性。抽象信息。这种融合策略不仅增强了模型对图像内容和细节的理解，还提高了整体特征表达的能力.

5 实验结果分析

本部分对实验所得结果进行分析，详细对实验内容进行说明，实验结果进行描述并分析。

5.1 数据集

实验使用的自监督前置训练数据集为 ImageNet。ImageNet 是一个广泛用于计算机视觉研究的大型图像数据库。这个数据集由斯坦福大学的研究人员开发，目的是用于各种视觉对象识别软件研究。最为人所熟知的是它的一个子集，即 ImageNet Large Scale Visual Recognition Challenge (ILSVRC)，通常简称为 ImageNet Challenge。ImageNet 数据集包含超过 1400 万张标记清晰的高分辨率图像，这些图像涵盖了超过 20,000 个类别。每个类别中的图像都是根据 WordNet（一种语义关系词典）中的同义词集（synsets）来分类的，每个类别大约包含数

百到数千张图像。ImageNet 不仅在规模上非常庞大，而且在多样性、多标签和图像的深度上也很出色。由于其规模和多样性，ImageNet 成为了计算机视觉领域的一个基准数据集，尤其是在图像分类和对象识别领域。自监督前置训练通常使用这个数据集来训练模型学习丰富的特征表示，这些特征表示随后可以应用于各种下游任务，如图像识别、分类或者更复杂的视觉任务。实验使用的下游任务也使用了 ImageNet 本身进行带监督的分类训练。

5.2 模型参数

这里重点对前置任务训练时采用的参数进行介绍。

- 网络结构：AlexNet
- 输入图像上应用 Sobel 滤波器
- 聚类算法：K-means
- 聚类数量：10000
- 学习率：0.05
- 训练周期：500
- Batch 大小：256
- 随机种子：3407

5.3 实验结果

为了验证我们特征融合模块和不确定性度量模块的有效性，我们首先进行了消融实验，首先前置任务是在一个官方提供的 Tiny ImageNet 上进行训练，然后我们将训练好的预模型基于 Tiny ImageNet 进行下游分类任务训练，实验结果如表1所示。在表格中 Convn 表示前 n 层卷积层被迁移到了下游任务里。从表格的结果可以看到，在 Baseline 中这一模型作为对照组，没有使用我们提出的特征融合和不确定性度量模块。从 Conv1 到 Conv5，分类准确率分别为 17.0%，25.1%，28.1%，28.9% 和 23.9%。UAD (U) 表示只应用了不确定性度量模块。结果显示，UAD (U) 在所有卷积层上，性能都略有提升，特别是在 Conv2 至 Conv5 层，准确率相比基线模型有显著提高。这说明了我们提出了不确定性模块能够这种方法使得模型能够有效地从高置信度的伪标签中学习，同时最小化低置信度标签对学习不利视觉特征的影响。

接下来比较提出的 UAD 模型与已有方法的性能，实验中的方法分为两大类：非依赖于聚类的方法（Non-Cluster-Dependent）和依赖于聚类的方法。具体来说实验结果包括了多种自监督学习策略的对比，如上下文填补、颜色化、双向生成对抗网络（BiGANs）、计数等，以及聚类依赖方法，如 DeepCluster。我们在 ImageNet 的不同卷积层级别（C1 至 C5）对各方法的性能进行了评估。最佳性能的层级用红色标记，而每个配置下获得的最高准确率则以粗体显示。

在 ImageNet 数据集上，基于聚类的方法中，我们的 UAD 方法在高级层级（如 C4 和 C5）显示了显著的性能提升。这表明这些方法能够更有效地捕获高级特征表示，从而提高分

表 1. 为了评估我们提出的特征融合策略和不确定性估计机制的有效性，我们在 Tiny ImageNet 数据集上进行了线性分类实验。为此，我们使用了 AlexNet 卷积层得到的激活作为特征输入。评估指标是 10 Crop 上计算得出的分类准确率。

方法	Conv1	Conv2	Conv3	Conv4	Conv5
Baseline	17.0	25.1	28.1	28.9	23.9
UAD (U)	**17.3**	26.0	29.2	29.3	24.1
UAD (U+F)	**17.3**	**27.0**	**31.2**	**32.7**	**27.5**

类准确率,，尤其是在中后期卷积层级别（C3, C4, C5）。这一结果说明与一些基于原始图像本身构建的伪标签相比，通过聚类的方法生成伪类标的形式更符合数据集本身的特性。类别信息是这些图像数据集中天然的内在标签，如果通过自监督学习得到的图像特征也能够反应数据的类别差异的话，能够帮助我们更好地解决下游任务，因此基于聚类的自监督方法表现出了更加优异的性能。

此外，通过将提出的方法与其他方法的准确率进行比较，结果如表2所示。可以看到，提出的 UAD 在 ImageNet 上的特征学习效果要优于许多已有的传统自监督方法比如 Inpainting，Colorization 和 BiGANs。另外 UAD 的整体正确率要远远高于 DeepCluster 这一最传统的基于聚类自监督方法。这不仅证明了 UAD 方法的潜力，也强调了其在不同类型数据集上的适应性和可靠性。

总之，UAD 作为一种新颖的自监督学习策略，其在各种数据集上的表现证明了其在特征学习领域的有效性和潜力，尤其是在处理复杂和多样化图像数据方面。这些成果不仅展示了 UAD 方法的实用性，也为未来在自监督学习领域的研究提供了新的方向和思路。

6 总结与展望

本研究工作提出了一种新颖基于聚类的自监督学习方法——不确定性感知深度聚类(UAD)，旨在提高自监督学习在特征提取和图像分类方面的性能。与传统的基于聚类自监督方法相比，我们的方法通过结合低级和高级特征，并引入伪标签不确定性的量化。

在目前流行的基于聚类的自监督学习方法中，主流的趋势倾向于设计复杂的对抗样本策略和对比学习模式，以改善传统方法。然而，正如我们的研究所展示的，简洁而直接的方法同样可以产生显著的效果。我们提出的特征融合机制正是一个典型例证。这种机制简单地将低级特征与高级特征进行拼接，用于后续的聚类过程。令人惊喜的是，这样一个直观的方法能够有效地提高了自监督学习的性能。此外，伪标签由人为设计而成，因此受到主观因素的影响，其不确定性具有一定的变动性。因此，在进行自监督学习时，不仅需要明确定义伪标签的形式，还应重视对伪标签不确定性的量化评估。这一评估的重要性在于，通过将伪标签的不确定性纳入特征表示学习过程，我们能够训练出更加可靠和有效的特征提取模式，进而为下游任务提供更强的支持。简而言之，对伪标签不确定性的细致分析和利用，将为自监督学习提供更深层次的理解。

在 ImageNet 数据集上的实验结果证明，UAD 在提高分类准确率方面具有显著的优势，

表 2. 为了全面评估我们提出方法的性能，我们在 ImageNet 数据集上开展了一系列线性分类实验。在这些实验中，我们采用了 AlexNet 卷积层输出的激活值作为特征，并在 10 Crop 上计算了分类准确率。此外，我们用 * 标记来表示采用的是改良后尺寸更大的 AlexNet 变体。实验中表现最佳的层级用醒目的红色进行了突出显示。Cn 的表示方法则是指我们使用了 AlexNet 中的前 n 层卷积层来完成下游任务。为了更清晰地展示实验结果，我们在每个 Cn 设置下获得的最高准确率都用**粗体**字体进行了标注

		ImageNet				
	Method	C1	C2	C3	C4	C5
	Places labels, [30]	-	-	-	-	-
	ImageNet labels, [30]	19.3	36.3	44.2	48.3	50.5
	Random, [30]	11.6	17.1	16.9	16.3	14.1
Non-Cluster-Dependent	Inpainting, [22]	14.1	20.7	21.0	19.8	15.5
	Colorization, [29]	12.5	24.5	30.4	31.5	
	BiGANs, [6]	17.7	24.5	31.0	29.9	28.0
	CFN, [20]	18.2	28.8	34.0	33.9	27.1
	Counting, [21]	18.0	30.6	34.3	32.5	25.7
	Instance retrieval, [27]	16.8	26.5	31.8	34.1	35.6
	RotNet, [7]	**18.8**	31.7	**38.7**	38.2	35.5
	AND*, [11]	15.6	27.0	35.9	39.7	37.9
	CMC*, [24]	18.4	**33.5**	38.1	**40.4**	**42.6**
	Deepcluster, [2]	16.8	26.5	31.8	34.1	35.6
	UAD	**17.3**	**28.6**	**34.6**	**37.2**	**40.2**
	Promotion	+0.5	+2.1	+2.8	+3.1	+4.6

优于许多传统自监督学习方法，如 Inpainting、Colorization 和 BiGANs。这些成果不仅彰显了 UAD 方法的高效性和适用性，也为自监督学习领域提供了新的研究方向。

展望未来，可以进一步探索和优化 UAD 方法，特别是在不确定性度量和特征融合策略上。可以考虑引入更复杂的网络结构或更先进的优化算法，以提高模型的泛化能力和处理复杂数据的能力。此外，还需要在更广泛和多样化的数据集上测试 UAD 方法。这包括不仅限于图像分类的各种任务，如目标检测、图像分割等，以验证和扩展 UAD 方法的适用性。

参考文献

[1] Yuki Markus Asano, Christian Rupprecht, and Andrea Vedaldi. Self-labelling via simultaneous clustering and representation learning. *arXiv preprint arXiv:1911.05371*, 2019.

[2] Mathilde Caron, Piotr Bojanowski, Armand Joulin, and Matthijs Douze. Deep clustering for unsupervised learning of visual features. In *Proceedings of the European conference on computer vision (ECCV)*, pages 132–149, 2018.

[3] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, 40(4):834–848, 2017.

[4] Ting Chen, Xiaohua Zhai, Marvin Ritter, Mario Lucic, and Neil Houlsby. Self-supervised gans via auxiliary rotation loss. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12154–12163, 2019.

[5] Virginia R de Sa. Learning classification with unlabeled data. *Advances in neural information processing systems*, pages 112–112, 1994.

[6] Jeff Donahue, Philipp Krähenbühl, and Trevor Darrell. Adversarial feature learning. *arXiv preprint arXiv:1605.09782*, 2016.

[7] Spyros Gidaris, Praveer Singh, and Nikos Komodakis. Unsupervised representation learning by predicting image rotations. *arXiv preprint arXiv:1803.07728*, 2018.

[8] Bobby He and Mete Ozay. Exploring the gap between collapsed & whitened features in self-supervised learning. In *International Conference on Machine Learning*, pages 8613–8634. PMLR, 2022.

[9] Kaiming He, Ross Girshick, and Piotr Dollár. Rethinking imagenet pre-training. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4918–4927, 2019.

[10] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision*, pages 1026–1034, 2015.

[11] Jiabo Huang, Qi Dong, Shaogang Gong, and Xiatian Zhu. Unsupervised deep learning by neighbourhood discovery. In *International Conference on Machine Learning*, pages 2849–2858. PMLR, 2019.

[12] Eyke Hüllermeier and Willem Waegeman. Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods. *Machine Learning*, 110:457–506, 2021.

[13] Yang Jiao, Ning Xie, Yan Gao, Chien-Chih Wang, and Yi Sun. Fine-grained fashion representation learning by online deep clustering. In *European Conference on Computer Vision*, pages 19–35. Springer, 2022.

[14] Longlong Jing and Yingli Tian. Self-supervised visual feature learning with deep neural networks: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 43(11):4037–4058, 2020.

[15] Frank Hyneman Knight. *Risk, uncertainty and profit*, volume 31. Houghton Mifflin, 1921.

[16] Gustav Larsson, Michael Maire, and Gregory Shakhnarovich. Colorization as a proxy task for visual understanding. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 6874–6883, 2017.

[17] Ekdeep Singh Lubana, Chi Ian Tang, Fahim Kawsar, Robert P Dick, and Akhil Mathur. Orchestra: Unsupervised federated learning via globally consistent clustering. *arXiv preprint arXiv:2205.11506*, 2022.

[18] Ishan Misra and Laurens van der Maaten. Self-supervised learning of pretext-invariant representations. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 6707–6717, 2020.

[19] Sadaaki Miyamoto, Hodetomo Ichihashi, Katsuhiro Honda, and Hidetomo Ichihashi. *Algorithms for fuzzy clustering*, volume 10. Springer, 2008.

[20] Mehdi Noroozi and Paolo Favaro. Unsupervised learning of visual representations by solving jigsaw puzzles. In *European conference on computer vision*, pages 69–84. Springer, 2016.

[21] Mehdi Noroozi, Hamed Pirsiavash, and Paolo Favaro. Representation learning by learning to count. In *Proceedings of the IEEE international conference on computer vision*, pages 5898–5906, 2017.

[22] Deepak Pathak, Philipp Krahenbuhl, Jeff Donahue, Trevor Darrell, and Alexei A Efros. Context encoders: Feature learning by inpainting. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2536–2544, 2016.

[23] Ali Sharif Razavian, Hossein Azizpour, Josephine Sullivan, and Stefan Carlsson. Cnn features off-the-shelf: an astounding baseline for recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 806–813, 2014.

[24] Yonglong Tian, Dilip Krishnan, and Phillip Isola. Contrastive multiview coding. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XI 16*, pages 776–794. Springer, 2020.

[25] Xi-Zhao Wang, Ran Wang, and Chen Xu. Discovering the relationship between generalization and uncertainty by incorporating complexity of classification. *IEEE transactions on cybernetics*, 48(2):703–715, 2017.

[26] Xizhao Wang and Yulin He. Learning from uncertainty for big data: future analytical challenges and strategies. *IEEE Systems, Man, and Cybernetics Magazine*, 2(2):26–31, 2016.

[27] Zhirong Wu, Yuanjun Xiong, Stella X Yu, and Dahua Lin. Unsupervised feature learning via non-parametric instance discrimination. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3733–3742, 2018.

[28] Asano YM., Rupprecht C., and Vedaldi A. Self-labelling via simultaneous clustering and representation learning. In *International Conference on Learning Representations*, 2020.

[29] Richard Zhang, Phillip Isola, and Alexei A Efros. Colorful image colorization. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part III 14*, pages 649–666. Springer, 2016.

[30] Richard Zhang, Phillip Isola, and Alexei A Efros. Split-brain autoencoders: Unsupervised learning by cross-channel prediction. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1058–1067, 2017.

[31] Xinlei Zhou, Sudong Chen, Nianjiao Peng, Xinpeng Zhou, and Xizhao Wang. Uncertainty guided pruning of classification model tree. *Knowledge-Based Systems*, 259:110067, 2023.

[32] Lei Zhu, Zhanghan Ke, and Rynson Lau. Towards self-adaptive pseudo-label filtering for semi-supervised learning. *arXiv preprint arXiv:2309.09774*, 2023.