

置信度感知的聚类分类和分割

二维微观食品晶体图像^{*}

Xiaoyu Ji, Ali Shakouri, Fengqing Zhu

Elmore School of Electrical and Computer Engineering, Purdue University,
West Lafayette, IN 47907, USA

ABSTRACT

食品晶体聚集是在结晶过程中发生的一种现象，它会在晶体之间困住水分并影响食品产品的质量。在 2D 显微图像中对聚集进行手动标注特别困难，这是因为水分子结合的透明性以及仅聚焦于成像样本单个切片的有限视角。为了解决这一挑战，我们首先提出了一种监督基线模型来生成用于粗略标记分类数据集的分割伪标签。接下来，训练了一个实例分类模型，该模型同时进行像素级分割。这两个模型都在推理阶段被使用，以结合它们在分类和分割方面的各自优势。为了保持晶体属性，设计并包含了一个后处理模块到两个步骤中。与现有的其他方法相比，我们的方法提高了真正的聚集分类准确性以及大小分布预测。鉴于手动标注的置信水平存在变异性，我们在两种置信水平下评估了我们提出的方法，并成功地对潜在的聚集实例进行了分类。

Index Terms— 图像处理，实例分割

1. 介绍

晶体颗粒聚结行为是一个过程中，颗粒之间形成桥梁，在结晶过程中捕获水分，导致大颗粒聚集 [1]。为了分析这种行为及其对产品质量的影响，测量聚结程度至关重要，定义为被聚结的单个晶体所占的百分比。在这种情况下，显微图像分析已被广泛使用 [2, 3, 4]。

然而，食品晶体和水桥的透明性在单视图图像的手动标注中引入了模糊性。存在几种现有方法来应对这一挑战。一种方法是多视角成像，这涉及使用特别设计的双摄像头设置来捕获样本 [2] 的图像。另一种方法将所有重叠晶体视为伪聚集体，并将清晰度差异小的配对视为真实聚集体 [4]。该方法用于针状晶体分析。第三种方法涉及使用替代技术，例如超声波辐照，尚未证明其可行性 [5]。最后一种方法是稀疏分布的晶体图像提出监督深度学习模型训练 [3]，它使用 MaskRCNN[6] 架构和数据增强策略。然而，这种方法没有考虑手动标注过程中的置信度问题。总之，所讨论的方法要么采用复杂的成像设置，要么专门针对稀疏晶体数据。

我们的方法仅依赖于单视图显微图像，无需额外输入。晶体分布密集，使得难以区分团聚体，特别是当一个晶体被另一个晶体遮挡时。我们邀请领域专家只为具有高度信心的团聚体进行标注。然后，我们的模型预测潜在的团聚体，并旨在正确分类所有高信心的标注。

类似于现有的食品晶体分析工作 [3, 4, 7]，我们的方法基于训练深度学习模型。MaskRCNN[6]、Yolo[8] 和 Stardist[9] 是广泛用于显微图像实例分割的架构 [10, 7, 11]。Mask-scoring R-CNN[12] 是 MaskRCNN 的扩展工作，校准了预测实例掩码置信度分数。除了在重叠实例预测方面能力有限的 Stardist 模型外，其他所有模型都在本文中用作基线模型。

我们的方法包括三个步骤。首先，训练一个监督实例分割模型用于晶体分割。该模型用于为粗标签数据集生成伪分割标注，减少了人工努力的需求。在第

THIS MATERIAL IS BASED UPON WORK SUPPORTED BY THE NATIONAL SCIENCE FOUNDATION UNDER GRANT NO. 2134667.

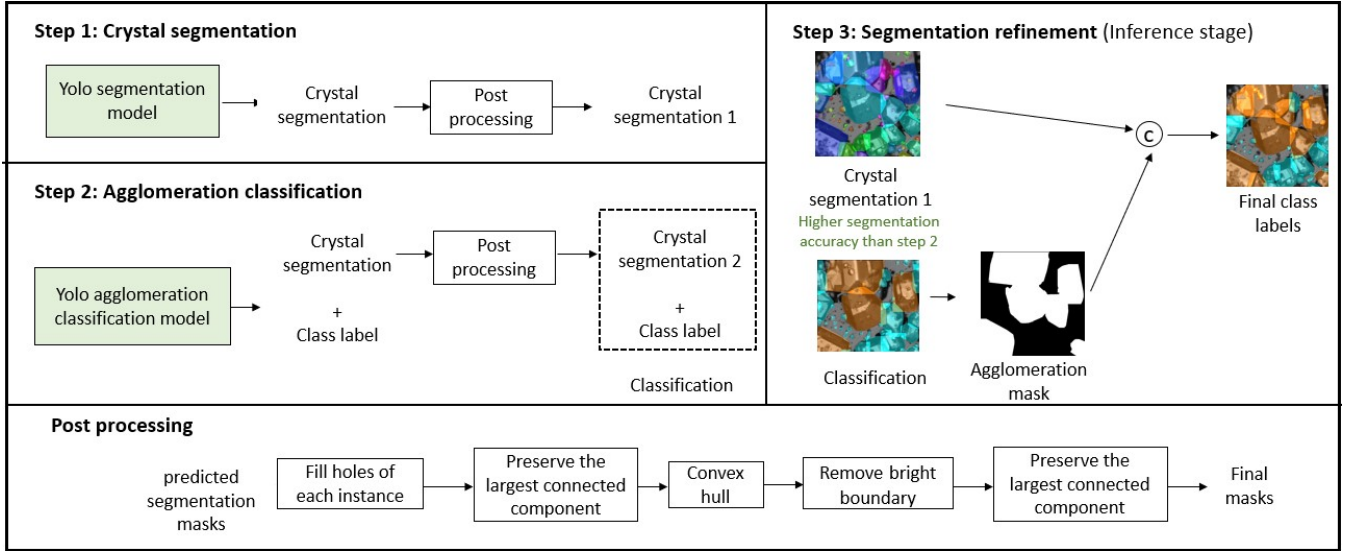


Fig. 1: 我们提出的聚集体分类方法概述。左上方的两个块代表晶体分割和聚集体分类步骤。绿色阴影的块代表正在训练的模型。第一步的模型用于在第二步中生成伪标签进行训练。第二步的模型预测实例分割及其对应类别标签。右上角的块说明了推理阶段的分割细化步骤，“c”符号表示基于二值掩模为分割后的实例分配类别标签。由于更高的准确性，晶体分割结果 1 比结果 2 更受欢迎。“后处理”块显示了详细步骤。

第二步中，我们训练了一个聚结分类模型，这是一个具有两个分支的分割和分类模型，预测每个分割实例的聚结分类标签。最后一步是推理阶段，它整合了第一步的分割输出和第二步的分类输出以提高分割准确性。

在评估过程中，我们在两个置信水平下测试分割准确性。我们准备带有手动标注置信水平的分割数据集，并且对每个置信水平分别进行评估。使用带有真实聚集体掩码的分类数据集来测量分类的真实阳性率。

2. 方法

在本节中，我们首先描述了两个训练步骤，包括晶体分割模型和聚结分类模型。然后我们介绍了推理过程中实现的细化。最后介绍了添加到分割预测中的后处理分割细化模块。

2.1. 晶体分割模型训练

如图 1 左上角步骤 1 块所示，Yolo 分割模型使用 YOLOv8 实例分割架构进行训练 [8]。主干网络结构由五层组成，分割头利用主干顶层的三个层次的输出特征。它通过全标注数据和部分标注数据进行监督训练，这

部分将在第 3 节中讨论。

2.2. 聚类分类模型训练

分类分割模型在图 1 的第 2 步中进行了说明。该模型预测实例聚合类别和像素级分割。训练数据集使用步骤 1 中的分割模型生成伪分割标签。此模型使用与步骤 1 相同的 YOLOv8 [8] 架构进行训练，并附加了一个分类头。损失函数如 (1) 所示。

$$loss = \alpha_1 loss_{segm} + \alpha_2 loss_{bbox} + \alpha_3 loss_{cls} + \alpha_4 loss_{DFL} \quad (1)$$

在公式 (1) 中，损失函数包括边界框误差 ($loss_{bbox}$)、分割损失 ($loss_{segm}$)、分类损失 ($loss_{cls}$) 和分布聚焦损失 (DFL) ($loss_{DFL}$)。详细定义可以在 [8] 中找到。分类模型 $\alpha_1, \alpha_2, \alpha_4$ 的损失权重与分割模型的设置不同，以平衡预训练知识和新分类信息的集成。微调的配置设置列在第 4.1 节中。

2.3. 分割细化

分割细化过程在推理阶段实现。我们在聚类分类模型训练过程中观察到分类准确率和分割准确率之间

的权衡。我们实现了良好的分类准确率，但比第 2.1 节中的分割模型的分割准确率要低。因此，我们将两个模型的结果结合起来：预测的分割模型实例与来自分类模型的预测聚类区域重叠超过 50% 的被归类为聚类。这种方法使我们能够同时在分割和分类预测中实现更好的性能。

2.4. 后处理

后处理管道如图 1 的下半部分所示。它包括五种形态学操作，而图表展示了每种对应处理结果的示例。前两种操作涉及填补预测掩码中的孔洞并保留最大的连通组件以确保晶体的包络属性。接下来，我们通过应用凸包操作来强调晶体的凸性。由于此操作可以扩展边界，我们将去除最亮颜色中前 15% 内的明亮边界，因为晶体边缘通常比背景暗。最后，在极少数情况下，如果边界移除将掩码分割成多个组件，我们会重复仅保留最大连通组件的步骤。

3. 数据集

3.1. 分割数据集

我们的数据集包含制造过程中间阶段的食物溶液样本的显微图像。该液体产品被放置在两片载玻片之间进行成像。

在这项研究中，我们使用了两个分割数据集。数据集 1 是一个完全像素级标注的数据集，包含 24 张训练图像、7 张验证图像和 12 张测试图像，每张图像的尺寸为 512×512 。所有的标注都由一名操作员完成，包括高置信度和低置信度水平。数据集 2 是一个部分标注的数据集，包含 156 张训练图像和 48 张验证图像。在这个数据集中，只有高置信度的晶体被领域专家标注，这位标注者与数据集 1 的标注者不同。为了避免模型将未标注的晶体误分类为背景，应用了一个窗口大小为 33×33 像素的模糊滤镜（示例如图 1 所示）。标注的置信度水平仅用于评估过程。

3.2. 分类数据集

由于难以区分团聚晶体，我们咨询了领域专家，手动圈出他们认为的团聚区域（示例如图 1 所示）。有

63 张图像被粗略标记，我们将此数据集命名为数据集 3。我们利用预训练的分割模型生成实例掩码。与团聚区域重叠率超过 50% 的实例被标记为团聚。此数据集分为 43 张训练图像、11 张验证图像和 9 张测试图像。

4. 实验

4.1. 实验设置

分割模型使用来自数据集 1 和数据集 2 的训练图像（180 张训练图像和 55 张验证图像）进行训练，给定真实实例掩码和置信度标签。训练和测试过程在一个 NVIDIA GeForce GTX 1080 Ti GPU 上执行。

我们使用来自 COCO 数据集的预训练权重 [13] 初始化 YOLOv8 模型，并最多训练 1,000 个周期，采用早停法，容忍度为 100 个周期。我们使用 Adam 优化器 [14]，带有 $\beta_1 = 0.937$ 、 $\beta_2 = 0.999$ 和 $w = 5 \times 10^{-4}$ 。学习率初始化为 0.01。损失权重 α_1 和 α_2 设置为 7.5， α_3 设置为 0.5， α_4 设置为 1.5。

分类模型使用数据集 3 进行训练。YOLOv8 模型参数设置与分割模型相同，只是将分割和边界框损失权重（方程 (1) 中的 α_1 和 α_2 ）微调至 9，DFL 损失权重 α_4 增加到 3。这种调整增强了对少数类（集聚类）分割特征的学习。该模型使用预训练的 COCO 分割模型权重进行初始化。比较用的 Mask RCNN 模型参数设置也进行了微调。

4.2. 在两个置信水平下的评估

分割预测的评估过程考虑了两个置信水平，如图 2 所示。左侧的两张图片分别是原始输入图像和真实掩码，其中红色表示高置信度而蓝色表示低置信度。右上角的图像是来自分割模型的预测掩码。对于每个预测实例，都会测量其与高置信度和低置信度的真实掩码的重叠比率。如果最大重叠比小于 50%，则将预测实例包含在残差掩码中。否则，将其分配给具有较高比率的置信水平。针对两个置信水平下的预测掩码以及残差掩码分别测量预测误差指标。

Table 1: 基线分割模型结果（有或无后处理）

Models	post process- ing	high conf corr (%) ↑	high conf χ^2 (%) ↓	low conf corr (%) ↑	low conf χ^2 (%) ↓	high conf mAP (%) ↑	high conf recall (%) ↑	Re- sErr (%) ↓
M-RCNN[6]	w/o	45.452	73.615	15.983	41.410	58.162	42.501	9.590
	w/	46.275	68.686	21.745	39.159	60.178	43.730	9.297
Ms-RCNN[12]	w/o	42.521	73.978	17.906	24.749	<u>54.817</u>	<u>42.908</u>	10.301
	w/	<u>48.932</u>	<u>69.189</u>	<u>23.768</u>	<u>21.717</u>	53.183	41.862	<u>9.523</u>
Yolov8[8]	w/o	96.471	9.115	70.683	<u>16.459</u>	72.139	65.976	7.920
	w/	<u>97.118</u>	<u>8.139</u>	<u>74.650</u>	16.666	<u>73.056</u>	<u>67.453</u>	<u>7.262</u>

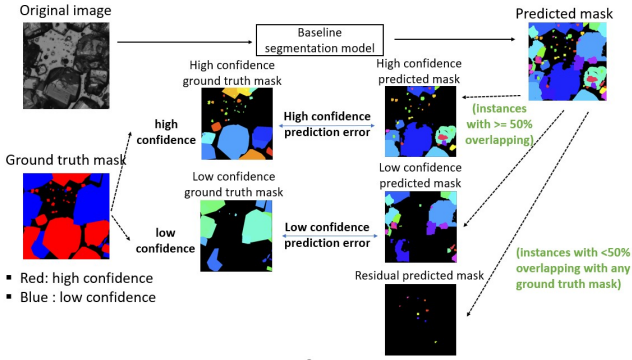


Fig. 2: 两个置信水平评估过程概述。右侧的预测掩码被分割为高置信度、低置信度和残差掩码，并与高置信度和低置信度的真实掩码配对。

4.3. 分割模型结果

如表 1 所示，基线模型 [6, 12, 8] 是否经过后处理与七个评估指标进行了比较。左侧的四个指标是晶体尺寸分布测量值。高置信度地面实况和预测掩码之间的指标表示为“高置信”，低置信度指标表示为“低置信”。“corr”代表地面实况和预测的晶体尺寸分布相关性， χ^2 代表每个区间内的地面实况与预测尺寸分布之间的均方晶粒尺寸误差（总共 10 个区间）。最后三个指标是实例分割精度指标，使用交集-并集 (IoU) 置信度阈值为 50% 的平均精度 (mAP) 和召回率。mAP 表示准确预测出的晶体比例，而召回率则表示被预测出来的地面实况实例的比例。“ResErr”代表残差掩码中预测出的晶体的比例，即实例级假阳性率。

表 1 中的粗体数字是每个指标的最佳值。下划线表示的数值是在比较有无后处理时模型表现更好的结果。大多数情况下，带有后处理的分割模型比没有后处理的有更好的性能，尤其是在高置信度指标方面。因为低置信度的真实值表明了人工标注过程中固有的不确定性，所以在低置信度指标上的准确率观察到低于高置信度指标的情况，这是意料之中的。

一般来说，Yolov8 模型在所有评估指标中都优于其他基线模型。后处理模块融合了食物晶体的内在特性，这些特性并未在模型中学习到，从而提高了尺寸分布预测和实例级预测精度。带有后处理的 Yolov8 模型被选为伪标签生成模型。

4.4. 分类模型结果

最终推理阶段预测了实例分割掩码和相应的类别标签。分类结果在数据集 3 的 9 张测试图像上进行评估。由于仅在真实标注中注释了高置信度的聚集晶体，我们只评估真实聚集区域与正确分类的预测掩码重叠的比例，即表 2 中所示的真阳性率 (TPR)。另外两个指标是同时用于分割评估的高置信度尺寸分布指标。

如表 2 所示，我们重构了集聚分类方法 [3]，结果如第一行所示。Yolov8 方法代表步骤 2 中模型的直接结果。我们可以观察到，第二和第三个尺寸分布指标的表现与表 1 的最后一行的结果相比较差。因此，我们将步骤 1 分割模型的分割结果与步骤 2 模型的分类结果结合起来，得到最终结果，即表中的最后一行。

如预期的那样，我们的方法在尺寸分布相关性和 χ^2 值方面优于其他方法。令人兴奋的是，观察到我们方法的分类准确率达到 91.079%，高于 YOLOv8 模型。这一观察结果表明，分割模型预测的掩码比分类模型覆盖了更大的真实聚集掩码部分。尽管准确性有所提高，但我们的方法的一个局限性是在推理过程中由于使用两个模型而导致时间复杂度增加。

Table 2: 聚类分类和分割度量。“corr”和 χ^2 度量均处于高置信水平。

Methods	TPR (%) $\uparrow$	high conf corr (%) $\uparrow$	high conf χ^2 (%) $\downarrow$
M-RCNN [6]	77.109	58.085	143.135
YOLOv8 [8]	90.285	94.272	17.035
Ours	91.079	97.118	8.139

4.5. 定性结果

在本节中，我们首先对不同基线方法的分割结果进行了可视化比较，如图 3 所示。图 3a 和图 3b 分别是原始图像和标注的真实标签图像。图 3b 中的每个叠加彩色蒙版代表一个晶体实例。图 3c、图 3d 和图 3e 是 Mask RCNN[6]、Mask-scoring RCNN[12] 和 YOLOv8[8] 分割模型经过后处理的预测结果。

有三个感兴趣区域被突出显示，用于比较带有后处理的三种模型。黄色矩形区域内包含大量重叠的晶体，在图 3e 中的 YOLOv8 预测结果显示与其它模型相比更准确地预测了实例。大型晶体的边界对齐不佳，这是由于训练数据集中手动标注聚集晶体时存在不确定性所致。两个黄色圆圈区域关注的是小晶体，在图 3c 和图 3d 中它们全部缺失或被错误地分割。YOLOv8 模型正确分割的小晶体数量比其他模型更多，这对应着较高的尺寸分布准确性。综上所述，带有后处理的 YOLOv8 模型具有比其他方法更高的分割准确性。

分类的视觉结果如图 4 所示。图 4b 中的地面真实掩码显示了用高置信度标记的聚集区域。请注意，未被归类为聚集状态的晶体在地面真实情况下仍可能具有聚集的潜力。这个示例中有两种晶体被分类为高

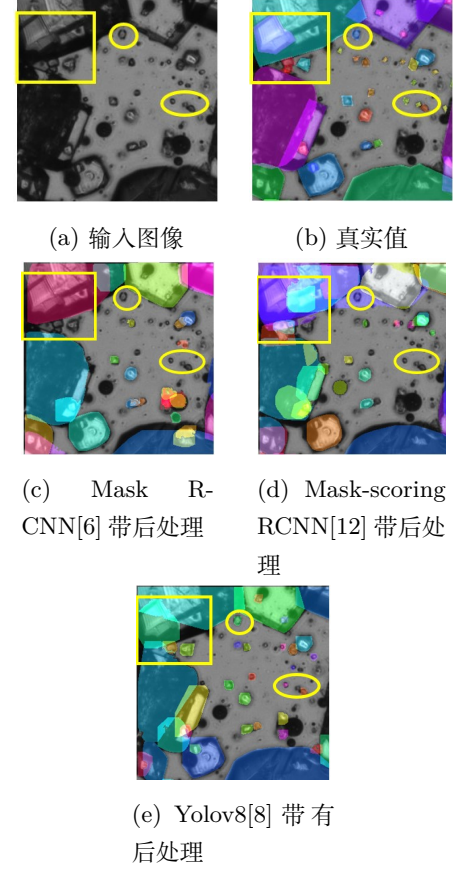


Fig. 3: 不同基线分割方法经后处理后的可视化比较，使用食品晶体数据集。子图 (a)-(e) 分别是输入图像、地面真实标签图像以及 Mask RCNN[6]、Mask-scoring RCNN[12] 和 YOLOv8[8] 的后处理结果。黄色突出显示了模型性能变化最明显的区域。

度自信的聚集物。采用数据增强策略方法 [3] 的 Mask RCNN 结果中缺失了一颗晶体，而在 YOLOv8 的结果图像（图 4d）中两者均正确分类。YOLOv8 模型还预测了其他可能具有聚集潜力的晶体。与 YOLOv8 模型相比，我们的方法有更好的分割效果，我们可以观察到大块聚集晶体上的小晶体，提供更多的细节。

5. 结论

本文提出了一种食品晶体显微图像的聚类分类和分割方法。我们的方法涉及一个两步训练过程，并在推理阶段整合两个模型的结果，以同时发挥它们在分割和分类方面的优势。我们设计了一个后处理模块来

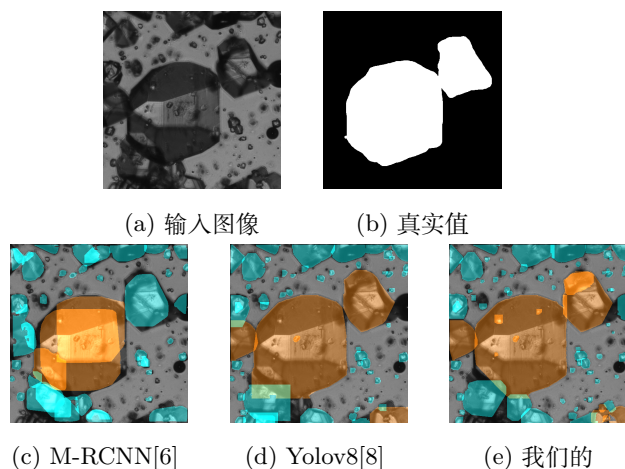


Fig. 4: 聚类分类的可视化比较。子图 (a)-(e) 分别是输入图像、高置信度聚类掩码的真实值以及对比方法 [3, 8] 和我们方法的结果。

保持预测中的晶体特性。鉴于手动标注的模糊性，分配了置信水平并将其纳入评估过程中。

6. REFERENCES

- [1] E.M. Ålander, M.S. Uusi-Penttilä, and A.C. Rasmuson, “Agglomeration of paracetamol during crystallization in pure and mixed solvents,” *Industrial & Engineering Chemistry Research*, vol. 43, pp. 629–637, 2004.
- [2] Y. Huo, T. Liu, X.Z. Wang, C.Y. Ma, and X. Ni, “Online detection of particle agglomeration during solution crystallization by microscopic double-view image analysis,” *Industrial & Engineering Chemistry Research*, vol. 56, pp. 11257–11269, 2017.
- [3] T.J. Hutagaol, J. Liu, M. Li, Z. Gao, and J. Gong, “Enhanced powder characteristics of succinic acid through crystallization techniques for food industry application,” *Journal of Food Engineering*, vol. 388, pp. 112376, 2025.
- [4] Z. Jiang, T. Liu, Y. Huo, and J. Fan, “Image analysis of crystal size distribution and agglomeration for β form l-glutamic acid crystallization based on yolov4 deep learning,” *China Automation Congress*, pp. 3017–3022, 2021.
- [5] S. Heisel, J. Holtkötter, and K. Wohlgemuth, “Measurement of agglomeration during crystallization: Is the differentiation of aggregates and agglomerates via ultrasonic irradiation possible?,” *Chemical Engineering Science*, vol. 210, pp. 115214, 2019.
- [6] K. He, G. Gkioxari, P. Dollár, and R. Girshick, “Mask R-CNN,” *IEEE International Conference on Computer Vision*, pp. 2961–2969, 2017.
- [7] X. Ji, J.P. Allebach, A. Shakouri, and F. Zhu, “Efficient microscopic image instance segmentation for food crystal quality control,” *2024 IEEE 26th International Workshop on Multimedia Signal Processing*, pp. 1–6, October 2024.
- [8] G. Jocher, A. Chaurasia, and J. Qiu, “Yolo by ultralytics,” *GitHub*, January 2023, Available online: <https://github.com/ultralytics/ultralytics> (accessed on 12 January 2023).
- [9] U. Schmidt, M. Weigert, C. Broaddus, and G. Myers, “Cell detection with star-convex polygons,” *International Conference on Medical Image Computing and Computer Assisted Intervention*, pp. 265–273, 2018.
- [10] T.H. Johansen, S.A. Sørensen, K. Møllersen, and F. Godtliebsen, “Instance segmentation of microscopic foraminifera,” *Applied Sciences*, vol. 11, pp. 14, 2021.
- [11] M. Weigert and U. Schmidt, “Nuclei instance segmentation and classification in histopathology images with stardist,” *2022 IEEE International Symposium on Biomedical Imaging Challenges (ISBIC)*, pp. 1–4, 2022.

- [12] Z. Huang, Licha L. Huang, Y. Gong, C. Huang, and X. Wang, “Mask scoring R-CNN,” IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 6409–6418, 2019.
- [13] T.Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C.L. Zitnick, “Microsoft coco: Common objects in context,” in Computer Vision – ECCV 2014. Springer International Publishing, 2014, pp. 740–755.
- [14] D.P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” International Conference for Learning Representations, p. 13, 2014.