arXiv:2301.00345v1 [cs.CV] 1 Jan 2023

MTNeuro：用于评估多个抽象层次下脑结构表征的基准测试

Jorge Quesada[1*], **Lakshmi Sathidevi**[1*], **Ran Liu**[1], **Nauman Ahad**[1], **Joy M. Jackson**[1], **Mehdi Azabou**[1], **Jingyun Xiao**[1], **ChristopherLiding**[1], **Matthew Jin**[1], **Carolina Urzay**[1], **William Gray-Roncal**[2], **ErikC.Johnson**[2,†], **Eva L.Dyer**[1,†]1-佐治亚理工学院 2-约翰斯·霍普金斯大学应用物理实验室

摘要

同一图像可以从多个抽象层次进行描述，具体取决于我们关注的是精细细节，还是图像更宏观的属性。在脑图谱绘制中，学习如何自动解析图像，从而构建对小尺度特征（例如细胞或血管的存在）和图像全局属性（例如图像来自哪个脑区）的表征，是一项至关重要且尚待解决的挑战。然而，现有大多数神经解剖学数据集和基准测试一次只考虑单个下游任务。为弥合这一差距，我们推出了一个新数据集及其标注，并设计了多个下游任务，以便从同一图像中以多种方式读取有关脑结构和构造的信息。我们的多任务神经影像基准测试（MTNeuro）基于小鼠脑大范围丘脑皮层切片的体积 X 射线显微断层成像图像，图像分辨率达到微米级，涵盖多个皮层和皮层下区域。我们设计了多种不同的预测挑战，并评估了若干有监督模型和自监督模型，用于脑区预测以及微观结构的像素级语义分割。实验不仅凸显了该数据集丰富的异质性，还揭示了自监督方法如何学习能够捕捉单幅图像多个属性的表征，并在各种下游任务中取得良好表现。数据集、代码和预训练基线模型见：https://mtneuro.github.io/。

1 引言

理解周围的自然环境，需要经过多个层次的感知处理：我们既能识别宏观物体（例如一棵树），也能辨认其中更精细的结构（例如树叶和树枝），还能够识别与情境相关的特征（例如枝叶繁茂程度、高度或季节）。这种多层次感知模式同样适用于医学图像领域：分析医学图像（无论由人类专家还是算法完成），都需要将宏观结构方面的认识（例如一个感兴趣区）与情境相关的微结构信息及人类可理解的特征（例如显微图像中某种细胞类型的密度）结合起来，从而得出诊断或描述目标样本。

*贡献相同。联系作者：（ELD、JQ、LS）{evadyer、jpacora3、lsathidevi3}@gatech.edu；（ECJ Erik。c.johnson@jhuapl.edu；†两位资深作者贡献相同。

第35届神经信息处理系统大会（NeurIPS 2022）数据集与基准赛道。

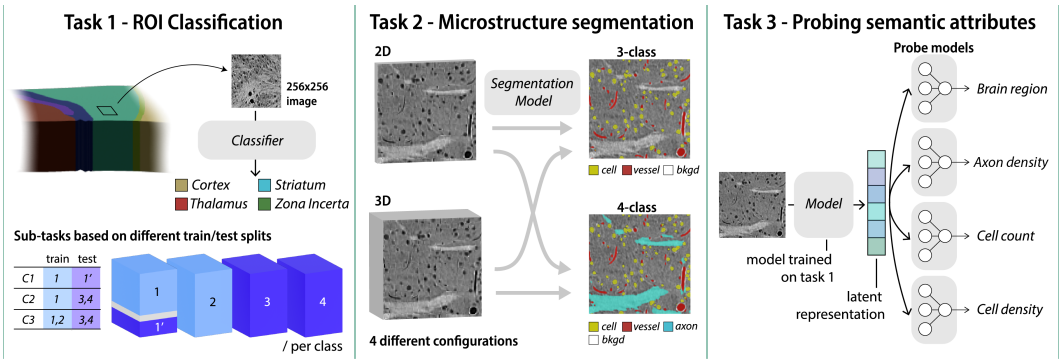


图1：概述 *MTNeuro* 基准测试。任务 1：脑区（宏观结构）分类（3 种配置，数据可用性和测试方案各不相同）；任务 2：像素级微观结构分割（4 种配置，样本维度、跨度和目标类别数量各不相同）；任务3：探究任务 1 所获图像级嵌入中的语义属性。

尤其是，要理解大脑的连接与动态变化，必须分析宏观层面的特性，例如区域级结构（1；2），以及细致的微观结构，例如特定细胞类型的大小或形状（3）。尽管通过多种成像模态和不同尺度的成像，研究人员在揭示大脑内部的特性与结构方面已取得显著进展（4；5；6；7），但现有的神经影像基准大多只针对单一空间尺度进行评估，或面向某一特定下游任务。这有多方面的原因，包括对跨多个尺度的数据进行人工标注成本高昂（8；9）、处理多尺度数据所需的计算成本，以及神经影像技术直到近期才发展出能够以高分辨率采集多区域体积数据的处理流程。

为弥补这一空白，我们推出 MTNeuro 基准（图2）：这是一个多任务、多尺度基准，基于一个大型三维 X 射线显微断层成像数据集，该数据集涵盖小鼠脑的多个区域。代码和数据访问方式见：https://mtneuro.github.io/。我们将数据集托管在脑观测存储服务与数据库（BossDB，一种专用交互式数据库（10））中，并提供集成的数据加载器，以便开展迁移实验和分析。该基准提供了一个统一框架，可用于评估在以下三个不同任务中产生的模型和表征：

- **任务 1 - 脑区图像级分类：**预测给定图像所属的脑区（躯体感觉皮层、纹状体、丘脑、未定带）。
- **任务 2 - 微观结构的像素级分割**：在数据集的四个核心脑区中，对神经微观结构（轴突、细胞体、血管、背景）进行像素级预测。
- **任务3 - 从学习到的图像级嵌入中探测多种语义特征**：从给定图像的表征中估计语义特征（人类可解释的特征，如平均细胞大小或轴突密度）；该表征是在“冻结”训练好的编码器的权重后获得的。

为了了解当前模型在这些不同任务上的表现，我们评估了一系列不同的有监督模型和自监督模型。我们在任务 1 和任务 3 中的结果凸显了自监督方法与监督式方法之间显著的泛化差距，这为进一步评估和开发用于这些任务的自监督学习（SSL）方法带来了有趣的机会。通过在一系列不同模型上针对多种任务进行测试，我们提出的基准测试不仅为评估自监督学习（SSL）方法提供了一个令人振奋的平台，也为努力提取关于微观和宏观尺度脑结构的基本洞见提供了丰富的工具。

2 背景与相关工作

2.1 脑图谱绘制和连接组学领域对基准测试的需求

在过去十年中，我们解析大脑中精细尺度神经解剖结构的能力取得了重大进展。随着这些进展，我们生成了大量涵盖多个空间尺度的脑数据，能够揭示大脑组织的不同特征。在纳米尺度上，电子显微镜提供了皮层小部分区域的详细连接图（11）。在微米尺度上，显微成像技术提供了细胞构筑的详细图像，即神经元和细胞如何组织排列（5）。在更大尺度上同时捕获多个脑区的工作，例如连接图谱和 X 射线显微断层成像数据集（12），提供了有关脑区之间的长程连接与细胞体密度等微观结构及其他脑结构形态特征之间相互作用的信息。

伴随这些新型数据生成工具而来的，是机器学习和计算方法的重大进展，这些方法可用于对这些数据集进行建模和分析，以解决目标检测、分割和分类等问题。虽然这些方法所提供的信息极其丰富，并且在多个尺度上具有丰富的结构，但任何一种方法通常只会针对某一特定尺度上的单项任务进行测试。标注和校对的成本可能相当高昂，而且标注人员通常需要具备扎实的神经解剖学知识（9）。此外，许多致力于提供高质量数据的工作（如（5））并未侧重于构建面向机器学习的基准测试，而是提供参考数据集和资源。

利用机器学习工具，在不同空间尺度上理解这些新兴脑数据集，既是一项挑战，也是一项日益迫切的需求。因此，人们正利用电子显微镜和X射线显微断层成像采集大型太尺度和拍尺度的连接组学数据集，其中包括果蝇 全脑数据（13；14）、小鼠脑的大面积数据（15；3），甚至还有一立方毫米的人类皮层数据（6）。成像技术的进步有望不断扩大成像的空间范围、涉及的物种数量以及成像个体数量。这些数据集具有亚细胞结构（如线粒体和突触）、微观结构（如神经胶质细胞、神经元和脉管系统）以及宏观结构（如脑区、皮层层状结构和长程白质投射）所需的亚微米或纳米级分辨率和广阔的空间覆盖范围。这些数据集具有多尺度特性且规模庞大，因此需要新的机器学习工具（16），也由此凸显了对基准测试的需求，以提取不同尺度上的神经结构表征。

2.2 用于解析脑结构的现有数据集和基准测试

由于空间尺度、神经解剖结构和成像模态种类繁多，针对神经影像数据已经提出了各种分割和分类问题。在宏观尺度上，长期以来，人们一直在为磁共振成像及弥散张量成像、功能性磁共振成像等相关模态中的不同数据集开发基准测试。例如，BraTS数据集（17）聚焦于脑肿瘤的磁共振成像，并推动了许多分割研究（18；19；20；21）。ADNI（22）和 MIRIAD（23）数据集提供了基于磁共振成像的阿尔茨海默病影像，重点追踪疾病进展（24；25；26；27；28；29；30）。英国生物样本库（31）收集了来自 50 万名人类参与者的大量脑部磁共振成像数据。然而，磁共振成像及相关方法的空间分辨率较低，限制了人们观察细胞或亚细胞层面微观结构的能力。

对于微尺度结构，由于标注和处理的规模与复杂性，相关基准测试工作较少。示例问题包括 CREMI 挑战赛（32）中的突触分割，该挑战赛提供了 5 立方微米非各向同性电子显微镜成像数据，突触图像具有纳米分辨率。其他问题定义包括使用文献（33）中的数据进行细胞膜二维图像分割（ISBI 2012 挑战赛），以及细胞三维分割（34；35）。此外，针对轴突实例分割（36）和线粒体分割（37），也开发了专门的基准测试。钙荧光显微成像等具有微米分辨率的其他显微成像方式适用于细胞体分割和功能时间序列数据提取，但缺乏其他微观结构信息。针对双光子钙荧光显微成像的功能轨迹估计，也已建立基准测试，其中包括 spikefinder（38）。这些基准测试对于推动特定空间尺度下特定问题的进展发挥了重要作用。然而总体而言，大多数针对微尺度的基准测试都聚焦于

相对较小的空间范围，且缺少大脑中同样存在的多尺度宏观结构。据我们所知，目前尚无公开可用的脑结构显微断层成像数据集，能够同时提供密集的微观结构标注和宏观尺度标注。

涵盖更大空间范围的项目，例如 BigBrain图谱，提供了小鼠脑的高分辨率切片及尼氏染色对比度（39），但其分辨率仅能分辨约 20μm 的细胞。大型电子显微镜数据集也被用于基准测试性能和分割神经元，并辅以反复的人工校对（40；41；6）；这些工作正产出规模庞大、标注日益复杂的分割数据集。然而，由于规模庞大且仍在持续完善，这些数据并不适合供机器学习领域的广泛研究群体用于大规模基准测试和算法开发。为了加快开发能够在同一高分辨率数据集中处理多个空间抽象层级的机器学习工具，需要涵盖大空间范围且具有高分辨率的基准测试。这将使更广泛的科学界能够将最先进的方法应用于这些重要领域。

3 数据集与任务

3.1 数据集概述

我们的基准测试基于一个大型开放获取的高分辨率（1.17μm各向同性）三维 X 射线显微断层成像数据集构建，该数据集提供了关于脑微观结构的精细信息，以及具有更鲜明全局属性的多种感兴趣区（15）。因此，该数据集独具丰富性，包含宏观尺度 （感兴趣区）和 微观尺度 （细胞、血管、轴突）结构，可在整个数据集中进行研究（42；16）。
完整的体积数据集提供了微米级分辨率，涵盖一个完整脑样本，共计 $5805 \times 1420 \times 720$ 个像素。该数据集覆盖四个感兴趣区：躯体感觉皮层（CTX）、纹状体（STR）、丘脑腹后区（VP）或未定带（ZI）（见图 2 A-B），并在分布于数据集中的二维切片上提供像素级微观结构（细胞、轴突、血管、背景）和宏观结构标签。

为提供更多训练和验证数据，我们将原始数据资源（15）中提供的像素级微观结构和宏观结构标签扩展到了数据中更大的连续体积区域。为扩展像素级标签，我们使用稀疏的二维标注训练了一个 4 类 U-Net模型，并由一位经过训练的专家校对标注，从而创建密集的像素级微观结构标签（见图 2B），将每个点识别为轴突、细胞、血管或背景的一部分。最终整理得到的像素级标签覆盖 4 个感兴趣区，每个感兴趣区包含 360 张 256x256 的密集标注图像。

为了考察不同图像的语义属性，在任务3中，我们利用密集的像素级标签，从重建结果中计算了若干语义特征：(i) 血管密度，(ii) 轴突密度，(iii) 细胞数量，(iv) 细胞大小，以及 (v) 每张切片中的平均细胞间距离。这些语义标签提供了细胞构筑不同特征的信息，可用于解读在该数据集上测试的模型所学习到的嵌入。除了这些微观结构标注外，我们还扩展了原始数据集中的宏观结构标注，将区域标签插值扩展到全部 720 张大小为 5805×1420 的切片。根据这些插值切片，我们提取了 12 个新的子体积（四个感兴趣区各三个），用于考察模型在任务 1 中的泛化能力。

3.2 数据访问

该数据集及其所有对应标签均存储在 BossDB（10）中，即脑观测存储服务与数据库。数据集项目页面见：http://bossdb.org/project/ prasad2020。BossDB 是一个专用于电子显微镜和 X 射线显微断层成像数据集的空间数据库，并可通过 Neuroglancer 无缝进行可视化；Neuroglancer 支持对大规模三维标注体积及其标注进行交互式可视化。所有数据均可公开获取，使用公开登录凭据即可（无需创建账户）。项目页面记录了项目元数据、引用说明，以及数据创建者和数据管理者的信息。

进行基准测试时，数据通过 Python intern API (43) 访问。此 API 可远程连接至 BossDB系统，包括按需下载任意数据三维裁剪数据块，

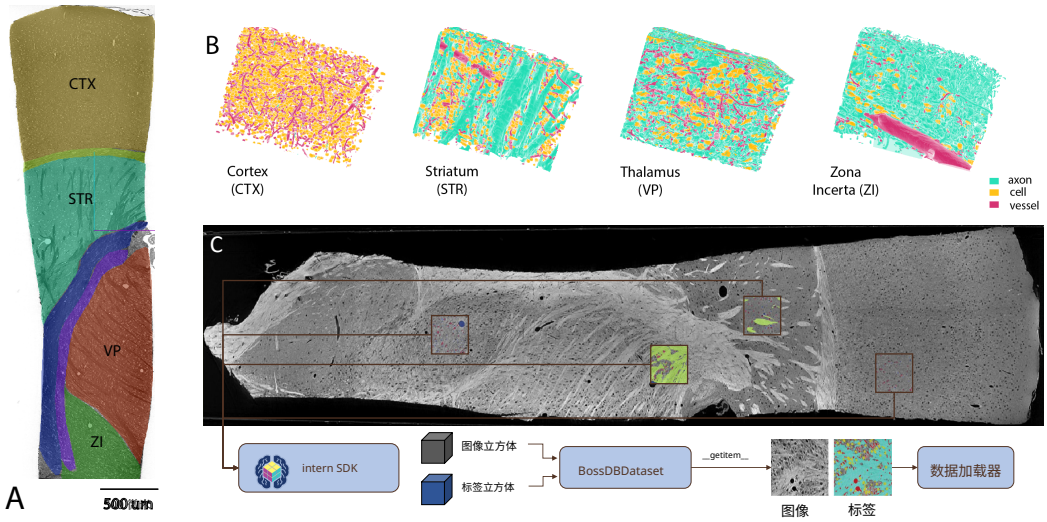


图 2：所用数据集和标注概览 *MTNeuro*。在 A 中，我们展示了该数据集涵盖的多个脑区，包括躯体感觉皮层（CTX）、纹状体（STR）、丘脑（腹侧苍白球，VP）和不确定带（ZI）。每个脑区都包含像素级微观结构的标注，例如轴突、血管和细胞；这些结构以密集的 X射线显微计算机断层扫描体积成像数据呈现（各向同性分辨率为 1.17 微米），如 B 所示。C 展示了处理流程和 BossDB 数据访问方式：通过 Intern API 访问通道，即可获取裁剪数据块，无需下载整个数据集。这使得我们能够为特定任务创建所需的裁剪数据块加载器。

其中包括原始图像和标注，无需将整个数据集下载到磁盘。为便于使用此 API，我们提供了一个 PyTorch数据加载器，以加快算法开发和测试。我们还提供了示例 Jupyter笔记本，演示如何下载任务裁剪数据块，并将其保存为 NumPy文件，以便在其他框架中进行开发。任务由特定任务的 JSON文件定义，其中指定了各任务的元数据。图 2C 展示了该数据加载器的用法，其中详细说明了如何通过 BossDB 高效访问数据集。
数据被组织为多个通道。原始图像、宏观结构标注和微观结构标注各自位于独立的通道中。详情如下。

原始图像：所有任务均使用相同的原始图像，它们是单色（灰度）图像，像素值为 8 位无符号整数。原始数据总体积为 $720 \times 1420 \times 5805$ 体素，分辨率为$1.17\mu m$。这些数据可通过原始图像通道获取：https://api.bossdb.io/v1/ mgmt/resources/prasad/prasad2020/image。

宏观结构标注：任务 1 和任务3使用不同脑区（宏观结构）的密集像素级标注：CTX（标签 0）；STR（标签 1）；腹侧苍白球（标签 2）；不确定带（标签 3）。在此尺度/层级下，各类别的样本数量相同，因此类别是均衡的。该标签数据以 64 位无符号整数表示，可通过宏观结构标注通道访问：https://api.bossdb.io/v1/mgmt/resources/prasad/ prasad_analysis/roi_labels。

微观结构标注：任务 2 和任务3使用大脑微观结构的像素级标注。 $256 \times 256 \times 360$ 的体积数据均进行了密集的微观结构标签标注。标签为：0：无标签（背景）；1：血管；2：细胞；3：有髓轴突。该数据以 64 位无符号整数表示，可通过微观结构标注通道访问：https://api. bossdb.io/v1/mgmt/resources/prasad/prasad_analysis/pixel_labels。
我们方法的一个关键特点是，数据访问和训练基础设施能够在 BossDB 中的神经影像数据集之间移植。这使得代码和基线方法能够轻松扩展到存储在 BossDB 中的新体积成像数据集。这些数据集包括来自新物种的数据、带有新微观结构标签（突触、膜、线粒体）的数据，以及带有新宏观结构标签（脑区、实验状态）的数据。通过修改数据加载器 JSON，开发者可以指定不同的 BossDB 数据集、空间区域和标注（标签）来源。这种灵活性满足了支持该数据集中不同任务的需求，也将助力这些基线模型进一步开展训练和部署

到新的数据集。这项工作为新兴的大规模神经影像数据集开发机器学习工具作出了重要贡献。

3.3 任务

3.3.1 任务 1：脑感兴趣区（ROI）的图像级分类

分析跨越多个不同脑区的成像数据时，一个重要问题是全局图像特征与样本来源脑区之间的相关程度（16）。因此，我们可以将其视为一个分类问题：从数据中提取一个小图像块（或小区域），并估计该样本来自 4 个脑区中的哪一个。鉴于整个体积数据中可用于此任务的样本数量丰富，我们可以将此任务细分为三种不同的训练方案（见图1），具体如下。这三种训练方案结合起来，使我们能够评估随着训练期间可用数据不断增加，不同类型的模型能够如何泛化。

ROI-C1. 对于此子任务，我们使用 4 个密集标注立方体（图1中任务 1以蓝色显示），每个立方体分别对应一个感兴趣区（CTX、STR、腹侧苍白球或不确定带）。我们将四个子体积分别划分为训练集和测试集：选取前 300 (256×256) 个图像切片用于训练，最后 50 张图像用于测试，并在训练数据和测试数据之间留出 10 张切片，以避免任何结构重叠。因此，此子任务的总样本量为 1400 张图像，其中训练集包含 1200 张图像，测试集包含 200 张图像。

ROI-C2. 在这个子任务中，我们评估模型在数据集更大三维上下文中的新区域上进行测试时的表现。我们为每个类别额外提取两个 $256 \times 256 \times 360$ 立方体（如图1、任务 1所示），作为模型的额外测试数据。该子任务使用 4080 张图像，其中训练集有 1200 张，测试集有 2880 张。

ROI-C3. 在这个子任务中，我们评估模型在能够利用更多数据进行学习时的表现。我们为每个类别额外提取一个立方体，作为模型的额外训练数据来源，并使用与 ROI-C2 相同的测试集。该子任务使用 5520 张图像，其中训练集有 2640 张，测试集有 2880 张。

3.3.2 任务 2：微观结构的像素级分割

全面绘制脑数据图谱的另一项重要要求，是正确识别细胞和血管等脑微观结构。在这项任务中，我们利用来自 4 个脑区感兴趣区（宏观结构；皮层、纹状体、腹侧苍白球和不确定带）的体积裁剪块的密集像素级微观结构标签，并评估不同基线方法在预测准确率方面的表现：将脑区测试体积中的每个像素分类为相应的脑微观结构类别（血管、细胞、轴突和背景）。

三类分割任务。我们首先考虑将图像进行像素级分割，分为三类：细胞体、血管或其他（背景和轴突）。我们在主要子体积中使用与任务 1 中 ROI-C1 相同的训练集和测试集划分，这些子体积均经过像素级密集标注（训练集 300 张图像，测试集 50 张图像，数据集之间间隔 10 个切片）。此子任务使用 1200 张图像进行训练，使用 200 张图像进行测试。

4 类分割任务。在此任务中，我们考虑将图像进行像素级分割，分为四类：细胞体、血管、背景或轴突。在进行轴突密集分割时，由于即使是人工标注者也难以在该子体积中可靠地分割轴突，因此我们从训练集和测试集中移除了不确定带区域。此子任务使用 900 张图像进行训练，使用 150 张图像进行测试。

3.3.3 任务3：从学习到的图像级嵌入中探测多种语义特征

在这项任务中，我们希望探索能否从任务 1 中学到的图像级表征解码出语义特征和人类可理解的特征。如果可行，就能提供一种方法，在无需使用脑图谱绘制中通常需要的、用于计算这些属性的昂贵像素级标签的情况下，构建包含微观结构信息的可解释的图像级特征图。

表 1： 脑区预测（任务 *1*）的图像分类准确率结果。

	感兴趣区 - *C1*	感兴趣区 - *C2*	感兴趣区 - *C3*
监督学习	0.88 ± 0.03	0.77 ± 0.03	0.88 ± 0.02
使用 Mixup 的监督式	0.90 ± 0.04	0.78 ± 0.03	0.90 ± 0.02
BYOL	0.88 ± 0.02	0.76 ± 0.02	0.97 ± 0.01
MYOW	0.90 ± 0.02	0.78 ± 0.05	0.98 ± 0.01
MYOW-m	0.94 ± 0.02	0.78 ± 0.03	0.98 ± 0.01
PCA	0.59	0.25	0.07
NMF	0.62	0.27	0.50

具体而言，我们尝试通过一个简单的线性读出，预测图像的以下全局属性：(i) 血管密度，(i) 细胞计数，(iii) 平均细胞大小，(iv) 轴突密度，以及 (v) 细胞间平均距离。对于监督学习模型，这些模型经过训练，可将图像分类到各自对应的脑区；然而，对于不使用脑区级标签来指导学习的自监督学习方法，模型的潜在空间中可能编码了图像的其他全局属性。

4 结果

4.1 任务 1：大脑感兴趣区图像级分类

实验设置。在此任务中，我们使用通过监督式方法和自监督方法学习到的表征，将不同图像分类到若干候选脑区（CTX、STR、VP 和 ZI）中。此任务中的所有模型均使用 Resnet18 编码器（44）进行训练。我们对两种监督学习模型进行了基准测试：一种采用标准正则化方法进行训练（丢弃率为 20%，权重衰减系数为 0.3），另一种则采用 Mixup（45）进行训练。鉴于体积数据中图像样本具有潜在的共享结构，我们还考虑了若干适用于此任务的自监督学习方法：（i）BYOL（46）、（ii）MYOW（47），以及（iii）我们测试的 MYOW 变体，该变体使用单个投影器和预测器（MYOW-merged，或 MYOW-m）。对于自监督学习模型，我们遵循标准流程：训练后冻结网络权重，然后在表征之上训练一个线性层。这可以让我们了解仅经过线性变换后，自监督学习损失对数据类别的捕捉效果如何。所有模型的潜在维度均为 256。作为额外的基线方法，我们还使用主成分分析（PCA）和非负矩阵分解（NMF）从数据中提取了 256 维嵌入，并在这些表征上训练了一个线性层。

在实验中，我们使用 5 个具有不同随机种子的训练实例评估所有方法，并报告总体准确率均值和标准差。所有模型均使用学习率为 0.03 的随机梯度下降优化器训练 100 个训练轮次。有关实验设置和模型的更多详情，请参阅附录第 3 节。

脑感兴趣区域分类结果。 我们在表 1 中报告了脑感兴趣区分类三个子任务的结果。在第一个子任务（ROI-C1）中，我们发现许多监督学习和自监督学习模型取得了相近的准确率，其中 MYOW-m 模型在这一有限训练条件下取得了最高准确率。在 ROI-C2 中，我们通过评估在 ROI-C1 中训练的模型在更大数据集其他部分的子体积上的表现，测试模型的泛化能力（表 1，ROI-C2）。脑区之间存在显著异质性，因此我们可以将其视为一种领域泛化。在这种情况下，我们观察到准确率显著下降，这可归因于数据中的领域偏移。在 ROI-C3 中，我们加入另一组训练数据（见图 2），并在与上一个子任务相同的体积数据上进行测试。在这种情况下，我们观察到自监督学习方法显著优于监督学习模型：自监督学习模型的准确率甚至高于小规模情形下的水平（97–98%），而有监督模型在新增训练数据后仅取得较为有限的提升（88–90%）。因此，我们可以观察到自监督学习模型与监督学习模型之间存在显著差距，这表明接触额外数据（在此指相对于任务 ROI-C1 新增的立方体）会大幅提升自监督学习模型对未见数据的泛化能力。

4.2 任务 2：微观结构的像素级分割

在任务 2 中，我们考虑像素级分割的四种不同变体。第一种变体使用二维模型进行像素级分割。第二种变体则使用三维模型。第三种

表 2：用于像素级分割任务（任务 *2*）的模型的 *F1* 分数和交并比得分。*I.* 二维像素级分割

方法	指标	三分类 背景 +轴突	血管	细胞	Avg.	4 类 Bg	血管	细胞	轴突	Avg.
二维 U-Net	F1	0.99	0.76	0.85	0.87 ± 0.012	0.97	0.82	0.87	0.94	0.90 ± 0.003
二维 U-Net	IoU	0.98	0.64	0.75	0.79 ± 0.014	0.89	0.70	0.77	0.60	0.74 ± 0.008
MA-Net	F1	0.99	0.79	0.87	0.88 ± 0.003	0.97	0.83	0.87	0.94	0.90 ± 0.002
MA-Net	IoU	0.98	0.68	0.78	0.81 ± 0.003	0.89	0.71	0.78	0.76	0.78 ± 0.011
FPN	F1	0.99	0.72	0.84	0.85 ± 0.01	0.96	0.73	0.84	0.93	0.86 ± 0.004
FPN	IoU	0.97	0.59	0.73	0.76 ± 0.015	0.87	0.59	0.72	0.72	0.72 ± 0.021
U-Net++	F1	0.99	0.79	0.87	0.89 ± 0.002	0.97	0.81	0.85	0.93	0.89 ± 0.015
U-Net++	IoU	0.98	0.68	0.78	0.81 ± 0.002	0.88	0.68	0.75	0.73	0.76 ± 0.036
PAN	F1	0.98	0.60	0.80	0.79 ± 0.035	0.95	0.69	0.80	0.93	0.84 ± 0.007
PAN	IoU	0.96	0.46	0.66	0.69 ± 0.039	0.85	0.53	0.67	0.76	0.70 ± 0.014
PSPNet	F1	0.97	0.48	0.74	0.73 ± 0.013	0.94	0.54	0.71	0.91	0.78 ± 0.012
PSPNet	IoU	0.94	0.39	0.61	0.65 ± 0.043	0.82	0.38	0.55	0.74	0.62 ± 0.015

II. 三维像素级分割

方法	指标	三分类 背景 +轴突	血管	细胞	Avg.	4 类 Bg	血管	细胞	轴突	Avg.
三维 U-Net	F1	0.99	0.77	0.87	0.88 ± 0.006	0.93	0.76	0.80	0.87	0.84 ± 0.032
三维 U-Net	IoU	0.98	0.65	0.76	0.80 ± 0.007	0.81	0.62	0.67	0.50	0.65 ± 0.045
VNetLight	F1	0.99	0.75	0.83	0.85 ± 0.012	0.90	0.65	0.73	0.76	0.76 ± 0.063
VNetLight	IoU	0.97	0.61	0.70	0.76 ± 0.013	0.78	0.46	0.58	0.43	0.56 ± 0.061
HighResNet	F1	0.99	0.74	0.84	0.85 ± 0.019	0.89	0.51	0.73	0.77	0.72 ± 0.083
HighResNet	IoU	0.97	0.61	0.72	0.77 ± 0.026	0.73	0.35	0.58	0.42	0.52 ± 0.075

种变体采用 4 类设置，即从涉及的脑区中移除不确定带（因为即使是人工标注者，也难以准确区分该区域中的轴突）。第四种变体采用三分类设置，使用全部 4 个脑区（皮层、纹状体、腹侧苍白球和不确定带），但仅考虑 3 个类别（血管、细胞、背景+轴突；不进行轴突分割）。

实验设置。 在我们的实验中，我们使用选定的一组二维和三维模型进行像素级分割。我们对每个模型分别进行超参数调优，通过在训练集上训练并在验证集上评估，寻找最优学习率和批量大小。每个模型使用其最优学习率和批量大小训练 20 轮，并在 5 次训练实例中进行评估（每次实例使用各自的随机种子）。我们报告每个模型的类别级 F1 分数、类别级交并比，以及这两项指标的总体均值和标准差（表 2）。我们不报告准确率，因为对于这一特定的像素级分割任务，由于类别不平衡，准确率无法恰当地反映模型性能（有关训练集和测试集中不同类别的细分情况，请参见附录第 4.2.3 节）。

我们在二维分割任务中使用的模型包括标准二维 U-Net 模型（48；49），以及 'segmentation models.pytorch' 库（50）中的部分模型：MAnet（51）、特征金字塔网络（52）、U-Net++ （53）、PAN（54）和 PSPNet（55）。我们在三维分割任务中使用的模型包括标准三维 U-Net 模型（48），以及 'MedicalZooPytorch'（56）中的部分模型：VNetLight（57；56）和 HighResNet（58）。更多详情请参见附录第 4 节。

二维像素级分割。 所选模型在我们的二维像素级分割任务上的结果列于表 2 的第 I 部分，并在图 3 中可视化展示。模型以单张切片作为输入，并在训练期间以批次方式送入。对于训练和评估，我们同时考察三分类和 4 类设置；所有模型均使用 EfficientNet-b7 编码器，因为在尝试过的 25 种编码器中，它的表现最佳。从结果可以看出，在二维模型中，MA-Net 的整体表现最佳，在 4 类设置下平均交并比为 0.78；其次是 U-Net++ ，平均交并比为 0.76（表 2）。从类别级交并比和类别级 F1 分数可以看出，对于大多数表现最佳的模型而言，细胞与血管是最难区分的组成部分；对于人工标注者来说，若没有额外的三维上下文，仅凭二维切片也很难辨认它们。此外，我们可以看到，从 4 类设置切换到三分类设置后，平均交并比（跨类别和模型计算）从 0.72 提高到 0.79。这表明性能如预期般有所提升，因为三分类设置包含更多切片（更多数据）和更少类别（难度更低）。

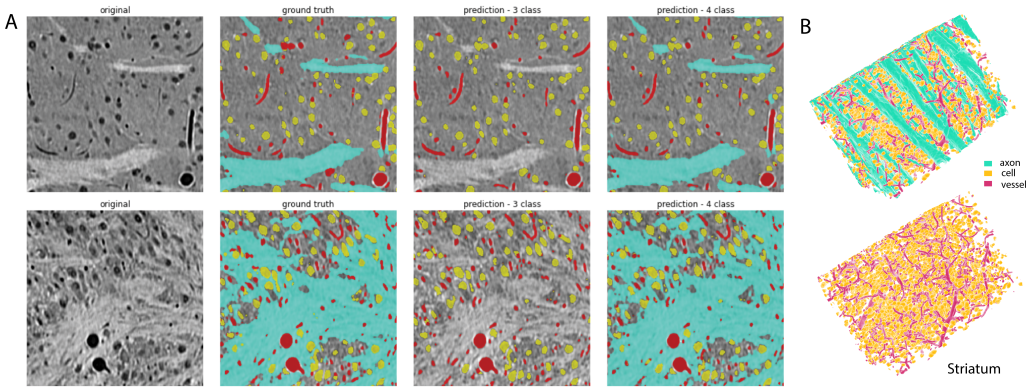


图 3：像素级分割任务（任务 2）的预测结果可视化。从左到右依次为：原始图像、叠加真实标注的图像，以及 U-Net 模型在三分类和 4 类设置下的预测结果（上排为纹状体，下排为腹侧苍白球）。轴突以青色显示，血管以红色显示，细胞体以黄色显示。B）4 类（上）和三分类（下）设置下纹状体预测结果的三维重建。

三维像素级分割。对于三维情况，我们采用与二维情况相同的分类细目和准确率度量（表 2 的第 II 部分）。为向模型提供三维输入，我们将连续的切片（8 张切片）作为一个子体积输入，并在三维空间中联合预测所有切片。U-Net 模型表现最佳，在三分类设置下交并比为 0.80，与最佳二维模型的表现相当。这可能是因为我们以不重叠的方式定义子体积，因此尽管三维上下文有所改善，模型看到的样本/输入也更少。出于同样的原因，从三分类设置切换到 4 类设置时，由于样本更少，性能有所下降。增加样本数量或允许重叠，都可能使三维模型表现得更好。

4.3 任务3：探测从学习得到的图像级嵌入中提取的多种语义特征

在任务 3 中，我们探索如何利用任务 1 中训练的模型所学习到的潜在空间，预测不同的语义属性（例如血管或细胞的密度）。

为此，我们利用高质量的密集微观结构标注，提取每张图像的属性信息，包括血管或轴突所占的像素比例、细胞计数和大小，以及细胞与其最近邻之间的平均距离。有关此任务实验设置的更多详细信息，请参见附录第5节。

我们使用任务 1 中训练好的模型，冻结其权重，并计算 ROI-C1 中所有已标注图像的表征（包括训练集和测试集，共 360 张切片）。从这些潜

Ta表 3： 任务*3*： R^2 在任务 *1* 中训练的监督学习模型和自监督学习模型的多任务特征读出得分。

I. 从单个子体积（*ROI-C1*）训练的模型中进行线性读出

方法	血管	轴突	细胞计数	细胞大小	距离（k=1）
监督学习	0.77 ± 0.06	0.94 ± 0.01	0.67 ± 0.06	0.61 ± 0.05	0.48 ± 0.05
使用 Mixup 的监督式	0.82 ± 0.02	0.95 ± 0.00	0.71 ± 0.02	0.67 ± 0.03	0.47 ± 0.02
BYOL	0.85 ± 0.01	0.94 ± 0.01	0.75 ± 0. 01	0.69 ± 0.01	0.49 ± 0.01
MYOW	0.85 ± 0.01	0.94 ± 0.01	0.74 ± 0.01	0.69 ± 0.01	0.50 ± 0.02
MYOW-m	0.87 ± 0.01	0.95 ± 0.01	0.77 ± 0.01	0.69 ± 0.01	0.51 ± 0.01
PCA	0.75	0.82	0.55	0.47	0.31
NMF	0.81	0.85	0.59	0.55	0.34

II. 在两个子体积（*ROI-C3*）上训练的模型的线性读出

方法	血管	轴突	细胞计数	细胞大小	距离 (k=1)
监督学习	0.79 ± 0.02	0.94 ± 0.02	0.73 ± 0.02	0.63 ± 0.04	0.49 ± 0.02
使用 Mixup 的监督式	0.75 ± 0.04	0.88 ± 0.04	0.64 ± 0.04	0.54 ± 0.07	0.37 ± 0.05
BYOL	0.88 ± 0.00	0.96 ± 0.00	0.79 ± 0.00	0.73 ± 0.01	0.53 ± 0.02
MYOW	0.88 ± 0.01	0.96 ± 0.00	0.79 ± 0.01	0.72 ± 0.01	0.52 ± 0.01
MYOW-m	0.87 ± 0.01	0.96 ± 0.01	0.78 ± 0.01	0.72 ± 0.01	0.53 ± 0.01
PCA	0.75	0.82	0.53	0.46	0.29
NMF	0.75	0.83	0.56	0.49	0.31

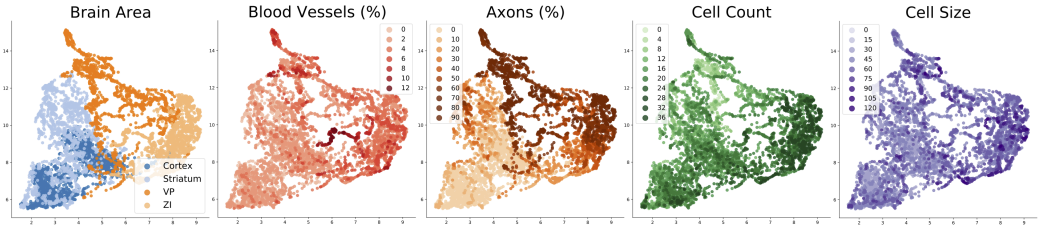


图 4： *MYOW-m* 在任务 *1*（*ROI-C1*）中学习到的表征可视化，每幅图像的全局语义属性以不同颜色显示。从左到右，潜变量依次按脑区（类别）、血管百分比、轴突百分比、细胞计数和细胞大小着色。

变量特征出发，我们在这些表征上拟合一个简单的线性回归模型，以预测每个目标语义属性。此任务的数据有限，因此我们报告所有潜变量上的 R^2 （此处不划分训练集和测试集）。

我们在表 3 中报告了 R^2 数值，涵盖任务 1 和任务 2 中使用的密集标注体积内全部 360 个切片；这些模型分别在 ROI-C1 上训练（表 3 I）以及在 ROI-C3 上使用更多数据训练（表 3 II）。在这两种情况下，我们发现自监督学习模型的表现都显著优于其余基线模型。在体积训练条件下，这一差距更为明显，凸显了监督学习与自监督学习方法在为广泛任务提供优质表征方面的泛化差异 (59)。在图 4 中，我们使用 UMAP (60) 投影 MYOW-m（ROI-C1）的嵌入，并叠加语义特征。

5 讨论

本文介绍了一项多任务基准测试，用于分析涵盖多个脑区的高分辨率神经影像数据中的脑结构。此外，我们还构建了通用基础设施，以便基于 BossDB 框架中的数据集训练模型，例如可适配不同数据集的数据加载器。这将推动大规模体积神经影像数据在机器学习工具开发中的应用。

局限性与未来工作。 在神经科学领域，获取高质量的带标签数据（尤其是任务 2 中提供的此类密集分割任务所需的数据）往往成本极高（61）。人工标注和数据校对过程十分繁重，因此限制了我们能够提供用于模型训练的带标签和已标注数据量，也限制了可评估的分布偏移范围。尽管这是本研究的一项局限，但它也准确反映了该领域面临的挑战，因此需要采用更具标签效率的学习方法，例如我们重点介绍的自监督学习方法。展望未来，我们希望利用本研究中测试的模型生成更多高质量的标注数据，以进一步提升模型性能和分割效果。

在设计当前的基准测试时，我们着重构建了一项多尺度挑战，使变化源于脑结构的差异，而非不同的制备方法或成像参数。因此，我们聚焦于单只动物，其完整的大脑体积较大，涵盖了许多具有异质性的脑区。虽然这可能会限制模型向新样本或数据集的泛化能力，但也回应了不同脑区异质性这一常被忽视的特性。我们的结果表明，即使在同一个大脑中，不同脑区之间也存在丰富的异质性，这使某些模型难以泛化。这也揭示了监督学习模型与自监督学习模型之间存在显著的泛化差距。

未来，我们希望将这项工作拓展到从其他非卷积架构中学习脑结构模型（例如，基于点云的神经结构模型（62）），并将我们的工具应用于新的多尺度脑数据集，或许可以采用新的光片技术（63）或全脑尺度扩展技术（64；65）。

更广泛的影响。 脑数据的高度异质性，加上该基准测试所涵盖任务在规模和性质上的差异，使其成为机器学习社区中一项颇具挑战性且实用的资源。此外，我们希望所提供的数据集与任务，以及配套的代码库和 BossDB 基础设施，能够帮助加速开发适用于社区正在收集的新兴高分辨率神经影像数据集的机器学习技术。

致谢

本项目获得了美国国家科学基金会项目 IIS-2039741、美国国立卫生研究院项目 1R01EB029852-01、美国国立卫生研究院项目 R01MH126684、美国国立卫生研究院项目 R24MH114785 的资助，此外还获得阿尔弗雷德·斯隆基金会、麦克奈特基金会和 CIFAR-阿兹列里全球学者项目的慷慨捐助。

参考文献

[1] Danielle S Bassett 和 Edward T Bullmore，“重新审视小世界脑网络”，《神经科学家》，第 23 卷，第 5 期，第 499–516 页，2017年。[2] Ran Liu、Cem Subakan、Aishwarya H. Balwani、Jennifer Whitesell、Julie Harris、Sanmi Koyejo 和 Eva L. Dyer，“一种用于解读脑结构群体水平变异的生成建模方法”，载于《医学图像计算与计算机辅助介入——*MICCAI 2020*》，Cham，2020年，第 257–266 页，Springer International Publishing。[3] Casey M Schneider-Mizell、Agnes L Bodor、Forrest Collman、Derrick Brittain、Adam A Bleck- ert、Sven Dorkenwald、Nicholas L Turner、Thomas Macrina、Kisuk Lee、Ran Lu 等，“吊灯细胞的解剖结构与功能揭示了一种分布多变但普遍存在的信号”，*bioRxiv*，2020年。[4] Timothy EJ Behrens 和 Olaf Sporns，“人类连接组学”，*Current Opinion in Neurobiology*，第 22 卷，第 1 期，第 144–153 页，2012年。[5] Michael Hawrylycz、Lydia Ng、David Feng、Susan Sunkin、Aaron Szafer 和 Chinh Dang，“艾伦脑图谱”，施普林格生物信息学与神经信息学手册，第 1111–1126 页，2014年。[6] Alexander Shapson-Coe、Michał Januszewski、Daniel R Berger、Art Pope、Yuelong Wu、Tim Blakely、Richard L Schalek、Peter H Li、Shuohong Wang、Jeremy Maitin-Shepard 等，“对人类大脑皮层拍字节级片段的连接组学研究”，*bioRxiv*，2021年。[7] Federico Scala、Dmitry Kobak、Matteo Bernabucci、Yves Bernaerts、Cathryn Ren´e Cadwell、Jesus Ramon Castro、Leonard Hartmanis、Xiaolong Jiang、Sophie Laturnus、Elanine Miranda 等，“小鼠运动皮层中转录组细胞类型的表型变异”，《自然》，第 598 卷，第 7879 期，第 144–150 页，2021年。[8] Jeff W Lichtman、Hanspeter Pfister 和 Nir Shavit，“连接组学的大数据挑战”，《自然·神经科学》，第 17 卷，第 11 期，第 1448–1454 页，2014年。[9] Alessandro Motta、Meike Schurr、Benedikt Staffler 和 Moritz Helmstaedter，“纳米尺度连接组学中的大数据与对训练标签的迫切需求”，*Current Opinion in Neurobiology*，第 55 卷，第 180–187 页，2019年。

[10] Robert Hider Jr, Dean Kleissas, Timothy Gion, Daniel Xenes, Jordan Matelsky, Derek Pryor, Luis Rodriguez, Erik C Johnson, William Gray-Roncal 和 Brock Wester，《大脑观测存储服务与数据库（bossdb）：一种面向拍字节级神经科学发现的云原生方法》，《神经信息学前沿》，第16卷，2022年。

[11] Narayanan Kasthuri, Kenneth Jeffrey Hayworth, Daniel Raimund Berger, Richard Lee Schalek, Jos´e Angel Conchello, Seymour Knowles-Barley, Dongil Lee, Amelio V´azquez-Reina, Verena Kaynig, Thouis Raymond Jones 等，《新皮层体积的饱和重建》，细胞，第162卷，第3期，第648–661页，2015年。

[12] Eva L Dyer, William Gray Roncal, Judy A Prasad, Hugo L Fernandes, Doga G¨ursoy, Vin-cent DeAndrade, Kamel Fezzaa, Xianghui Xiao, Joshua T Vogelstein, Chris Jacobsen 等，《使用X射线显微断层成像量化中尺度神经解剖学》，*eNeuro*，第4卷，第5期，2017年。

[13] C Shan Xu、Michal Januszewski、Zhiyuan Lu、Shin-ya Takemura、Kenneth J Hayworth、Gary Huang、Kazunori Shinomiya、Jeremy Maitin-Shepard、David Ackerman、StuartBerg 等，《成年果蝇中央脑的连接组》，*bioRxiv*，2020年。

[14] Zhihao Zheng、J Scott Lauritzen、Eric Perlman、Camenzind G Robinson、Matthew Nichols、Daniel Milkie、Omar Torrens、John Price、Corey B Fisher、Nadiya Sharifi 等，《成年黑腹果蝇大脑的完整电子显微镜体积》，细胞，第174卷，第3期，第730–743页，2018年。[15] Judy A Prasad、Aishwarya H Balwani、Erik C Johnson、Joseph D Miano、Vandana Sampathku- mar、Vincent De Andrade、Kamel Fezzaa、Ming Du、Rafael Vescovi、Chris Jacobsen 等，《用于表征脑异质性的三维丘脑皮层数据集》，*Scientific Data*，第7卷，第1期，第1–7页，2020年。[16] Aishwarya Balwani、Joseph Miano、Ran Liu、Lindsey Kitchell、Judy A Prasad、Erik C Johnson、William Gray-Roncal 和 Eva L Dyer，《X射线图像中神经结构的多尺度建模》，载于*2021*年*IEEE*国际图像处理会议 *(ICIP)*。IEEE，2021年，第141–145页。[17] Bjoern H Menze、Andras Jakab、Stefan Bauer、Jayashree Kalpathy-Cramer、Keyvan Farahani、Justin Kirby、Yuliya Burren、Nicole Porz、Johannes Slotboom、Roland Wiest 等，《多模态脑肿瘤图像分割基准（BRATS）》，*IEEE*医学影像学汇刊，第34卷，第10期，第1993–2024页，2014年。[18] Liang Chen、Paul Bentley、Kensaku Mori、Kazunari Misawa、Michitaka Fujiwara 和 Daniel Rueckert，《利用图像上下文恢复进行医学图像分析的自监督学习》，医学图像分析，第58卷，第101539页，2019年。[19]Dheerendranath Battalapalli、B. Rao、P. Yogeeswari、Chandrasekharan Kesavadas 和 Venkateswaran Rajagopalan，《一种适用于低分辨率、非连续切片临床MRI数据集的最优脑肿瘤分割算法》，*BMCMedical Imaging*，第22卷，2022年05月。[20] Javeria Amin、Muhammad Anjum、Muhammad Sharif、Saima Jabeen、Seifedine Kadry 和 Pablo Ger，《一种结合集成迁移学习与量子变分分类器的脑肿瘤检测新模型》，*Computational Intelligence andNeuroscience*，第2022卷，第1–13页，2022年04月。[21] Ghazanfar Latif、Ghassen Ben Brahim、D. N. F. Awang Iskandar、Abul Bashar 和 Jaafar Al- ghazo，《利用基于深度神经网络的特征与支持向量机分类器进行胶质瘤分类》，*Diagnostics*，第12卷，第4期，2022年。[22] C.R. Jack Jr.、M.A. Bernstein、N.C. Fox、P. Thompson、G. Alexander、D. Harvey、B. Borowski、P.J. Britson、J.L. Whitwell、C. Ward、A.M. Dale、J.P. Felmlee、J.L. Gunter、D.L.G. Hill、R. Kil- liany、N. Schuff、S. Fox-Bosetti、C. Lin、C. Studholme、C.S. DeCarli、G. Krueger、H.A. Ward、G.J. Metzger、K.T. Scott、R. Mallozzi、D. Blezek、J. Levy、J.P. Debbins、A.S. Fleisher、M. Al- bert、R. Green、G. Bartzokis、G. Glover、J. Mugler 和 M.W. Weiner，《阿尔茨海默病神经影像学倡议（ADNI）：MRI方法》，*JournalofMagneticResonance Imaging*，第27卷，第4期，第685–691页，2008年。[23] Ian B Malone、David Cash、Gerard R Ridgway、David G MacManus、Sebastien Ourselin、Nick C Fox 和 Jonathan M Schott，《MIRIAD——公开发布的多个时间点阿尔茨海默病磁共振成像数据集》，*NeuroImage*，第70卷，第33–36页，2013年。[24] Manhua Liu、Danni Cheng、Kundong Wang 和 Yaping Wang，《用于阿尔茨海默病诊断的多模态级联卷积神经网络》，*Neuroinformatics*，第16卷，第295–308页，2018年。[25] Donghuan Lu、Karteek Popuri、Gavin Weiguang Ding、Rakesh Balachandar 和 Mirza Faisal Beg，《基于多尺度深度神经网络的FDG-PET图像分析，用于阿尔茨海默病早期诊断》，医学图像分析，第46卷，第26–34页，2018年。[26] Jyoti Islam 和 Yanqing Zhang，《利用深度卷积神经网络集成系统进行阿尔茨海默病诊断的脑部MRI分析》，*Brain informatics*，第5卷，第2期，第1–14页，2018年。

[27] Mingxia Liu、Jun Zhang、Ehsan Adeli 和 Dinggang Shen，《通过深度多任务多通道学习联合分类与回归以诊断阿尔茨海默病》，*IEEE* 生物医学工程汇刊，第 66 卷第 5 期，第 1195–1206 页，2019年。[28]Alejandro Puente-Castro、Enrique Fernandez-Blanco、Alejandro Pazos 和 Cristian R. Munteanu，《基于深度学习技术自动评估阿尔茨海默病诊断》，*Computers in Biology and Medicine*，第 120 卷，第 103764 页，2020年。[29] Eman N. Marzban、Ayman M. Eldeib、Inas A. Yassine、Yasser M. Kadah 及阿尔茨海默病神经退行性疾病倡议，《使用卷积神经网络根据扩散张量图像诊断阿尔茨海默病》，*PLOS ONE*，第 15 卷第 3 期，第 1–16 页，2020年3月。[30] N. Deepa 和 S.P. Chokkalingam，《利用算术优化算法优化 VGG16，以早期检测阿尔茨海默病》，*Biomedical Signal ProcessingandControl*，第 74 卷，第 103455 页，2022年。[31] Sudlow C、Gallacher J、Allen N、Beral V、Burton P、Danesh J 等，《英国生物样本库：用于确定中老年人多种复杂疾病病因的开放获取资源》，*PLOS Med 12(3): e1001779.*，2015年。[32] Zhihao Zheng、Cam G. Robinson、Daniel Milkie、Eric Perlman、John Price、Davi Bock、Misha Kazhdan、Khaled Khairy、Bill Karsh、Eric Trautman、Peter Li、Chris Ordish、Chong Zhang、Ju- lia Buhmann、Jan Funke 和 Stephan Saalfeld，《MICCAI 电子显微镜图像神经回路重建挑战赛》，https://cremi.org/data/，2022年。[33] Albert Cardona、Stephan Saalfeld、Stephan Preibisch、Benjamin Schmid、Anchi Cheng、Jim Pulokas、Pavel Tomancak 和 Volker Hartenstein，《通过计算机辅助连续切片电子显微镜对果蝇脑进行微观与宏观结构的综合分析》，*PLOSBiology*，第 8 卷第 10 期，第 e1000502 页，2010年。[34] Kisuk Lee、Jonathan Zung、Peter Li、Viren Jain 和 H Sebastian Seung，《SNEMI3D 连接组学挑战赛中的超人准确率》，*arXiv* 预印本 *arXiv:1706.00120*，2017年。[35] Step hen M Plaza，《通过有针对性的校对，从电子显微镜图像大规模重建神经连接组》，载于*Deep Learning and Data Labeling for Medical Applications*，第 249–258 页。施普林格，2016年。[36] Donglai Wei、Kisuk Lee、Hanyu Li、Ran Lu、J Alexander Bae、Zequan Liu、Lifu Zhang、M´arcia dos Santos、Zudi Lin、Thomas Uram 等，《AxonEM 数据集：脑皮层区域的三维轴突实例分割》，载于医学图像计算与计算机辅助干预国际会议。施普林格，2021年，第 175–185 页。[37] Donglai Wei、Zudi Lin、Daniel Franco-Barranco、Nils Wendt、Xingyu Liu、Wenjie Yin、Xin Huang、Aarush Gupta、Won-Dong Jang、Xueying Wang 等，《MitoEM 数据集：基于电子显微镜图像的大规模三维线粒体实例分割》，载于医学图像计算与计算机辅助干预国际会议。施普林格，2020年，第 66–76 页。[38] Philipp Berens、Jeremy Freeman、Thomas Deneux、Nikolay Chenkov、Thomas McColgan、Ar- tur Speiser、Jakob H Macke、Srinivas C Turaga、Patrick Mineault、Peter Rupprecht 等，《基于社区的基准测试提升了双光子钙成像数据的尖峰率推断》，*PLOS Computational Biology*，第 14 卷第 5 期，第 e1006157 页，2018年。[39] Katrin Amunts、Claude Lepage、Louis Borgeat、Hartmut Mohlberg、Timo Dickscheid、Marc-´Etienne Rousseau、Sebastian Bludau、Pierre-Louis Bazin、Lindsay B Lewis、Ana-Maria Oros- Peusquens 等，《BigBrain：超高分辨率三维人脑模型》，*Science*，第 340 卷第 6139 期，第 1472–1475 页，2013年。[40] Michał Januszewski、J¨orgen Kornfeld、Peter H Li、Art Pope、Tim Blakely、Larry Lindsey、Jeremy Maitin-Shepard、Mike Tyka、Winfried Denk 和 Viren Jain，《利用泛洪填充网络高精度自动重建神经元》，*Nature Methods*，第 15 卷第 8 期，第 605–610 页，2018年。

[41] Peter H Li、Larry F Lindsey、Michał Januszewski、Mike Tyka、Jeremy Maitin-Shepard、Tim Blakely 和 Viren Jain，《利用泛洪填充网络和局部重新对齐自动重建连续切片电子显微镜果蝇脑》，*Microscopy and Microanalysis*，第 25 卷 S2 期，第 1364–1365 页，2019年。[42] Aishwarya H Balwani 和 Eva L Dyer，《利用深度特征学习建模脑结构变异》，载于*2019*年第*53*届阿西洛马会议：信号、系统与计算机。IEEE，2019年，第 1186–1191 页。[43] Jordan Matelsky、Luis Rodriguez、Daniel Xenes、Timothy Gion、Robert Hider、Brock Wester 和 William Gray-Roncal，《Intern：用于可扩展、可复现神经科学的集成工具包》，*bioRxiv*，2020年。[44] 何恺明、张祥雨、任少卿和孙剑，《用于图像识别的深度残差学习》，载于*IEEE/CVF*计算机视觉与模式识别会议（*CVPR*）论文集，2016年，第 770–778 页。[45] Hongyi Zhang、Moustapha Cisse、Yann N Dauphin 和 David Lopez-Paz，《Mixup：超越经验风险最小化》，*arXiv* 预印本 *arXiv:1710.09412*，2017年。[46] Jean-Bastien Grill、Florian Strub、Florent Altch´e、Corentin Tallec、Pierre Richemond、Elena Buchatskaya、Carl Doersch、Bernardo Avila Pires、Zhaohan Guo、Mohammad Ghesh- laghi Azar 等，《自举潜变量（BYOL）——一种新的自监督学习方法》，神经信息处理系统进展，第 33 卷，第 21271–21284 页，2020年。[47] Mehdi Azabou、Mohammad Gheshlaghi Azar、Ran Liu、Chi-Heng Lin、Erik C Johnson、Kiran Bhaskaran-Nair、Max Dabagia、Bernardo Avila-Pires、Lindsey Kitchell、Keith B Hengen 等，《挖掘自己的视图：通过跨样本预测进行自监督学习》，*arXiv* 预印本 *arXiv:2102.10106*，2021年。[48] Olaf Ronneberger、Philipp Fischer 和 Thomas Brox，《U-Net：用于生物医学图像分割的卷积网络》，*CoRR*，第 abs/1505.04597 卷，2015年。[49] Johannes Schmidt，《PyTorch-2D-3D-U-Net 教程》，https://github.com/johschmidt42/PyTorch-2D-3D-UNet-Tutorial，2021年。[50]Pavel Yakubovskiy，《segmentation models pytorch》，https://github.com/qubvel/segmentation_models.pytorch，2020年。[51] Tongle Fan、Guanglei Wang、Yan Li 和 Hongrui Wang，《MA-Net：用于肝脏和肿瘤分割的多尺度注意力网络》，*IEEE Access*，第 8 卷，第 179656–179665 页，2020年。[52] Tsung-Yi Lin、Piotr Doll´ar、Ross Girshick、何恺明、Bharath Hariharan 和 Serge Be- longie，《用于目标检测的特征金字塔网络》，载于*IEEE/CVF*计算机视觉与模式识别会议（*CVPR*）论文集，2017年，第 936–944 页。[53] Zongwei Zhou、Md Mahfuzur Rahman Siddiquee、Nima Tajbakhsh 和 Jianming Liang，《UNet：一种用于医学图像分割的嵌套式 U-Net 架构》，载于医学图像分析中的深度学习与临床决策支持的多模态学习，Danail Stoy- anov、Zeike Taylor、Gustavo Carneiro、Tanveer Syeda-Mahmood、Anne Martel、Lena Maier- Hein、Jo˜ao Manuel R.S. Tavares、Andrew Bradley、Jo˜ao Paulo Papa、Vasileios Belagiannis、Jacinto C. Nascimento、Zhi Lu、Sailesh Conjeti、Mehdi Moradi、Hayit Greenspan 和 Anant Madabhushi 编，Cham，2018年，第 3–11 页，Springer International Publishing。[54] Hanchao Li、Pengfei Xiong、Jie An 和 Lingxue Wang，《用于语义分割的金字塔注意力网络》，*CoRR*，第 abs/1805.10180 卷，2018年。[55] Hengshuang Zhao、Jianping Shi、Xiaojuan Qi、Xiaogang Wang 和 Jiaya Jia，《金字塔场景解析网络》，*CoRR*，第 abs/1612.01105 卷，2016年。[56] Adaloglou Nikolaos，《医学图像分析中的深度学习：使用三维深度神经网络对多模态脑 MRI 分割进行比较分析》，硕士论文，帕特雷大学，2019年，https://github.com/black0017/MedicalZooPytorch。

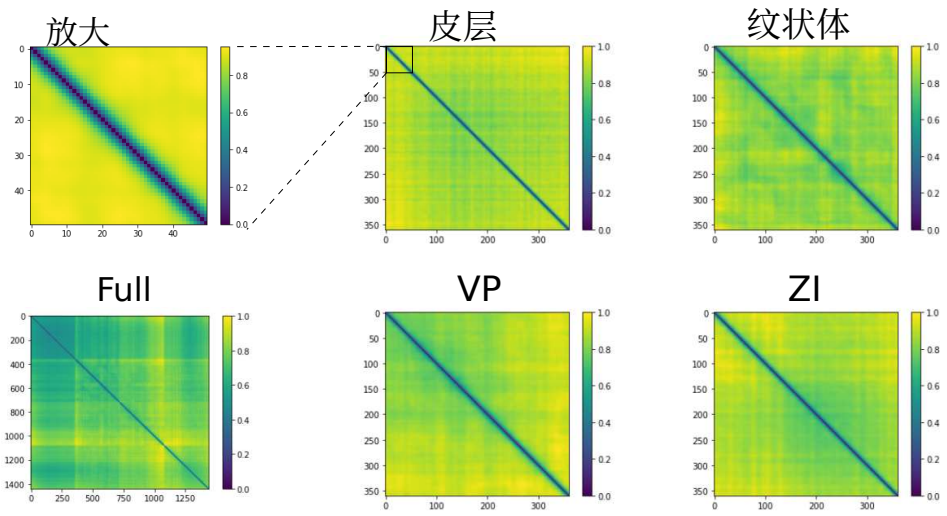
[57] Fausto Milletari、Nassir Navab 和 Seyed-Ahmad Ahmadi，“V-net：用于体积医学图像分割的全卷积神经网络”，*CoRR*，第 abs/1606.04797 卷，2016年。[58] Wenqi Li、Guotai Wang、Lucas Fidon、S´ebastien Ourselin、M. Jorge Cardoso 和 Tom Ver- cauteren，“关于三维卷积网络的紧凑性、效率与表征：将脑区划分作为前置任务”，*CoRR*，第 abs/1707.01992 卷，2017年。[59] Atharva Tendle 和 Mohammad Rashedul Hasan，“自监督表征泛化能力研究”，*MachineLearning with Applications*，第 6 卷，第 100124 页，2021年。[60] Leland McInnes、John Healy 和 James Melville，“UMAP：用于降维的统一流形近似与投影”，*arXivpreprint arXiv:1802.03426*，2018年。[61] David Rolnick 和 Eva L Dyer，“大规模神经解剖学数据集的生成模型与抽象”，*Current Opinion in Neurobiology*，第 55 卷，第 112–120 页，2019年，《机器学习、大数据与神经科学》。[62] Joy M Jackson、Ran Liu 和 Eva L Dyer，“通过分层点云网络构建不同脑区的表征”，载于*Medical Imaging with Deep Learning*，2022年。[63] Elizabeth M.C. Hillman、Venkatakaushik Voleti、Wenze Li 和 Hang Yu，“神经科学中的光片显微镜”，*AnnualReview ofNeuroscience*，第 42 卷，第 1 期，第 295–313 页，2019年，PMID：31283896。[64] Scott Trinkle、Sean Foxley、Narayanan Kasthuri 和 Patrick La Rivi`ere，“同步辐射X射线显微CT作为小鼠全脑扩散磁共振成像的验证数据集”，*MagneticResonanceinMedicine*，第 86 卷，第 2 期，第 1067–1076 页，2021年。[65] Sean Foxley、Vandana Sampathkumar、Vincent De Andrade、Scott Trinkle、Anastasia Sorokina、Katrina Norwood、Patrick La Riviere 和 Narayanan Kasthuri，“以五个数量级的分辨率对单个小鼠脑进行多模态成像”，*NeuroImage*，第 238 卷，第 118250 页，2021年。[66] Paul A. Yushkevich、Joseph Piven、Heather Cody Hazlett、Rachel Gimpel Smith、Sean Ho、James C. Gee 和 Guido Gerig，“用户引导的解剖结构三维主动轮廓分割：显著提高效率与可靠性”，*Neuroimage*，第 31 卷，第 3 期，第 1116–1128 页，2006年。

附录

1 数据集详情

1.1 图像之间的空间自相关

我们计算了主要子体积中图像的互协方差，以考察相邻或邻近图像切片之间存在多大程度的相关性。分析结果证实，经过 5–7 张切片后，图像之间的互相关会持续下降（见图 A1）。根据这项分析，我们在所有子体积的训练集与测试集之间裁去 10 张切片。



Fig图 A1：互协方差矩阵。所有 4 个所考察区域的类别互协方差（右），放大图 - 视图中 皮层交叉协方差（左上）和整个数据集的交叉协方差（左下）。

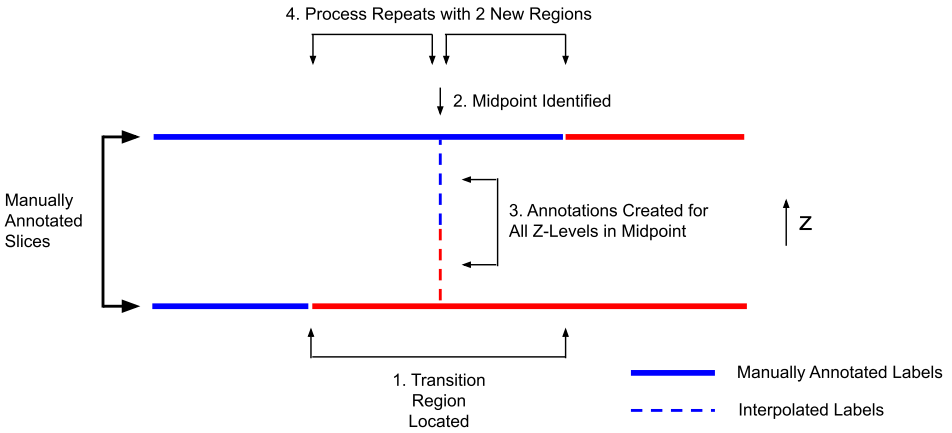
1.2 密集像素级标注

利用文献 (15) 中的稀疏标注，我们训练了一个二维 U-Net，将图像数据分割为 4 个类别（血管、细胞体、轴突和背景），并将其应用于四个体积数据中尚未标注的切片。随后，校对员使用 ITK-Snap (66) 在 y-z 平面上逐一检查体积中每张切片的标注。在校对每张切片时，首先更正标签被分到不同类别的结构（U-Net 的错误）。然后，检查标签未被拆分的结构是否正确，并更正错误标注。最后，标注员检查下一张 Z 切片，以确保各组成部分连续。对体积中的每张切片都重复这一流程。为确保标注一致性，同一位标注员随后再次校对该体积，这次沿 x-z 平面逐一检查。校对完成后，我们生成了四个感兴趣区域的密集标注体积，每个体积的大小为 256x256x360。

1.3 插值感兴趣区标注

从手动标注的 z 层级切片插值得到该 X 射线显微断层扫描数据集的完整感兴趣区域（ROI）标注，这些切片大约相隔 50 个像素。为生成插值层，研究人员实现了一种递归算法，将两个标注类别之间的边界从一个已有层线性移动到下一层：找到移动的边界区域后，确定移动的中点，并沿该中点创建完整的单体素宽标注，使其跨越所有必要的 z 层级。此过程将过渡区域分成大致相等的两部分，并在这两个过渡空间中分别重复递归过程，直至整个过渡区域都完成标注。

然而，此过程无法处理所有边界过渡。因此，约 10% 的边界过渡是通过可视化软件应用程序 ITK-SNAP 手动标注的。z 层级 109 至 459（含）均提供完整的插值 ROI 标注。



图A2：*ROI*标注插值的递归过程。我们采用四步递归过程，根据手动标注的图层生成经过插值的ROI标注层。该过程递归重复，直到整个过渡区域都完成ROI标注插值。

2 数据访问

为便于公众访问数据，数据集存储在BossDB系统中。值得注意的是，该数据库支持高效访问大型体积数据集中的任意裁剪数据块。项目页面位于 https://bossdb.org/project/prasad2020。无需用户账户。使用公共凭据即可只读访问数据，无需创建用户账户。

原始图像的资源标识符为 bossdb://prasad/prasad2020/image，对应“prasad”实验、“prasad2020”实验和“image”通道。原始图像数据采用无符号8位整数格式，建议使用 ZYX 格式进行索引。无效数据区域（例如图像体积之外的区域）赋值为 0。最大和最小索引分别为 $z = [0, 720]$、$y = [0, 1420]$ 和 $x = [0, 5805]$。数据具有 $1.17\mu m$ 各向同性分辨率。

标注图像可通过资源标识符 bossdb://prasad/prasad_analysis/pixel_labels 和 bossdb://prasad/prasad_analysis/roi_labels 获取，它们对应“prasad”集合、“prasad analysis”实验，以及“pixel labels”和“roi labels”通道。标注数据采用无符号64位整数格式，首选使用 ZYX格式进行索引。无效数据区域（例如图像体积之外）赋值为 0。至于原始图像通道，最大和最小索引分别为 $z = [0, 720]$、$y = [0, 1420]$ 和 $x = [0, 5805]$。数据具有 $1.17\mu m$ 各向同性分辨率。

可通过 Python intern数组API 访问数据，该 API 支持类似 NumPy 的引用方式。使用此库，用户可以创建一个 intern数组，即 image_array = array(boss_url)，其中 boss_url 是上文列出的资源标识符之一。随后，用户可以使用类似 NumPy 的语法访问通道中的数据。下面的代码示例展示了如何下载对应于皮层的切块。

从 Intern 导入数组 图像 =数组(boss_网址) 数据
=图像[110:379, 900:1156, 4600:4856]

我们提供了一个数据加载器，用于快速开发 PyTorch 模型；本研究中测试的所有模型均使用该数据加载器。创建时，数据加载器会加载任务配置“.json”文件，其中指定了任务的参数。支持两种运行模式：download=true 和 download=false。对于前一种模式，创建数据加载器时，数据会被下载到本地并存储

为 NumPy 文件。如果 NumPy 文件已存在，则改为从磁盘加载。在后一种情况下，数据会被下载到内存中，而不会写入磁盘。数据存储在一个 NumPy 张量中，其中包含来自皮层、纹状体、腹侧苍白球、不确定带的拼接切片（用于四类别问题）。根据配置，项目检索函数可以提供整数区域标签（用于任务 1）或微观结构掩膜（用于任务 2）。基本的输入数据配置参数如下：

- **图像通道** - 从中提取原始图像数据的 BossDB 通道。
- **标注通道** - 从中提取图像标注（宏观或微观结构）的 BossDB 通道。

- **cor 的 x 范围** - 来自皮层切片在 x 轴（完整数据上）的范围。
- **cor 的 y 范围** - 来自皮层切片在 y 轴（完整数据上）的范围。
- **stri 的 x 范围**- 来自纹状体切片在 x 轴（完整数据上）的范围。
- **stri 的 y 范围**- 来自纹状体切片在 y 轴（完整数据上）的范围。
- **VP 范围**- VP 切片在完整数据 x 轴方向上的范围。
- **VP 范围**- VP 切片在完整数据 y 轴方向上的范围。
- **ZI 范围**- ZI 切片在完整数据 x 轴方向上的范围。
- **ZI 范围**- ZI 切片在完整数据 y 轴方向上的范围。
- **Z 训练集** - 用作训练集的 z 轴方向范围（切片）。
- **Z 验证集** - 用作验证集的 z 轴方向范围（切片）。
- **Z 测试集**——用于测试集的 z 轴范围（切片）。建议在训练/验证集与测试集之间留出 10 个切片的缓冲区。
- **体积z值**——体积切片中包含的切片数量。对于二维情况，单个切片是模型的输入（一个体积切片包含 1 个切片）。对于三维情况，体积切片是模型的输入。

代码仓库 https://github.com/MTNeuro/MTNeuro 包含用于数据访问和运行模型的脚本及 Python 笔记本。对于不使用 PyTorch 的用户，我们提供了用于下载图像和标注 NumPy 文件的 Python 笔记本。这些笔记本基于为该数据集开发的原始数据访问笔记本：https://github.com/nerdslab/xray-thc。可以通过 pip包管理器安装代码仓库及其依赖项。代码仓库的结构如下：

- MTNeuro：用于运行 PyTorch数据加载器和 PyTorch模型的核心代码
- 笔记本：为不使用 PyTorch 的用户提供访问笔记本的途径
- 脚本：用于运行各任务 PyTorch模型的示例脚本，可作为开发新算法的基础

3 任务 1 的更多详情

3.1 模型

3.1.1 有监督模型

- Resnet18：该模型的 18 层版本，它将层重新定义为相对于层输入学习残差函数，从而高效训练更深层架构 (44)。

- Resnet18 + Mixup：采用上文所述的相同模型架构（Resnet18），但通过 Mixup 进行训练。Mixup 是一种数据增强方法，它通过凸组合混合输入及其对应的标签，已被证明能显著提升监督分类模型的性能（45）

3.1.2 自监督模型

我们测试了几种无需负样本即可生成良好数据表征的基于对比学习的自监督学习方法（46）。所有自监督学习模型均使用 Resnet18 编码器主干网络进行训练，以便与接受评估的监督学习模型进行比较。

我们测试了以下自监督学习模型：

- BYOL：BYOL 采用在线网络和目标网络的镜像结构，二者相互学习（46）。具体而言，在线网络通过最小化潜在空间中目标图像视图 x 与增强视图 x_a之间的距离，尝试预测目标图像视图的编码。

- MYOW：MYOW 在 BYOL 的基础上，也最小化视图之间在潜在空间中的距离，但采用视图挖掘方法，在数据集中搜索表征空间中的邻近样本。随后，通过额外的预测器网络，将增强样本和挖掘样本整合到统一的潜在空间中（47）。

- MYOW-m：MYOW-m 是 MYOW 的扩展，整合了 MYOW 的样本挖掘流程，但对增强视图和挖掘视图使用同一个预测器，而非 MYOW 提出的级联投影器设计。

由于所有这些自监督学习方法都依赖某种变换来处理训练过程中的视图，因此我们为这三种方法选择了尽可能简单的数据增强：随机裁剪图像样本中宽度和高度各为一半的图像块，以生成增强视图和挖掘视图（见图 A3）。为统一对这些方法的评估，我们评估训练后编码器对图像样本四个角分别所属类别的分类表现。

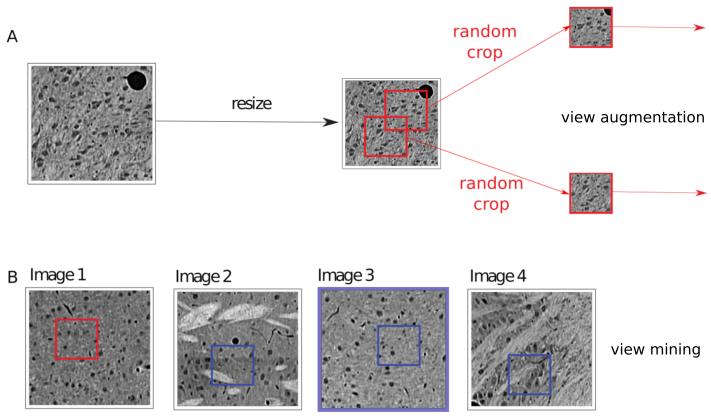


图 A3：自监督方法的视图生成。 A 部分展示了如何从给定的图像样本中提取随机裁剪，以生成增强视图。B 部分示意了视图挖掘过程：将样本池中的不同随机裁剪与目标视图（红色）进行比较，并选择表征空间中距离最近的图像作为挖掘得到的视图（图像 3，已突出显示）。

3.2 训练的更多细节

3.2.1 配置文件

训练时提供两种类型的配置文件作为输入：

- **网络配置** -一个“.json”文件，其中包含优化器设置、要使用的数据增强、相应监督式或自监督模型的参数，以及训练时使用的随机种子。

- **任务配置** - 一个“.json”文件，指定训练所需访问的数据切片相关信息。包括要访问的数据数据库通道和区域级标注、每个区域切片的 x 和 y 范围、切片大小（对应所应用的下采样级别）、训练集-验证集-测试集划分，以及每个切片样本的深度（为 1，因为此任务侧重于二维处理）。

设置	值立方体集合 A
cor 的 x 范围	[4600,4856]
cor 的 y 范围	[900,1156]
stri 的 x 范围	[3700,3956]
stri 的 y 范围	[500,756]
VP 范围	[3063,3319]
VP 范围	[850,1106]
ZI 范围	[1543,1799]
ZI 范围	[650,906]
Z_训练集	[110, 379]
Z_验证集	[380, 409]
Z_测试集	[420, 470]

表 A1： 任务 *1*（*ROI-C1*）和任务 *2* 使用的数据集配置

设置	Values 立方体集合 B	Values 立方体集合 C	Values 立方体集合 D
cor 的 x 范围	[5312,5568]	[5056,5312]	[4600,4856]
cor 的 y 范围	[388,644]	[644,900]	[400,656]
stri 的 x 范围	[3828,4084]	[3344,3600]	[3800,4056]
stri 的 y 范围	[912,1168]	[400,656]	[244,500]
VP 范围	[2551,2807]	[2151,2407]	[2295,2551]
VP 范围	[850,1106]	[950,1206]	[694,950]
ZI 范围	[1287,1543]	[1031,1287]	[1799,2055]
ZI 范围	[906,1162]	[906:1162]	[650:906]

表 A2: *ROI-C2*（*C* 和 *D*）及 *ROI-C3*（*B*、*C* 和 *D*）中使用的附加立方体组的坐标 *)*

3.2.2 模型设置

基本模型设置如下:

- **模型**- 用于训练分类任务的模型（监督学习时为分类器，自监督学习时为编码器主干网络）。
- **类别**- 模型的输出类别数量。
- **方法**-（仅适用于自监督学习）要训练的自监督方法。

3.2.3 输入数据配置

表 1 中 ROI-C1 列结果所使用的数据集配置列于表 A1。ROI-C2 和 ROI-C3 中使用的额外立方体坐标见表 A2。

3.2.4 训练与评估设置

我们使用 0.03 的学习率、256 的批量大小（为确保线性层评估稳定而选择）以及 5 个不同的随机种子值：1、100、350、631 和 872，训练所有有监督和自监督模型。我们汇总这 5 次结果，计算每种所考虑方法的分类准确率均值和标准差。对于 Resnet18-Mixup 设置，我们评估训练过程中应用数据增强的 5 种不同概率：0.01、0.1、0.3、0.5 和 0.7；并报告均值准确率表现最佳的设置。

3.3 补充实验

我们在表 A3 中报告了额外的测试结果：在 ROI-C1 训练设置下评估所有纳入考察的模型，但对图像样本采用不同的下采样因子：4 倍、2 倍和全分辨率（不进行下采样）。我们观察到，分辨率越高，性能越好（因为模型能够获取更多信息）。令人意外的是，与监督学习方法相比，自监督学习方法从 2 倍下采样切换到全分辨率测试时，性能提升幅度显著更大；而监督学习方法的性能仅略有提升。这进一步

表 A3： 脑区预测的图像分类结果（任务 *1*）。

	ROI-C1		
	下采样 *=4*	下采样 *=2*	全分辨率
Resnet18	0.83 ± 0.06	0.88 ± 0.03	0.87 ± 0.1
Resnet18-Mixup	0.85 ± 0.06	0.90 ± 0.04	0.91 ± 0.03
BYOL	0.83 ± 0.04	0.88 ± 0.02	0.98 ± 0.01
MYOW	0.84 ± 0.04	0.90 ± 0.02	0.96 ± 0.03
MYOW-m	0.84 ± 0.05	0.94 ± 0.02	0.99 ± 0.01
PCA	0.59	0.59	0.59
NMF	0.55	0.62	0.54

支持了我们的观察：与监督学习模型相比，自监督学习模型从接触更多数据中获益更多。在 4 倍下采样设置下，每个自监督学习训练实例平均需要 30 分钟训练；在 2 倍下采样设置下，需要 1.5 小时；在全分辨率设置下，需要 4 小时（单个随机种子的运行时间，训练使用 RTX 3090 图形处理器节点）。在任务 1 的其余实验中，我们选择使用 2 倍下采样，因为它在性能和运行时间之间取得了良好的平衡。

4 关于任务 2 的更多详细信息

4.1 基线方法

4.1.1 二维模型

我们用于二维分割任务的模型如下：

- 标准的二**维 U-Net**模型（48；49）
- 从“segmentation models.pytorch”库（50）中选取的模型：**– MAnet** - 一种采用多尺度注意力机制的模型，最初为肝脏和肝肿瘤分割而设计（51），**– FPN** - 将用于目标检测的特征金字塔网络架构修改后用于图像分割（52），**– U-Net++**– 一种为医学图像分割开发的嵌套式 U-Net 架构（53），**– PAN** （金字塔注意力网络）- 一种融入空间金字塔注意力结构的模型，旨在利用语义分割中的全局上下文信息（54），**– PSPNet** （金字塔场景解析网络） - 一种采用金字塔池化和场景解析网络以更好地学习全局上下文信息的模型（55）。

4.1.2 三维模型

我们用于三维分割任务的模型如下：

- 标准的 三**维 U-Net**模型（48），
- 从“MedicalZooPytorch”中选取的模型（56）：**– VNetLight**——一种更轻量的 V-Net 卷积网络架构，专为体积分割而开发（57；56），**– HighResNet**——一种紧凑的高分辨率卷积网络，用于体积分割，最初展示于脑部磁共振图像脑区划分的前置任务中（58）。

4.2 训练详情

4.2.1 配置文件

训练时需要提供两种类型的配置文件：

• **网络配置——**一个包含优化器、数据增强设置、模型设置、评估设置、输出保存设置以及训练所用随机种子的'.json'文件。

• **任务配置** - 一个".json"文件，用于指定训练过程中要访问的数据切片相关信息。内容包括要访问的数据和标注所对应的数据库通道、各区域切片的 x 和 y 范围、切片大小、训练集-验证集-测试集划分、体积切片的大小（三维），以及训练采用 3 类还是 4 类设置。表 2 中结果所使用的任务配置列于表 A1。

这两个文件共同完整地指定了训练运行所需的配置。

4.2.2 模型设置

基本模型设置如下：

• **编码器名称** - 与模型搭配使用的编码器。此项仅适用于二维模型和"UNet 3D"模型。更多信息请参阅"segmentation models.pytorch"库（（50））。表-2 中训练结果所用的编码器为"efficientnet-b7"，因为在尝试的多个编码器中，它的性能最佳。

• **编码器权重** - 模型要使用的预训练权重。表 2 中二维模型和「UNet 3D」的结果使用了在 ImageNet 上训练的权重。

• **输入通道**- 输入通道的数量。

• **类别**- 输出类别的数量。

4.2.3 类别比例详情

在微观结构层级/尺度上，各类别在不同脑区中的频率如下：

• 皮层 -背景：93%；血管：2.33%；细胞：4.64%；轴突：0%
• 纹状体 - 背景：72.63%；血管：2.5%；细胞：5.66%；轴突：19.22%
• 腹侧苍白球 - 背景：22.75%；血管：4.73%；细胞：5.73%；轴突：66.8%
• 不确定带 - 背景：32.15%；血管：5.4%；细胞：9.41%；轴突：53.04%
• 总计 - 背景：55.13%；血管：3.74%；细胞：6.36%；轴突：34.77%

4.2.4 超参数和随机种子

为了为每个模型选择最优超参数，我们通过使用以下学习率训练模型来进行超参数搜索：0.1、0.05、0.01、0.005、0.001；以及以下批量大小：2、4、6、8、10。随后选出表现最佳的学习率和批量大小，并使用该最优学习率和批量大小，分别以随机种子 1、100、350、631 和 872 训练每个模型的 5 个独立实例。这 5 次结果（针对每个模型）用于计算性能指标的均值和标准差（$Mean \pm SD$）（结果见表 2）。此外，在多个模型的比较中发现，Adam优化器的表现优于随机梯度下降，因此所有训练均使用 Adam优化器。各模型在每种设置下找到的最优超参数列于表 A4。

5 任务3的更多详情

为了从 4 个密集连接的立方体中的微观结构标注提取语义特征，我们首先分离出相关的像素级标签（细胞、血管或轴突）。提取出所需类别并将其他所有内容标记为背景后，我们可以进行连通分量分析，为每个图像样本计算所需的语义特征（细胞计数、大小和距离，以及轴突和血管密度）。

计算出不同任务的语义特征后，我们使用在任务 1 ROI-C1（表 3，上）和 ROI-C3 上训练的模型，提取 4 个立方体中所有样本的表征

表 A4： 表 *2* 中任务 *2* 模型使用的最佳超参数 *I.* 二维像素级分割

模型	三分类 学习率	三分类 批量大小	不含不确定带的四分类 学习率	不含不确定带的四分类 批量大小
二维 U-Net	0.1	8	0.01	4
MA-Net	0.01	2	0.001	8
FPN	0.01	2	0.01	4
U-Net++	0.001	8	0.01	8
PAN	0.01	10	0.001	8
PSPNet	0.01	4	0.1	2

II. 3D 像素级分割

模型	三分类 学习率	三分类 批量大小	不含 *ZI* 的四分类 学习率	不含 *ZI* 的四分类 批量大小
三维 U-Net	0.005	1	0.01	1
VNetLight	0.01	1	0.01	1
HighResNet	0.005	1	0.001	1

（表 3，下），并使用 scikit-learn 拟合线性回归，以预测语义特征，并报告整个感兴趣子体积上的 R^2 。此外，我们在图 A4 中展示了使用在两种不同数据集条件（ROI-C1 和 ROI-C3）下训练的 BYOL 编码器得到的 4 个感兴趣立方体的表征及其对应语义特征的可视化结果。

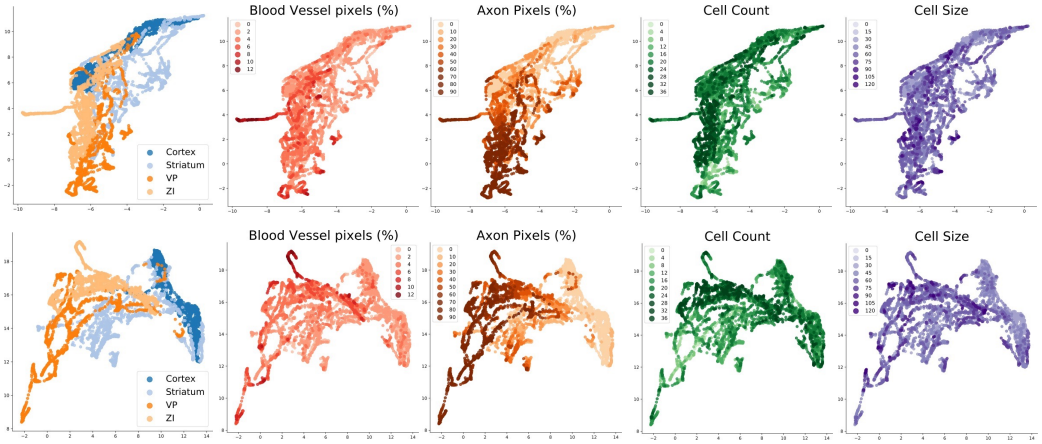


图 A4：语义特征叠加于二维学习表征上的可视化。此处，第一行展示了学习得到的 BYOL（ROI-C1）嵌入的二维 U-map 投影，并在潜变量上叠加了不同的语义属性。从左到右，我们依次按照脑区（类别）、血管占比、轴突占比、细胞计数和细胞大小对嵌入进行着色。第二行展示了学习得到的 BYOL（ROI-C3）嵌入的二维投影，其中同样叠加了不同的语义属性。

6 维护、许可与伦理考量

该数据集、数据访问 API 和数据加载器由 BossDB 团队（bossdb.org）维护。该团队受托担任 BRAIN 倡议纳米尺度连接组学数据集的正式存档机构。团队正在依据 FAIR 原则（https://www.go-fair. org/fair-principles/）制定标准，以确保永久标识符和广泛的可访问性。数据采用 CC-BY-4.0 许可提供。基线代码以开源形式托管在 GitHub 组织中；欢迎社区创建分支并提交拉取请求，供代码库维护者审核。随着未来挑战的基线代码库不断改进，相关变更将作为新的带版本号的发布推送至 MTNeuro 代码库。

所用数据集包含动物数据，这些数据是在获批的 IACUC 方案下收集的，详情见原始论文。本数据集中不包含人类来源数据。我们强调，这一

数据集旨在开发用于结构神经科学数据基础分析的算法，但未来也可能有人不当使用这些数据来开发具有临床相关性的算法，从而造成负面的社会影响。由于高分辨率 X射线显微断层成像向临床成像方式的泛化情况尚不明确，因此不鼓励这种用途。