

病理CLIP常见测试方法及数据集分享

常见测试方法

- 零样本的图块/全扫描图像分类

旨在测试模型没有经过特别训练的情况下是否能够正确识别这些类别的图像，具体步骤如下：

- a. 数据集准备

选择一个包含多种病理类别，并且类别标签明确的病理图像数据集。例如公开的 TCGA数据集。对于零样本测试，不需要使用训练集，而是直接使用测试集或验证集。

从全扫描图像中使用规则网格或随机采样等方法提取图块（通常为224 x 224像素至512 x 512像素大小）。

- b. 零样本预测

为每个病理类别构建描述性的文本提示（如“肿瘤组织”、“炎症”等）。将图块图像和文本提示分别输入到 CLIP 模型的图像编码器和文本编码器中，得到对应的图像嵌入和文本嵌入并计算他们的相似度。选择相似度最高的文本提示对应的类别作为图块的预测类别。

对于全扫描图像，使用多实例学习方法根据图块预测结果生成全图像级别的预测结果。

- c. 评估

通常使用平衡准确率和加权F1得分。

- 少样本的图块/全扫描图像分类

旨在评估模型在少量标注样本的情况下，快速适应新类别的能力，具体步骤如下：

- a. 数据集准备

数据集要求同零样本分类任务，只是这里需要将数据集划分为训练集与测试集，训练集仅包含少量样本（如每个类别1个、5个、10个）。

- b. 微调 & 少样本预测 / 根据类原型分类

- 使用少样本训练集对预训练的病理学CLIP模型进行微调。随后在测试集上使用微调或调整后的模型进行分类预测。
 - 也可以使用类原型辅助分类。将少样本中的每个样本计算特征向量，随后计算每个类别的平均值。根据距离判断样本属于哪个类别。

- c. 评估

通常使用平衡准确率和加权F1得分。

- 零样本跨模态检索

测试模型使用文字检索相关图片以及图片检索对应文字的能力，流程如下：

a. 准备

需要一个包含病理图像和对应文本描述的数据集。其中：

- 病理图像库：可以是图块或全扫描图像
- 文本描述库：可以是病理诊断报告、结构化的报告/标签、相关的自然语言描述等

b. 零样本检索过程

将文本或图像输入CLIP对应的编码器，得到文本嵌入或图像嵌入。计算嵌入和库中实例的相似度，按相似度排序检索出最相关的若干实例。

c. 评估

Recall@K（Top-K召回率）：在检索结果列表的前 K 个结果中，有多少比例包含了与查询相关的正确答案。常用的 K 值包括 1, 5, 10。Recall@K 越高，说明检索性能越好。

• 零样本分割

测试模型区分全扫描图像中不同类型组织结构的能力，流程如下：

a. 准备

准备一个包含全扫描图像以及对应掩码的数据集，例如SICAPv2以及DigestPath。

b. 分割

将全扫描图像切分为图块，对每个图块进行零样本分类。

c. 评估

使用Dice得分、召回率和精度进行评估。

数据集

宫颈癌特辑

无监督

• TCGA-TILS中的宫颈癌部分

57张WSI

• TCGA-CESC

914张WSI

• GTEx

GTEx Cervix Ectocervix 24张无标签WSI

GTEx Cervix Endocervix 23张无标签WSI

有监督

- **TissueNet**: Detect Lesions in Cervical Biopsies

<https://www.data.gouv.fr/en/datasets/tissuenet-detect-lesions-in-uterine-cervix-specimens-open-data-set/>

<https://www.drivendata.org/competitions/67/competition-cervical-biopsy/page/254/>

464张0-4分级label的WSI，295张多边形区域label的WSI

- **PanNuke**中的cervix片子

https://warwick.ac.uk/fac/cross_fac/tia/data/pannuke

293个tiles，256x256，带细胞核类别注释，用于分割任务

一般

- **CRC100K**

人类结直肠癌和正常组织的H&E染色的组织学图像。共100,000张224x224像素的图块，标记为脂肪组织、背景、碎片、淋巴细胞、粘液、平滑肌、正常结肠粘膜、癌症相关基质以及结直肠腺癌上皮。

- **WSSS4LUAD**

肺腺癌组织的H&E染色全扫描图像。包含来自广东省人民医院的67张图像以及来自TCGA的20张图像，标记为癌症、基质以及正常。

- **SICAPv2**

带有格里森等级的前列腺组织图像。包含18,784张512 x 512像素的图像块，标记为非癌性、格里森评分3分、格里森评分4分、格里森评分5分。数据集包含掩码，可用于测试分割任务。

- **PanNuke**

来自19种组织的H&E染色图块，共7901张256 x 256像素的图片，标记为良性和恶性组织。

- **BACH**

乳腺组织学显微镜和全扫描图像。包含400个显微镜图像，标记为正常、良性、原位癌以及浸润性癌；30张全扫描图像，每张图可具有多个标记为正常、良性、原位癌以及浸润性癌的区域。

- **LC25000**

结肠腺癌、良性结肠组织、肺腺癌、肺鳞状细胞癌和良性肺组织的病理学图像。每个类别5000张，所有图像均为768 x 768像素。

- **SkinCancer**

皮肤组织、基底细胞癌，鳞状细胞癌和黑色素瘤的病理学图像。包含129,364张395 x 395像素的图像块，标记为鳞状细胞癌、基底细胞癌、黑色素瘤、痣、表皮、软骨组织、真皮、神经、坏死、骨骼肌、毛囊、汗腺/小汗腺、皮脂腺、血管、皮下组织和弹性组织变性。

- **TCGA TILS**

来自13个肿瘤类型的肿瘤浸润淋巴细胞的H&E图像。共4759张图片，标记为膀胱尿路上皮癌、乳腺浸润性癌、宫颈鳞状细胞癌和宫颈管腺癌、结肠腺癌、肺腺癌、肺鳞状细胞癌、胰腺腺癌、前列腺腺癌、直肠腺癌、皮肤黑色素瘤、胃腺癌、子宫内膜癌、葡萄膜黑色素瘤。

- **CRC-MSI**

来自TCGA队列中结直肠癌和胃癌患者的组织学图像。包含411,890个独特的图像块，224 x 224像素，标记为“MSS”（微卫星稳定）或“MSIMUT”（微卫星不稳定或高度突变）。

- **RenalCell**

来自TCGA图像档案和赫尔辛基数据集的肾细胞癌H&E染色组织学图像块。包含两个子集：**【1】** 52,713个300x300像素的图像块，标记为：肾癌(24.8%)、正常肾组织(16.4%)、基质(10.4%)、红细胞(1.9%)、空白背景(30.4%)、其他(包括坏死、撕裂和脂肪组织, 16.2%) **【2】** 25,095个图像块(多数为256x256像素，也有更小的)，标记为：低淋巴细胞密度(80.1%)或高淋巴细胞密度(19.9%)。

- **DigestPath**

包含660张全扫描图像的结肠镜检查H&E组织切片数据集。这些图像提供了结肠直肠癌组织的像素级病变注释，可用于分割任务。

- **DHMC LUAD**

143张来自DHMC的肺腺癌H&E染色全扫描图像，标记为附壁生长型、腺泡型、乳头状型、微乳头状型、实性型

- **EBRAINS**

126种脑肿瘤的H&E染色全扫描图像数据集，包含3115张图像。

- **TCGA BRCA**

TCGA浸润性乳腺癌数据集，包含两种类型的全扫描图像，包括浸润性导管癌(IDC)和浸润性小叶癌(ILC)。全扫描图像总数为1048，其中837个为IDC，211个为ILC。

- **TCGA RCC**

肾细胞癌的TCGA数据集，包含三种类型的全扫描图像，包括透明细胞肾细胞癌(CCRCC)、乳头状肾细胞癌(PRCC)和嫌色性肾细胞癌(CHRCC)。图像总数为922个，其中519个为CCRCC、294个为PRCC、109个为CHRCC。

- **TCGA NSCLC**

非小细胞肺癌(NSCLC)的TCGA数据集，包含两种类型的全扫描图像，包括肺腺癌(LUAD)和肺鳞状细胞癌(LUSC)病例。全扫描图像总数为1041个，其中529个为LUAD，512个为LUSC。

- **TCGA LUAD**

CONCH使用的肺腺癌的TCGA数据集，包含165个图文对，来自49张肺腺癌H&E染色的全扫描图像。用于图文检索任务。

- **CAMELYON16**

进行淋巴结转移检测的乳腺癌全扫描图像数据集。图像总数为400张，仅提供全扫描图像级标签。

- **BRACS**

乳腺癌的H&E染色全扫描图像数据集。547张全扫描图像，被标记为7种类别：正常组织、病理性良性、寻常导管增生、导管非典型增生、平坦上皮非典型增生、导管原位癌、浸润性癌。

模型表现对比

对我们的模型以及业内领先的模型进行了零样本的图块分类测试；所有数据均为加权F1得分；每个数据集上的最好表现通过**加粗**呈现；浅蓝色背景为我们的模型。

单一提示词：

使用单一的文本提示作为真实分类结果，如“肺腺癌”。

[illegible]

KEP(Quilt)[2]	0.4421	0.6423	0.7607	0.7614	0.2899	0.4779	0.5915	0.3606	0.3583	0.3234	0.5008
CONCH[4]	0.6689	0.5619	0.8215	0.5984	0.1343	0.4247	0.3780	0.3722	0.3615	0.0947	0.4416

聚合提示词：

为每个类别创建一个包含多个相似文本提示的集合，如“这张图展示了肺腺癌”和“这是肺腺癌组织”。

	WSSS4LUAD	SICAPv2	CRC-100K	Avg.
VirCLIP	0.762	0.439	0.909	0.703
ConText-L-14-336px	0.738	0.668	0.748	0.718
LIPBaC-B-16	0.693	0.451	0.623	0.589
ConText-B-32-Knowledge	0.647	0.485	0.767	0.633
KEP(Quilt)ViT-B-32[2]	0.421	0.526	0.609	0.519
PathgenCLIP[3]	0.606	0.506	0.639	0.584
KEP(Quilt)[2]	0.527	0.577	0.656	0.587
CONCH[4]	0.608	0.412	0.774	0.598

参考文献：

[1] Jongho Park, HyukJun Kwon, Seowoo Kim, Junyoung Lee, Minho Ha, Euicheol Lim, Mohsen Imani, and Yeseong Kim. 2022. QuiltNet: efficient deep learning inference on multi-chip accelerators using model partitioning. In Proceedings of the 59th ACM/IEEE Design Automation Conference (DAC '22). Association for Computing Machinery, New York, NY, USA, 1159–1164. <https://doi.org/10.1145/3489517.3530589>

[2] Zhou, X., Zhang, X., Wu, C., Zhang, Y., Xie, W., Wang, Y. (2025). Knowledge-Enhanced Visual-Language Pretraining for Computational Pathology. In: Leonardis, A., Ricci, E., Roth, S.,

Russakovsky, O., Sattler, T., Varol, G. (eds) Computer Vision – ECCV 2024. ECCV 2024. Lecture Notes in Computer Science, vol 15110. Springer, Cham. https://doi.org/10.1007/978-3-031-72943-0_20

[3] Y. Sun, Y. Zhang, Y. Si, C. Zhu, Z. Shui, K. Zhang, J. Li, X. Lyu, T. Lin, and L. Yang. 2024. PathGen-1.6M: 1.6 Million Pathology Image-text Pairs Generation through Multi-agent Collaboration. arXiv.

<https://doi.org/10.48550/arXiv.2407.00203>

[4] Lu, M.Y., Chen, B., Williamson, D.F.K. *et al.* A visual-language foundation model for computational pathology. *Nat Med* **30**, 863–874 (2024). <https://doi.org/10.1038/s41591-024-02856-4>