# MatSSL: 金相图像分割的稳健自监督表征学习

Hoang Hai Nam Nguyen<sup>1,2†</sup>, Phan Nguyen Duc Hieu<sup>1,2†</sup>, Ho Won Lee<sup>1,2</sup> <sup>1</sup>Korea Institute of Materials Science <sup>2</sup>University of Science and Technology

namhhn@ust.ac.kr, hieupnd@ust.ac.kr, h.lee@kims.re.kr

<sup>†</sup> Contribute equally

### Abstract

MatSSL 是一种简化的自监督学习 (SSL) 架构, 它在 骨干网络的每个阶段采用门控特征融合,以有效整合 多层次的特征表示。目前金属材料的显微图像分析依 赖于监督方法,这些方法需要为每个新数据集重新训 练,并且在只有少量标记样本的情况下表现往往不一 致。虽然 SSL 通过利用未标记数据提供了一个有前景 的替代方案,但大多数现有方法仍然依赖于大规模数 据集以达到有效。MatSSL 旨在克服这一限制。我们首 先在一个小规模未标记的数据集上进行自监督预训练, 然后在多个基准数据集上微调模型。所得的分割模型 在 MetalDAM 上实现了 69.13 % 的 mIoU, 优于使用 ImageNet 预训练编码器所达到的 66.73 %, 并且相比 于使用 MicroNet 预训练的模型, 在环境障碍涂层基准 数据集 (EBC) 上平均 mIoU 的提升一直达到将近 40 %的幅度。这表明 MatSSL 能够有效适应冶金领域, 仅 需少量未标记数据,同时保留了从自然图片大规模预 训练中学习到的丰富且可迁移的特征。源代码完全可 用1。

# 1. 介绍

准确的微观结构分割是从合金设计到失效分析等材料 工程关键任务的基础,但由于像素级注释极度稀缺,仍 然具有挑战性。传统的监督深度学习方法 [1,2],即使 利用自 ImageNet [3]的迁移学习,在训练时仍会出现 不一致和低准确性,因为在小型、专家标注的金相数据 集上训练时,这些数据集通常每种材料系统仅限于几 十幅显微图。

无监督和半监督流程(例如,聚类、伪标签)可以 利用未标记的数据[4,5],但通常无法达到可靠表征所 需的细粒度准确性。自监督学习(SSL)通过使用对比、 生成、预测或自标记等先验任务,在无需真实标签的 情况下学习有用的表示,提供了一种有吸引力的替代 方案[6-9]。大多数最先进的 SSL 模型(例如, MoCo, SimCLR)需要在数十万到数百万来自相关领域的图像 上进行预训练,以实现良好的性能。这在金相学中构成 了一个主要挑战,因为通常只有几千张未标记的显微 图可用。

在这项工作中,我们提出了 MatSSL,这是一个简 单且轻量级的对比自监督学习框架,专为不同材料的 金相图像设计。MatSSL 使用一种新颖的门控特征融合 机制调整单个 ResNet-50 [10] 编码器,而无需 MoCo [7,11] 和 SimCLR [6] 中使用的双编码器或大型内存队 列,然后将调整后的编码器转移到 U-Net++ [12] 分割 网络中,以便在有限的标记数据上进行微调。

我们在多个金相基准数据集上评估 MatSSL: MetalDAM [13], Aachen-Heerlen 注释的钢材微结构数据 集 [14],以及环境阻挡涂层基准(EBC)少样本数据 集 [1]。MatSSL 的结果表明:

- 在 MetalDAM 上, mIoU 相比于使用 ImageNet 预 训练模型的基线取得了高达 3.2 % 的提升。
- 在 EBC 少样本数据集上平均 mIoU 稳定增加接近 40 %。
- 在小规模金相数据集上优于标准自监督学习方法, 在金相基准测试中表现超越 MoCoV2 高达 2.76 %。 这项工作表明,通过适当的架构修改,SSL 即使在 极低的数据环境中也能有效和实用,从而能够在各种 冶金系统中实现稳健的微结构分割。

### 2. 相关工作

在金属材料中微观结构的分割已经通过多种技术进行研究。早期的研究探索了经典的图像处理方法,如阈值分割和区域生长 [15],以及无监督的方法,如聚类和传统神经网络 [13]。尽管这些方法在基础性贡献上具有重要意义,然而其通常缺乏现代深度学习模型所达到的稳健性和准确性。在大规模标注数据集上训练的卷积神经网络(CNN)在识别复杂的微观结构模式方面显示出了显著的改进。为了在这种情况下提高分割性能,除了扩展标注数据的数量和质量 [16]或开发更复杂的网络结构 [17],还需要整合先进的训练策略。例如,[18]展示了结合多种色彩空间转换(RGB、HSV、YUV)可以显著提高分割精度,而迁移学习、数据增强和损失函数优化(例如焦点损失或多项式衰减)也被用

<sup>&</sup>lt;sup>1</sup>代码可在 https://github.com/duchieuphan2k1/MatSSL 下载

来提高结果。最近, MicroNet [1], 在一个包含 100,000 多幅图像并跨越 54 种微观结构类别的显微镜语料库上 进行有监督地预训练, 显示出在特定领域的数据上进 行分类器预训练能够实质性地提升下游少类分割任务 的表现。

然而,该领域面临的一个持久挑战是高质量、像素级标注数据集的稀缺及不一致共享。准确的标注既在技术上具有挑战性又具有主观性,通常需要专家解读[14]。尽管已提出了无监督学习的流程来生成伪标签[5],这些通常仍需要人类的细化以便于准确分割。存在多个公共金相数据库[19],但大多数都集中在分类上而非分割,或者只提供部分标注的数据[14]。超高碳钢(UHCS)数据库[20]是可用的最大数据库之一,提供了 900多个样本,但只有 24 个具有详细的分割标注。MetalDAM [13] 提供了 42 个标注和 126 个未标注的金相图像。尽管有这些努力,与标注数据相比,未标注数据的丰富性突显出迫切需要能够利用二者的方法,以实现稳健的分割。

自监督学习(SSL)[11,21]已经在面临类似数据稀 缺挑战的领域(尤其是在医学成像)中成为一种强大 的方法。在这个领域,大量的无标签数据和高昂的专 家注释成本使得 SSL 在分割和诊断任务中特别有价值。 综合基准测试[22]显示,在特定领域的大规模无标签 图像集上进行 SSL 预训练,可以产生在各种下游任务 [11,23-25]中性能超过 ImageNet 预训练和随机初始化 网络的模型。

然而,重要的是要注意,几乎所有医学影像中的 SSL 研究都利用大规模无标签数据集进行预训练,通常使 用成千上万甚至几十万张图像。例如,[22]使用了诸如 TCGA [24]、CheXpert [26]和 MIMIC-CXR [27]等数 据集,每个数据集都包含超过 100,000 张图像,而 [28] 则在数千张 MRI 图像上进行了预训练。SSL 框架如 SimCLR 和 SimSiam ([23];[11])也依赖于这些庞大 的资源来最大化表征学习。

相比之下,材料科学中的金相图像分析通常缺乏如此庞大的数据集。可用的带标签和未标记的组合图像总数仅限于几千张。这种根本性的区别为将 SSL 应用于材料微观结构分割带来了独特的挑战,并强调了在极低数据环境中开发有效的 SSL 方法的重要性。

## 3. 方法论

我们的方法包括两个主要步骤,如图 1 所示。首先,我 们使用在 ImageNet 上预训练的 ResNet-50 主干,并通 过在未标记的钢微观图像上进行自监督学习(SSL)进 一步将其适应于目标域。我们将这个适应后的编码器 称为 MatSSL。其次,我们将 MatSSL 预训练的编码器 插入到 U-Net++ [12] 分割架构中,并在标记数据上进 行端到端微调。

#### 3.1. MatSSL 架构

我们在自监督学习(SSL)框架中使用 ResNet50 作为骨干网络,以便在 SSL 训练过程中利用预训练的 ImageNet 权重。所采用的 SSL 方法是对比学习,具体



Figure 1. 训练流程概述。MatSSL 在无标签图像上的预训练 随后是使用预训练的 ResNet 编码器和 U-Net++ 进行的监督分割。

来说是使用 NTXentLoss [29]。尽管此方法在各种任务 中被证明是有效的,但将标准的对比学习直接应用于 像 ResNet 这样的多阶段骨干网络存在一个显著的缺 点:在 ImageNet 预训练期间从数百万图像中学习到的 重要特征可能会在信息经过多个连续阶段时丢失。在 我们的环境中,这种特征退化尤其成问题,因为用于金 相图像的可用 SSL 数据有限。因此,如果这些预训练 的特征没有被保留,下游分割准确性的显著下降可能 会减弱 SSL 预训练的效果。

为了解决这个问题,我们通过在 ResNet 的各个阶段引入门控特征融合来修改传统的对比学习架构。这个机制使用可学习的门控连接每个 ResNet 阶段的中间特征,旨在保留并选择性更新在预训练过程中获取的重要表示。具体来说,设*Fi*为第*i*个 ResNet 阶段中所有特征图经过全局平均池化(GAP)后得到的平均特征图,*Pi*为一个可学习的门控参数。我们计算一个门控特征

$$G_i = \sigma(F_i) \times P_i \tag{1}$$

,其中  $\sigma(\cdot)$  是 sigmoid 激活函数。来自所有阶段的  $G_i$  被串联起来,并通过一个投影 MLP 进行处理,以产生 嵌入 z。

这种门控机制使得网络能够学习哪些特征是需要保留的,以及哪些特征需要适应金相领域,从而在数据通过主干网络时最大限度地减少有价值的预训练信息的损失。在门控融合之后,更新后的特征会通过投影头(如 SimCLR 中一样),然后用于计算 *z<sub>i</sub>* 与其正样本 *z<sub>j</sub>*的对比训练的 NTXentLoss:

$$\ell_{i,j} = -\log \frac{\exp\left(\sin(z_i, z_j)/\tau\right)}{\sum_{k=1}^{2N} \mathbf{1}_{[k \neq i]} \exp\left(\sin(z_i, z_k)/\tau\right)}, \quad (2)$$

其中 sim(·,·) 是余弦相似性,  $\tau$  是温度, 2N 是对比 对的批量大小。

#### 3.2. 使用标记数据进行微调

在自监督预训练之后,我们将 MatSSL 编码器集成到 U-Net++ [12] 语义分割模型中。解码器和分割头是随



Figure 2. 我们提出的 MatSSL 架构概述。每个输入金相图像的两个增强视图通过 ResNet50 骨干网络进行处理。在每个阶段 应用了门控特征融合,以保留和适应重要的预训练特征。融合后的表示通过投影头并在对比学习框架中使用 NTXentLoss 进行 优化。

机初始化的,而编码器的权重来自 SSL 阶段。然后我 们使用 Dice 损失对带标签的金相图像进行整个网络的 微调:

$$\mathcal{L}_{\text{dice}} = 1 - \frac{2\sum_{i=1}^{N} y_i \, \hat{y}_i}{\sum_{i=1}^{N} y_i + \sum_{i=1}^{N} \hat{y}_i}, \qquad (3)$$

其中  $y_i$  是像素 i 的真实 (二元) 标签,  $\hat{y}_i$  是其预测概 率, N 是总像素数。

这个微调步骤利用了 SSL 的领域自适应特征和像素 级注释,以产生精确的微观结构分割。

### 4. 实验

我们探索了这些真实世界的钢铁微观结构数据集:

- 超高碳钢显微图像数据集 [20] (简称 UHCS): 956 张未标记的超高碳钢微观结构的扫描电子显微镜图 像。
- 亚琛-赫伦注释的钢微观结构数据集 [14] (简称为亚 琛): 1705 张注释图像(类别:背景、马氏体/奥氏 体)。

- 来自增材制造 [13] 的金相学数据集(简称为 Metal-DAM): 178 张无标签 +42 张有标签的图像(4 类: 基体、奥氏体、马氏体/奥氏体、缺陷; 33 张用于训 练/9 张用于测试)。
- 环境屏障涂层基准数据集 [1] (简称 EBC): 我们 使用了来自 EBC (环境屏障涂层)基准数据集的 41 张标记图像 (分为 EBC-1-3)。

我们从 UHCS 和 MetalDAM 中的未标记图像以及 Aachen 中的标记图像中构建 SSL 预训练集,并排除用 于微调测试集的图像。每张图像被裁剪为 256×256 个 补丁,其重叠率分别为 0.6 (UHCS)、0.7 (MetalDAM) 和 0.0 (Aachen),分别产生 15,296、14,566 和 16,836 个补丁。这些补丁被组合成三个 SSL 预训练集(表 1)。

下游语义分割在标记过的 Aachen-Heerlen 和 MetalDAM 数据集以及 EBC 上进行评估。对于 Aachen-Heerlen 和 MetalDAM,图像被分割成 8:2 的训练集和 测试集比例,然后被裁剪成不重叠的 256 × 256 块,以 保留微观结构细节(见表 2)。

通过 EBC,我们使用基准数据集的原始分割和图像 尺寸(512×512)来训练和评估模型,该数据集由训

Table 1. SSL 预训练集 (256×256 个补丁)。

Pretraining set	# Patches
UHCS + MetalDAM	29.862
Aachen-Heerlen + UHCS	32.132
Aachen-Heerlen + UHCS + MetalDAM	46.698

Table 2. 亚琛和 MetalDAM 的数据集用于训练/测试的划分。

Dataset	# Train	# Test
Aachen MetalDAM	$\begin{array}{c} 1.403 \\ 33 \end{array}$	$302 \\ 9$

练集、验证集和测试集组成(见表3)。我们使用验证 集来选择最佳模型,并在完全未见过的测试集上计算 准确率。

Table 3. EBC 上的数据集分割。

Subset	# Train	#Val	# Test
$\begin{array}{c} \text{EBC-1} \\ \text{EBC-2} \\ \text{EBC-3} \end{array}$	$\begin{array}{c} 18 \\ 4 \\ 1 \end{array}$	3 3 3	3 3 3

#### 4.1. 实验设置

所有实验均在一台带有两个 NVIDIA A100 GPU 的 Intel 系统上使用 PyTorch v2.7.0+cu126 实现。我们将 随机种子固定为 0,并启用了确定性行为以实现完全可 重现性。

在 SSL 预训练过程中,我们遵循 SimCLR [6] 增强 管道:随机调整大小裁剪、随机水平翻转、颜色抖动 (亮度/对比度/饱和度 ±0.1)、随机灰度转换、高斯模 糊和 ImageNet 标准化。

我们训练 50 个周期, 批量大小为 128, 使用 SGD (初始学习率为 0.1, 动量为 0.9, 权重衰减 10<sup>-6</sup>) 和 余弦学习率调度, 衰减至 10<sup>-4</sup>。NT-Xent 损失使用温 度  $\tau = 0.07$ 。

分割模型 U-Net++ 由 [30] 实现后,在 MetalDAM 和 EBC 上使用 Adam 优化器 (学习率为  $10^{-4}$ , 权重 衰减为  $10^{-5}$ ) 微调 200 个周期,而在 Aachen 上训练 了 50 个周期。我们将批处理大小固定为 128,并记录 了平均交并比 (mIoU)。

#### 4.2. 结果

表格 4 总结了使用不同预训练编码器初始化的 U-Net++ 分割头的平均交并比 (mIoU)。在 MetalDAM 和 Aachen-Heerlen 两个数据集中, MatSSL 始终优于 ImageNet 基线和两种流行的 SSL 方法 (DenseCL [21], MoCoV2 [11]),特别是在未标记的数据集较小时。



Figure 3. 在 MetalDAM 上的每个 epoch 的验证 mIoU (Aachen+UHCS SSL 对比 ImageNet)。



Figure 4. 在 MetalDAM 上每个时期的验证 mIoU (所有三 个 SSL 对比 ImageNet)。

MetalDAM ImageNet 预训练的编码器在 Metal-DAM 上达到 66.73 % mIoU。在 Aachen+UHCS 上 预训练的 MatSSL 将其提升到 69.95 % (提高了 3.22 %)(表 4)。图 3显示 MatSSL 相比于 ImageNet 基线收敛得更快且达到更高的水平,而图 4证实了 使用所有三个 SSL 集时获得了类似的提升。与相同训 练配置下的最先进 SSL 方法比较, MatSSL 分别超过 DenseCL 和 MocoV2 1.19 % 和 2.77 % mIoU。



Figure 5. 在 Aachen-Heerlen 上的每个 epoch 的验证 mIoU (UHCS+MetalDAM SSL vs. ImageNet)。

**亚琛-赫尔伦** Aachen 是一个大型标注数据集,包含数 千张标记图片。ImageNet 基线实现了 65.59 % mIoU。

Finetune Dataset	Pretrain	Pretrain Dataset	mIoU ( $\%$ )
MetalDAM	super. ImageNet	-	66.73
	DenseCL	Aachen + UHCS	68.76
	MocoV2	Aachen + UHCS	67.18
	MatSSL	Aachen + UHCS	69.95
	DenseCL	Aachen + UHCS + MetalDAM	68.34
	MocoV2	Aachen + UHCS + MetalDAM	68.40
	MatSSL	Aachen + UHCS + MetalDAM	69.02
Aachen	super. ImageNet	-	65.59
	DenseCL	UHCS + MetalDAM	65.82
	MocoV2	UHCS + MetalDAM	65.90
	MatSSL	UHCS + MetalDAM	65.98
	DenseCL	Aachen + UHCS + MetalDAM	65.56
	MocoV2	Aachen + UHCS + MetalDAM	65.65
	MatSSL	Aachen + UHCS + MetalDAM	65.86

Table 4. 不同预训练策略和微调数据集的 mIoU (%) 比较。

Table 5. 在 NASA EBC 基准测试集上的分割性能比较。MatSSL (在 Aachen + UHCS + MetalDAM 上训练)在所有测试 集上均取得了最佳的 mIoU,并在 EBC-3 上的改进尤为显著。

Finetune Dataset	Finetuning Method	Pretraining Method	Average mIoU (%)
EBC-1	Unet++	MicroNet	95.17
	Transformer	MicroLite	93.01
	CS-UNet	MicroNet and MicroLite	95.98
	Unet++	MatSSL (Aachen + UHCS + MetalDAM)	96.79
EBC-2	Unet++	MicroNet	84.6
	Transformer	MicroLite	84.3
	CS-UNet	MicroNet and MicroLite	86.73
	Unet++	MatSSL (Aachen + UHCS + MetalDAM)	94.70
EBC-3	Unet++	MicroNet	42.58
	Transformer	MicroLite	56.72
	CS-UNet	MicroNet and MicroLite	45.69
	Unet++	MatSSL (Aachen + UHCS + MetalDAM)	84.53



Figure 6. 在 Aachen-Heerlen 上的每个 epoch 的验证 mIoU (所有三种 SSL 与 ImageNet)。

预训练在 UHCS+MetalDAM 的 MatSSL 将其提升至 65.98 % (+0.39 %), 并且在预训练中加入 Aachen 补

丁获得 65.86 %。DenseCL 和 MoCoV2 均保持在 65.8-65.9 % 左右,表明 MatSSL 的门控融合带来了微小但一致的提升(表 4)。图 5 和图 6 中的训练曲线展示了 MatSSL 稳定的优势。结果表明,MatSSL 不仅对小型或少样本数据集有效,还在大规模微调数据集上取得改进。

图 7展示了使用 MatSSL 预训练的 U-Net++ 的样本预测。与使用 ImageNet 预训练相比,这种方法有效地减轻了过度预测,并减少了在 Aachen 和 MetalDAM 数据集中的遗漏。

**少样本微网**(EBC)。 表 5 总结了在 EBC 基准测试 中三个不同测试集上的分割性能。我们将使用 MatSSL 预训练的 Unet++ 与 Transformer [31]、CS-UNet [17] 和使用其他预训练的 Unet++ 进行了比较。MatSSL 预训练模型在每个 EBC 切片中持续表现优于所有其他



Sample Prediction in MetalDAM test set

Sample Prediction in Aachen test set

Figure 7. 使用 MatSSL 作为 Unet++ 的预训练与 ImageNet 预训练在 MetalDAM 和 Aachen 测试集上的样本预测,以及对 应的真实情况 (GT) 图像



Figure 8. 使用通过 MatSSL 预训练的 U-Net++ 对 EBC 数据集(EBC-1、EBC-2、EBC-3) 样本预测的可视化。真阳性以 绿色显示,假阳性以红色显示,假阴性以蓝色显示。

### 方法。

具体来说,在 EBC-1 上,MatSSL 达到了最高的 平均 mIoU 为 96.79%,略微超过了强基线 CS-UNet (95.98%)并且优于 MicroNet 或 MicroLite 预训练基 线。对于具有更复杂损伤模式的 EBC-2,MatSSL 继续 保持领先,以 94.70%的成绩领先 CS-UNet 显著提升 了 7.97%。最值得注意的是,在具有挑战性的 EBC-3 分组中,仅有 1 张训练图像,其他模型的平均 mIoU 波 动剧烈且低于 57%的情况下,MatSSL 达到了显著的 84.53%,超出排名第二的几乎 40 个百分点。图 8 展示 了在 EBC 数据集上使用 MatSSL 预训练的 U-Net++ 的一个具有代表性的预测结果。

这些结果不仅反映了 MatSSL 在跨域转移中的强大 泛化能力,还突显了其在训练过程中性能的稳定性,尤 其是在像 EBC-3 这样的低数据和高方差设置中。在 EBC-3 上的 mIoU 表现持续提升至 +38.84 %,强调了 MatSSL 学习高可迁移性和损伤敏感特征的能力,为现 实世界中有限监督的结构分割任务提供了有力的优势。

### 5. 结论

在这项工作中,我们提出了 MatSSL,这是一种针对低数据量情况下金相图像分割设计的自监督学习 (SSL) 框架。通过在轻量级对比学习流程中引入一个门控特 征融合模块,MatSSL 能够有效地将 ResNet-50 编码器 适应金相显微镜图像领域,而无需依赖大型未标记数 据集。

在真实世界的数据集(MetalDAM, Aachen-Heerlen) 和少样本 EBC 分割基准上的广泛评估显示, MatSSL 相对于 ImageNet 预训练和其他 SSL 基线具有优越性。 该方法在分割精度上取得了显著的提升,特别是在少 样本和低标注场景中,验证了其稳健性和泛化能力。

对于未来的工作,整合生成方法来合成额外的金相 样本可以进一步提高预训练的多样性和性能,特别是 在获得带注释的甚至是未标记的数据仍然是一个根本 限制的领域。

#### References

- J. Stuckner, B. Harder, and T. M. Smith, "Microstructure segmentation with deep learning encoders pretrained on a large microscopy dataset," npj Computational Materials, vol. 8, p. 200, Sept. 2022.
- [2] J. Luengo, R. Moreno, I. Sevillano, D. Charte, A. Peláez-Vegas, M. Fernández-Moreno, P. Mesejo, and F. Herrera, "A tutorial on the segmentation of metallographic images: Taxonomy, new metaldam dataset, deep learning-based ensemble model, experimental analysis and challenges," Information Fusion, vol. 78, pp. 232–253, 2022.
- [3] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 248–255, 2009.
- [4] R. Balestriero, M. Ibrahim, V. Sobal, A. Morcos, S. Shekhar, T. Goldstein, F. Bordes, A. Bardes, G. Mialon, Y. Tian, A. Schwarzschild, A. G. Wilson, J. Geiping, Q. Garrido, P. Fernandez, A. Bar, H. Pirsiavash, Y. LeCun, and M. Goldblum, "A cookbook of selfsupervised learning," 2023.
- [5] H. Kim, J. Inoue, and T. Kasuya, "Unsupervised microstructure segmentation by mimicking metallurgists" approach to pattern recognition," Scientific Reports, vol. 10, no. 1, p. 17835, 2020. Published: 2020/10/20.
- [6] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," 2020.
- [7] K. He, H. Fan, Y. Wu, S. Xie, and R. Girshick, "Momentum contrast for unsupervised visual representation learning," 2020.
- [8] V. Rani, S. T. Nabi, M. Kumar, A. Mittal, and K. Kumar, "Self-supervised learning: A succinct review," Archives of Computational Methods in Engineering, vol. 30, pp. 2761–2775, May 2023.
- [9] Z. Wen, O. Pizarro, and S. Williams, "Active self-semisupervised learning for few labeled samples," Neurocomputing, vol. 614, p. 128772, Jan. 2025.
- [10] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2015.
- [11] X. Chen, H. Fan, R. Girshick, and K. He, "Improved baselines with momentum contrastive learning," 2020.
- [12] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: Redesigning skip connections to exploit multiscale features in image segmentation," 2020.
- [13] J. Luengo, R. Moreno, I. Sevillano, D. Charte, A. Peláez-Vegas, M. Fernández-Moreno, P. Mesejo, and F. Herrera, "A tutorial on the segmentation of metallographic images: Taxonomy, new metaldam dataset, deep learning-based ensemble model, experimental analysis and challenges," Information Fusion, vol. 78, pp. 232–253, 2022.
- [14] D. Iren, M. Ackermann, J. Gorfer, G. Pujar, S. Wesselmecking, U. Krupp, and S. Bromuri, "Aachenheerlen annotated steel microstructure dataset," Sci-

entific Data, vol. 8, no. 1, p. 140, 2021. Published: 2021/05/26.

- [15] R. Gonzalez and R. Woods, Digital Image Processing. Prentice-Hall, 3rd ed., 2006.
- [16] M. Fernández-Moreno, B. Lei, E. A. Holm, P. Mesejo, and R. Moreno, "Exploring the trade-off between performance and annotation complexity in semantic segmentation," Engineering Applications of Artificial Intelligence, vol. 123, p. 106299, 2023.
- [17] K. Alrfou, T. Zhao, and A. Kordijaz, "Cs-unet: A flexible segmentation algorithm for microscopy images," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, pp. 8114–8119, June 2024.
- [18] M. Biswas, R. Pramanik, S. Sen, A. Sinitca, D. Kaplun, and R. Sarkar, "Microstructural segmentation using a union of attention guided u-net models with different color transformed images," Scientific Reports, vol. 13, no. 1, p. 5737, 2023. Published: 2023/04/07.
- [19] K. A. Rozman, Ömer N. Doğan, R. Chinn, P. D. Jablonksi, M. Detrois, and M. C. Gao, "Dataset for machine learning of microstructures for 9Data in Brief, vol. 45, p. 108714, 2022.
- [20] B. L. DeCost, M. D. Hecht, T. Francis, B. A. Webler, Y. N. Picard, and E. A. Holm, "Uhcsdb: Ultrahigh carbon steel micrograph database," Integrating Materials and Manufacturing Innovation, vol. 6, no. 2, pp. 197–205, 2017. Published: 2017/06/01.
- [21] X. Wang, R. Zhang, C. Shen, T. Kong, and L. Li, "Dense contrastive learning for self-supervised visual pre-training," 2021.
- [22] M. Kang, H. Song, S. Park, D. Yoo, and S. Pereira, "Benchmarking self-supervised learning on diverse pathology datasets," 2023.
- [23] J. Zbontar, L. Jing, I. Misra, Y. LeCun, and S. Deny, "Barlow twins: Self-supervised learning via redundancy reduction," 2021.
- [24] K. Chang, C. J. Creighton, C. Davis, ..., et al., "The cancer genome atlas pan-cancer analysis project," Nature Genetics, vol. 45, pp. 1113–1120, Oct. 2013.
- [25] K. Kim, S. Lyu, S. Mantri, and T. W. Dunn, "Tulip: Multi-camera 3d precision assessment of parkinson's disease," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 22551–22562, June 2024.
- [26] J. Irvin, P. Rajpurkar, M. Ko, Y. Yu, S. Ciurea-Ilcus, C. Chute, H. Marklund, B. Haghgoo, R. Ball, K. Shpanskaya, J. Seekins, D. A. Mong, S. S. Halabi, J. K. Sandberg, R. Jones, D. B. Larson, C. P. Langlotz, B. N. Patel, M. P. Lungren, and A. Y. Ng, "Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison," 2019.
- [27] A. E. W. Johnson, T. J. Pollard, N. R. Greenbaum, M. P. Lungren, C. ying Deng, Y. Peng, Z. Lu, R. G. Mark, S. J. Berkowitz, and S. Horng, "Mimic-cxr-jpg, a large publicly available database of labeled chest radiographs," 2019.

- [28] A. Kalapos and B. Gyires-Tóth, "Self-supervised pretraining for 2d medical image segmentation," 2022.
- [29] K. Sohn, "Improved deep metric learning with multiclass n-pair loss objective," in Advances in Neural Information Processing Systems (D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett, eds.), vol. 29, Curran Associates, Inc., 2016.
- [30] P. Iakubovskii, "Segmentation models pytorch," 2019.
- [31] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale," 2021.