

澳洲幸运10稳赢图片

EMCm7DuGMf9IBRLV

澳洲幸运10稳赢图片视觉学习新突破：Meta开源WebSSL，摆脱语言束缚

IT之家 4 月 25 日消息，科技媒体 marktechpost 昨日（4 月 24 日）发布博文，报道称 Meta 公司发布 WebSSL 系列模型，参数规模从 3 亿到 70 亿，基于纯图像数据训练，旨在探索无语言监督的视觉自监督学习（SSL）的潜力。

以 OpenAI 的 CLIP 为代表，对比语言-图像模型已成为学习视觉表征的默认选择，在视觉问答（VQA）和文档理解等多模态任务中表现突出。不过受到数据集获取的复杂性和数据规模的限制，语言依赖面临诸多挑战。

Meta 公司针对上述痛点，在 Hugging Face 平台上发布了 WebSSL 系列模型，涵盖 DINO 和 Vision Transformer（ViT）架构，参数规模从 3 亿到 70 亿不等。

这些模型仅使用 MetaCLIP 数据集（MC-2B）中的 20 亿张图像子集进行训练，排除了语言监督的影响。Meta 的目标并非取代 CLIP，而是通过控制变量，深入评估在不受数据和模型规模限制下，纯视觉自监督学习（SSL）的表现潜力。

WebSSL 模型采用两种视觉自监督学习范式：联合嵌入学习（DINOv2）和掩码建模（MAE）。训练统一使用 224×224 分辨率图像，并冻结视觉编码器以确保结果差异仅源于预训练策略。

模型在五个容量层级（ViT-1B 至 ViT-7B）上训练，评估基于 Cambrian-1 基准测试，覆盖通用视觉理解、知识推理、OCR 和图表解读等 16 个 VQA 任务。此外，模型无缝集成于 Hugging Face 的 transformers 库，便于研究和应用。

实验揭示了多项关键发现：随着参数规模增加，WebSSL 模型在 VQA 任务上的表现接近对数线性提升，而 CLIP 在超过 30 亿参数后性能趋于饱和。

WebSSL 在 OCR 和图表任务中表现尤为突出，尤其在数据筛选后，仅用 1.3% 的富文本图像训练即超越 CLIP，在 OCRBench 和 ChartQA 任务中提升高达 13.6%。

此外，高分辨率（518px）微调进一步缩小了与 SigLIP 等高分辨率模型的差距，在文档任务中表现尤为出色。

WebSSL 模型在无语言监督下仍展现出与预训练语言模型（如 LLaMA-3）的良好对齐性，表明大规模视觉模型能隐式学习与文本语义相关的特征。

同时，WebSSL 在传统基准测试（如 ImageNet-1k 分类、ADE20K 分割）上保持强劲表现，部分场景甚至优于 MetaCLIP 和 DINOv2。

IT之家附上参考地址

黑马在线人工计划

澳洲10定位计划

澳洲幸运10三码玩法

极速快三彩票一秒一开

澳洲幸运10冠军计划规律图片

168网澳洲幸运10开奖结果

澳洲10官方精准计划

腾讯分分彩注册平台app下载

飞艇6码稳定公式图

极速飞艇最牛稳赚计划

168澳洲幸运10正规官网开奖直播

澳洲10点计划网页版

7码计划王100%成功

168澳洲10开奖官方纪录

飞艇精准冠军计划全天免费

澳洲10稳定计划

2025澳洲幸运5官网开奖直播

澳洲幸运10开奖结果直播官网

澳彩大数据分析软件