

中图分类号: TP391

单位代号: 10280

密 级: 公开

学 号: 23721590

上海大学



专业硕士学位论文

SHANGHAI UNIVERSITY
PROFESSIONAL MASTER'S
DISSERTATION

题 目	基于形状空间理论的特征增强 方法研究
--------	-----------------------

作 者 李睿杰

学科专业 软件工程

导 师 韩越兴

完成日期 二〇二六年六月

姓 名：李睿杰

学号：23721590

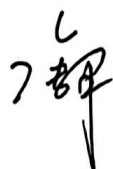
论文题目：基于形状空间理论的特征增强方法研究

上海大学

本论文经答辩委员会全体委员审查，确认符合上海大学专业硕士学位论文质量要求。

答辩委员会签名：

主 席：



委 员：



导 师：



答辩日期：2026 年 6 月 9 日

姓 名：李睿杰

学号：23721590

论文题目：基于形状空间理论的特征增强方法研究

上海大学学位论文原创性声明

本人郑重声明：所呈交的学位论文是本人在导师指导下，独立进行研究工作所取得的成果。除了文中特别加以标注和致谢的内容外，论文中不包含其他人已发表或撰写过的研究成果。参与同一工作的其他研究者对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

学位论文作者签名：李睿杰

日期：2026 年 6 月 9 日

上海大学学位论文使用授权说明

本人完全了解上海大学有关保留、使用学位论文的规定，即：学校有权保留论文及送交论文复印件，允许论文被查阅和借阅；学校可以公布论文的全部或部分内容。

(保密的论文在解密后应遵守此规定)

学位论文作者签名：李睿杰

导师签名：李伟达

日期：2026 年 6 月 9 日

日期：2026 年 6 月 9 日

上海大学工学专业硕士学位论文

基于形状空间理论的特征增强
方法研究

作者: 李睿杰

导师: 韩越兴

学科专业: 软件工程

计算机工程与科学学院

上海大学

2026 年 6 月

A Dissertation Submitted to Shanghai University for the
Degree of Professional Master in Engineering

Feature Enhancement Based on the Shape Space Theory

Candidate: Ruijie Li

Supervisor: Yuexing Han

Major: Software Engineering

School of Computer Engineering and Science

Shanghai University

JUNE, 2026

摘 要

在数据受限的场景中，深度学习模型提取的样本特征分布极为稀疏，不足以充分刻画类内特征的分布多样性，导致下游任务难以取得良好表现。特征级数据增强方法是一类在预训练模型提取的特征上操作的重要数据增强手段，然而现有特征增强方法大多依赖样本自身或者样本之间在欧氏空间中的线性插值假设，忽视了特征向量在高维空间中所呈现的非线性流形结构。形状空间理论为此提供了新的视角，通过将深度模型提取的特征向量投影至预形状空间，利用该流形的测地线性质实现预形状空间下的非线性特征增强。然而，已有的基于测地线的特征增强方法受限于迭代搜索式的优化方式，采取大量点采样与测地距离计算，导致增强样本拟合程度较差，且优化效率存在明显瓶颈，并在单样本等极端场景下无法运行。

针对上述不足，本文开展了以下研究工作：

(1) 针对已有测地线增强方法效率低下和迭代搜索式优化效果不佳的问题，提出了基于形状空间理论的自适应测地线特征增强（Feature Augmentation on Adaptive Geodesic Curve, FAAGC）算法模块。该模块将深度模型提取的特征向量投影至预形状空间，通过对预形状空间分布拟合的分析确立优化目标，重新设计包含样本相似性损失与分布差异损失的联合损失函数，并采用梯度下降方法在预形状空间中拟合能够反映类内特征分布的测地线，最后沿测地线进行参数化采样以生成增强特征。可视化与实验证明，与已有的测地线增强方法相比，FAAGC 在计算效率上实现了大幅提升，在限制训练集数量的 CIFAR-10@5 数据集上实现了约 131 倍的加速，能耗降低至 FAGC 的约 0.76%，同时在多个基准数据集的小样本条件下取得了更优的分类精度。

(2) 针对测地线增强方法在单样本等极端场景下无法有效运行的问题，构建了可扩展测地线特征增强（Scalable Feature Augmentation on Adaptive Geodesic Curve, SFAAGC）框架。该框架通过引入图像增强驱动的特征扩充技术，在输入层面生成多样化的增强图像并提取特征，为测地线拟合提供充足的初始化样本，克服了单样本场景下测地线拟合无法执行的限制；同时采用小批次梯度下降与基于样本来源的加权机制，有效缓解了大规模增强样本下的计算瓶颈。SFAAGC 框架得到的拟合测地

线具有更强泛化性，并且增强特征在多个数据集上的下游任务中取得了显著性能提升，在效率与性能之间实现了良好的平衡。

本文对上述两种方法构建了实验平台，并开展了大量实验。实验结果表明，所提方法在数据受限条件下稳定优于多种特征级增强基线方法，且在不同骨干网络架构和不同数据规模下均表现出良好的适用性与鲁棒性。消融实验进一步验证了框架各组件的必要性以及方法对超参数的鲁棒性。本文的研究将预形状空间的流形几何结构引入特征增强，从增强模块到完整框架构建了一套面向数据受限场景的系统性方案，在保持流形性质的前提下显著提升了测地线拟合效率与对数据规模的适应能力，为小样本条件下深度模型的训练提供了新的技术路径。

关键词：形状空间理论；预形状空间；测地线；特征增强；小样本学习

ABSTRACT

In data-limited scenarios, the feature distributions extracted by deep learning models are extremely sparse, insufficient to fully characterize the intra-class distribution diversity of features and consequently leading to poor performance on downstream tasks. Feature-level data augmentation methods constitute an important class of data augmentation techniques that operate on features extracted by pre-trained models. However, most existing feature augmentation methods rely on the linear interpolation assumption over individual samples or between sample pairs in Euclidean space, overlooking the nonlinear manifold structure exhibited by feature vectors in high-dimensional space. Shape space theory offers a new perspective: by projecting the feature vectors extracted by deep models into the pre-shape space, the geodesic properties of this manifold can be exploited to achieve nonlinear feature augmentation within the pre-shape space. Nevertheless, existing geodesic-based feature augmentation methods are constrained by an iterative search-based optimization scheme that involves extensive point sampling and geodesic distance computation, leading to inferior fitting quality of augmented samples and noticeable efficiency bottlenecks, and rendering these methods inapplicable to extreme scenarios such as single-shot settings.

To address these limitations, this thesis carries out the following research:

(1) To address the problems of low efficiency and the suboptimal results of iterative search-based optimization in existing geodesic-based augmentation methods, an adaptive geodesic feature augmentation module based on shape space theory, namely Feature Augmentation on Adaptive Geodesic Curve (FAAGC), is proposed. This module projects the feature vectors extracted by deep models into the pre-shape space, establishes the optimization objective through an analysis of distribution fitting in the pre-shape space, redesigns a joint loss function comprising a sample similarity loss and a distribution divergence loss, and employs gradient descent to fit a geodesic curve in the pre-shape space that reflects the intra-class feature distribution. Augmented features are then generated by parameterized sampling along the fitted geodesic curve. Visualization and experimental results demonstrate that,

compared with existing geodesic-based augmentation methods, FAAGC achieves substantial improvements in computational efficiency—about 131x speedup on the size-restricted CIFAR-10@5 dataset, with energy consumption reduced to approximately 0.76% of that of FAGC—while attaining superior classification accuracy on multiple benchmark datasets under few-shot conditions.

(2) To overcome the limitation that geodesic-based augmentation methods cannot operate effectively in extreme scenarios such as single-shot settings, a Scalable Feature Augmentation on Adaptive Geodesic Curve (SFAAGC) framework is constructed. By incorporating an image-augmentation-driven feature oversampling technique, the framework generates diversified augmented images at the input level and extracts their features, thereby providing sufficient initialization samples for geodesic fitting and overcoming the restriction that geodesic fitting cannot be executed under single-shot scenarios. In addition, mini-batch gradient descent combined with a source-aware sample weighting mechanism is adopted to effectively alleviate the computational bottleneck arising from large-scale augmented samples. Geodesic curves fitted within the SFAAGC framework exhibit stronger generalization, and the resulting augmented features yield significant performance gains on downstream tasks across multiple datasets, achieving a favorable balance between efficiency and performance.

An experimental platform is built for the two proposed methods, on which extensive experiments are conducted. Experimental results show that the proposed methods consistently outperform various feature-level augmentation baselines under data-limited conditions, and exhibit strong applicability and robustness across different backbone architectures and data scales. Ablation experiments further validate the necessity of each framework component and the robustness of the methods to hyperparameter choices. By introducing the manifold geometry of the pre-shape space into feature augmentation, this work develops a systematic solution for data-limited scenarios that progresses from a single augmentation module to a complete framework. While preserving the manifold properties, the proposed methods substantially improve geodesic fitting efficiency and adaptability to varying data scales, offering a new technical pathway for training deep models under few-shot conditions.

Keywords: Shape Space Theory; Pre-shape Space; Geodesic Curve; Feature Augmenta-

tion; Few-shot Learning

目 录

摘 要	I
ABSTRACT	III
第一章 绪论	1
1.1 课题来源	1
1.2 研究背景	1
1.3 研究意义	3
1.4 国内外相关工作	3
1.4.1 图像域数据增强方法	3
1.4.2 特征域数据增强方法	5
1.4.3 基于形状空间理论的特征增强方法	6
1.5 论文主要工作	7
1.6 论文结构	8
第二章 相关理论与技术基础	10
2.1 深度学习理论基础	10
2.2 形状空间理论	13
2.2.1 预形状空间的基本结构	13
2.2.2 形状空间与形状空间距离	14
2.2.3 预形状空间中的测地线与类内采样	14
2.2.4 基于形状空间理论进行特征增强	15
2.3 评价指标与可视化方法	15
2.3.1 分类准确率	16
2.3.2 运行时间与能耗	16
2.3.3 t-SNE 可视化	17
2.4 本章小结	17
第三章 基于形状空间理论的自适应测地线特征增强算法模块	18
3.1 整体流程	18

3.2	特征提取与预形状空间投影	18
3.3	基于梯度下降的测地线拟合算法	20
3.3.1	优化目标及可学习参数	20
3.3.2	损失函数构造与优化	21
3.3.3	损失函数的分布意义	23
3.4	基于测地线的采样增强与数据可视化	24
3.5	实验分析	24
3.5.1	实验设置	24
3.5.2	测地线增强算法的运行时间与准确率对比实验	27
3.5.3	测地线拟合增强模块消融实验	28
3.6	本章小结	31
第四章	可扩展的测地线特征增强框架	34
4.1	总体框架	34
4.2	基于图像增强的特征扩充方法	35
4.3	基于样本加权与小批次梯度下降的测地线拟合算法	35
4.4	增强样本数据可视化	37
4.5	实验分析	39
4.5.1	实验设置	39
4.5.2	对比分析	40
4.5.3	增强框架效率对比实验	43
4.5.4	结合特征扩充的测地线增强框架消融实验	44
4.6	本章小结	48
第五章	测地线特征增强实验平台设计与实现	50
5.1	平台概述与需求分析	50
5.2	平台总体架构设计	51
5.3	功能模块设计与实现	52
5.3.1	数据管理模块	53
5.3.2	特征提取与特征扩充模块	54
5.3.3	测地线增强模块	55

5.3.4 评估与可视化模块	56
5.3.5 交互界面模块	56
5.4 本章小结	58
第六章 总结与展望	59
6.1 研究总结	59
6.2 未来工作	60
插图索引	62
表格索引	63
参考文献	64
攻读专业硕士学位期间取得的研究成果	71
致 谢	72
附录 A 测地线形式与插值形式的等价性证明	73
附录 B 升维策略对分类准确率的影响	75

第一章 绪论

1.1 课题来源

本课题受到国家自然科学基金面上项目“基于小样本机器学习的 Mg-Gd 系合金图像-组织-性能内禀关系研究”（编号：52273228）以及云南省科技重点项目新材料专项“基于高分子材料应用的锡基功能助剂制备关键技术研发”（编号：202302AB080022）的资助。

1.2 研究背景

在机器学习领域，利用深度神经网络提取特征并开展各类应用任务，已成为研究和实践中广泛采用的方法。在训练数据充足且训练流程规范的前提下，深度神经网络能够展现出强大的特征提取和特征构建能力。然而，当作用于规模较小的数据集时，深度神经网络经常受困于过拟合问题^[1-2]，即模型过于关注训练集样本的个别特征，而缺乏对样本整体特征的关注，这种情况随着模型规模的增大而变得更为严重。最直接的解决过拟合问题的方法是扩大数据集规模，然而在一些特定领域中，扩大数据集是十分耗时和昂贵的，如材料图像或医学图像领域。

数据增强是在实际训练数据集数量有限的情况下防止过拟合的常用方法。数据增强在保持样本原有分布的前提下，通过某种方法提升训练集样本的数量。在计算机视觉领域，最常用的图像数据增强方法包括基于数字图像处理技术的图像翻转、图像旋转、图像裁剪等，在深度神经网络训练之前采用这些图像增强方法对数据进行变换，已成为一种常用且标准的数据集扩充手段，但这类增强方法存在增强形式有限、多样性不足等问题。近年来，生成式模型（如生成对抗网络、扩散模型等）的发展为数据增强带来了新的方向，但这类方法通常对数据量要求较高，或存在生成成本高、结果偏离原始分布等问题^[3]。

虽然数据增强对于深度神经网络的训练通常是有益的，但是在应用之前必须将数据集的实际情况和下游任务情况纳入考量，有时需要领域的专业人士进行判断。例如，对于以生活中图像为主的 CIFAR-10 数据集^[4]，使用随机翻转和旋转进行数据增强，并将其应用于图像分类任务中可能比较合理，但是在医学影像任务中采用这种

操作可能导致错误模式的出现。同时,在数据量十分受限的情况下,采用图像增强得到的数据依然不能展现出足够的多样性,导致训练或微调得到的深度模型仍产生过拟合现象。在这种情况下,在特征空间中对数据进行增强有着非常重要的意义。例如,DeVries 和 Taylor 提出了基于特征空间的特征级数据增强方法^[5],并发现沿流形遍历时,在特征空间中遇到真实样本的概率要高于在输入空间中遇到的概率。大多数深度学习中的特征级数据增强方法,如 Manifold Mixup 和 Feature-level SMOTE,通常通过线性插值、特征混合或分布采样来生成合成特征。然而,这类做法往往忽略了特征空间中潜在的结构,可能无法保持特征的内在语义一致性,也无法充分利用特征分布的潜在信息。

形状空间(shape space)理论^[6-7]为研究对象的几何形态提供了严格的流形结构,在该理论中,所有仅因平移、缩放与旋转不同而产生差异的形状实例被视为同一形状的等价类,并在形状空间中对对应同一点。其上层流形——预形状空间(pre-shape space)通过消除平移与缩放,使具有相同形状但不同位置与尺度的对象映射为同一点,而同一形状在不同旋转角度下的实例则在预形状空间的单位超球面上形成一条大圆,即一段测地线。由于形状空间能够刻画对象几何结构差异,该理论已被用于目标识别等图像处理任务^[8]。

传统形状空间理论通常以对象轮廓、骨架或关键部位上的一组标志点作为研究对象,通过去除平移、尺度和旋转等无关变换,刻画关键点之间的关系。对于深度神经网络提取的图像特征而言,虽然其并非直接来源于人工标注的物体轮廓关键点,但特征向量的各维元素或特征图同样包含图像样本的语义、局部结构和类内变化信息。因此,可以将深度特征抽象为一组特征点,并映射到预形状空间中。基于这一对应关系,深度特征向量可以借助形状空间中的测地距离和测地线路径进行度量、插值与增强,从而比欧氏空间中的线性插值更有利于保持特征分布的非线性结构。

基于上述假设,研究者提出测地线特征增强(Feature Augmentation on Geodesic Curves, FAGC)方法,利用形状空间理论中的预形状空间和测地线距离等概念,将深度学习提取的特征投影至预形状空间,在其上拟合能反映特征分布的测地线,并在该测地线上进行参数化采样以生成新的特征向量^[9]。最终将这些采样特征并入训练特征集中参与训练,以提升下游任务的性能。然而,该方法存在以下缺陷:首先,其优化流程依赖迭代式更新且优化目标迭代搜索式设计,难以保证测地线的最优拟合结果,单次运行耗时也较长;其次,测地线构建至少需要两个样本,单样本场景无法

适用；第三，在较大样本数量下计算效率显著下降，限制了方法在大规模数据集上的实用性。

1.3 研究意义

虽然当前基于形状空间理论的测地线特征增强方法存在上述缺陷，但是其在样本数量受限情况下对于特征的优秀增强表现展现了该类增强方法的显著优势。利用形状空间理论对特征空间中的样本进行更加高效与有效的增强，并构建图像数据集的增强框架，有着以下重要意义：

在理论层面，形状空间理论为特征增强提供了一种有效的数学框架。传统特征增强方法大多在欧氏空间中执行线性插值或随机扰动，而忽视了特征向量在高维空间中所呈现的非线性流形结构。将特征向量投影至预形状空间并沿测地线进行采样，能够在保持流形一致性的前提下生成非线性关系的增强样本，为理解和利用特征空间的内在几何结构开辟了新的研究思路。

在应用层面，基于形状空间理论增强的特征显著区别于利用线性插值或者利用随机扰动的方法，其非线性生成特征的特性使新生成的数据具有更强的多样性；而形状空间理论确保新生成的样本向量与原样本向量落在相同的预形状空间上，从而保证样本生成的一致性。另一方面，随着机器学习技术在大量行业中的应用，出现了各种特定领域的小规模数据集，为了提升基于这些数据集开展的机器学习任务的表现，各领域的工作者亟需一类能够在样本数量有限的情况下有效增强数据集，进而提升下游各类任务表现的特征增强方法。本文所探讨并实现的数据集增强模块，能够有效地增强特征数据并提升下游任务表现，同时还可以通过简单的修改适配各类不同数据集，具有较好的适用性。此外，本文基于图像数据集，还提出了基于特征扩充与加权小批次优化的测地线特征增强框架，能够进一步提升在计算机视觉任务中生成特征对样本实际分布的拟合程度，以及图像分类任务的表现。

1.4 国内外相关工作

1.4.1 图像域数据增强方法

在样本数量受限的情况下，为了提升模型性能和泛化能力，研究者提出了多种基于图像域的数据增强方法。

在计算机视觉领域,常见的图像数据增强技术包括翻转、旋转、平移、缩放、添加噪声、遮挡以及颜色扰动等。这些操作通常根据预先设定的概率、顺序和强度,被单独或组合应用于输入图像,从而增加训练样本的多样性并提升模型的泛化能力^[10-11]。另一类颇具影响力的图像增强策略由 Zhang 等人提出,通过对两个训练样本及其标签进行凸组合,构造虚拟训练样本,称作混合训练 (Mixup)^[12],其简洁有效的思路在后续研究中得到广泛拓展与应用。

虽然上述方法取得了一定成效,但仍需依赖人工经验来设计符合特定领域先验知识的增强策略。为此,有研究通过将数据增强策略的搜索过程建模为离散优化问题,并利用强化学习寻找能够最大化验证集准确率的变换策略,称为自动数据增强 (AutoAugment)^[13]。随后又出现了更为简便的方法,通过随机选择增强操作并使用统一幅度,使策略搜索更加灵活与高效,称作随机增强 (RandAugment)^[14]。此外,也有名为增强混合 (AugMix)^[15]的研究通过混合多个随机采样的增强链来提升多样性,从而进一步提高模型的稳健性和泛化性。

生成式图像增强方法可以在医学影像等高精度任务中应用^[16]。有工作通过使用生成对抗网络 (Generative Adversarial Networks, GANs)^[17-18]来生成具有真实病灶特征的合成样本,并对 CT 和 MRI 数据进行特定切片增强。然而,生成对抗网络的有效训练高度依赖充足的数据,当样本数量有限时,往往会出现模式崩塌或过拟合等问题,难以生成高质量的合成样本^[19-20]。在医学等专业领域中,通用和领域特定的数据增强方法均可被采用^[21],但需要通过针对特定数据集的实验验证,并获得领域专家的认可,以确保其适用性和有效性。近年,也有基于扩散模型 (Diffusion Models)^[22-23]的图像生成方法被应用于图像数据增强中,比如 Trabucco 等通过利用扩散模型的变换增强图像数据集^[24],Islam 等提出 DiffuseMix 和 GenMix^[25-26],这些方法都基于扩散模型和引导提示词,结合掩码混合插值技术生成增强图像^[27]。

尽管图像域数据增强方法取得了丰富的成果,但其本质上仍在输入空间进行操作,所产生的变换可能未必能有效模拟特征空间中的深层语义变化,且在数据量不足时难以产生足够的多样性。因此,研究者逐渐将增强策略从输入空间转移到特征空间。

1.4.2 特征域数据增强方法

与上述图像域方法不同，另一类方法完全在图像模型提取的特征层面进行数据增强，从而提升深度学习模型的泛化能力。Goodfellow 等人^[28]的研究揭示了深度学习模型对对抗样本的脆弱性，并基于此提出在特征空间中加入扰动以提升深度学习模型的稳健性。也有研究利用变分自编码器 (Variational Autoencoders, VAEs)^[29] 将数据映射到潜在空间，并在其中进行外推、插值与扰动，以生成多样化的增强样本^[5]。类似地，Chen 等提出了一种单样本数据增强方法，通过自编码器将图像特征投影到语义特征空间，在语义空间中检索近邻概念后，再由解码器将其映射回图像特征空间，以合成新的增强特征，实现单样本图像增强^[30]。随后的工作在 Mixup 思想的基础上，将插值操作从输入空间延伸至模型的隐层特征空间，以提高模型对抗攻击的鲁棒性和准确性，称作流形混合 (Manifold Mixup)^[31]。

一些简单但有效的策略也相继出现，例如李盼等发现，通过在样本特征中加入特定噪声即可显著提升迁移学习场景下的泛化能力和鲁棒性，称为简易特征增强 (Simple Feature Augment, SFA)^[32]。此外，李盼等还提出通过交换不同样本之间特征的均值和方差来生成新样本，称作矩交换 (Moment Exchange, MoEx)^[33]。也有研究探索基于流形的生成方法，例如在流形约束的潜在空间中进行采样以改进增强效果^[34]。此外，Chu 等人提出针对长尾数据分布的特征级增强方法^[35]，Chawla 等人提出的合成少数类过采样技术 (Synthetic Minority Over-sampling Technique, SMOTE)^[36-37] 通过在少数类样本的特征近邻之间进行插值合成新样本，有效缓解了类别不平衡问题。在此基础上，Liu 等人将过采样思路进一步延伸至深度学习模型的特征空间，以增加故障诊断任务中少数类的样本数量，该方法称为 Feature-level SMOTE^[38]。Bär 等人基于 Vision Transformer 模型在特征提取过程中不改变分块结构的特点，将部分图像增强方法应用于特征数据的增强，并在此过程中保持骨干网络参数不变，该方法称作冻结特征增强 (Frozen Feature Augmentation)^[39]。

此外，也有研究在图像域数据增强的基础上，进一步对深度神经网络的中间层特征进行扰动，以提升模型的稳健性。Sandru 等提出特征级增强方法 (Feature-Level Augmentation, FLA)，在对输入图像施加常规数据增强的同时，随机选取网络中间层，并对其激活图施加平移、旋转与缩放等仿射变换，通过多尺度特征扰动增强模型对微小仿射变换的鲁棒性^[40]。再比如，Kumar 等提出 DiSMHA 框架，将 Stable Diffusion

图像生成与特征级 SMOTE 相结合：首先利用扩散模型在像素层面合成少数类图像，再经深度学习模型提取特征后，在特征空间中利用 SMOTE，从而系统性地缓解极端类别不平衡问题^[41]。

总体而言，特征域增强方法直接在模型提取的特征层面进行操作，可在主干网络参数固定的前提下作用于分类器训练阶段，避免了在图像层面逐一生成增强样本的需要。然而，上述特征域增强方法大多依赖欧氏空间中的线性插值或随机扰动，难以充分捕捉特征向量在高维空间中所呈现的非线性流形结构。为了更好地利用特征的内在几何性质，部分研究者开始探索在非欧氏流形空间中进行特征增强。

1.4.3 基于形状空间理论的特征增强方法

深度模型提取的高维特征通常包含样本之间的语义相似性和类内变化信息，但这些变化未必能够用欧氏空间中的直线关系充分描述。尤其在样本数量有限时，若仅依赖线性插值或随机扰动生成新特征，可能会忽略样本之间更细致的几何关系。已有研究表明，将数据特征约束到单位超球面或其他具有明确几何性质的空间中，有助于更稳定地刻画样本之间的距离和变化关系^[42-43]。因此，从具有明确流形约束的空间中构造增强样本，为小样本条件下的特征增强提供了新的思路。

部分研究将数据增强思想从欧氏空间扩展到非欧氏流形空间，其中具有代表性的理论之一是 Kendall 提出的形状空间理论^[6-7]。该理论通过消除样本在位置、尺度与方向上的差异，将对象的形状表示为位于预形状空间和形状空间中的点，从而能够利用流形上的测地线距离衡量不同对象之间的相似程度。该理论最初被用于统计形状分析，随后逐渐扩展到目标识别等图像处理任务中，用于衡量不同物体形状之间的差异^[8]。

随着深度学习的发展，研究者开始尝试将形状空间理论与深度学习特征结合，用于改进表示学习和数据增强方法。例如，有研究通过对变分自编码器进行改造，使其能够将输入图像映射到一个分离旋转因素的潜在空间中，从而生成具有不同角度变化的新样本^[44]。也有研究利用具有几何约束的低维潜在空间，通过在潜在流形上进行随机扰动，不断扩充三维模型数据集，从而获得大量高质量的训练样本^[45]。Hu 等通过将深度学习模型提取的特征投影至预形状空间，并在此基础上与可学习的注意力模块相结合，改善了模型在样本数量有限条件下的迁移学习表现^[46]。

也有研究进一步探索如何在不改变图像本身的情况下，仅在特征层面利用形状

空间进行数据增强。测地线特征增强方法（Feature Augmentation on Geodesic Curves, FAGC）在预形状空间中拟合深度特征的测地线路径，并沿该路径采样生成新的特征样本，用于增强低数据量场景下的分类任务^[9]。FAGC 在一些数据集上的增强效果较好，在许多经典分类器上能够显著提升分类准确率。然而，由于其需要在每一次增强迭代中计算大量数据点之间的几何距离，当数据规模扩大时，整体计算量会迅速增加，影响其实用性。

Han 等人将测地线特征增强思想扩展到零样本文本引导的扩散风格迁移任务中，提出 FAGStyle 方法。该方法通过滑动窗口裁剪获得局部图像块，并利用测地曲面特征增强（Feature Augmentation on Geodesic Surface, FAGS）加强图像块之间的特征交互，同时引入预形状自相关一致性约束以保持内容结构一致性^[47]。这表明基于预形状空间的几何特征增强不仅可用于小样本分类任务，也具备向生成式视觉任务扩展的潜力。

总体来看，上述基于形状空间理论的研究展示了在流形上进行数据集特征增强的潜力，为低资源条件下的模型训练提供了新的方向。

1.5 论文主要工作

为探索在小样本环境下实现有效且高效的特征增强，本论文基于形状空间理论提出并研究了以下两项核心工作，如图 1.1 所示。

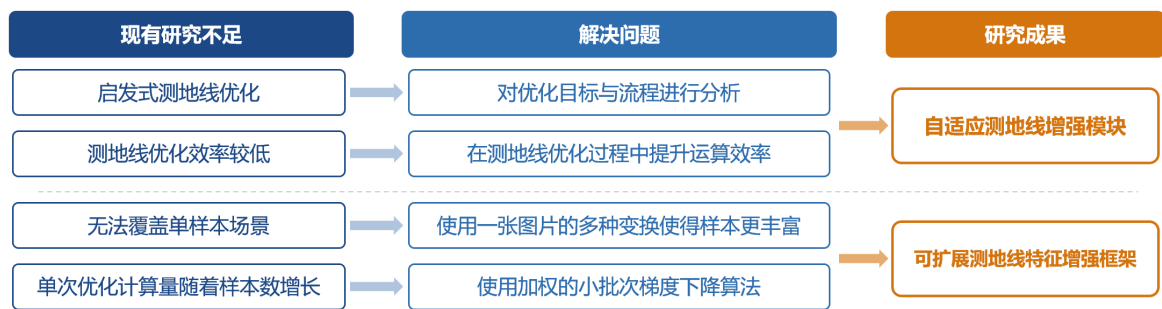


图 1.1 本文主要工作示意图

(1) 提出一种基于形状空间理论的自适应测地线特征增强（Feature Augmentation on Adaptive Geodesic Curve, FAAGC）算法模块，以解决现有的基于形状空间理论的测地线拟合优化过程中，测地线优化目标不明确及优化过程效率较低的问题。通过重新设计预形状空间中的测地线拟合算法，建立优化目标损失函数，并借助梯度下

降及矩阵化运算进行高效优化，使训练样本特征向量在预形状空间中的测地线拟合更加准确、收敛更加稳定。通过对优化目标的理论分析可证明，在特定条件下所生成的测地线增强样本相较其他测地线插值得到的样本更加接近训练特征向量，从而在远低于以往方法计算开销的前提下实现了良好的增强效果。

(2) 提出一种可扩展测地线特征增强 (Scalable Feature Augmentation on Adaptive Geodesic Curve, SFAAGC) 框架，以解决现有测地线算法无法对单样本图像数据集进行拟合与增强、且单次计算量随训练样本特征数量线性上升的问题。将基于图像增强技术的特征扩充算法与 FAAGC 相结合，并将测地线增强模块中的梯度下降流程改进为小批次加权梯度下降，使本框架能够适用于包括单样本场景在内的各种规模的数据增强任务。与 FAAGC 相比，SFAAGC 在损失函数设计上引入基于样本来源的加权相似性损失与按批次权重缩放的分布差异损失，以调节原始样本与扩充样本对测地线拟合的贡献；在计算效率上，单次迭代的计算量由随训练样本总数线性增长转变为仅与固定批次大小相关，从而显著降低了大规模样本下的拟合开销。通过数据可视化，以及在多个标准图像数据集与不同骨干网络上进行下游分类任务的对比实验与消融实验，系统验证了所提出框架的有效性。

此外，本文构建了测地线特征增强实验平台，将所提出方法封装为配置驱动的自动化实验系统，基于可视化交互界面支持实验配置与结果展示，以支撑对比实验、消融实验及可视化分析的系统性开展。

1.6 论文结构

针对现有特征增强方法缺少特征空间非线性关系关注以及基于形状空间的测地线增强方法计算效率不足等问题，本文提出了一种基于形状空间理论的自适应测地线特征增强算法模块及其可扩展框架，旨在实现高效与有效的特征增强，并在多种数据规模与骨干网络下验证其有效性。本文共分为六章，各章内容安排如下：

第一章为绪论，介绍课题来源、研究背景与意义，综述国内外相关工作，概述论文主要工作与论文结构。

第二章介绍相关理论与技术基础，包括深度学习理论基础，形状空间理论、预形状空间中的几何性质以及评价指标和可视化方法。

第三章提出基于形状空间理论的 FAAGC 模块，给出测地线拟合与采样的关键

步骤与损失函数分布意义，并通过效率对比与模块消融实验验证方法的有效性与效率。

第四章在前述方法基础上，构建 SFAAGC 框架，并通过对比实验、增强框架效率对比实验与框架消融实验验证其在不同数据规模下的适用性与计算效率。

第五章介绍测地线特征增强实验平台的设计与实现，将所提出方法封装为配置驱动的自动化实验系统，并基于可视化交互界面支持实验配置与结果展示。

第六章对全文工作进行总结，并展望未来研究方向。

第二章 相关理论与技术基础

本章旨在为后续章节所提出的测地线特征增强方法提供必要的理论基础。首先介绍深度学习的基本理论，包括骨干网络与特征提取、分类任务中的损失函数，以及面向小样本场景的数据增强策略；随后系统阐述形状空间理论，涵盖预形状空间的基本结构、测地线的构造方法以及基于测地线采样的特征增强原理；最后介绍用于评估增强效果的分​​类准确率、运行时间与能耗等定量指标，以及 t-SNE 可视化方法，为理解后续方法的动机、算法设计与实验评估奠定基础。

2.1 深度学习理论基础

机器学习旨在从数据中学习映射关系或决策规则，从而在分类、回归与结构化预测等任务中实现自动化推断，并已广泛应用于计算机视觉、语音识别与自然语言处理等领域。深度学习是机器学习的重要分支，其核心思想是利用多层神经网络在端到端训练过程中逐级抽象输入，自动学习从低级模式到高级语义的层次化表示^[48]。相较于依赖人工设计特征的传统方法，深度学习的优势主要体现在：能够在大规模数据与计算资源支持下学习更强的表示能力、能够更好地进行任务迁移与特征复用，并能通过统一的优化框架将特征学习与任务学习协同完成。

为了在不同任务之间复用已验证有效的表示学习能力，计算机视觉模型通常采用“骨干网络（backbone）+ 任务头（head）”的设计。其中，骨干网络指负责将原始输入映射为通用特征表示的一段网络结构，可被视为特征提取器；而任务头则根据具体下游任务（分类、检测、分割等）对特征进行适配与输出。骨干网络之所以被广泛使用，一方面是因为其能够在大规模数据上预训练并迁移到下游任务^[49]，降低针对特定任务训练的难度；另一方面是因为不同方法往往可以在相同骨干网络上进行公平比较，从而将改进点聚焦在特征增强、损失函数或优化策略等关键环节。

其中，最关键的部分是特征提取部分，即骨干网络（backbone）。骨干网络负责将原始输入（例如图像）变换为高维特征表示，使得后续的分类器或预测头能够更容易地完成下游任务。目前，主流的骨干网络架构主要包含以深度卷积神经网络（CNN）为基础的以及以 Transformer 为基础的两​​大类。

卷积神经网络（convolutional neural network, CNN）是图像领域应用最广泛的特征提取结构之一。其设计动机源自图像数据的两个基本特性：一是像素之间的相关性主要集中于局部邻域，二是同一视觉模式可能出现在图像的任意位置。若采用全连接网络直接处理图像，参数量会随输入分辨率的平方增长，且不显式利用空间结构；CNN 通过局部连接与权值共享，使同一卷积核能够在不同空间位置检测相似的局部模式，在保留二维空间结构的同时显著降低参数规模。CNN 对图像的处理由卷积、非线性激活、归一化与下采样四类基本操作级联构成。卷积层使用若干二维卷积核在输入特征图上滑动，计算局部感受野内的加权求和，提取边缘、纹理等局部模式；非线性激活函数（如 ReLU）逐元素施加非线性变换，使网络能够拟合非线性映射；批归一化对小批次特征进行均值与方差归一化以稳定训练；下采样则通过池化或步幅大于 1 的卷积压缩空间分辨率。从特征张量的形状变化看，记输入为 $x \in \mathbb{R}^{H_0 \times W_0 \times C_0}$ ，经过第 ℓ 层后得到 $x^{(\ell)} \in \mathbb{R}^{H_\ell \times W_\ell \times C_\ell}$ ，通道数 C_ℓ 随深度递增以容纳更多种类的模式，空间尺寸 $H_\ell \times W_\ell$ 随下采样递减以扩大单个单元的感受野；与此同时，特征所表达的语义层级也由低层视觉模式逐步过渡到与类别相关的高层语义^[50]。在网络末端，池化层将空间维度上的信息聚合为紧凑的向量表示：池化操作不引入可学习参数，通过取最大值或平均值的方式聚合邻域信息，既降低了特征冗余与计算开销，又增强了表示对小幅平移与形变的鲁棒性；常用的全局平均池化将最后一层特征图 $\mathbb{R}^{H_L \times W_L \times C_L}$ 沿空间维度平均，得到 C_L 维向量作为送入分类头的全局特征。代表性的深度 CNN 结构可追溯至 AlexNet^[50]，其在大规模图像分类任务中验证了深层卷积网络的有效性；随后，残差网络通过残差连接缓解深层网络训练中的退化问题，使更深的 CNN 能够稳定训练并获得更强的特征表达能力^[51-52]。

视觉 Transformer（Vision Transformer, ViT）是 Transformer 架构在计算机视觉领域的代表性应用，采用与卷积不同的特征提取方式。卷积的局部感受野使长距离依赖只能通过堆叠多层间接建模，对全局结构的刻画效率有限；自注意力机制在单层内即可对任意两个空间位置直接建立相关性，因此更适合捕获跨区域依赖^[53]。ViT 主要由分块嵌入、位置编码、多头自注意力、前馈网络与层归一化组成。给定输入图像 $x \in \mathbb{R}^{H \times W \times 3}$ ，ViT 首先将其划分为 $N = HW/r^2$ 个大小为 $r \times r$ 的不重叠图像块，并通过线性投影将每个图像块映射为 d 维向量，得到 N 个 token；随后在序列首部拼接一个可学习的分类 token（记为 [CLS]），并对全部 $N + 1$ 个 token 加入位置编码以保留空间位置信息，形成形状为 $(N + 1) \times d$ 的输入序列。多头自注意力将输入

序列分别投影为查询、键与值，按缩放点积计算 token 之间的相关性权重，并在多个头上并行执行后拼接，使模型能够在不同子空间中捕获不同类型的依赖关系。在多层 Transformer 编码器中，序列形状始终保持为 $(N + 1) \times d$ ，但 token 间的语义关系不断被自注意力重组并融合全局上下文。在序列末端，取 [CLS] 对应的输出向量，或对全部图像块 token 取平均，得到 d 维全局特征向量作为分类头的输入^[54-55]；该步骤在功能上与 CNN 中的全局平均池化相对应。

池化或聚合后的全局特征会被输入到任务头 (head) 中以完成下游任务。由于本文后续实验主要面向分类任务，这里对分类头作更具体说明：设骨干网络经过池化操作输出的全局特征向量为 $v \in \mathbb{R}^d$ ，类别数为 N_c ，则最常见的线性分类器可写为一个全连接线性层

$$z = Wv + b, \quad (2.1)$$

其中 $W \in \mathbb{R}^{N_c \times d}$ 、 $b \in \mathbb{R}^{N_c}$ ，输出 $z \in \mathbb{R}^{N_c}$ 通常称为 logits，表示样本属于各类别的未归一化得分。进一步经过 softmax 函数归一化后，可得到类别概率分布 $p \in \mathbb{R}^{N_c}$ ，满足 $\sum_{c=1}^{N_c} p_c = 1$ 。监督学习中常用的真实标签表示方式之一是 One-Hot 编码：对每个样本，用向量 $y \in \{0, 1\}^{N_c}$ 表示其类别，其中正确类别对应的位置为 1，其余位置为 0（且 $\sum_{c=1}^{N_c} y_c = 1$ ）。在此表示下，交叉熵损失可用于度量预测分布 p 与真实分布 y 的差异，从而驱动骨干网络与分类头共同优化^[56]。

更具体地，单个样本的交叉熵损失可写为

$$\mathcal{L}_{\text{CE}}(p, y) = - \sum_{c=1}^{N_c} y_c \log p_c, \quad (2.2)$$

当 y 为 One-Hot 向量时，上式等价于对真实类别对应的预测概率取负对数，从而鼓励模型在真实类别上给出更高的置信度。深度学习模型以小批次随机梯度下降的方式最小化交叉熵损失，从而提升模型的分类性能。

除分类外，在回归任务中，任务头输出连续值；在检测、分割等密集预测任务中，则会结合多尺度特征与解码结构进一步产生空间对齐的预测结果。

在机器学习训练过程中，常因欠拟合或过拟合导致训练模型效果不及预期。欠拟合指模型规模或训练不足导致无法有效拟合训练数据，训练误差与测试误差均较高；过拟合则指模型过度适配训练集中的特定模式或噪声，导致训练误差较低但泛

化性能较差^[56]。在样本稀少的情况下，深度模型由于参数量大、表示能力强，更容易出现过拟合。在不改变任务定义与数据分布前提下，通过数据增强扩充训练样本、提升有效数据量，成为小样本与低资源场景下的缓解过拟合常用手段^[10]。如 1.4 所述，在计算机视觉领域，数据增强可大致分为图像域增强（包括传统的几何/外观变换以及基于生成模型的生成式增强）与特征域增强^[5,10]。图像域增强直接对输入图像施加几何或外观变换，是实践中的常用手段；生成式增强作为图像域增强的一个分支，利用生成模型合成新样本；特征域增强则在深度模型提取的表示空间中扩充样本，可以避免训练主干网络和生成图像样本等操作。在特征域增强的研究中，形状空间理论因其流形几何结构，被引入到特征层面的数据增强中，尤其在样本数量有限的场景下展现出独特的优势。

2.2 形状空间理论

2.2.1 预形状空间的基本结构

形状空间（shape space）由 Kendall 等系统提出，为几何对象的不变性分析提供了严格的流形框架^[6-7]。根据形状理论，通过消除平移、缩放和旋转，可以将具有相同形状的实例映射到形状空间中的同一点。其上层流形为预形状空间（pre-shape space），通过对数据进行中心化与归一化，使具有相同形状但位置与尺度不同的实例映射到单位超球面上的同一点，而不同旋转角度的实例在该空间中形成大圆曲线，即流形上的测地线^[7]。由于从图像数据中提取的特征通常也包含一定的内在结构信息，将这些特征投影至预形状空间，并利用测地线生成新的预形状特征，可以作为一种有效的数据集扩充手段。

在基于关键点的形状表示中，目标对象通常由欧氏空间中的若干关键点坐标描述，并可在消除无关几何变换后用于对象识别与形状比较^[8,57]。对象 v 用欧氏空间中的 d 个关键点表示为：

$$v = [(x_1, y_1), \dots, (x_i, y_i), \dots, (x_d, y_d)] \quad (2.3)$$

为了消除平移影响，对所有坐标进行中心化，得到：

$$v' = [(x_1 - \bar{x}, y_1 - \bar{y}), \dots, (x_i - \bar{x}, y_i - \bar{y}), \dots, (x_d - \bar{x}, y_d - \bar{y})] \quad (2.4)$$

随后，为消除尺度影响，将对象投影到单位超球面，得到预形状向量：

$$\tau = \frac{v'}{\|v'\|} \quad (2.5)$$

该向量位于单位超球面 S_*^{2d-3} 上，具有 $2d-3$ 个自由度。

2.2.2 形状空间与形状空间距离

对对象进行旋转变换所得到的一组预形状向量构成一个轨道 (orbit)，记作 $O(\tau)$ 。所有轨道构成形状空间，其集合形式为：

$$\Sigma_2^d = \{O(\tau) : \tau \in S_*^{2d-3}\} \quad (2.6)$$

预形状空间中两个点 τ_1 和 τ_2 之间的测地距离为：

$$d(\tau_1, \tau_2) = \arccos\langle \tau_1, \tau_2 \rangle \quad (2.7)$$

其中 $\langle \cdot, \cdot \rangle$ 为内积。形状空间中两形状的距离定义为其轨道间的最小测地距离：

$$d_p[O(\tau_1), O(\tau_2)] = \inf\{d(\alpha, \beta) : \alpha \in O(\tau_1), \beta \in O(\tau_2)\} \quad (2.8)$$

上述距离为形状判别和特征比较提供了度量基础。

2.2.3 预形状空间中的测地线与类内采样

在预形状空间中，两个预形状向量 τ_s 与 τ_e 之间存在唯一的大圆测地线，其参数化形式可表示为^[6-7]：

$$\begin{aligned} \Gamma_{\tau_s, \tau_e}(s) &= (\cos s) \tau_s + (\sin s) \frac{\tau_e - \tau_s \cos \theta}{\sin \theta}, \\ 0 &\leq s \leq \theta \end{aligned} \quad (2.9)$$

其中角度

$$\theta = \arccos\langle \tau_s, \tau_e \rangle \quad (2.10)$$

描述了从起点 τ_s 到终点 τ_e 的大圆路径，通过在区间 $[0, \theta]$ 上对参数 s 进行采样，即可得到位于测地线上的任意中间点。本研究利用该测地线，在特征空间中拟合深度特征变化的方向，并沿测地线进行采样以获得新特征，实现特征级数据增强^[9]。

当同一类别中包含多个训练样本时，仅连接任意两个样本点所得到的大圆路径只能描述这两个样本之间的变化，难以代表该类别样本在预形状空间中的整体分布。因此，在将形状空间理论用于特征增强时，需要根据同类样本集合确定一条能够反映类内主要样本的测地线，再在该路径上进行采样。这样生成的增强特征既保留在预形状空间中，又能够更充分地覆盖原始样本之间的连续变化，为后续分类器训练提供更丰富的特征样本。

2.2.4 基于形状空间理论进行特征增强

基于形状空间理论，将深度特征视为关键点（landmarks）并投影到预形状空间，可以利用特征向量的流形结构进行数据增强。测地线特征增强（Feature Augmentation on Geodesic Curves, FAGC）^[9] 是一类典型方法：它在预形状空间中根据样本分布拟合测地线，并沿测地线采样生成新的特征表示。

具体而言，FAGC 首先将训练样本特征投影到预形状空间，并选取测地距离最远的两点作为大圆的初始端点；随后在该测地线上均匀采样多个候选点，通过比较候选点到训练样本的测地距离之和来评估拟合程度，并以迭代方式最小化该测地线距离和；迭代结束后，算法使用式 (2.9) 的测地线插值公式生成增强特征。该方法能够显式利用预形状空间中样本间的非线性关系，使生成的增强样本在几何结构上与原始样本保持一致，可在样本数量有限的图像数据集中对各类经典分类器带来性能提升。总体而言，FAGC 展示了在预形状空间中利用测地线进行特征增强的理论价值与应用潜力。

需要指出的是，FAGC 在测地线优化过程中通常以测地距离下降幅度的阈值作为停止准则，这种优化方法难以从理论上保证收敛到最优解或近似最优解。此外，FAGC 依赖大量采样与距离计算，计算开销较大，使得该方法在大规模数据集上难以直接应用，在优化目标与计算效率等方面仍需要进一步改进。

2.3 评价指标与可视化方法

为评估特征增强方法的有效性，本文通过在原训练集与增强数据集合并后的数据集上训练分类器，并以测试集上的分类准确率作为核心指标，同时报告测地线拟合阶段的实际运行时间与能耗，以兼顾增强效果与算法效率。此外，本文利用 t-SNE 将高维特征投影至二维空间，通过可视化直观地观察增强前后特征分布的变化。

2.3.1 分类准确率

分类准确率（Accuracy）是分类模型评价指标，定义为模型在给定测试集上正确分类的样本数占总样本数的比例：

$$\text{Accuracy} = \frac{1}{N_{\text{test}}} \sum_{i=1}^{N_{\text{test}}} \mathbb{I}[\hat{y}_i = y_i], \quad (2.11)$$

其中 N_{test} 为测试集样本总数， $\hat{y}_i$ 为模型对第 i 个样本的预测类别， y_i 为其真实类别， $\mathbb{I}[\cdot]$ 为指示函数，当条件成立时取 1 否则取 0^[56]。

在分类任务中，准确率直接反映了模型的泛化能力——准确率越高，说明模型在未见样本上的预测能力越强^[50]。在数据增强的评估场景下，首先在原始训练集上应用特征增强方法生成增强特征，将增强特征与原始特征合并构成扩展训练集；随后在该扩展训练集上训练分类器，利用测试集上的分类准确率作为增强效果的定量度量。若某种增强方法能够持续带来准确率提升，则说明其生成的增强特征有助于模型更好地拟合数据的真实分布^[10]。

2.3.2 运行时间与能耗

对于具有解析步骤数的经典算法，时间复杂度（ $\mathcal{O}(\cdot)$ ）通常被视为衡量计算代价的标准指标。然而，对于迭代式优化算法而言，不同方法的“一轮迭代”在语义与计算粒度上存在差异，仅凭迭代次数或理论复杂度难以反映算法之间的真实计算代价；同时，时间复杂度模型默认串行执行，并未刻画 GPU、张量并行等硬件特性对实际耗时的影响。因此，体系结构与高性能计算领域的通行做法是以统一硬件下的实际运行时间作为算法效率的评价基准^[58]。基于上述考虑，本文在统一深度学习框架与硬件环境下，以实际运行时间作为优化阶段计算开销的度量。

除运行时间外，能耗亦是衡量算法运行效率的实用指标。在深度学习训练规模持续增长的背景下，能耗已被纳入算法效率的一等评价维度——较低的能耗不仅意味着更低的运行成本与更宽松的硬件门槛，也对应更小的碳足迹与环境影响^[59-60]。本文采用 CodeCarbon 工具^[61]，在算法运行过程中持续监测 CPU 与 GPU 的利用率与电源功率，并据此估算出优化阶段的累计能耗（kWh），作为该指标的具体测量手段。

2.3.3 t-SNE 可视化

t-SNE (t-distributed Stochastic Neighbor Embedding) 是由 van der Maaten 和 Hinton 提出的一种非线性降维与可视化方法^[62]。其核心思想是在高维空间中定义样本间的相似度（通常基于高斯分布），在低维映射空间中定义另一相似度（基于 t 分布），并通过最小化两个分布之间的 KL 散度来寻找最优的低维嵌入。

相较于传统的线性降维方法（如 PCA），t-SNE 能更好地在二维或三维空间中保留数据在高维流形上的局部邻近结构，使属于同一类别的样本点在低维空间中更容易形成聚簇，而不同类别的样本点则趋于分离。

在特征增强的评估场景下，本文将不同增强方法生成的特征与原始特征一同投影至二维空间进行可视化对比。通过观察不同方法生成的增强特征对原始样本分布的拟合程度与覆盖范围，可以直观地判断各增强方法所生成样本的优劣：若增强特征紧密围绕原特征分布并在其邻域内形成合理的扩展，则表明该增强方法能够较好地捕捉类内变化模式；反之，若增强特征偏离原始分布或仅集中在少数孤立区域，则说明增强质量较低。t-SNE 作为可视化工具，其分析结果通常与分类准确率等定量指标配合使用。

2.4 本章小结

本章回顾了深度学习中的特征表示基础与评价方法，概述了形状空间与预形状空间的几何结构，以及在该流形上构造测地线并进行特征采样的基本原理。此外，本章介绍了用于评估增强效果的分类型准确率、运行时间与能耗等定量指标，以及 t-SNE 可视化方法。最后，结合现有的测地线特征增强方法，讨论了其优势与局限性，为后续章节提出基于形状空间与测地线拟合的特征增强方法奠定了理论与评估基础。

第三章 基于形状空间理论的自适应测地线特征增强算法模块

针对训练样本稀缺时深度学习模型易过拟合的问题，本章提出一种基于形状空间理论的自适应测地线增强（Feature Augmentation on Adaptive Geodesic Curve, FAAGC）算法模块。该方法利用预形状空间中来自同一形状的向量位于同一条大圆测地线上这一几何性质，将深度特征投影至预形状空间并拟合测地线，在测地线上进行参数化采样以生成增强特征，从而在特征层面捕获样本之间的非线性结构关系。本章依次介绍方法的总体框架、特征提取与预形状空间投影、基于梯度下降的测地线拟合算法及损失函数的分布意义，以及基于测地线的采样增强与可视化结果，并通过实验分析验证该方法在准确率与效率方面的优势。

3.1 整体流程

FAAGC 的整体流程如图 3.1 所示，共包含三个阶段：首先，使用深度神经网络模型从训练集图像中提取特征，并将其投影到预形状空间；随后，通过设计损失函数，优化测地线的起点、终点以及训练阶段的采样参数，使增强特征向量尽可能接近训练样本向量的分布；最后，在该测地线上进行参数化采样，生成增强特征。

3.2 特征提取与预形状空间投影

对于图像数据集中的一张图像，可将其表示为形状为 $H \times W \times C$ 的张量。由于大规模预训练视觉模型能够提取高度判别性的图像特征，本研究采用 ViT 等预训练深度神经网络作为特征提取主干。具体而言，图像经骨干网络前向传播后，在池化层或特征输出层截取 d 维特征向量。由此，可实现从原图像空间 $\mathbb{R}^{H \times W \times C}$ 到特征空间 $\mathbb{R}^d$ 的映射。其中， d 表示特征向量中的元素个数，随不同模型结构而变化。

对于带有标签的训练数据，本文将提取到的特征与其标签按照类别 c 进行配对，得到类内特征集合：

$$\mathcal{D}_{\text{train},c} = \{(v_i, c) \mid i = 1, \dots, m_c\}, \quad c = 1, \dots, N_c, \quad (3.1)$$

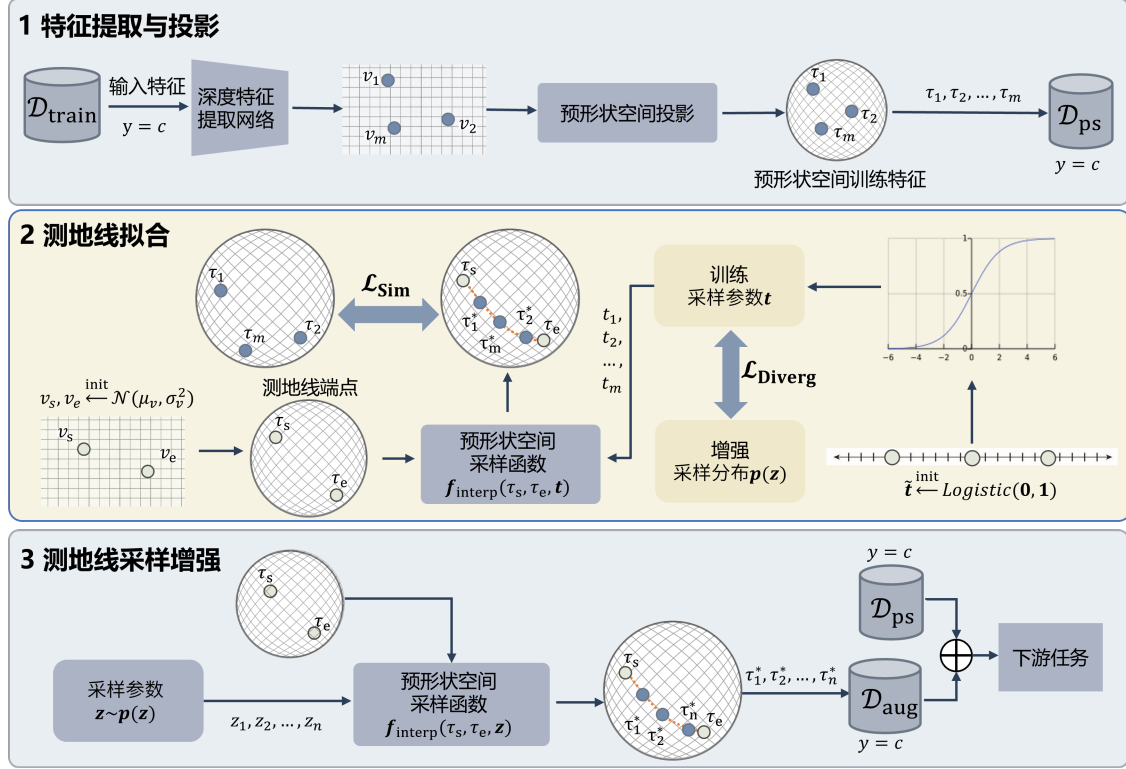


图 3.1 测地线特征增强流程示意图

其中 $v_i \in \mathbb{R}^d$ 为第 i 个样本的特征向量， m_c 为该类别中的样本数， N_c 为类别总数。由于不同类别的样本在预形状空间应由不同测地线表示，以下增强步骤均对于同一类的样本特征展开，因此可将数据集记为 $\mathcal{D}_{\text{train}}$ 。为简洁起见，以下将每类样本数 m_c 统一记为 m 。

根据形状空间理论，形状结构的关键点（landmarks）通常为二维或三维坐标，而深度特征向量中的每个元素是一维特征值。为使特征能够在预形状空间中表示，本文将每个一维特征元素复制一次，对于第 i 个特征向量 $v_i = (v_i[1], \dots, v_i[d])$ ，其升维方式为：

$$\bar{v}_i = [(v_i[1], v_i[1]), \dots, (v_i[d], v_i[d])], \quad (3.2)$$

其中 $v_i[k]$ 表示特征向量第 k 个元素。该映射将原始一维特征组织为由 d 个二维特征点构成的点集。不同升维策略的对比实验见附录 B，结果表明其对分类准确率的影响差异不显著。

随后，如公式 2.4 和公式 2.5 所示，对升维后的特征向量执行中心化与归一化，将其投影至预形状空间，最终得到的预形状空间特征集合为：

$$\mathcal{D}_{\text{ps}} = \{(\tau_i, c) \mid i = 1, \dots, m\}, \quad (3.3)$$

其中 $\tau_i \in S_*^{2d-3}$ 为样本在预形状空间上的向量，由于后续增强操作均针对类内样本特征展开，后续公式省略了类别标识 c 。

综上，本文将特征向量 v 经升维、中心化与归一化后得到的预形状表示记为

$$\tau = P(v), \quad (3.4)$$

其中， $P(\cdot)$ 表示由升维、中心化和归一化操作共同构成的预形状空间投影映射。

3.3 基于梯度下降的测地线拟合算法

3.3.1 优化目标及可学习参数

测地线拟合过程中需要反复在测地线上采样插值点。公式 2.9 中决定插值点位置的参数 s ，其取值范围 $[0, \theta]$ 由端点 τ_s, τ_e 决定，若直接将 s 作为可学习参数，其取值范围会随 τ_s, τ_e 变化。为此，本文将重参数化为 $z = s/\theta \in [0, 1]$ ，使可学习参数的范围与端点解耦，调整后的插值公式如下

$$\tau^* = f_{\text{interp}}(\tau_s, \tau_e, z) = \frac{\sin[(1-z)\theta]}{\sin \theta} \tau_s + \frac{\sin(z\theta)}{\sin \theta} \tau_e, \quad (3.5)$$

其中 $\theta = \arccos\langle \tau_s, \tau_e \rangle$ 为两端点在预形状空间中的测地距离（夹角）， τ^* 为经此插值得到的增强特征向量。该形式与公式 2.9 之间关系的推导过程见附录 A。若一次取 n 个采样参数并组成向量 $\mathbf{z} = (z_1, \dots, z_n)^\top$ ，则对应得到 n 个插值特征向量，记为 $\boldsymbol{\tau}^* = f_{\text{interp}}(\tau_s, \tau_e, \mathbf{z}) = [\tau_1^*, \dots, \tau_n^*]$ 。

FAAGC 通过两类损失约束测地线的拟合组成。样本相似性损失 $\mathcal{L}_{\text{Sim}}$ 使测地线上的增强特征逼近训练集合 $\mathcal{D}_{\text{ps}}$ 的特征分布。分布差异损失 $\mathcal{L}_{\text{Diverg}}$ 使训练阶段采样参数的分布逼近增强阶段所用的 $p(z) = U(0, 1)$ ，如图 3.1 中步骤 2 所示。

FAAGC 拟合算法需要优化的参数包括两个端点可学习参数 $v_s, v_e \in \mathbb{R}^d$ ，用于决定测地线在预形状空间中的走向，以及一组可学习采样参数 $\tilde{\mathbf{t}} = (\tilde{t}_1, \tilde{t}_2, \dots, \tilde{t}_m)^\top$ ，其中每个 $\tilde{t}_i \in \mathbb{R}$ ，用于控制生成插值点沿测地线的分布规律。

对于 $v_s, v_e \in \mathbb{R}^d$ 的初始化，从被深度模型提取的特征向量 $\{v_1, \dots, v_m\}$ 中统计均

值与标准差，并以此为参数从正态分布中随机采样两个向量，分别记为 v_s 与 v_e 。为在预形状空间中进行测地线插值，需获得 v_s, v_e 在预形状空间的投影

$$\tau_s = P(v_s), \quad \tau_e = P(v_e), \quad \tau_s, \tau_e \in S_*^{2d-3}, \quad (3.6)$$

作为后续大圆拟合优化的初值。通过高斯初始化的方式，可减少优化迭代轮数，实验验证见第 3.5.3 节。

图 3.2 展示了可学习参数 $\tilde{t}$ 的必要性。若优化端点时采样点由固定分布 $z \sim U(0, 1)$ 决定，每个 τ_i^* 在测地线上的相对位置即被 z_i 固定，只能随端点整体平移而无法单独调整。这导致 τ_i^* 向 τ_i 靠拢时，其他 τ_j^* ($j \neq i$) 也会被一同移动，难以使所有采样点同时贴近各自的目标样本。将 $\tilde{t}$ 设为可学习参数后，每个采样点可沿测地线独立滑动，在端点固定的前提下即可与各自目标逐一对齐，从而降低样本相似性损失，如图 3.2 所示。

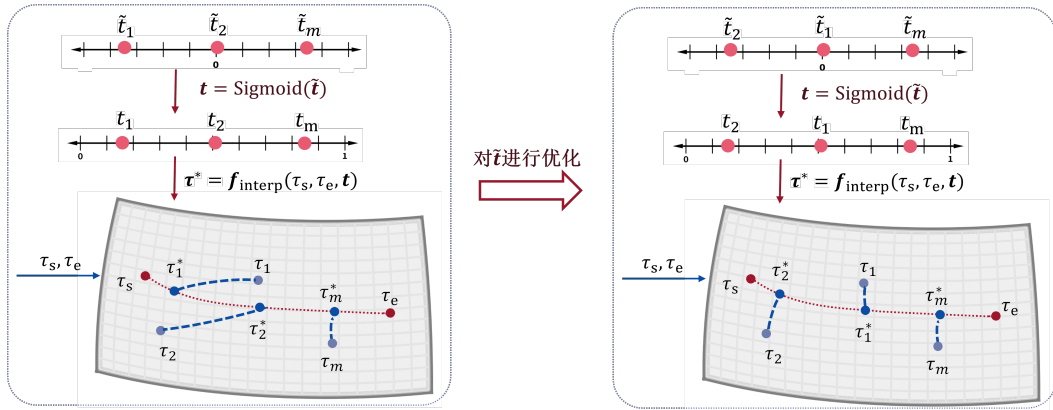


图 3.2 可学习的采样参数 $\tilde{t}$ 能够自适应地对齐采样点与原始样本，从而降低损失函数

因为测地线插值要求插值系数位于 $[0, 1]$ ，本文通过 Sigmoid 函数将 $\tilde{t}$ 映射为 $t = \text{Sigmoid}(\tilde{t}) = 1/(1 + e^{-\tilde{t}}) \in (0, 1)^m$ 。综合以上，测地线拟合过程中，测地线采样向量的计算如公式 3.7 所示。

$$\tau^* = f_{\text{interp}}(P(v_s), P(v_e), \text{Sigmoid}(\tilde{t})). \quad (3.7)$$

3.3.2 损失函数构造与优化

在参数与采样表达式确定之后，下面给出两类损失的具体形式。样本相似性损失 $\mathcal{L}_{\text{Sim}}$ 用于约束增强样本 τ_i^* 接近对应的训练样本 τ_i 。本文采用与测地线距离等价的

余弦相似度误差平方和作为损失：

$$\mathcal{L}_{\text{Sim}} = \sum_{i=1}^m \left(1 - \tau_i^{*\top} \tau_i\right)^2. \quad (3.8)$$

当训练样本数 m 较大时，逐项求和的开销较大，因此可改写为下式的矩阵形式，便于 PyTorch^[63] 等机器学习框架并行执行。记插值过程为 $\tau^* = f_{\text{interp}}(\tau_s, \tau_e, \mathbf{t})$ ，则

$$\mathcal{L}_{\text{Sim}} = \left\| \mathbf{1}_m - (\tau^* \odot \tau)^\top \mathbf{1}_{2d} \right\|^2, \quad (3.9)$$

其中 $\odot$ 表示 Hadamard 乘积（逐元素相乘）， $\mathbf{1}_m$ 与 $\mathbf{1}_{2d}$ 分别为长度为 m 和 $2d$ 的全 1 列向量。 $(\tau^* \odot \tau)^\top \mathbf{1}_{2d}$ 等价于对每一列并行计算内积，再与 $\mathbf{1}_m$ 相减并取向量模的平方，即得到与逐项求和形式等价的损失值。代入式 3.7 所述的参数替换，矩阵化的形式为

$$\mathcal{L}_{\text{Sim}}(v_s, v_e, \tilde{\mathbf{t}}) = \left\| \mathbf{1}_m - \left(f_{\text{interp}}(P(v_s), P(v_e), \text{Sigmoid}(\tilde{\mathbf{t}})) \odot \tau\right)^\top \mathbf{1}_{2d} \right\|^2. \quad (3.10)$$

分布差异损失 $\mathcal{L}_{\text{Diverg}}$ 用于衡量 $\mathbf{t}$ 所对应的分布与 $p(z)$ 之间的差异。计算分布差异的常用方式包括 KL 散度、JS 散度与 Wasserstein 距离等。由于 $\mathbf{t}$ 本质上是有限样本集，无法直接获得概率密度，如果采用需要显式概率密度函数的 KL 与 JS 散度，需对 $\mathbf{t}$ 额外做密度估计；相比之下，一维 Wasserstein 距离^[64] 仅需对样本排序即可构造有效的最优传输距离，能够在保持计算简洁的同时提供稳定梯度。在计算损失时，将 $\mathbf{t}$ 与从 $p(z)$ 采样所得的 $\mathbf{z}$ 分别排序，按一维 Wasserstein 距离进行度量：

$$\mathcal{L}_{\text{Diverg}}(\tilde{\mathbf{t}}) = D_W(\text{Sigmoid}(\tilde{\mathbf{t}}), \mathbf{z}) = \frac{1}{m} \sum_{j=1}^m |\text{Sigmoid}(\tilde{\mathbf{t}})_{(j)} - z_{(j)}|, \quad (3.11)$$

其中 D_W 表示一维 Wasserstein 距离，下标 (j) 表示从小到大排序后的第 j 个元素， m 为采样参数的数量。

为平衡损失函数，引入权重系数 β ，总损失写为

$$\mathcal{L}_{\text{train}}(v_s, v_e, \tilde{\mathbf{t}}) = \mathcal{L}_{\text{Sim}}(v_s, v_e, \tilde{\mathbf{t}}) + \beta \mathcal{L}_{\text{Diverg}}(\tilde{\mathbf{t}}). \quad (3.12)$$

公式 3.12 中的可学习参数 v_s, v_e 与 $\tilde{\mathbf{t}}$ 均无范围约束，借助可微映射 $P(\cdot)$ 与 $\text{Sigmoid}(\cdot)$ ，对它们的优化等价地完成了对预形状空间端点 $\tau_s, \tau_e \in S_*^{2d-3}$ 与采样参数 $\mathbf{t} \in (0, 1)^m$ 的优化。这一形式避免了直接在预形状空间或者 $[0, 1]$ 区间上进行带

约束的优化,使得整个拟合过程可直接套用标准无约束梯度优化器,便于集成到 PyTorch^[63] 等深度学习框架的自动求导流程中。

具体优化中,本文采用 Adam 优化器^[65]对 v_s, v_e 与 $\tilde{t}$ 进行迭代更新,共训练 E 轮。为灵活调控不同类型参数的更新速度以提升训练稳定性,端点参数 τ_s, τ_e 学习率设为 η_p , 采样参数 $\tilde{t}$ 学习率设为 η_t 。

基于上述优化流程,FAAGC 拟合算法的时间复杂度为 $\mathcal{O}(E \times m \times d)$,而 FAGC^[9] 的时间复杂度为 $\mathcal{O}(m^2 \times d + E' \times m \times S \times d)$,其中 E' 为 FAGC 流程中的优化轮数, S 对应 FAGC 中的超参数 steps。由于两种算法采取不同的优化流程, E 与 E' 无法直接比较,因此不能仅凭时间复杂度判定 FAAGC 快于 FAGC。但上述复杂度分析基于串行计算,未体现现代硬件对矩阵与张量运算的并行加速能力。因此,FAAGC 的实际效率优势需结合运行时间与能耗评估,相关结果见第 3.5.2 节。

3.3.3 损失函数的分布意义

为从分布角度解释第 3.3.1 小节与第 3.3.2 小节给出的两类损失,本小节将训练与增强两阶段放在同一概率框架下进行比较。记 $q(\tau)$ 为训练集合 $\mathcal{D}_{ps}$ 中特征的分布,也即特征增强希望逼近的目标分布。训练阶段学得参数集合 $\mathbf{t} = (t_1, \dots, t_m)^\top$ 自身形成一个分布,以 t 表示从该分布中任取的单个分量。增强阶段则从分布 $p(z) = U(0, 1)$ 中抽取单个采样参数 z 。据此, $p_{\tau_s, \tau_e}(\tau^* | t)$ 表示训练阶段以 t 经 f_{interp} 得到的增强特征分布, $p_{\tau_s, \tau_e}(\tau^* | z)$ 表示增强阶段以 z 经 f_{interp} 得到的增强特征分布。

在上述记号下,两类损失各自对应一项分布对齐约束。最小化 $\mathcal{L}_{\text{Sim}}$ 使条件分布 $p_{\tau_s, \tau_e}(\tau^* | t)$ 逼近目标分布 $q(\tau)$,从而使以训练阶段参数生成的特征尽可能接近目标分布。最小化 $\mathcal{L}_{\text{Diverg}}$ 使 $\mathbf{t}$ 的分布逼近 $p(z)$,也即 t 与 z 同分布。当两项约束同时成立时,条件 t 可被同分布的 z 替换,因而增强阶段实际生成的特征分布 $p_{\tau_s, \tau_e}(\tau^* | z)$ 同样逼近 $q(\tau)$,这便从分布层面保证了 FAAGC 在增强阶段的有效性。直观地说, $\mathcal{L}_{\text{Sim}}$ 负责让训练阶段插值得到的 τ^* 尽可能贴近真实训练特征, $\mathcal{L}_{\text{Diverg}}$ 则负责让训练阶段的参数分布与增强阶段所用的采样分布对齐,两者合在一起,等价于要求增强阶段生成的特征整体上与训练特征足够接近。

3.4 基于测地线的采样增强与数据可视化

如图 3.1 所示, 当针对初始样本得到最优的起点向量 τ_s 与终点向量 τ_e 后, 从区间 $(0, 1)$ 的均匀分布 $U(0, 1)$ 中采样参数集合 $\{z_1, \dots, z_n\}$ 。将这些采样参数代入插值函数 $f_{\text{interp}}(\tau_s, \tau_e, z)$ 中, 即可沿测地线生成增强样本, 从而获得增强后的预形状向量集合 $\{\tau_1^*, \dots, \tau_n^*\}$ 。与训练阶段不同, 数据增强阶段的采样过程允许根据需要任意指定采样规模 n 。对于每个类别, 通过上述算法生成 n 个增强向量, 并为其赋予与该类别特征相同的标签 c 。原始数据因而得到扩增, 增强数据集表示为

$$\mathcal{D}_{\text{aug}} = \{(\tau_i^*) \mid i = 1, \dots, n\}. \quad (3.13)$$

最后, 将预形状空间训练特征集合 $\mathcal{D}_{\text{ps}}$ (见第 3.2 节) 与增强数据集 $\mathcal{D}_{\text{aug}}$ 一同应用于下游任务。

为了直观展示 FAAGC 在特征增强任务中的有效性, 采用 t-SNE 数据降维方法^[62]对 CIFAR-10 数据集子集的特征进行降维与可视化。训练样本、使用 FAGC 算法进行增强的样本, 以及使用 FAAGC 进行增强的样本的可视化结果如图 3.3 所示。可以明显看到, 相较于 FAGC, FAAGC 能够更准确地拟合训练样本投影到预形状空间后的特征分布。

3.5 实验分析

为进一步验证 FAAGC 模块在数据受限场景下的有效性与效率, 本节首先介绍实验设置, 随后将 FAAGC 与 FAGC 在效率和性能上进行对比, 并对模块各组成部分进行消融分析。FAAGC 与更多特征级增强方法的系统对比见第四章第 4.5.2 节。

3.5.1 实验设置

在本节和下一章开展的实验 (第 4.5.1 节) 中使用到了以下视觉基准数据集:

- **CIFAR-10**^[4]: 包含 10 个类别的 50000 张训练图像和 10000 张测试图像 (分辨率为 32×32)。
- **CIFAR-100**^[4]: 包含 100 个类别的 50000 张训练图像和 10000 张测试图像 (分辨率为 32×32)。

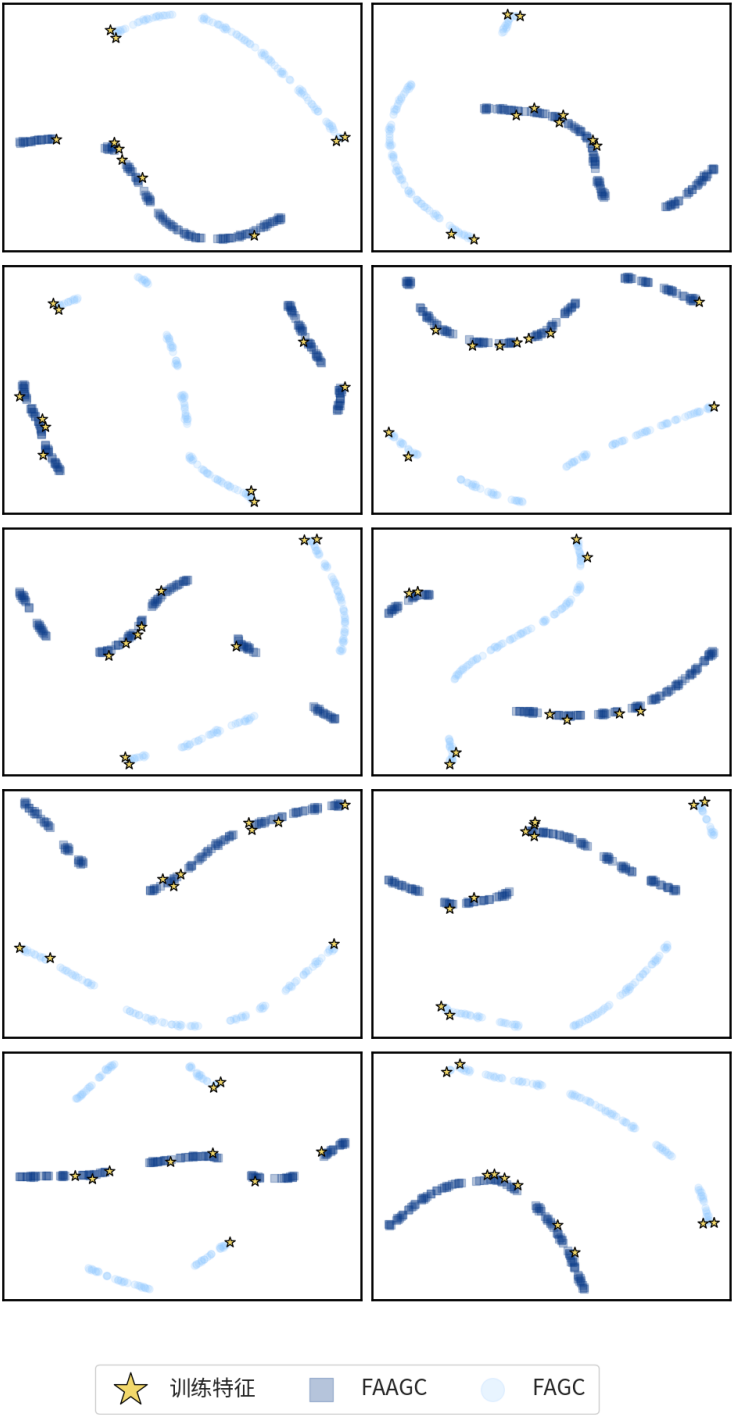


图 3.3 从 CIFAR-10 子集提取的预形状空间原始特征、FAGC 增强特征与 FAAGC 增强特征

- **CUB-200-2011**^[66]: 包含 200 个鸟类类别, 约 5900 张训练图像和约 5700 张测试图像, 图像分辨率各异 (下文简称 CUB-200)。
- **Fashion-MNIST**^[67]: 包含 10 个类别的 60000 张训练图像和 10000 张测试灰度图像 (分辨率为 28×28)。
- **Flowers102**^[68]: 包含 102 个花卉类别, 共 8189 张图像, 其中训练集、验证集和测试集分别包含 1020、1020 和 6149 张图像, 图像具有较大的尺度、姿态和光照变化。
- **Caltech101**^[69]: 包含 101 个类别的物体数据集, 约 9000 张图像, 带有手工标注 (分辨率约为 300×200)。

为了模拟数据受限的场景, 本文通过随机选取将每个数据集的训练集缩减为每类 m 个样本, 并记为“数据集名 @ m ”的形式, 同时保留完整的测试集用于评估。例如, 当每类仅选取 5 个样本时, 分别记为 CIFAR-10@5、CIFAR-100@5、CUB-200@5、Fashion-MNIST@5、Flowers102@5 和 Caltech101@5; 当每类选取 10 个样本时, 记作 CIFAR-10@10、CIFAR-100@10 等, 以此类推。特别地, 对于 Caltech101 数据集, 由于其未预先划分训练集与测试集, 本文首先从每类中随机选取 20 个样本组成测试集, 再从剩余样本中每类随机抽取 m 个样本作为训练集, 记为 Caltech101@ m 。

在骨干模型方面, 本文在大多数实验中使用来自 timm 库^[70]、输出特征维度为 $d = 192$ 的预训练 ViT-Tiny 来提取特征。输入图像被调整为 224×224 大小并进行了归一化处理。

在特征提取之前, 本文首先使用 Adam 优化器进行 50 个 epoch 的微调, 批次大小为 128。微调时采用差异化学习率, 将分类头的学习率设置为 1×10^{-3} , 骨干网络的学习率设置为 1×10^{-5} , 这是因为骨干网络参数来自 ImageNet 数据集预训练, 而分类头由于类别数不同导致形状不匹配, 采用随机初始化, 因此需要更大的学习率以完成收敛。

本文 FAAGC 实验可通过第 5 章所述的配置驱动实验平台执行。该平台统一完成数据加载、骨干网络微调、预形状空间投影、测地线拟合、分类评估以及结果记录, 从而保证不同实验配置下流程的一致性与可复现性。

在下游任务评估方面, 所有实验均使用完整的测试集进行分类准确率评估。FAGC^[9] 与本章提出的 FAAGC 均在预形状空间 S_*^{2d-3} 上进行特征增强, 测试集特征按第 3.2 节所述流程处理: 先由微调后的骨干模型提取特征向量, 再投影至 S_*^{2d-3} ,

并在该空间中完成分类评估。

为了公平地比较各方法的增强效果，所有特征增强方法在进入下游分类任务时，增强数据集 $\mathcal{D}_{\text{aug}}$ 的规模统一为每类 m 个增强特征，其中 m 为每类原始样本数。

分类器训练时，训练特征与增强特征共同参与监督学习。为平衡二者对分类器优化的影响，本文采用加权分类损失：

$$\mathcal{L}_{\text{cls}} = \mathcal{L}_{\text{CE}}(p, y) + \lambda_{\text{aug}} \mathcal{L}_{\text{CE}}(\tilde{p}, \tilde{y}), \quad (3.14)$$

其中， p 和 $\tilde{p}$ 分别表示分类器对训练特征与增强特征输出的类别概率分布， y 和 $\tilde{y}$ 分别表示对应的 One-Hot 标签； $\mathcal{L}_{\text{CE}}$ 为第 2.1 节定义的交叉熵损失， λ_{aug} 用于控制增强特征损失项的权重。本文实验中取 $\lambda_{\text{aug}} = 0.5$ 。

对于 FAAGC 模块，测地线拟合的超参数设置如下：总损失函数 (3.12) 中的平衡系数 $\beta = 0.3$ ；使用 Adam 优化器迭代 $E = 100$ 个 epoch；端点参数学习率 $\eta_p = 3 \times 10^{-4}$ ，采样参数学习率 $\eta_t = 3 \times 10^{-3}$ 。上述超参数在所有 FAAGC 对比实验中保持一致。

3.5.2 测地线增强算法的运行时间与准确率对比实验

与 FAGC^[9] 类似，FAAGC 同样将特征投影至预形状空间，在该空间中构建测地线并沿测地线采样以生成增强特征。因此，FAAGC 可以视为对 FAGC 的改进。具体而言，两者在最优端点向量的求解方式上存在差异：FAGC 通过迭代计算测地距离来获得端点，而 FAAGC 通过设计损失函数并采用梯度下降策略直接优化端点向量。该设计能够显著提升数据增强阶段的计算效率，并带来更好的增强性能。

由第 3.3 节的分析可知，FAGC 和 FAAGC 算法的串行时间复杂度不能直观体现两者计算代价的差异，因此在本节中，为对比两种方法的运行时间差异，在 CIFAR-10@5 与 CIFAR-10@10 两个子集上进行测地线拟合、样本增强以及增强数据集分类准确率评估实验。原始样本由 ViT-Tiny 模型提取特征表示。随后分别应用 FAGC 与 FAAGC 进行特征增强，并记录两种方法在同一硬件环境下的训练耗时。增强后使用分类头评估分类准确率。除了实际运行时间，还利用 CodeCarbon 库^[61] 计算两种算法在运行时 CPU 和 RAM 的能耗总和。两种方法的特征增强步骤均在 Intel(R) Xeon(R) Silver 4210R CPU 上完成。

从表 3.1 可以看出，FAAGC 在特征增强后取得更高的分类准确率，同时显著降低了计算耗时与能耗。这主要得益于 FAAGC 在测地线拟合过程中采用了更高效的

表 3.1 FAGC 与 FAAGC 的训练时间与分类准确率对比

方法	每类样本数 m	训练时间 (s)	算法运行能耗 (kWh)	准确率 (%)
FAGC	5	292.73	7.75×10^{-3}	85.20
FAAGC	5	2.23	5.90×10^{-5}	85.84
FAGC	10	665.45	1.51×10^{-2}	86.30
FAAGC	10	9.12	2.40×10^{-4}	88.34

端点优化策略与损失计算方式（见第 3.3 节）。相比之下，FAGC 依赖大量点采样与测地距离计算等计算开销较大的操作，从而导致整体运行时间更长。表 3.1 中的能耗计算仅覆盖了测地线拟合与增强采样阶段，而未将特征提取阶段纳入统计。这是因为在上述实验中，FAGC 与 FAAGC 均对相同的原始训练集进行特征提取，操作步骤与计算量完全一致，因此特征提取阶段的能耗差异可以忽略。

3.5.3 测地线拟合增强模块消融实验

为考察测地线拟合增强 (FAAGC) 模块在微调阶段与不同图像增强策略得到的微调模型的兼容性，在 CIFAR-100@5 数据集上开展特征增强和图像分类实验，在微调过程中分别采用如下五种图像增强设置：（1）不使用图像增强，仅进行图像缩放与归一化处理，该设置与前述实验保持一致；（2）采用 ViT 默认的数据增强策略，在缩放与归一化基础上额外加入中心裁剪与随机水平翻转，记作 DefaultAug；（3）在设置（2）的基础上，引入 RandAugment^[14] 进行微调阶段的数据增强；（4）在设置（2）的基础上，引入 AugMix^[15] 风格的数据增强策略；（5）在设置（2）的基础上，引入 DiffuseMix^[25] 风格的数据增强。

上述不同图像增强设置所提取的特征，分别直接输入分类器，或先应用 FAAGC 进行特征增强后再送入分类器进行训练评估。

如图 3.4 所示，在数据受限条件下，无论微调时是否采用图像级增强，FAAGC 模块均能够对特征进行有效增强并提升分类性能。进一步地，当 FAAGC 处理经过有效图像增强微调后的模型提取特征时，分类准确率还可得到进一步提升，表明 FAAGC 与各类图像增强方法得到的微调模型具有良好的兼容性。

为评估 FAAGC 流程中各阶段处理步骤的必要性，本文在 CIFAR-100@5 上进行消融实验。使用 ViT-Tiny 提取输入特征，并比较三种策略：（1）不进行任何处理，直接在提取到的特征组成的训练集 $\mathcal{D}_{\text{train}}$ 上训练分类器并评估，作为基线；（2）加入预

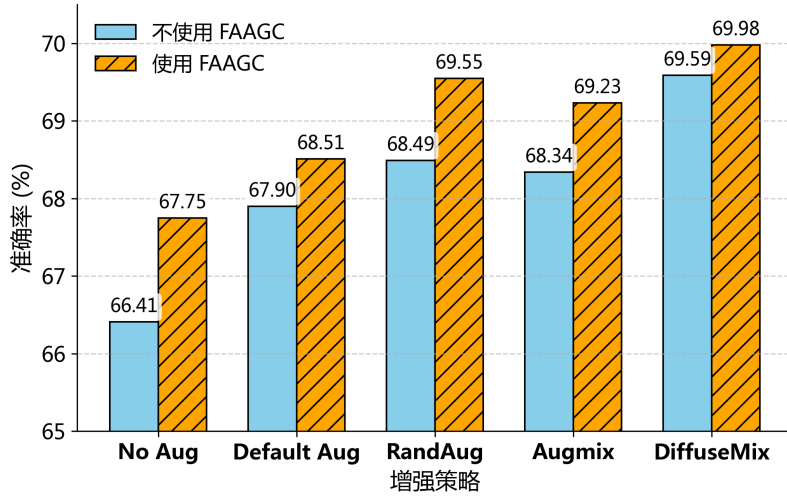


图 3.4 不同图像增强的微调策略下 FAAGC 带来的准确率提升

形状空间投影，仅使用投影后的预形状空间数据 $\mathcal{D}_{ps}$ 进行训练并评估，用以单独考察投影步骤的贡献；（3）进一步加入测地线增强，即采用完整的 FAAGC 流程，在 $\mathcal{D}_{ps}$ 与增强集合 $\mathcal{D}_{aug}$ 上联合训练并评估，用以考察测地线增强相对于仅投影方案的额外提升。FAAGC 相关超参数与前述实验保持一致，并使用 ViT-Tiny 的分类头在测试集上进行评估。三种数据配置下的测试集分类准确率结果如表 3.2 所示：

表 3.2 FAAGC 各阶段增量带来的分类准确率提升

投影	测地线增强	准确率 (%)
否	否	66.71
是	否	67.44
是	是	68.15

表 3.2 表明，完整 FAAGC 流程取得最高准确率（68.15%），优于基线（66.71%）以及仅投影不增强的策略（67.44%）。本组消融实验中的基线结果可能与第 4 章对比实验略有差异，源于骨干网络每次微调输出特征的随机性（参数保持一致）。

针对 FAAGC 测地线拟合算法的端点参数 v_s, v_e 的初始化方法，本节设计了系统性对比实验。在固定 $\tilde{t}$ 采用 Logistic(0, 1) 初始化的前提下，比较三种 v_s, v_e 初始化策略：（1）随机两点初始化，不使用任何类内统计量，直接从类内训练特征集合 $\{v_1, \dots, v_m\}$ 中随机选取两个样本作为初始值，以此作为基线；（2）为 2-近似最远点对初始化，先计算类内特征均值 μ_v ，选取距均值最远的样本 $v_s = \arg \max_i \|v_i - \mu_v\|_2$ ，再取距 v_s 最远的样本 $v_e = \arg \max_j \|v_j - v_s\|_2$ ，所得端点为类内特征集合的近似直径

端点,使初始测地线尽可能跨越类内分布的主要展向;(3)为高斯随机初始化(即本文所采用的方法),计算类内特征的均值 μ_v 与各维标准差 σ_v ,从正态分布 $\mathcal{N}(\mu_v, \sigma_v^2)$ 中随机采样两个向量作为初始值,从而以类内整体统计分布为先验,使初始端点既贴合特征空间的位置与尺度,又保留一定随机性。为考察训练轮数对各初始化方法的影响,将 FAAGC 测地线拟合的训练轮数分别设为 100、200、500、1000 与 2000。每组实验独立运行 5 次,报告最佳验证准确率的均值与标准差。其余实验设置均与第 3.5.1 节相同,数据集为 CIFAR-100@5。实验结果如表 3.3 所示。

表 3.3 v_s, v_e 不同初始化方法在各训练轮数下的分类准确率

初始化方法	$E = 100$	$E = 200$	$E = 500$	$E = 1000$	$E = 2000$
随机两点	66.88±0.45	67.00±0.34	65.18±0.41	66.71±0.24	67.87±0.38
2-近似最远点对	67.26±0.13	67.15±0.27	67.33±0.30	67.31±0.11	67.94±0.18
高斯随机(本文方法)	67.75±0.20	67.57±0.33	67.48±0.19	67.78±0.30	67.76±0.17

由表 3.3 可见,三种初始化方法呈现出不同的收敛特性。随机两点初始化波动较大, $E = 500$ 时准确率降至 65.18%,这种不稳定的结果原因可能在于初始端点由随机抽取的样本决定,而不同的样本对可能把初始测地线引向较差的优化方向。2-近似最远点对初始化表现稳定,且在 $E = 2000$ 时达到最高准确率 67.94%,但其需要更多优化步数方能完成收敛,在低训练轮数下不及高斯随机初始化。高斯随机初始化在 $E = 100$ 时即达 67.75%,体现出最优的收敛速度与准确率平衡。高斯随机初始化以类内统计分布为先验,使初始端点与特征空间的整体结构相匹配,兼顾了收敛效率与稳定性,故本文采用该策略。

接下来进一步考察插值参数 $\tilde{t}$ 的初始化方法对最终分类性能的影响,对比两种策略:(1)高斯随机初始化,从标准正态分布 $\mathcal{N}(0, 1)$ 中随机采样得到 $\tilde{t}$,再经 Sigmoid 变换映射至 $(0, 1)$ 区间,此时 t 集中在 0.5 附近,与增强阶段的 $U(0, 1)$ 存在一定分布偏差;(2)Logistic 初始化(即本文所采用的方法),从 Logistic(0, 1) 分布中采样得到 $\tilde{t}$,再经 Sigmoid 映射得到 t ,由概率积分变换可知初始 t 恰好服从 $U(0, 1)$,从而在训练初始与增强阶段采样分布一致^[71]。实验中固定 v_s, v_e 采用高斯随机初始化,将 FAAGC 测地线拟合的训练轮数分别设为 100、200、500、1000 与 2000,每组实验独立运行 5 次,报告各组最佳验证准确率的均值与标准差。其余实验设置均与第 3.5.1 节相同,数据集为 CIFAR-100@5。实验结果如表 3.4 所示。

表 3.4 $\tilde{t}$ 不同初始化方法在各训练轮数下的最佳分类准确率

初始化方法	$E = 100$	$E = 200$	$E = 500$	$E = 1000$	$E = 2000$
高斯随机	67.82 ± 0.13	67.61 ± 0.15	67.50 ± 0.13	67.62 ± 0.20	67.69 ± 0.12
Logistic (本文方法)	67.75 ± 0.16	67.71 ± 0.19	67.73 ± 0.21	67.66 ± 0.08	67.65 ± 0.13

由表 3.4 可见, 两种初始化方法的准确率差距较小, 且随训练轮数增大差距进一步收窄, 表明 $\tilde{t}$ 的初始化选择对最终性能影响有限。

为进一步验证 FAAGC 对训练超参数的鲁棒性, 本研究在 CIFAR-100@5 与 CIFAR-100@10 两个基准上对端点学习率 η_p 与采样参数学习率 η_t 进行敏感性分析。两者均在同一候选集合 $\{10^{-4}, 3 \times 10^{-4}, 10^{-3}, 3 \times 10^{-3}, 10^{-2}, 3 \times 10^{-2}\}$ 内取值, 对 (η_p, η_t) 进行网格搜索, 每组的分类性能均通过 ViT 分类头在测试集上进行评估。本实验与其他实验的特征提取模型采用相同微调流程, 但模型轮次不同, 故准确率可能略有差异。

从图 3.5 的热力图可以看出, FAAGC 在两个基准上均对学习率设置表现出良好的鲁棒性: 在 CIFAR-100@5 上, 各学习率组合下的分类准确率波动不超过 0.5 个百分点, 在 CIFAR-100@10 上波动约为 0.4 个百分点, 且无明显的超参数敏感区域。在 CIFAR-100@10 上, 分类准确率随 η_p 的增大呈现出一定的上升趋势, 但整体波动仍在 0.4 个百分点以内, 并不构成显著的超参数依赖。这表明 FAAGC 对训练超参数的选择具有较强的适应性, 在不同样本规模下均能稳定发挥增强效果, 降低了实际部署中的调参成本。

3.6 本章小结

本章提出了基于形状空间理论的自适应测地线增强算法模块 FAAGC。该方法将深度特征投影至预形状空间, 以可学习的端点向量与采样参数构造测地线, 通过样本相似性损失与分布差异损失联合优化, 并在测地线上参数化采样以生成增强特征。

FAAGC 在三个层面共同保证了所生成增强样本的合理性。从几何层面看, 测地线插值公式确保所有增强特征严格位于预形状空间的测地线上; 进一步地, 从分布层面看, 第 3.3.3 节的分析表明, 相似性损失与分布差异损失的同时优化使增强阶段的条件特征分布 $p_{\tau_s, \tau_e}(\tau^* | z)$ 逼近原始类内分布 $q(\tau)$; 在此基础上, 从实际任务层面, 可视化与下游分类实验一致表明 FAAGC 较过去的测地线特征增强方法在样本覆盖、

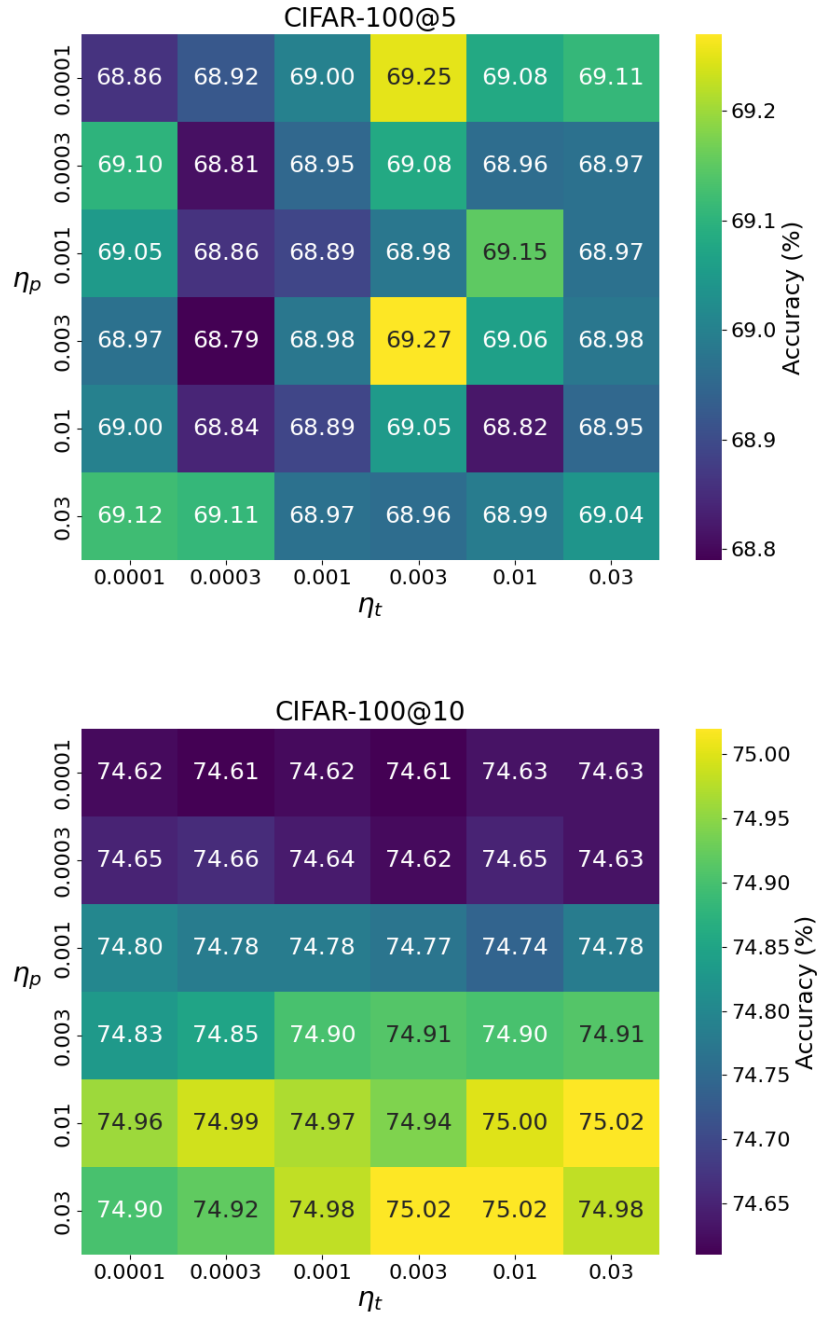


图 3.5 学习率 η_p 与 η_t 对分类准确率的影响

分类准确率、运行时间与能耗方面均具有优势，消融实验亦验证了各步骤的必要性，且整体方法对初始化策略与学习率取值表现出良好稳定性。

第四章 可扩展的测地线特征增强框架

上一章提出的 FAAGC 方法将特征投影至预形状空间，通过测地线拟合与采样生成增强特征，在样本稀缺场景下展现了良好的增强效果。然而，该方法在实际数据增强任务中存在两点局限：一是测地线拟合至少需要两个样本点才能运行，无法适用于单样本（One-shot）场景；二是第 3.3 节的损失函数依赖全样本矩阵运算，单次迭代计算量随样本数线性增长，难以扩展至较大规模的数据集。为此，本章提出可扩展测地线特征增强（Scalable Feature Augmentation on Adaptive Geodesic Curve, SFAAGC）框架，从特征扩充与加权小批次梯度下降两方面对 FAAGC 进行改进，并将其融入第三章的特征增强流程，使测地线特征增强方法能够适用于从单样本到大规模数据集的各种场景。本章依次介绍框架的总体设计、特征扩充方法、改进的测地线拟合算法、完整的算法流程，以及增强样本的可视化分析，并通过实验分析验证该框架在不同数据规模与不同骨干网络下的有效性与稳定性。

4.1 总体框架

SFAAGC 框架的整体结构如图 4.1 所示，主要由以下两个部分组成：

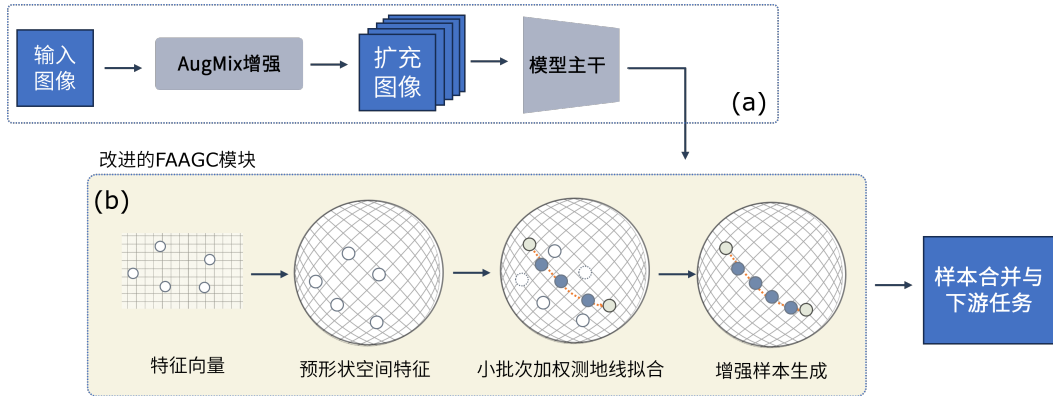


图 4.1 可扩展测地线特征增强框架总体结构

(1) 基于图像增强方法的特征提取与特征扩充模块（对应图 4.1 (a)）。该模块负责在输入图像层面通过图像增强技术生成多样化的扩充图像，随后经骨干网络提取特征向量，从而在特征空间中构建包含原始样本与增强样本的扩充训练集 $\mathcal{D}_{\text{train}}$ 。该模块解决了单样本场景下测地线拟合算法无法初始化的问题，同时为后续拟合提供

了更密集的预形状空间样本。

(2) **改进的测地线拟合算法** (对应图 4.1 (b) 中的 FAAGC 模块)。该模块在第 3 章 FAAGC 算法的基础上, 通过引入小批次优化和样本来源加权机制, 降低单次迭代对总样本规模的依赖, 并调节原始样本与增强样本在测地线拟合中的贡献。整体流程包括预形状空间投影、小批次加权测地线拟合和测地线采样增强特征生成。

通过以上两个模块的协同工作, 使用者能够根据自身计算资源调整计算量, 使基于 FAAGC 的数据增强在不同规模数据集上均能取得良好效果。

4.2 基于图像增强的特征扩充方法

本研究方法先在输入层面对图像进行增强以扩充数据集, 再经深度学习模型提取特征并投影至预形状空间, 最后由 FAAGC 完成测地线拟合, 如图 4.1 (a) 所示。实验中采用 AugMix^[15] 作为增强方法。

测地线拟合算法的预形状空间投影步骤 (见第 3.2 节) 中的特征训练集 $\mathcal{D}_{\text{train}}$ 与预形状空间投影特征训练集 $\mathcal{D}_{\text{ps}}$ 定义为

$$\mathcal{D}_{\text{train}} = \{(v_i, a_i) \mid i = 1, \dots, m'\}, \quad (4.1)$$

$$\mathcal{D}_{\text{ps}} = \{(\tau_i, a_i) \mid i = 1, \dots, m'\}. \quad (4.2)$$

其中, 当 $a_i = 0$ 时, 表示原始样本; 当 $a_i = 1$ 时, 表示图像增强扩充样本, m' 表示扩充后的总样本数量。与第三章一致, 由于后续增强操作均针对类内样本特征展开, 本章公式省略了类别标记 c 。

4.3 基于样本加权与小批次梯度下降的测地线拟合算法

本工作在第 3.3.2 节 FAAGC 测地线拟合目标的基础上进行两点改进: 首先, 将原方法中基于全部预形状空间样本计算 $\mathcal{L}_{\text{Sim}}$ 、 $\mathcal{L}_{\text{Diverg}}$ 并更新 v_s, v_e 与 $\tilde{t}$ 的全量梯度下降, 改为小批次梯度下降; 其次, 引入基于样本来源的加权机制, 以区分原始样本与图像增强样本对测地线拟合的贡献。

与第 3.3.1 节和第 3.3.2 节所述过程一致，首先初始化可学习端点参数 v_s, v_e 与采样参数 $\tilde{t}$ ，共迭代 E 轮，通过最小化样本相似性损失 $\mathcal{L}_{\text{Sim}}$ 与分布差异损失 $\mathcal{L}_{\text{Diverg}}$ 来优化上述参数；端点参数与采样参数分别以学习率 η_p 和 η_t 进行更新。

在每次迭代中，从当前类别的预形状空间样本中选取 B 个样本组成一个小批次，记其下标集合为 $I = \{i_1, \dots, i_B\}$ ，对应的批次特征矩阵为 $\boldsymbol{\tau}_{\text{batch}} = [\tau_{i_1}, \dots, \tau_{i_B}]$ 。为区分原始样本与图像增强样本，本文为批次内每个样本设置来源权重 λ_{i_j} ：若样本为原始样本，即 $a_{i_j} = 0$ ，则令 $\lambda_{i_j} = 1$ ；若样本为增强样本，即 $a_{i_j} = 1$ ，则令 $\lambda_{i_j} = \lambda_{os}$ 。其中， λ_{os} 为增强样本权重超参数，本文默认取 0.6。

在进行测地线采样步骤时，记批次采样参数为 $\mathbf{t}_{\text{batch}} = (t_{i_1}, t_{i_2}, \dots, t_{i_B})^\top$ ，其中 $t_{i_j} = \text{Sigmoid}(\tilde{t}_{i_j})$ 。使用该批次采样参数进行插值，得到在测地线上的插值向量 $\boldsymbol{\tau}_{\text{batch}}^* = [\tau_{i_1}^*, \dots, \tau_{i_B}^*]$ 。

随后，在计算相似性损失函数 $\mathcal{L}_{\text{Sim}}$ 时，按样本来源进行加权，以控制不同来源样本对测地线的影响：

$$\mathcal{L}_{\text{Sim}} = \frac{1}{B} \sum_{j=1}^B \lambda_{i_j} (1 - \tau_{i_j}^{*\top} \tau_{i_j})^2. \quad (4.3)$$

与第 3.3.2 节同理，上式可改写为便于并行执行的矩阵形式。记批次权重向量为 $\boldsymbol{\lambda}_{\text{batch}} = (\lambda_{i_1}, \dots, \lambda_{i_B})^\top$ ，首先计算

$$\mathbf{r}_{\text{batch}} = \mathbf{1}_B - (\boldsymbol{\tau}_{\text{batch}}^* \odot \boldsymbol{\tau}_{\text{batch}})^\top \mathbf{1}_{2d}, \quad (4.4)$$

则相似性损失可表示为

$$\mathcal{L}_{\text{Sim}} = \frac{1}{B} \boldsymbol{\lambda}_{\text{batch}}^\top (\mathbf{r}_{\text{batch}} \odot \mathbf{r}_{\text{batch}}). \quad (4.5)$$

在计算分布差异损失时，使用批次权重的均值 $\bar{\lambda} = \frac{1}{B} \sum_{j=1}^B \lambda_{i_j}$ 对分布距离进行缩放。对 $\mathbf{t}_{\text{batch}}$ 与从均匀分布采样得到的 $\mathbf{z} = (z_1, \dots, z_B)^\top$ 分别升序排序，得到排序后的采样参数 $t_{(j)}$ 和排序后的均匀分布采样 $z_{(j)}$ ，计算：

$$\mathcal{L}_{\text{Diverg}} = \bar{\lambda} \cdot \frac{1}{B} \sum_{j=1}^B |t_{(j)} - z_{(j)}|. \quad (4.6)$$

与第 3.3.2 节一致，总损失为 $\mathcal{L} = \mathcal{L}_{\text{Sim}} + \beta \cdot \mathcal{L}_{\text{Diverg}}$ ，在每个小批次上执行梯度下

降。当不进行特征扩充 ($m' = m$) 且 $B = m$ 时, 算法退化为全量梯度下降, 其优化目标与第 3.3 节所述方法等价。本文取 $B = 512$, 实际可按硬件调整。由于 $\mathcal{L}_{\text{Sim}}$ 引入了 $\frac{1}{B}$ 归一化系数, 其值域不再随样本数增长, 而 $\mathcal{L}_{\text{Diverg}}$ 的量级基本不变, 两项损失的相对大小较第三章发生变化, 因此平衡系数 β 需相应减小 (见第 4.5.1 节)。

在完成测地线拟合后, 与第 3.3 节方法一致, 按照增强需要从 $U(0, 1)$ 中采样 n 个采样参数 $\{z_1, \dots, z_n\}$, 通过插值公式 3.5 沿测地线生成增强特征, 得到增强数据集 $\mathcal{D}_{\text{aug}} = \{(\tau_i^* \mid i = 1, \dots, n)\}$ 。最后将预形状空间训练特征集合 $\mathcal{D}_{\text{ps}}$ 与增强数据集 $\mathcal{D}_{\text{aug}}$ 一同应用于下游任务。

在样本数量较多时, 小批次梯度下降相当于每次仅随机选取部分预形状空间向量参与拟合, 因此算法的单次计算量得到显著降低; 在样本数量较少时, 算法能够有效利用本章第 4.2 小节中的特征扩充技术提升增强效果。

综合以上改进, SFAAGC 框架的完整算法流程如算法 4.1 所示:

4.4 增强样本数据可视化

为了直观地展示 SFAAGC 框架的增强效果, 本节采用 t-SNE 数据降维方法^[62]将高维特征空间中的样本投影到二维平面上, 对增强前后的特征分布进行可视化分析。

本文基于 CIFAR-10 数据集的训练集子集, 对各类别分别进行了 t-SNE 可视化。对于各个类别, 展示以下四类样本的特征分布: (1) CIFAR-10 子集的训练集样本; (2) CIFAR-10 测试集样本; (3) 第三章 FAAGC 算法基于 CIFAR-10 子集训练特征生成的测地线增强特征; (4) 本章 SFAAGC 框架基于 CIFAR-10 子集训练样本生成的测地线增强特征。本文选取两个具有代表性的类别可视化结果进行展示, 如图 4.2 所示。

通过可视化结果可以观察到: 在图 4.2(a) 所展示的类别中, SFAAGC 增强样本在特征空间中的覆盖范围明显优于 FAAGC, 不仅更密集地分布在训练样本附近, 也在测试集样本所在区域形成了较好的覆盖, 表明特征扩充使拟合测地线更准确地刻画了类内分布结构。在图 4.2(b) 所展示的类别中, 尽管 SFAAGC 在部分训练样本区域的覆盖密度略低于 FAAGC, 但其与测试集样本的重叠程度明显更高, 说明 SFAAGC 生成的增强特征具有更好的泛化能力。综合来看, SFAAGC 框架使增强样本在特征空间中兼具更高的多样性与更优的泛化覆盖, 验证了本章所提改进方法的有效性。

算法 4.1: 基于特征扩充与加权小批次梯度下降的测地线特征增强框架

输入: 图像集合 $\mathcal{I}_c$ (类别 c 的所有图像), 骨干网络 M , 投影算子 P , 批次大小 B , 迭代轮数 E , 学习率 η_p, η_t , 平衡系数 β , 扩充权重 λ_{os} , 扩充倍数 n_{os} , 增强数量 n

输出: 增强后的预形状特征集合 $\mathcal{D}_{aug}$

// 模块 (a): 特征提取与特征扩充

$\mathcal{D}_{train} \leftarrow \emptyset$

foreach $x \in \mathcal{I}_c$ **do**

$v \leftarrow M(x)$

$\mathcal{D}_{train} \leftarrow \mathcal{D}_{train} \cup \{(v, 0)\}$ // $a = 0$: 原始样本

$\mathcal{D}_{train} \leftarrow \mathcal{D}_{train} \cup \{(M(\text{AugMix}(x)), 1)\}_{k=1}^{n_{os}}$ // $a = 1$: 增强样本 (共 n_{os} 份)

end

// 模块 (b) 步骤一: 预形状空间投影

$\mathcal{D}_{ps} \leftarrow \emptyset$

foreach $(v_i, a_i) \in \mathcal{D}_{train}$ **do**

$\tau_i \leftarrow P(v_i)$

$\mathcal{D}_{ps} \leftarrow \mathcal{D}_{ps} \cup \{(\tau_i, a_i)\}$

end

$m' \leftarrow |\mathcal{D}_{ps}|$

// 模块 (b) 步骤二: 小批次加权测地线拟合

$\mu_v \leftarrow \frac{1}{m'} \sum_{i=1}^{m'} v_i, \quad \sigma_v \leftarrow \text{std}(\{v_1, \dots, v_{m'}\})$

$v_s \sim \mathcal{N}(\mu_v, \sigma_v^2), \quad v_e \sim \mathcal{N}(\mu_v, \sigma_v^2)$

$\tilde{t} \sim \text{Logistic}(0, 1)$ with size m'

for epoch $\leftarrow 1$ **to** E **do**

Randomly partition $\{1, \dots, m'\}$ into mini-batches of size B

foreach mini-batch index set $I = \{i_1, \dots, i_B\}$ **do**

$\tau_{batch} \leftarrow [\tau_{i_1}, \dots, \tau_{i_B}]$

$\lambda_{batch} \leftarrow (\lambda_{i_1}, \dots, \lambda_{i_B})^\top$, where $\lambda_{i_j} = 1$ if $a_{i_j} = 0$ else λ_{os}

$\tilde{t}_{batch} \leftarrow (\tilde{t}_{i_1}, \dots, \tilde{t}_{i_B})^\top$

$t_{batch} \leftarrow \text{Sigmoid}(\tilde{t}_{batch})$

$\tau_s \leftarrow P(v_s), \quad \tau_e \leftarrow P(v_e)$

$\tau_{batch}^* \leftarrow f_{\text{interp}}(\tau_s, \tau_e, t_{batch})$

$\mathbf{r}_{batch} \leftarrow \mathbf{1}_B - (\tau_{batch}^* \odot \tau_{batch})^\top \mathbf{1}_{2d}$

$\mathcal{L}_{\text{Sim}} \leftarrow \frac{1}{B} \lambda_{batch}^\top (\mathbf{r}_{batch} \odot \mathbf{r}_{batch})$

$\bar{\lambda} \leftarrow \frac{1}{B} \lambda_{batch}^\top \mathbf{1}_B$

Sample $\mathbf{z} \in \mathbb{R}^B$ from $U(0, 1)$

Sort t_{batch} and $\mathbf{z}$ in ascending order; denote sorted entries as $t_{(j)}, z_{(j)}$

$\mathcal{L}_{\text{Diverg}} \leftarrow \bar{\lambda} \cdot \frac{1}{B} \sum_{j=1}^B |t_{(j)} - z_{(j)}|$

$\mathcal{L} \leftarrow \mathcal{L}_{\text{Sim}} + \beta \cdot \mathcal{L}_{\text{Diverg}}$

Update v_s, v_e with lr η_p and $\tilde{t}_{batch}$ with lr η_t by gradient descent on $\mathcal{L}$

end

end

// 模块 (b) 步骤三: 测地线采样生成增强特征

$\tau_s \leftarrow P(v_s), \quad \tau_e \leftarrow P(v_e)$

$\mathcal{D}_{aug} \leftarrow \emptyset$

Sample $\{z_1, \dots, z_n\}$ from $U(0, 1)$

for $k \leftarrow 1$ **to** n **do**

$\tau_k^* \leftarrow f_{\text{interp}}(\tau_s, \tau_e, z_k)$

$\mathcal{D}_{aug} \leftarrow \mathcal{D}_{aug} \cup \{(\tau_k^*)\}$

end

return $\mathcal{D}_{aug}$

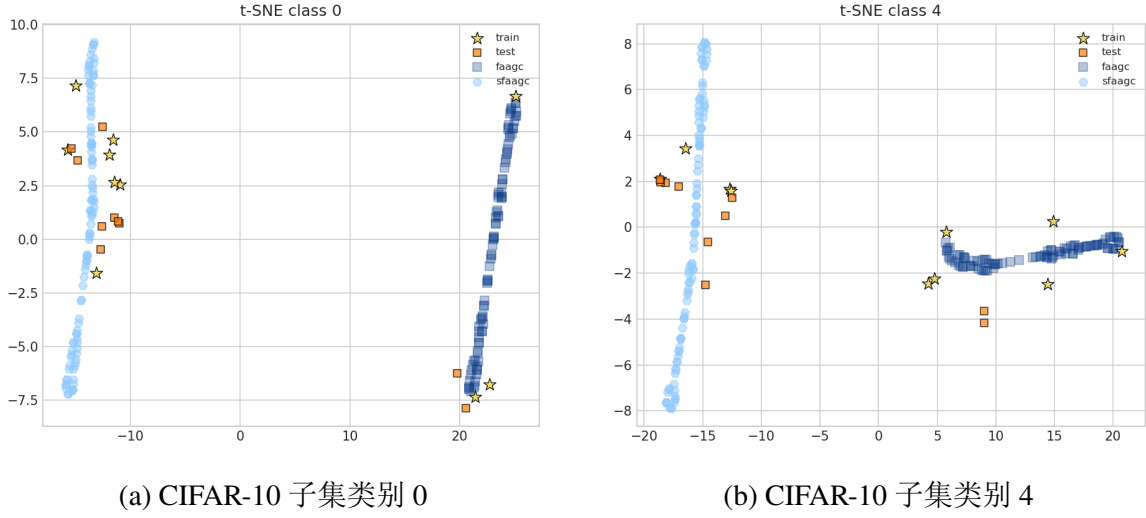


图 4.2 CIFAR-10 两个代表类别的增强样本 t-SNE 可视化

4.5 实验分析

本节在多个数据集上,从对比实验、效率实验和框架消融实验等角度,对 SFAAGC 框架的实际效果进行系统分析。

4.5.1 实验设置

本章实验沿用第 3.5.1 节的数据集划分方法,将训练集减少到每类 m 样本的数据集记作 $\text{dataset}@m$, 并采用相同的骨干模型与特征提取设置。实验可通过第 5 章所述的配置驱动实验平台执行,从而保证不同实验配置下流程的一致性与可复现性。

在下游任务评估方面,所有实验均使用完整的测试集进行分类准确率评估。对于在预形状空间上进行增强的方法(包括 FAGC^[9]、第 3 章提出的 FAAGC 与本章提出的 SFAAGC 框架),测试集特征按照第 3.2 节所述流程处理:先由微调后的骨干模型提取特征向量,再投影至 S_*^{2d-3} ,并在该空间中完成分类评估。对于其他特征增强方法(FGSM、Manifold Mixup、SFA-S、MoEx、Feature-level SMOTE 与 FroFA)以及基于图像增强的对比方法(DefaultAug、AugMix、RandAugment 与 DiffuseMix),测试集评估基于同一骨干模型提取的特征向量 $v \in \mathbb{R}^d$,直接在欧氏特征空间中完成。

与第 3.5.1 节不同的是, $\mathcal{D}_{\text{aug}}$ 规模统一为每类 $9 \times m$, 其中 m 为每类原始样本数。对于 SFAAGC 框架,由于在特征扩充阶段已经生成了 $4 \times m$ 个扩充特征,测地线增强阶段只需额外生成 $n = 5 \times m$ 个增强特征,两者合计同样为 $9 \times m$ 。关于图像扩充倍数 n_{os} 与测地线增强样本数量 n 对最终分类性能的影响,见表 4.8 所示的消融实

验。

对于 SFAAGC 框架，各阶段的超参数设置如下。在特征扩充阶段，默认采用 AugMix 方法生成增强图像，扩充倍数 $n_{os} = 4$ ，即扩充后每类样本总数为 $(1 + n_{os}) \times m = 5 \times m$ ，在测地线拟合阶段（见第 4.3 节），使用 Adam 优化器迭代 100 个 epoch；端点参数学习率 $\eta_p = 3 \times 10^{-4}$ ，采样参数学习率 $\eta_t = 3 \times 10^{-3}$ ；扩充权重 $\lambda_{os} = 0.6$ ，平衡系数 $\beta = 0.1$ 。

4.5.2 对比分析

为验证 SFAAGC 框架在数据受限场景下的有效性，本节在 CIFAR-10@5、CIFAR-100@5、CUB-200@5、Fashion-MNIST@5、Flowers102@5 与 Caltech101@5 六个数据受限子集上比较不同增强方法的分类表现。下游分类器训练损失沿用第 3.5.1 节，采用式 3.14 所示的加权分类损失。对于 SFAAGC 框架，增强特征包括扩充特征与测地线增强特征，二者均作为增强特征以权重 λ_{aug} 参与分类器损失计算。

参与对比的方法包括 FAAGC 特征增强模块、可扩展测地线特征增强（SFAAGC）框架，以及 FGSM^[28]、Manifold Mixup^[31]、SFA-S^[32]、MoEx^[33]、Feature-level SMOTE^[38]、FroFA^[39]（采用论文中的 cFroFA Brightness）和 FAGC^[9]。此外，还引入四种基于图像增强的对比方法：ViT 默认采用的中心裁剪与随机水平翻转（记作 DefaultAug）、AugMix^[15]、RandAugment^[14] 以及 DiffuseMix^[25]。对于这四种图像增强方法，先对训练图像施加相应增强策略，再通过微调后的骨干网络提取特征，并以每类 $9 \times m$ 个增强特征的规模参与分类器训练。除非另有说明，后续实验均使用 6 个随机种子运行，并报告平均性能。

如表 4.1 所示，以 FAAGC 与 SFAAGC 为代表的测地线增强方法在整体上取得了最优的分类表现。下面分别从不同类型方法的角度进行分析。

在特征增强基线方法中，FGSM、Manifold Mixup、SFA-S、MoEx 与 Feature-level SMOTE 五种方法的分类准确率与无增强基线相比差异甚微，在大多数数据集上的提升幅度不超过 0.3 个百分点，表明这些方法在每类仅 5 个样本的极端数据受限场景下，难以为特征空间引入足够有效的多样性。

FroFA 则在所有数据集上均显著低于无增强基线，其原因可能在于 FroFA 依赖 ViT 模型的多头注意力池化模块，而该模块需要在较大数据集上进行预训练才能有较好效果；其余增强方法均直接以 ViT 输出的首个 token（即 [CLS] token）作为特征

表 4.1 不同特征增强方法在各数据集上的分类准确率

方法	CIFAR-10@5	CIFAR-100@5	CUB-200@5	Fashion-MNIST@5	Flowers102@5	Caltech101@5
无增强	84.82	66.41	70.70	74.74	96.80	87.28
FGSM	84.84	66.33	70.62	74.52	96.89	87.97
Manifold Mixup	83.25	66.63	70.64	74.72	96.83	87.82
SFA-S	84.84	65.26	70.04	74.81	96.83	87.08
MoEx	84.86	66.25	70.49	75.00	96.85	88.02
Feature-level SMOTE	84.81	66.41	70.52	74.99	96.81	87.32
FroFA	65.07	51.17	53.99	66.39	93.46	73.83
DefaultAug	84.65	66.32	70.76	75.00	97.11	87.30
AugMix	79.73	65.51	70.64	75.05	96.68	87.41
RandAugment	79.51	65.90	70.26	74.74	97.02	87.06
DiffuseMix	77.68	63.70	69.11	73.97	96.40	86.77
FAGC	85.20	66.39	70.78	75.43	96.84	88.72
FAAGC	85.92	68.01	71.23	75.96	96.96	88.12
SFAAGC	88.90	74.29	71.93	76.30	97.32	89.61

向量，无需额外的池化模块。由于无法获取 FroFA 所需的预训练数据集，本研究未进行该预训练。

在四种基于图像增强的对比方法中，DefaultAug、AugMix 与 RandAugment 在 CUB-200@5、Fashion-MNIST@5、Flowers102@5 和 Caltech101@5 数据集上令部分数据集得到更好的准确率，例如 DefaultAug 在 Flowers102@5 上取得了 97.11% 的较好成绩；然而，这三种方法在两个 CIFAR 数据集上反而导致准确率下降，尤其是 AugMix 和 RandAugment 在 CIFAR-10@5 上分别降至 79.73% 和 79.51%，相较无增强基线有明显准确率下降。这表明将图像级语义变换转换为特征增强并不总能带来语义的丰富化和下游任务表现的提升。DiffuseMix 在所有数据集上的准确率均低于无增强基线，表明基于扩散模型的图像增强方法在样本极度稀缺时反而引入了较大的噪声干扰。

FAAGC 在大部分数据集上取得了次优性能（仅次于 SFAAGC），表明基于梯度下降的测地线拟合方法在数据受限场景下是可行且高效的。

对于 SFAAGC 框架，其在全部六个数据集上均取得最高分类准确率，分别为 88.90%、74.29%、71.93%、76.30%、97.32% 和 89.61%。其中，在 CIFAR-10@5 和 CIFAR-100@5 上较 FAAGC 分别提升 3.06 和 6.42 个百分点。

为考察骨干网络差异对增强效果的影响，本研究在 CIFAR-100@5 上使用不同类型和规模的骨干网络提取特征，并在相同设置下比较多种特征增强方法。具体考虑的骨干网络包括 ResNet-50^[51]、EfficientNet-B0^[72]、ViT-Tiny 与 ViT-Base^[54]，以及

Swin Transformer-Tiny 与 Swin Transformer-Base^[73]。其中, ResNet-50 与 EfficientNet-B0 在 ImageNet-1k 上预训练, ViT-Tiny 与 ViT-Base、Swin Transformer-Tiny 与 Swin Transformer-Base 均在 ImageNet-21k 上预训练。除 FAAGC 与 SFAAGC 外, 本研究同时评估 FGSM、Manifold Mixup、SFA-S、MoEx、Feature-level SMOTE 与 FAGC, 所有方法均作用于相应骨干网络提取的特征并采用一致的分类评估方式。

表 4.2 不同骨干网络下各特征增强方法的分类准确率

方法	ResNet-50	EfficientNet-B0	ViT-Tiny	ViT-Base	Swin-Tiny	Swin-Base
无增强	42.07	38.28	66.41	79.37	73.01	82.13
FGSM	42.09	38.38	66.33	79.44	73.11	82.08
Manifold Mixup	42.07	38.22	66.63	79.31	73.34	82.05
SFA-S	42.02	38.36	65.26	80.07	68.01	81.90
MoEx	41.94	38.13	66.25	79.38	73.03	81.89
Feature-level SMOTE	41.96	38.27	66.41	79.29	72.91	82.08
FAGC	41.89	38.59	66.39	80.14	73.17	82.18
FAAGC	41.99	38.54	68.01	80.36	73.26	82.29
SFAAGC	42.05	38.96	74.29	82.90	76.85	83.50

由表 4.2 可以观察到以下几点。首先, 对于在较小数据集 ImageNet-1k 上预训练的模型 (ResNet-50 与 EfficientNet-B0), 由于其预训练数据集规模有限, 模型的特征提取能力相对较弱, 因而无增强基线准确率明显低于其他骨干网络 (分别为 42.07% 与 38.28%)。同时本研究注意到, 在这两个骨干网络上, 表 4.2 所列全部特征增强方法均未能带来显著的性能提升, 准确率与无增强基线相比差异甚微。这表明当骨干网络本身的特征表达能力不足时, 仅在特征层面进行增强难以弥补底层表征质量的缺陷。

其次, 对于在更大规模数据集上预训练的模型 (ViT 系列与 Swin Transformer 系列均在 ImageNet-21k 上预训练), 其无增强基线准确率显著更高 (ViT-Tiny 为 66.41%, ViT-Base 为 79.37%, Swin-Tiny 为 73.01%, Swin-Base 为 82.13%)。并且, 经过 FAAGC 模块或 SFAAGC 框架进行特征增强后, 分类准确率获得了显著提升。

上述结果表明, 在训练样本受限的场景下, 应优先选择在较大规模数据集上进行预训练的图像模型作为特征提取器。这类模型不仅能够提供更高质量的基础特征表示, 还能与 FAAGC 及 SFAAGC 等测地线特征增强方法产生更好的协同效果, 从而最大化特征增强带来的性能收益。

进一步地, 本研究考察 FAAGC 与 SFAAGC 在不同训练集规模 m 下的数据增强效果。对于 CIFAR-100 与 Fashion-MNIST, 本研究从原始训练集中抽取每类 $m \in \{1, 3, 10, 30, 100\}$ 个样本构成子集; 对于 Flowers102, 因样本数量限制, $m \in \{1, 3, 10\}$ 。若方法无法在单样本条件下完成增强, 则结果记为 N/A。为排除骨干网络微调过程中随机性对结果的干扰, 本实验直接使用预训练 ViT-Tiny 提取特征, 不进行微调。

表 4.3 不同每类训练样本数下 FAAGC 与 SFAAGC 的分类准确率

m	1	3	10	30	100
CIFAR-100	34.26	53.94	64.89	70.44	72.61
+FAAGC	N/A	56.27	66.30	71.43	72.64
+SFAAGC	39.86	57.85	67.96	73.27	74.79
Fashion-MNIST	56.95	71.04	76.10	80.86	82.92
+FAAGC	N/A	72.90	77.91	81.67	82.94
+SFAAGC	58.38	73.71	79.06	82.11	83.55
Flowers102	82.76	93.53	97.85	—	—
+FAAGC	N/A	93.84	97.58	—	—
+SFAAGC	83.72	94.15	97.97	—	—

从表 4.3 可以看到, FAAGC 在单样本条件 ($m = 1$) 下无法运行, 因为测地线拟合至少需要两个样本作为端点初始化; 相比之下, SFAAGC 借助特征扩充仍可在该条件下完成增强。同时, SFAAGC 在各数据集和各样本规模下均优于 FAAGC 与无增强基线, 说明特征扩充能够稳定提升测地线增强的效果。最后, 随着每类样本数 m 的增大, FAAGC 的增强效果逐渐消失, 而 SFAAGC 增强依然可以给任务带来表现提升; 在 Flowers102 的 $m = 10$ 设置下, FAAGC 略低于无增强基线, 而 SFAAGC 仍保持优势。

4.5.3 增强框架效率对比实验

对于 SFAAGC, 由于特征扩充会增加特征提取数量, 仅比较测地线拟合与采样阶段已不足以反映完整开销。因此, 本研究在 CIFAR-10@5 与 CIFAR-10@10 上进一步比较 FAGC 与 SFAAGC 特征提取阶段和测地线拟合阶段的运行时间与能耗。其中特征提取和扩充阶段在 NVIDIA GeForce RTX 4090 GPU 上完成, 能耗统计通过 CodeCarbon 库记录 CPU、RAM 与 GPU 的能耗总和; 测地线增强阶段在 Intel(R) Xeon(R) Silver 4210R CPU 上完成, CodeCarbon 库记录 CPU 与 RAM 的能耗总和。为与第三

章效率对比实验（第 3.5.2 节）保持一致，扩充特征数量与增强特征数量均固定为 m 。实验结果如表 4.4 所示。

表 4.4 FAGC 与 SFAAGC 两阶段的分阶段运行时间、能耗与分类准确率对比

方法	m	特征提取阶段		测地线拟合阶段		合计		准确率 (%)
		时间 (s)	能耗 (kWh)	时间 (s)	能耗 (kWh)	时间 (s)	能耗 (kWh)	
FAGC	5	3.12	3.19×10^{-4}	292.73	7.75×10^{-3}	295.85	8.07×10^{-3}	85.20
SFAAGC	5	3.53	2.38×10^{-4}	4.27	1.04×10^{-4}	7.80	3.42×10^{-4}	88.51
FAGC	10	3.64	3.64×10^{-4}	665.45	1.51×10^{-2}	669.09	1.55×10^{-2}	86.30
SFAAGC	10	5.08	4.24×10^{-4}	4.54	1.20×10^{-4}	9.62	5.44×10^{-4}	92.76

从表 4.4 可以看出，尽管 SFAAGC 在特征提取阶段因扩充而引入了额外的 GPU 计算开销（这个开销在 CodeCarbon 库的统计误差内），但其测地线拟合阶段采用小批次梯度下降，计算效率显著高于 FAGC 的迭代式测地距离优化。综合来看，SFAAGC 完整流程的总能耗较 FAGC 降低约一个数量级，同时在分类准确率上取得了明显的提升。

4.5.4 结合特征扩充的测地线增强框架消融实验

为了验证结合特征扩充的测地线增强框架中各个部分的必要性，本节将以下三类预形状空间特征向量以不同组合输入图像分类器进行训练，以评估各组件对最终分类性能的贡献。为避免不同组合之间因样本数量差异造成影响，实验中每类每种特征均取 m 个样本。三种特征分别是：（1）由原图经 ViT-Tiny 模型提取并投影至预形状空间得到的原始特征向量；（2）由 AugMix 增强图像经 ViT-Tiny 模型提取并投影至预形状空间得到的图像增强特征向量；（3）经过 SFAAGC 框架进行测地线拟合以及测地线插值得到的测地线增强特征向量。

从实验结果中可以看到，当三种特征联合使用时，分类模型取得了最优的分类性能。值得注意的是，与第 4.5.2 节中第一个实验使用相同图像增强方法的情形相比，本实验在组合原特征与图像增强特征后，分类准确率有了明显的提升（本实验准确率为 68.83%，而对比实验中使用 AugMix 结果为 65.51%）。两项准确率存在显著差异的原因，除不同实验轮次带来的误差外，主要在于本实验中图像增强特征被投影到了预形状空间中，该操作已在第 3.5.3 节中得到证实，有利于在样本数量受限时提升分类准确率。

表 4.5 不同特征组合输入下的分类准确率

原特征	图像增强特征	测地线增强特征	准确率 (%)
是	否	否	67.71
否	是	否	67.44
否	否	是	67.69
是	是	否	68.83
是	否	是	69.00
否	是	是	69.56
是	是	是	70.07

为了验证不同图像增强方法对扩充特征质量及最终分类效果的影响，本文采用多种图像增强方法生成扩充特征，分别是：（1）不采用任何图像增强得到的扩充特征；（2）采用简单的中心裁剪和随机翻转得到的扩充特征；（3）采用 RandAugment 得到的扩充特征；（4）采用 AugMix 得到的扩充特征；（5）采用 DiffuseMix 得到的扩充特征。

表 4.6 不同图像增强扩充方法对分类准确率的影响

扩充方法	准确率 (%)
无图像增强	68.08
中心裁剪 + 随机翻转	73.74
RandAugment	73.77
AugMix	74.29
DiffuseMix	68.62

从表 4.6 的实验结果可以看出，采用 AugMix 得到的扩充特征进行测地线拟合与增强时，最终分类准确率最高，而使用 DiffuseMix 方法增强得到的特征准确率明显不如其他扩充方法。

为进一步验证 SFAAGC 框架中特征扩充模块的通用性，本研究在 CIFAR-100@5 上考察该模块与不同特征级增强方法相结合时的下游分类性能，实验按 5 种特征增强方法与 2 种特征扩充设置的组合构成 10 组配置。所考察的 5 种特征级增强方法包括：不进行任何特征级增强（基线）、Manifold Mixup、SFA-S、FAGC 测地线增强以及 FAAGC 测地线增强。2 种特征扩充设置分别为无扩充与使用扩充：前者仅对原始训练集提取特征后直接施加特征级增强；后者先通过 AugMix 图像增强对训练集进行扩充，由微调后的 ViT-Tiny 统一提取特征，再在此基础上施加特征级增强。特别

地，基于扩充使用 FAAGC 组合时，实际采用 SFAAGC 框架进行增强。对于除“无特征级增强”外的各方法，每类扩充特征数量与特征级增强特征数量在无扩充设置下分别为 0 和 $9m$ ，在使用扩充设置下分别为 $4m$ 和 $5m$ ；对于“无特征级增强”基线，两种设置下均不生成特征级增强特征，对应的扩充特征数量与特征级增强特征数量分别为 0 和 0、 $4m$ 和 0。实验结果如表 4.7 所示。

表 4.7 特征扩充与不同特征增强方法结合的分类准确率

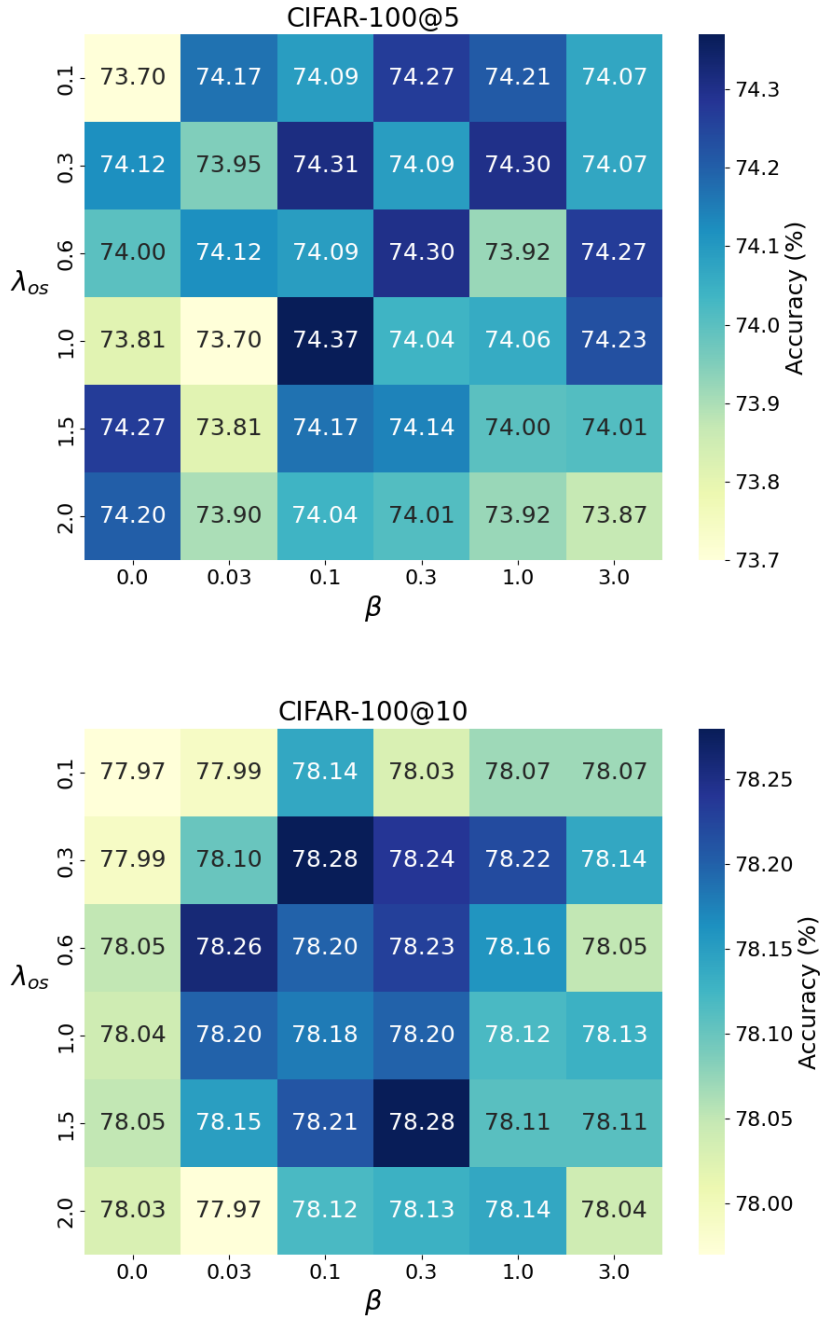
特征增强方法	无扩充	使用扩充
无特征级增强	66.41	65.50
Manifold Mixup	66.63	65.74
SFA-S	65.26	70.38
FAGC	66.39	70.44
FAAGC	68.01	74.29

表 4.7 的结果表明，引入特征扩充后各方法的性能变化存在明显差异。Manifold Mixup 施加特征扩充后，准确率不如仅施加特征扩充，这可能是因为扩充引入的图像增强特征与原始特征之间存在分布偏移，而该特征增强方法缺乏有效的机制来利用或修正这种偏移，直接将扩充特征混入训练反而引入了噪声。相比之下，SFA-S、FAGC 与 FAAGC 在引入扩充后准确率均有所提升，说明这些方法能够从更丰富的输入特征中受益：SFA-S 通过特征统计量的增强缓解了分布偏移的影响，而 FAGC 与 FAAGC 则通过在预形状空间中构建测地线，天然地对投影后的扩充特征具有更强的整合能力。此外，完整的 SFAAGC 框架取得了最优的分类性能（74.29%），较无扩充的 FAAGC 提升了 6.42 个百分点，验证了该框架各组件协同工作的有效性。

为进一步验证 SFAAGC 对损失函数权重设置的鲁棒性，本文在 CIFAR-100@5 与 CIFAR-100@10 两个基准上进行了超参数敏感性分析。对于总损失中的平衡系数 β 与特征扩充样本权重 λ_{os} ，本文在 $\beta \in \{0, 0.03, 0.1, 0.3, 1, 3\}$ 、 $\lambda_{os} \in \{0.1, 0.3, 0.6, 1.0, 1.5, 2.0\}$ 的组合上评估分类准确率，结果如图 4.3 所示：

从图 4.3 可以看出，对于 CIFAR-100@5 与 CIFAR-100@10 两个不同规模的数据集，SFAAGC 对超参数变化表现出较好的稳定性：当 $\lambda_{os} \in [0.3, 1.5]$ 且 $\beta \in [0.1, 1.0]$ 时，两种数据规模下的下游分类准确率整体均保持在较高水平。

为了确定图像增强扩充倍数与测地线特征增强数量对最终分类准确率的影响，本节在 CIFAR-100@5 上进行了实验。扩充倍数 n_{os} 分别设置为 $[1, 2, 4, 6, 10]$ ，测地线

图 4.3 不同 β 与 λ_{os} 组合的分类准确率热力图

特征样本增强数量分别设置为 $[m, 3 \times m, 5 \times m, 7 \times m, 10 \times m, 15 \times m]$ 。其中 $m = 5$ 表示训练集中每类样本的个数，结果如表 4.8 所示。

表 4.8 不同扩充倍数与测地线特征增强数量对分类准确率的影响

扩充倍数 n_{os}	测地线特征增强数					
	m	$3 \times m$	$5 \times m$	$7 \times m$	$10 \times m$	$15 \times m$
1	70.07	70.10	70.21	70.29	70.11	69.78
2	72.15	72.29	72.38	72.47	72.25	72.62
4	74.39	74.14	74.29	74.27	74.14	74.32
6	74.91	75.26	75.17	75.32	75.25	75.36
10	75.47	75.74	75.93	75.83	75.81	75.93

从实验结果可以看到，扩充倍数的增加对测地线拟合及后续分类任务有正面作用；而随着测地线特征增强数量的增多，分类准确率基本呈现提升的趋势，但在增强数量达到 $5 \times m$ 后，继续增加特征增强样本对分类性能的提升已十分有限。

4.6 本章小结

本章提出了面向图像数据集增强场景的可扩展测地线特征增强（SFAAGC）框架。该框架在 FAAGC 的基础上引入特征扩充与小批次加权测地线拟合，前者通过图像增强、特征提取与投影生成扩充特征，解决单样本场景下测地线拟合难以初始化的问题，使拟合的测地线更能反映类内特征的真实分布；后者通过小批次梯度下降与样本来源权重降低大样本量下的计算开销，并调节原始样本与图像增强样本对拟合过程的贡献。本章的可视化结果表明，SFAAGC 能够提升增强特征的覆盖性；进一步的实验分析表明，SFAAGC 在不同样本规模和不同骨干网络下均能稳定提升分类性能，并在运行时间与能耗方面保持较高效率。同时，消融实验验证了特征扩充、小批次加权拟合与测地线增强的有效性，说明 SFAAGC 将 FAAGC 从小规模特征增强模块扩展为适用于单样本至大规模图像数据集的完整增强框架。

需要说明的是，本章所述框架的“可扩展”主要指数据规模（样本数量）维度上的可扩展性，即框架能够适应从单样本到大规模数据集的不同样本量场景。在其他维度上，该框架亦具备一定的扩展潜力：在类别维度上，SFAAGC 对每个类别独立地进行特征扩充与测地线拟合，类别之间不存在耦合，计算开销随类别数量线性

增长，且各类别的拟合过程可并行执行，因此能够自然地扩展至类别数更多的数据集；在数据模态维度上，框架的测地线拟合模块以特征向量为操作对象，与特征的来源模态无关，只需将特征扩充模块中的图像增强方法替换为文本、音频等模态对应的数据增强手段，即可将框架迁移至其他模态的数据增强任务。相关探索方向将在第 六 章的展望中进一步讨论。

第五章 测地线特征增强实验平台设计与实现

在前几章中，本文分别提出了测地线特征增强模块 FAAGC（第 三 章）和可扩展测地线特征增强框架 SFAAGC（第 四 章）。然而，上述方法的实验涉及特征提取、预形状空间投影、测地线拟合、采样增强以及分类评估等多阶段操作，存在数据依赖，手动逐步执行易出错且难以复现。同时，实际研究中需要频繁切换数据集、骨干网络与超参数组合，缺乏实验管理机制会影响实验效率与结果可复现性。

为此，本章设计并实现了一个测地线特征增强实验平台，将上述多阶段流水线封装为配置驱动的自动化系统。本章首先进行需求分析（第 5.1 节），随后介绍平台的总体架构设计（第 5.2 节），最后说明各功能模块的设计与实现（第 5.3 节）。

5.1 平台概述与需求分析

根据本文实验流程的特点，平台的功能需求可归纳为五点：

（1）数据管理，支持 CIFAR-10、CIFAR-100、CUB-200-2011、Fashion-MNIST、Flowers102、Caltech101 等多个视觉基准数据集的自动下载与统一加载，并支持按类别进行 m -shot 子集采样；

（2）模型管理，支持 ResNet 系列、EfficientNet 系列、Vision Transformer 系列和 Swin Transformer 系列等多种骨干网络的加载与特征提取，可通过配置文件灵活切换；

（3）增强方法执行与扩展，支持 FAAGC 全批次测地线拟合模式、SFAAGC 小批次加权拟合模式以及多种对比增强方法，平台对所选方法依次执行增强流程；同时提供可配置的参数初始化策略，并支持通过统一接口接入新的增强方法；

（4）评估输出，对所选每一种增强方法分别计算分类准确率，并将多种增强方法生成的特征通过 t-SNE 降维进行可视化；同时记录运行时间与能耗；

（5）实验参数管理，通过配置文件统一保存数据集、骨干网络、增强方法、优化超参数与随机种子等实验参数，并支持使用交互界面修改默认配置。

在非功能需求方面，平台不绑定特定的 Python 解释器或环境管理工具，支持本机运行与远程服务器部署，用户可通过浏览器远程执行测地线增强任务；新增数据集或增强算法时仅需实现相应接口函数并完成注册，已被模型库支持的骨干网络则

可通过配置面板直接切换；所有实验过程由全局随机种子统一管理，确保在相同环境下运行结果一致。

5.2 平台总体架构设计

为支持上述功能与非功能需求，平台总体架构如图 5.1 所示，依次为运行环境、存储层、应用层、传输层与展示层。

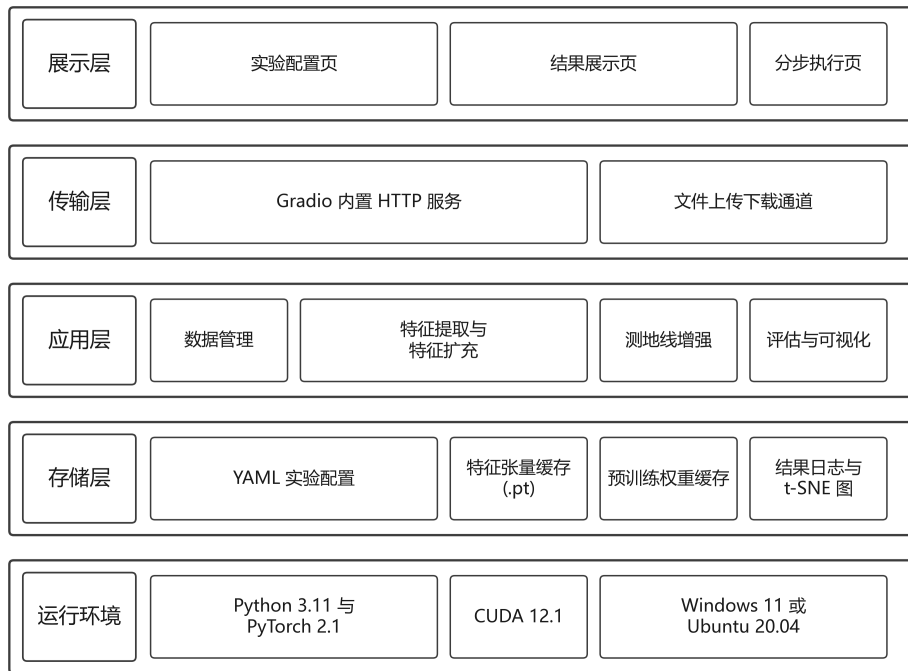


图 5.1 系统总体架构图

运行环境提供平台执行所需的软件与硬件条件：平台基于 Python 3.11 与 PyTorch 2.1 开发，运行环境包含 CUDA 12.1，可在 Windows 与 Ubuntu 上运行。存储层基于本地文件系统，保存 YAML 实验配置、提取后的特征张量缓存（.pt 文件）、预训练骨干网络权重以及实验结果日志与 t-SNE 可视化输出。应用层由数据管理、特征提取与特征扩充、测地线增强以及评估与可视化四个模块组成，模块间通过张量、字典与配置对象传递数据。传输层基于 Gradio 内置的 HTTP 服务实现浏览器与应用层之间的请求转发与文件上传下载，使平台既可在本机运行，也可部署至远程服务器供浏览器访问。展示层基于 Gradio 构建了包含实验配置、结果展示与分步执行在内的交互页面，将运行参数映射为图形化控件，使用户在不修改源代码的前提下组织实验流程。

平台一次完整实验的业务流程如图 5.2 所示。

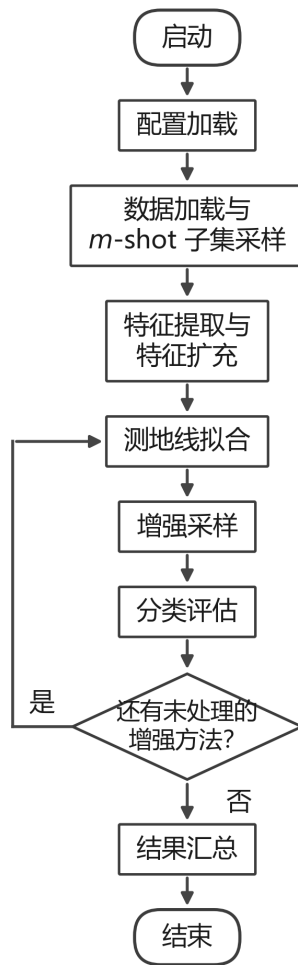


图 5.2 平台业务流程图

用户在界面完成配置后，平台依次执行数据加载与 m -shot 子集采样、特征提取与特征扩充；随后对用户勾选的每一种增强方法循环执行特征增强与分类评估；循环结束后，多种增强方法的评估结果与生成特征统一汇总至结果展示页。多种增强方法共享同一份特征提取结果，避免重复计算。

5.3 功能模块设计与实现

平台的功能架构如图 5.3 所示，由数据管理模块、特征提取与特征扩充模块、测地线增强模块、评估与可视化模块以及交互界面模块五个部分构成。

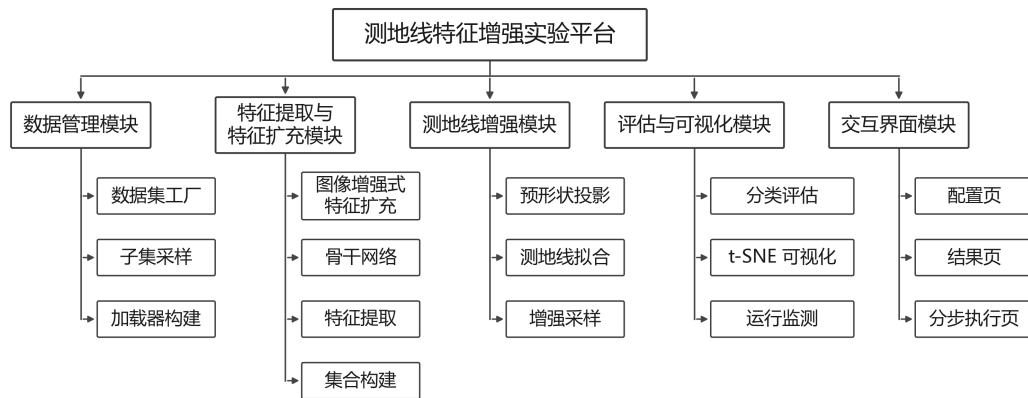


图 5.3 系统功能架构图

5.3.1 数据管理模块

数据管理模块负责接收配置数据并生成训练集与测试集的数据加载接口，其内部结构如图 5.4 所示。

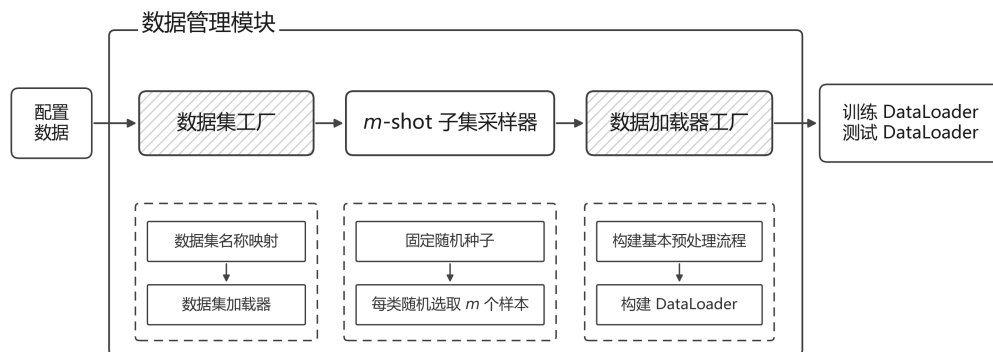


图 5.4 数据管理模块结构

该模块由三个组件构成。数据集工厂根据配置中的数据集名称完成名称映射，并实例化对应的数据集加载器。平台通过统一的数据集适配接口封装不同数据集的读取逻辑。

m -shot 子集采样器用于模拟数据受限场景，首先固定配置文件中给定的随机种子，再从完整训练集中按类别随机选取 m 个样本，从而保证不同实验设置下训练子集的可复现性。当 m 大于等于类别样本数量时，采样器只起到随机扰动样本顺序的作用。

数据加载器工厂根据配置构建基本预处理流程，在默认情况下对图像尺寸进行

调整，并按通道进行归一化。在此基础上，分别为训练集与测试集创建 `DataLoader`，以供后续特征提取模块直接调用。

5.3.2 特征提取与特征扩充模块

特征提取与特征扩充模块对应第 4 章提出的 SFAAGC 框架中的基于图像增强的特征扩充方法（见第 4.2 小节），负责将输入图像转换为欧氏空间中的特征向量，并构建包含样本来源标记的训练特征集合 $\mathcal{D}_{\text{train}}$ 。如图 5.5 所示，该模块依次经历图像增强式特征扩充、骨干网络加载、特征提取和特征集合构建四个步骤。其中，图中的斜线底纹模块“图像增强式特征扩充器”与“骨干网络管理器”表示增强策略和骨干网络均可通过配置文件直接调整。

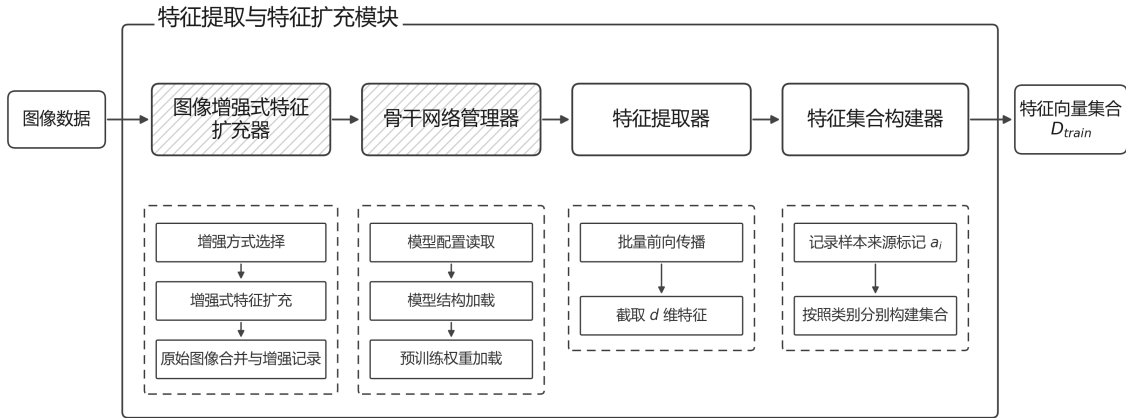


图 5.5 特征提取与特征扩充模块流程

图像增强式特征扩充器根据配置选择具体的图像增强策略，并在 SFAAGC 模式下对每张原始训练图像生成 n_{os} 份增强副本。增强后的图像与原始图像一并保留，并记录样本来源，作为后续特征提取的输入。当前平台支持 AugMix、RandAugment 与 DefaultAug 等增强方式；该步骤也可被略过，使整个模块退化为常规的特征提取流程。

骨干网络管理器读取模型配置，完成模型结构加载与预训练权重载入，并统一管理不同骨干网络的接口形式，使后续特征提取流程能够以一致方式调用。目前支持包括 ResNet 系列、EfficientNet 系列、Vision Transformer 系列和 Swin Transformer 系列在内的骨干网络。

特征提取器对原始图像与增强图像执行批量前向传播，并在全局平均池化层、

[CLS] token 或模型指定的其他表征位置截取 d 维特征向量 $v_i \in \mathbb{R}^d$ ，从而将输入图像映射到欧氏特征空间。

特征集合构建器将提取到的特征按类别组织为统一格式，并记录样本来源标记 a_i ，其中原始样本记为 $a_i = 0$ ，增强样本记为 $a_i = 1$ ，最终构建得到供后续测地线增强模块使用的训练特征集合 $\mathcal{D}_{\text{train}}$ 。

5.3.3 测地线增强模块

测地线增强模块实现了本文提出的 FAAGC 与 SFAAGC 测地线特征拟合与增强算法。该模块内部结构如图 5.6 所示，由预形状空间投影、测地线拟合与增强采样三个环节构成。

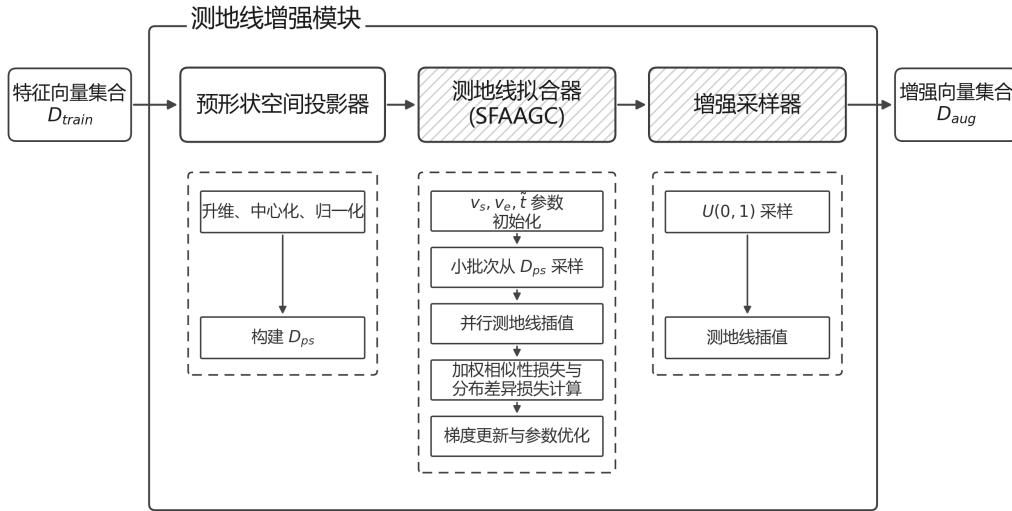


图 5.6 测地线增强模块内部结构

预形状空间投影器将训练特征集合 $\mathcal{D}_{\text{train}}$ 中的欧氏空间特征向量 $v \in \mathbb{R}^d$ 按照第 3.2 节描述的步骤执行升维、中心化与归一化，构建预形状空间特征集合 $\mathcal{D}_{\text{ps}}$ 。

测地线拟合与增强采样均以类别为单位独立执行：模块依据样本的类别标签将 $\mathcal{D}_{\text{ps}}$ 划分为各类别特征子集，并对每个类别依次完成下述拟合与采样流程。

测地线拟合器首先完成端点参数 v_s, v_e 与采样参数 $\tilde{t}$ 的初始化，随后从当前类别的特征子集中按小批次采样，接着执行测地线插值，计算加权相似性损失与分布差异损失，并通过梯度更新完成参数优化。拟合模式、批次大小、损失权重以及相关优化超参数均可通过配置文件调整；当增强模式设置为 FAAGC 时，不会执行加权小批次机制。

增强采样器在拟合完成后从 $U(0, 1)$ 中采样参数，并沿优化后测地线执行插值，生成增强向量集合 $\mathcal{D}_{\text{aug}}$ 。采样规模等参数可通过配置文件指定，从而控制增强特征的数量。

该模块还支持自定义增强算法的执行：用户实现统一接口的增强函数并完成注册后，即可在交互界面中调用。当用户在交互界面同时勾选多种增强方法时，平台依次调用对应的增强函数，并将各方法生成的增强向量集合 $\mathcal{D}_{\text{aug}}^{(k)}$ 分别保留，供评估模块统一处理。

5.3.4 评估与可视化模块

评估与可视化模块负责对增强效果进行定量评估与定性分析。该模块为当次实验所选的每一种增强方法分别训练一个线性分类头，并在完整测试集上计算各自的分类准确率；t-SNE 可视化组件以类别为单位绘制散点图，图中同时绘制该类别的训练特征、各增强方法生成的增强特征以及测试集特征，并以不同颜色或标记区分不同来源，直观对比不同增强方法对训练特征的拟合程度与泛化性差异。运行监测组件自动记录 RAM、CPU 与 GPU 的能耗，并记录各阶段的运行时间，输出至交互界面的日志栏。

5.3.5 交互界面模块

为降低使用门槛，平台基于 Gradio 构建可视化交互界面，使用户可通过图形化操作完成实验配置、结果查看与分步执行。Gradio^[74] 是面向机器学习应用的 Python 交互界面库，通过声明式的组件定义即可生成包含输入控件与输出展示区域的界面，平台启动后在本地浏览器中呈现，主要包含以下三组功能页面：

实验配置页如图 5.7 所示，用户在该页一次性完成数据集、骨干网络、增强方法及其超参数、随机种子等所有实验配置。其中增强方法支持多选，用户可勾选一项或多项增强方法纳入本次实验。参数修改完成后点击“开始实验”按钮即可启动流水线，平台将复用同一份特征提取结果，依次为每种被选中的增强方法执行增强与评估流程。

结果展示页如图 5.8 所示，在实验完成后集中呈现评估结果：所选每一种增强方法的分类准确率以表格形式呈现，便于横向对比；t-SNE 散点图按类别输出，显示训练特征与增强特征的分布；运行时间与能耗统计以文本形式输出。用户可在界面中

方法	Final Mean	Best Mean	Baseline
不增强 (Baseline)	60.43%	60.43%	—
FAAGC	66.51%	66.87%	+6.08
SFAAGC	67.87%	68.12%	+7.44

前缀	耗时	CPU (s)	GPU (s)	RAM (MB)	合计 (s)
S01 数据管理-加载	4m29.51s	0.61837	0.61393	0.60284	0.62724
S02 特征提取	4m11.15s	0.59906	0.61189	0.60355	0.62348
S03 测地线增强	6.87s	0.00009	0.00009	0.00009	0.00009
S04 测地线增强	12.77s	0.00042	0.00046	0.00014	0.00122
S05 分类评估	6.73s	0.00021	0.00034	0.00007	0.00062
总计	8m58.23s	0.82086	0.82673	0.80548	0.85238

图 5.7 交互界面实验配置页

查看与比较各增强方法在相同数据集与骨干网络下的表现。

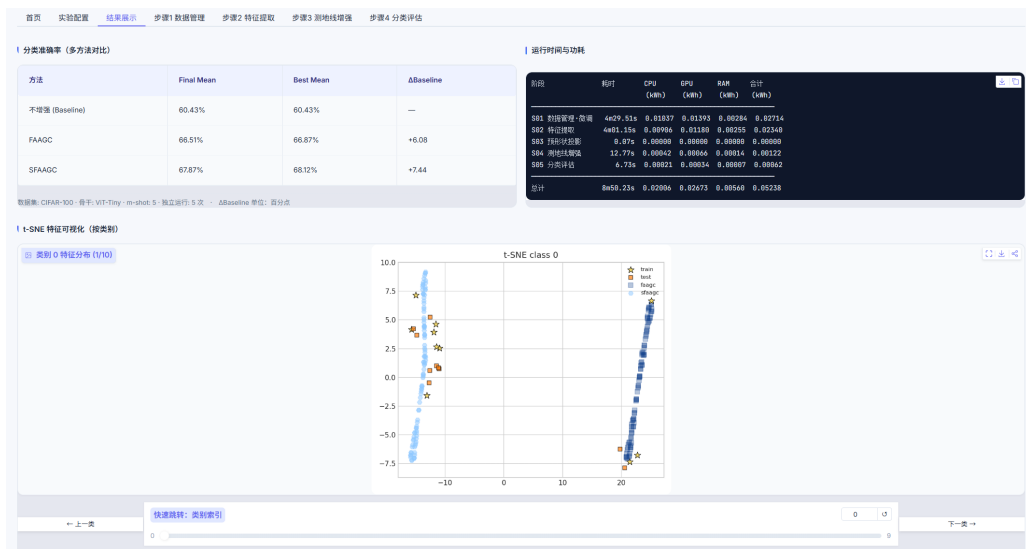


图 5.8 交互界面结果展示页

分步执行页为便于用户逐步调试和观察各阶段的中间结果而设计，包含数据管理、特征提取、测地线增强与分类评估四个子页面，分别对应流水线的四个步骤。在每个子页面中，用户将上一步骤的输出文件作为输入，调整超参数后执行该步骤，并在完成后下载输出文件以供后续步骤使用或离线分析。通过分步执行，用户可在不重新运行完整流水线的前提下，针对特定阶段反复调整参数并观察其对最终结果的影响。图 5.9 展示了分步执行页中的测地线增强步骤子页面。

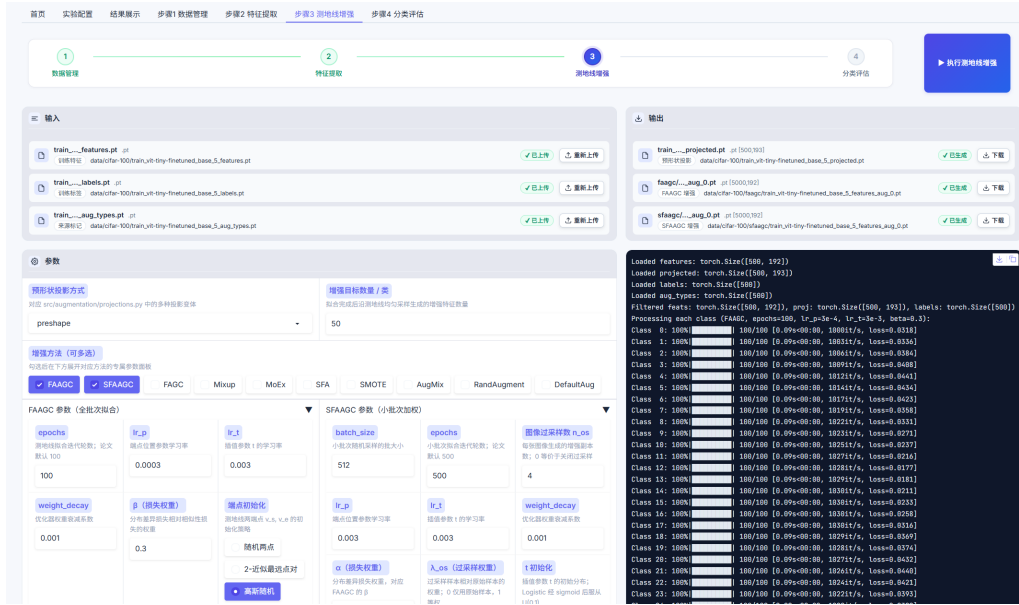


图 5.9 测地线增强分步执行页

交互界面负责参数收集与结果展示，不包含实验计算逻辑。用户在界面中设置的参数会被转换为配置对象，传递给应用模块执行。

5.4 本章小结

本章设计并实现了测地线特征增强实验平台，将本文提出的 FAAGC 与 SFAAGC 方法封装为配置驱动的自动化实验系统。平台由运行环境、存储、应用、传输与展示五层组成，应用层包含数据管理、特征提取与特征扩充、测地线增强以及评估与可视化四个模块，并通过基于 Gradio 的交互界面完成参数配置、结果展示与分步执行。平台支持在一次实验中批量运行多种增强方法，并以表格与 t-SNE 形式直观对比与定量分析各方法的结果，降低了在各类特征增强任务中运用 FAAGC 与 SFAAGC 的门槛。同时，目前平台仍存在一定不足：新增数据集和自定义增强算法仍需在源代码中注册，后续可将该流程纳入交互界面，进一步提升平台的可扩展性。

第六章 总结与展望

6.1 研究总结

数据受限场景是深度学习应用中的普遍挑战。在每类训练样本仅有数个甚至一个的极端条件下，深度神经网络提取的特征分布极为稀疏，难以充分描述类内的结构多样性。现有的特征级数据增强方法大多依赖特征自身或欧氏空间中的线性插值假设，忽视了特征向量在高维空间中所呈现的非线性流形结构。已有工作（如 FAGC）虽尝试将预形状空间的测地线结构引入特征增强，但受限于大量点采样与测地距离计算，其优化效率存在明显瓶颈，且在单样本等极端场景下缺乏有效扩展机制。

针对上述问题，本文以形状空间理论为核心工具，提出了一种基于预形状空间测地线结构的特征级数据增强方案。其核心思想是：将深度神经网络提取的特征向量视为高维空间中的几何对象，通过升维、中心化与归一化将其投影至预形状空间，在去除尺度与位置等无关变换后保留特征的内在形态结构，进而在预形状空间中构建拟合类内分布的测地线，并沿测地线进行采样生成新的增强特征。该方法为数据受限场景下深度模型的训练提供了具有实用性的基于形状空间理论的增强手段。本文的主要研究工作与贡献总结如下：

（1）提出了特征级数据增强模块 FAAGC。针对预形状空间中测地线拟合效率低下的问题，FAAGC 重新设计了优化目标与损失函数，以端点参数和采样参数为优化变量，采用梯度下降方式替代 FAGC 中的大量点采样与测地线距离计算操作，在保持或超越 FAGC 增强效果的同时，大幅提升了拟合效率。在 CIFAR-10@5 上，FAAGC 相较于 FAGC 实现了约 131 倍的加速（2.23s vs 292.73s），能耗仅为 FAGC 的约 0.76%，同时分类准确率提升至 85.84%。在六个视觉基准数据集的小样本条件下，FAAGC 均取得了优于或相当于 FAGC 的分类准确率，验证了其在效率与有效性两方面的优势。

（2）构建了可扩展测地线特征增强框架 SFAAGC。为将测地线增强方法推广至更广泛的数据规模（包括单样本场景），SFAAGC 引入了图像增强驱动的特征扩充技术，为测地线拟合提供充足的训练特征；同时采用小批次梯度下降与样本加权机制，有效缓解了大规模增强样本下的计算瓶颈。实验结果表明，相较无增强基线和其他特征增强方法，SFAAGC 框架在 CIFAR-10@5 和 CIFAR-100@5 上分别将分类准

准确率提升至 88.90% 和 74.29%，取得了显著的性能增益，且完整流程总能耗显著低于 FAGC，验证了该框架在效率与性能之间的良好平衡。

(3) 通过系统的实验，验证了所提方法在数据受限场景下的有效性与适用性。实验在 CIFAR-10、CIFAR-100、CUB-200-2011、Fashion-MNIST、Flowers102 和 Caltech101 六个视觉基准数据集上展开，涵盖与多种特征级增强方法的对比分析、与 FAGC 的运行时间及能耗对比、各组件的消融研究以及超参数灵敏度分析。实验结果从定性（t-SNE 可视化）与定量（分类准确率）两个角度一致表明，FAAGC 与 SFAAGC 在数据受限条件下能够有效提升特征空间的类内多样性，在多数场景下超越现有特征级增强基线，消融研究也确认了流程中各组件的必要性。

(4) 设计并实现了测地线特征增强实验平台。为降低上述多阶段实验流水线的操作复杂度，本文将 FAAGC 与 SFAAGC 方法封装为配置驱动的自动化实验系统。平台采用分层架构，通过配置文件统一管理所有实验参数，并基于 Gradio 构建可视化交互界面，支持实验配置、分步执行与结果展示。

综上所述，本文将预形状空间的流形几何结构引入深度特征增强，从算法模块，到完整框架，再到工程平台，构建了一套系统性的数据受限场景特征增强解决方案，为小样本条件下深度学习模型的训练提供了兼顾几何合理性与工程可行性的新思路。

6.2 未来工作

尽管本文提出的方法具备上述创新与优势，但仍存在若干有待进一步研究的方向。首先，本方法在沿测地线采样的过程中，尚未考虑样本沿测地线的分布可能并非均匀这一情况。当训练样本数量受限时，将均匀分布作为某类样本沿测地线的先验分布较为合理，然而当样本数量增加时，这种假设存在一定局限性；其次，由于目前的测地线增强以经深度神经网络提取并全局池化后的特征向量作为输入，无法捕捉图像的细粒度空间信息，其应用场景目前受限于图像分类等以整体语义为基础的任务，在目标检测、语义分割等需要空间感知的细粒度任务中尚缺乏有效的应用机制。最后，在实验结果中可以发现，基于特征扩充与加权小批次梯度下降的测地线特征增强框架在应用中，当采用 DiffuseMix 这类基于生成模型的方法时，效果不如基于 AugMix、RandAugment 等基础图像增强方法，这说明在引入图像增强方法进行特征扩充时，不同的图像增强方法可能会对测地线拟合的效果产生不同程度的影响，

图像增强方法对图像特征在预形状空间中分布的影响是一个值得进一步研究的问题。

因此，未来的研究可以从以下几个方面进行改进：一是引入更为多样的测地线分布假设，以更好地适应不同样本数量和分布情况；二是探索在骨干网络全局池化之前的空间特征图中，提取代表局部区域语义的特征向量进行测地线增强，将该技术拓展至目标检测、语义分割等细粒度任务，以充分利用其在特征空间中的增强能力；三是系统地分析和评估经不同类型图像增强方法处理后的图像所提取的特征对测地线拟合的影响，以指导实际应用中将图像层面增强方法与特征层面增强方法更好地融合；四是将 SFAAGC 框架向更大类别规模以及图像以外的其他数据模态扩展，系统评估其在数据规模之外的维度上的可扩展性。

此外，随着跨模态大模型的快速发展，基于形状空间理论的自适应测地线特征增强算法模块在跨模态理解与生成等新兴应用方向中同样具有广阔的应用前景。以 CLIP^[75] 为代表的视觉-语言对齐模型将图像特征与文本特征映射至共享的嵌入空间。在数据稀缺的少样本条件下，可利用测地线拟合与采样技术对稀缺模态的特征进行增强，有望缓解特定领域中跨模态数据不均衡问题，提升少样本跨模态检索与分类的性能。此外，由于测地线采样具有几何连续性，沿测地线插值所得特征序列天然保持平滑过渡，这一特性使其同样适用于视频生成等连续内容生成任务^[76]，为帧间特征的平滑插值提供流形约束。上述探索将有助于拓展基于形状空间理论的数据增强方法的适用边界，推动其在更广泛场景中发挥作用。

插图索引

图 1.1	本文主要工作示意图	7
图 3.1	测地线特征增强流程示意图	19
图 3.2	可学习的采样参数 $\tilde{t}$ 能够自适应地对齐采样点与原始样本，从而降低 损失函数	21
图 3.3	从 CIFAR-10 子集提取的预形状空间原始特征、FAGC 增强特征与 FAAGC 增强特征	25
图 3.4	不同图像增强的微调策略下 FAAGC 带来的准确率提升	29
图 3.5	学习率 η_p 与 η_t 对分类准确率的影响	32
图 4.1	可扩展测地线特征增强框架总体结构	34
图 4.2	CIFAR-10 两个代表类别的增强样本 t-SNE 可视化	39
图 4.3	不同 β 与 λ_{os} 组合的分类准确率热力图	47
图 5.1	系统总体架构图	51
图 5.2	平台业务流程图	52
图 5.3	系统功能架构图	53
图 5.4	数据管理模块结构	53
图 5.5	特征提取与特征扩充模块流程	54
图 5.6	测地线增强模块内部结构	55
图 5.7	交互界面实验配置页	57
图 5.8	交互界面结果展示页	57
图 5.9	测地线增强分步执行页	58

表格索引

表 3.1	FAGC 与 FAAGC 的训练时间与分类准确率对比	28
表 3.2	FAAGC 各阶段增量带来的分类准确率提升	29
表 3.3	v_s, v_e 不同初始化方法在各训练轮数下的分类准确率	30
表 3.4	$\tilde{t}$ 不同初始化方法在各训练轮数下的最佳分类准确率	31
表 4.1	不同特征增强方法在各数据集上的分类准确率	41
表 4.2	不同骨干网络下各特征增强方法的分类准确率	42
表 4.3	不同每类训练样本数下 FAAGC 与 SFAAGC 的分类准确率	43
表 4.4	FAGC 与 SFAAGC 两阶段的分阶段运行时间、能耗与分类准确率对比 ..	44
表 4.5	不同特征组合输入下的分类准确率	45
表 4.6	不同图像增强扩充方法对分类准确率的影响	45
表 4.7	特征扩充与不同特征增强方法结合的分类准确率	46
表 4.8	不同扩充倍数与测地线特征增强数量对分类准确率的影响	48
表 B.1	不同升维策略的分类准确率	75

参考文献

- [1] 刘颖, 雷研博, 范九伦, 等. 基于小样本学习的图像分类技术综述[J]. 自动化学报, 2021, 47(2): 297-315.
- [2] 孙景浩, 聂凯. 小样本图像识别研究综述[J/OL]. 计算机测量与控制, 2026, 34(01): 1-8. DOI: [10.16526/j.cnki.11-4762/tp.2026.01.001](https://doi.org/10.16526/j.cnki.11-4762/tp.2026.01.001).
- [3] 余凯, 宾焱, 郑自强, 等. 基于条件语义增强的文本到图像生成[J/OL]. 软件学报, 2024, 35(05): 2150-2164. DOI: [10.13328/j.cnki.jos.007024](https://doi.org/10.13328/j.cnki.jos.007024).
- [4] KRIZHEVSKY A, HINTON G, et al. Learning multiple layers of features from tiny images[M]. Toronto, ON, Canada, 2009.
- [5] DEVRIES T, TAYLOR G W. Dataset augmentation in feature space[A]. 2017.
- [6] KENDALL D G. Shape manifolds, procrustean metrics, and complex projective spaces[J]. Bulletin of the London mathematical society, 1984, 16(2): 81-121.
- [7] KENDALL D G, BARDEN D, CARNE T K, et al. Shape and shape theory[M]. John Wiley & Sons, 2009.
- [8] HAN Y, WANG B, IDESAWA M, et al. Recognition of multiple configurations of objects with limited data[J]. Pattern Recognition, 2010, 43(4): 1467-1475.
- [9] HAN Y, HU G, WAN G, et al. Fagc:feature augmentation on geodesic curve in the pre-shape space[A/OL]. 2025. arXiv: [2312.03325](https://arxiv.org/abs/2312.03325). <https://arxiv.org/abs/2312.03325>.
- [10] SHORTEN C, KHOSHGOFTAAR T M. A survey on image data augmentation for deep learning[J]. Journal of big data, 2019, 6(1): 1-48.
- [11] 杨锁荣, 杨洪朝, 申富饶, 等. 面向深度学习的图像数据增强综述[J]. 软件学报, 2025, 36(03): 78-79.
- [12] ZHANG H, CISSE M, DAUPHIN Y N, et al. mixup: Beyond empirical risk minimization[C]//International Conference on Learning Representations. 2018.

- [13] CUBUK E D, ZOPH B, MANE D, et al. Autoaugment: Learning augmentation strategies from data[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 113-123.
- [14] CUBUK E D, ZOPH B, SHLENS J, et al. Randaugment: Practical automated data augmentation with a reduced search space[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops. 2020: 702-703.
- [15] HENDRYCKS D, MU N, CUBUK E D, et al. Augmix: A simple data processing method to improve robustness and uncertainty[A]. 2019.
- [16] FRID-ADAR M, KLANG E, AMITAI M, et al. Synthetic data augmentation using gan for improved liver lesion classification[C]//2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018). IEEE, 2018: 289-293.
- [17] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets[J]. Advances in neural information processing systems, 2014, 27.
- [18] 王坤峰, 苟超, 段艳杰, 等. 生成式对抗网络 GAN 的研究进展与展望[J]. 自动化学报, 2017, 43(3): 321-332.
- [19] KARRAS T, AITTALA M, HELLSTEN J, et al. Training generative adversarial networks with limited data[J]. Advances in neural information processing systems, 2020, 33: 12104-12114.
- [20] CHLAP P, MIN H, VANDENBERG N, et al. A review of medical image data augmentation techniques for deep learning applications[J]. Journal of medical imaging and radiation oncology, 2021, 65(5): 545-563.
- [21] ATHALYE C, ARNAOUT R. Domain-guided data augmentation for deep learning on medical imaging[J]. PloS one, 2023, 18(3): e0282532.
- [22] HO J, JAIN A, ABBEEL P. Denoising diffusion probabilistic models[J]. Advances in neural information processing systems, 2020, 33: 6840-6851.
- [23] ROMBACH R, BLATTMANN A, LORENZ D, et al. High-resolution image synthesis with latent diffusion models[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022: 10684-10695.

- [24] TRABUCCO B, DOHERTY K, GURINAS M, et al. Effective data augmentation with diffusion models[C]//2023.
- [25] ISLAM K, ZAHEER M Z, MAHMOOD A, et al. Diffusemix: Label-preserving data augmentation with diffusion models[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024: 27621-27630.
- [26] ISLAM K, ZAHEER M Z, MAHMOOD A, et al. Genmix: Effective data augmentation with generative diffusion model image editing[A/OL]. 2025. arXiv: [2412.02366](https://arxiv.org/abs/2412.02366).
<https://arxiv.org/abs/2412.02366>.
- [27] 韦炎炎, 毛天一, 李柏昂, 等. 视觉模型及多模态大模型推进图像复原增强研究进展[J]. 中国图象图形学报, 2025, 30(05): 1197-1219.
- [28] GOODFELLOW I J, SHLENS J, SZEGEDY C. Explaining and harnessing adversarial examples[A]. 2014.
- [29] KINGMA D P, WELLING M. Auto-encoding variational bayes[A]. 2013.
- [30] CHEN Z, FU Y, ZHANG Y, et al. Multi-level semantic feature augmentation for one-shot learning[J]. IEEE Transactions on Image Processing, 2019, 28(9): 4594-4605.
- [31] VERMA V, LAMB A, BECKHAM C, et al. Manifold mixup: Better representations by interpolating hidden states[C]//International conference on machine learning. PMLR, 2019: 6438-6447.
- [32] LI P, LI D, LI W, et al. A simple feature augmentation for domain generalization[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 8886-8895.
- [33] LI B, WU F, LIM S N, et al. On feature normalization and data augmentation[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 12383-12392.
- [34] CHADEBEC C, ALLASSONNIÈRE S. Data augmentation with variational autoencoders and manifold sampling[C]//Deep Generative Models, and Data Augmentation, Labelling, and Imperfections: First Workshop, DGM4MICCAI 2021, and First Workshop, DALI 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, October 1, 2021, Proceedings 1. Springer, 2021: 184-192.

- [35] CHU P, BIAN X, LIU S, et al. Feature space augmentation for long-tailed data[C]//Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXIX 16. Springer, 2020: 694-710.
- [36] CHAWLA N V, BOWYER K W, HALL L O, et al. Smote: synthetic minority over-sampling technique[J]. Journal of artificial intelligence research, 2002, 16: 321-357.
- [37] 石洪波, 陈雨文, 陈鑫. SMOTE 过采样及其改进算法研究综述[J]. 智能系统学报, 2019, 14(6): 1073-1083.
- [38] LIU D, ZHONG S, LIN L, et al. Feature-level smote: Augmenting fault samples in learnable feature space for imbalanced fault diagnosis of gas turbines[J]. Expert Systems with Applications, 2024, 238: 122023.
- [39] BÄR A, HOULSBY N, DEHGHANI M, et al. Frozen feature augmentation for few-shot image classification[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024: 16046-16057.
- [40] SANDRU A, GEORGESCU M I, IONESCU R T. Feature-level augmentation to improve robustness of deep neural networks to affine transformations[C]//European Conference on Computer Vision. Springer, 2022: 332-341.
- [41] KUMAR R, KIM Y W, BYUN Y C. Hybrid framework combining diffusion based image augmentation and feature level smote for addressing extreme class imbalance [J]. IEEE Access, 2025.
- [42] DAVIDSON T R, FALORSI L, DE CAO N, et al. Hyperspherical variational auto-encoders[A]. 2018.
- [43] BOUMAL N. An introduction to optimization on smooth manifolds[M]. Cambridge University Press, 2023.
- [44] VADGAMA S, TOMCZAK J M, BEKKERS E J. Kendall shape-vae: Learning shapes in a generative framework[C]//NeurIPS 2022 Workshop on Symmetry and Geometry in Neural Representations. 2022.
- [45] MURALIKRISHNAN S, CHAUDHURI S, AIGERMAN N, et al. Glass: Geometric latent augmentation for shape spaces[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 18552-18561.

- [46] HU G, HAN Y. A projection module based on the shape space theory for small-sample image processing[J/OL]. Expert Systems with Applications, 2025, 293: 128668. <https://www.sciencedirect.com/science/article/pii/S0957417425022869>. DOI: <https://doi.org/10.1016/j.eswa.2025.128668>.
- [47] HAN Y, RUAN L, WANG B. Geodesic feature augmentation for zero-shot text-guided diffusion style transfer[J]. The Visual Computer, 2025, 41(13): 11239-11259.
- [48] 刘建伟, 刘媛, 罗雄麟, 等. 深度学习研究进展[J]. 计算机应用研究, 2014, 31(7): 1921-1930.
- [49] 庄福振, 罗平, 何清, 等. 迁移学习研究进展[J]. 软件学报, 2015, 26(1): 26-39.
- [50] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks[J]. Advances in neural information processing systems, 2012, 25.
- [51] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]// Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.
- [52] 林景栋, 吴欣怡, 柴毅, 等. 卷积神经网络结构优化综述[J]. 自动化学报, 2020, 46(1): 24-37.
- [53] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. Advances in neural information processing systems, 2017, 30.
- [54] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16x16 words: Transformers for image recognition at scale[J]. ICLR, 2021.
- [55] 田永林, 王雨桐, 王建功, 等. 视觉 Transformer 研究的关键问题: 现状及展望[J]. 自动化学报, 2022, 48(4): 957-979.
- [56] 周志华. 机器学习[M]. 北京: 清华大学出版社, 2016.
- [57] HAN Y. Recognize objects with three kinds of information in landmarks[J]. Pattern Recognition, 2013, 46(11): 2860-2873.
- [58] HENNESSY J L, PATTERSON D A. Computer architecture: a quantitative approach [M]. Elsevier, 2011.

- [59] SCHWARTZ R, DODGE J, SMITH N A, et al. Green ai[J]. Communications of the ACM, 2020, 63(12): 54-63.
- [60] STRUBELL E, GANESH A, MCCALLUM A. Energy and policy considerations for deep learning in nlp[C]//Proceedings of the 57th annual meeting of the association for computational linguistics. 2019: 3645-3650.
- [61] COURTY B, SCHMIDT V, LUCCIONI S, et al. mlco2/codecarbon: v2.4.1[CP/OL]. Zenodo, 2024. <https://doi.org/10.5281/zenodo.11171501>.
- [62] VAN DER MAATEN L, HINTON G. Visualizing data using t-sne.[J]. Journal of machine learning research, 2008, 9(11).
- [63] PASZKE A, GROSS S, MASSA F, et al. Pytorch: An imperative style, high-performance deep learning library[J]. Advances in neural information processing systems, 2019, 32.
- [64] FROGNER C, ZHANG C, MOBAHI H, et al. Learning with a wasserstein loss[J]. Advances in neural information processing systems, 2015, 28.
- [65] KINGMA D P, BA J. Adam: A method for stochastic optimization[A]. 2014.
- [66] WAH C, BRANSON S, WELINDER P, et al. The caltech-ucsd birds-200-2011 dataset [M]. California Institute of Technology, 2011.
- [67] XIAO H. Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms[A]. 2017.
- [68] NILSBACK M E, ZISSERMAN A. Automated flower classification over a large number of classes[C]//2008 Sixth Indian conference on computer vision, graphics & image processing. IEEE, 2008: 722-729.
- [69] FEI-FEI L, FERGUS R, PERONA P. Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories [C]//2004 conference on computer vision and pattern recognition workshop. IEEE, 2004: 178-178.
- [70] WIGHTMAN R. Pytorch image models[J/OL]. GitHub repository, 2019. <https://github.com/huggingface/pytorch-image-models>. DOI: 10.5281/zenodo.4414861.

- [71] ANGUS J E. The probability integral transform and related results[J/OL]. SIAM Review, 1994, 36(4): 652-654. <https://doi.org/10.1137/1036146>.
- [72] TAN M, LE Q. Efficientnet: Rethinking model scaling for convolutional neural networks[C]//International conference on machine learning. PMLR, 2019: 6105-6114.
- [73] LIU Z, LIN Y, CAO Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). 2021.
- [74] ABID A, ABDALLA A, ABID A, et al. Gradio: Hassle-free sharing and testing of ml models in the wild[A]. 2019.
- [75] RADFORD A, KIM J W, HALLACY C, et al. Learning transferable visual models from natural language supervision[C]//International conference on machine learning. PmLR, 2021: 8748-8763.
- [76] CAO Y, SONG Z, YANG C. Video latent flow matching: Optimal polynomial projections for video interpolation and extrapolation[A]. 2025.

攻读专业硕士学位期间取得的科研成果

一、论文

- [1] Li R, Han Y. Fast-FAAGC: Accelerating Geodesic Feature Augmentation via Distribution-Aligned Initialization[C]//2026 2nd International Conference on Computer Vision and Augmented Reality (CVAR 2026). Changsha, China, 2026. (EI Compendex, Scopus, 已录用, 待检索; 本人一作, 导师通讯作者)

二、专利

- [1] 专利名称: 一种用于含锡基热稳定剂高分子材料配方设计的时序数据增强方法, 发明人: 韩越兴、李睿杰、宋娜、丁鹏、陈侨川、李凌。申请号或专利号: 202511467290.5, 申请公布号: CN121393662A, 申请日: 2025 年 10 月 14 日, 公开日: 2026 年 01 月 23 日。(导师一作, 本人二作)

三、软件著作权

- [1] 韩越兴, 李睿杰. 基于形状空间理论的自适应测地线特征增强平台 V1.0[CP]. 著作权人: 上海大学. 登记号: 2025SR2177835. 证书号: 软著登字第 16834033 号. 权利取得方式: 原始取得. 权利范围: 全部权利. 登记日期: 2025-11-10. (导师第一作者, 本人第二作者)

四、项目

- [1] 国家重点研发计划项目“共性关键技术与空间应用”(编号: 2022YFB3707800), 课题“树脂基复合材料数据挖掘及专题数据库建设”, 参与。
- [2] 云南省科技重点项目, 项目类别: 新材料专项, 项目编号: 202302AB080022, 项目名称: 基于高分子材料应用的锡基功能助剂制备关键技术研发, 参与。

致 谢

过去在上海大学的多年学习生涯即将告一段落，回首这段求学之路，心中充满感激。

首先，衷心感谢我的导师韩越兴教授。从论文选题、方法设计到撰写修改，韩老师在每一个阶段都倾注了大量心血，给予了我耐心细致的指导。感谢陈侨川老师、孙妍老师对本文提出的宝贵意见与建议，你们的点拨让论文更加完善。

感谢课题组的胡干师兄和万冠新师兄，你们在学术讨论中给予了我大量启发，在论文构思与算法实现的关键阶段提供了无私帮助。感谢课题组所有老师和同门的陪伴与支持，与你们在学术上交流、在课题上共同努力，朝夕相处的日子是我硕士期间最温暖的回忆。感谢上海大学计算机工程与科学学院全体教师，你们的精彩教学为我打下了坚实的专业基础。

特别感谢我的父母，你们始终无条件地支持我的学业，在我遇到困难时给予我最大的鼓励与宽容，是我最坚实的后盾。

虽已历经数年求学，眼前依然是渺无边际的知识海洋。未来的路上，我将铭记师恩，心怀梦想，脚踏实地。命运会眷顾勇敢之人——*Audentes fortuna iuvat*。

李睿杰

上海

2026年6月

附录 A 测地线形式与插值形式的等价性证明

本附录给出一个详细推导，证明预形状空间中端点为 τ_s, τ_e 的测地线表达式与正文中使用的插值形式在数学上等价。

设两端点之间的测地距离为 $\theta = d(\tau_s, \tau_e)$ 。测地线参数化形式可写为

$$\Gamma_{(\tau_s, \tau_e)}(s) = (\cos s) \tau_s + (\sin s) \frac{\tau_e - \tau_s \cos \theta}{\sin \theta}, \quad (\text{A.1})$$

其中 $s \in [0, \theta]$ 。而插值形式写为

$$f_{\text{interp}}(\tau_s, \tau_e, z) = \frac{\sin[(1-z)\theta]}{\sin \theta} \tau_s + \frac{\sin(z\theta)}{\sin \theta} \tau_e, \quad (\text{A.2})$$

其中 $z \in [0, 1]$ ，并令 $s = z\theta$ 。

从测地线公式 (A.1) 出发，将第二项展开可得

$$\begin{aligned} \Gamma_{(\tau_s, \tau_e)}(s) &= (\cos s) \tau_s + \frac{\sin s}{\sin \theta} \tau_e - \frac{\sin s}{\sin \theta} \cos \theta \tau_s \\ &= \left(\cos s - \frac{\cos \theta \sin s}{\sin \theta} \right) \tau_s + \frac{\sin s}{\sin \theta} \tau_e. \end{aligned} \quad (\text{A.3})$$

接下来化简 τ_s 的系数。利用三角恒等式

$$\sin(\theta - s) = \sin \theta \cos s - \cos \theta \sin s, \quad (\text{A.4})$$

可得

$$\cos s - \frac{\cos \theta \sin s}{\sin \theta} = \frac{\sin \theta \cos s - \cos \theta \sin s}{\sin \theta} = \frac{\sin(\theta - s)}{\sin \theta}. \quad (\text{A.5})$$

将 (A.5) 代回 (A.3)，得到

$$\Gamma_{(\tau_s, \tau_e)}(s) = \frac{\sin(\theta - s)}{\sin \theta} \tau_s + \frac{\sin s}{\sin \theta} \tau_e. \quad (\text{A.6})$$

最后，代入 $s = z\theta$ ($z \in [0, 1]$)，便有

$$\Gamma_{(\tau_s, \tau_e)}(z\theta) = \frac{\sin[(1-z)\theta]}{\sin \theta} \tau_s + \frac{\sin(z\theta)}{\sin \theta} \tau_e, \quad (\text{A.7})$$

这与插值函数 $f_{\text{interp}}(\tau_s, \tau_e, z)$ 的表达式 (A.2) 完全一致。因此，测地线参数化形式与插值形式在数学上等价。

附录 B 升维策略对分类准确率的影响

本附录对第 3.2 节中所采用的预形状空间投影步骤中的升维策略进行系统对比，以验证选取复制升维方式的合理性。实验使用的骨干模型、特征提取流程、下游分类器训练与评估方式均与第 3.5.1 节所述相同，数据集为 CIFAR-10@5，骨干网络为预训练 ViT-Tiny。

设训练特征向量为 $v_i \in \mathbb{R}^d$ ，各升维策略通过不同方式将每个标量特征 $v_i[k]$ 扩展为二维关键点坐标 $(v_i[k], g(v_i[k]))$ ，从而生成 $\bar{v}_i \in \mathbb{R}^{2d}$ ，再经中心化与归一化投影至预形状空间 S_*^{2d-3} 。此外，本文还测试了将特征向量前 $d/2$ 维与后 $d/2$ 维分别配对为关键点横、纵坐标的策略（关键点数为 $d/2$ ，对应预形状空间 S_*^{d-3} ）。各升维策略的实验结果如表 B.1 所示。

表 B.1 不同升维策略的分类准确率

升维策略	准确率 (%)
$g(x) = x$ （复制升维，本文方法）	87.92
前 $d/2$ 维与后 $d/2$ 维配对	87.76
$g(x) = \text{BatchNorm}(x)$	87.72
$g(x) = \sin(x)$	87.71

由表可见，四种策略的分类准确率最大差距不超过 0.21 个百分点，升维方式的选择对最终性能的影响较小。这一现象可能源于以下原因：在将升维后的向量投影至预形状空间时，中心化与归一化操作会对不同升维方式所引入的坐标值差异进行整体压缩，使各策略在预形状空间中形成相近的结构，从而导致下游分类性能趋于一致。基于以上结果，本文选取实现最简单且与已有测地线增强方法保持一致的复制升维方式 $g(x) = x$ 。