




推进实例级别识别 (ILR)研究

本文由Cam Askew and André Araujo, Software Engineers, Google Research发布。

实例级识别 (ILR) 是识别一个物体的特定实例而不是简单识别出所属类别的计算机视觉任务。例如，我们不会把一张图像标注为“后印象派绘画”，我们真正感兴趣的就是实例级别的标签，例如“文森特梵高的罗纳河上的星空”或者“法国巴黎的凯旋门”而不是简单的拱门。实例级识别的难题存在很多领域，例如地标，艺术品，商品，或者标志 (logo)，而且实例级识别在很多领域多有很多的应用，例如视觉搜索软件，个人相册识别，购物和其他更多的应用。在过去的几年中，谷歌在实例级识别的研究中做出了重大的贡献，例如[Google 地标数据集](#)和[谷歌地标数据集V2](#)和[DELG](#)与[检测到检索](#)的新模型。

			
Basic-level	Painting	Building	Hat
Fine-grained	Post-Impressionist Painting	Arch	Cap
Instance-level	Starry Night Over the Rhone by Vincent van Gogh	Arc de Triomphe de l'Étoile, Paris, France	Noogler Cap

三种图像识别问题，这三种问题对于来自艺术品，地标和商品域有不同的标签粒度的层级（基本的，细粒度，实例级别）。我们主要聚焦于实例级的识别。

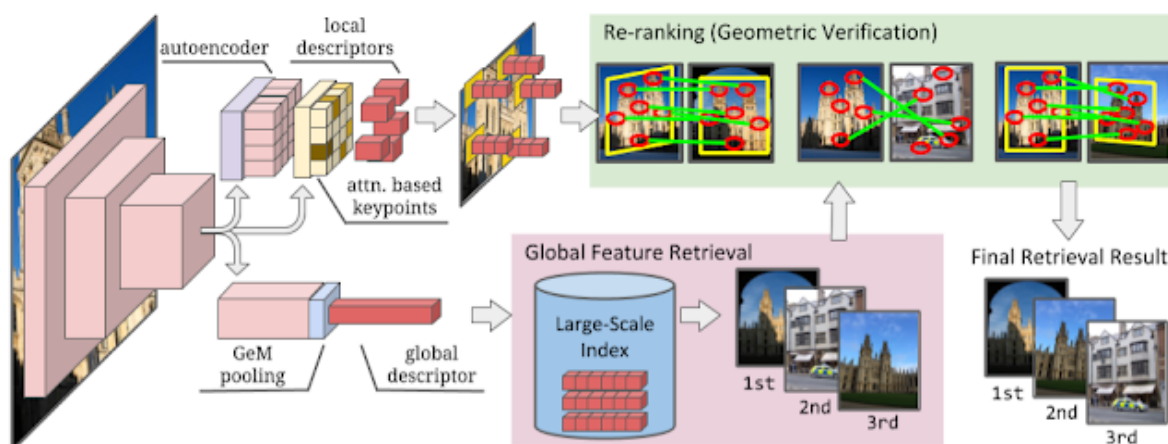
今天，我们强调在[ECCV20](#)的[实例级识别研讨会](#)上的一些结果。这个研讨会聚集了在这个领域的专家和爱好者，这学多新鲜有意思的讨论中，包含了我们的ECCV20的论文“[DEep Local and Global features](#)” (DELG)，这是一个目前最先进的实例级识别图像特征模型，还包含了一个DELG支持的[开源代码](#)和其他的实例级识别技术。在这个研讨会上也提出了一个基于GLDV2的两个新的地标挑战赛（在识别与检索任务中）。未来ILR挑战赛也会扩展到其他领域：艺术品识别，商品检索。这个研讨会的长期目标是去促进这个领域的进步，并且通过整合来自不同领域的研究团队来追求最先进的模型，在很多不同的领域中的任务到目前为止已经获得了很好的解决。

DELG: DEep Local and Global Features（深度局部与全局特征）

高效的图像表达是实例级图像识别的主要内容。通常局部与全局两种类型的表达是必要的。一个全局图像特征的总结会得到一个紧凑的表达，但是会丢失关于视觉元素空间组织的信息，这些信息往往是样本独特的个性化特征。另一方面，局部特征会包含关于特定图像区域的描述与几何信息。他们对于匹配图像中描绘相同的物体是非常有用的。

现在，大多数依赖于这两种类型的特征的系统都需要使用不同的模型单独的应用他们中的每一个，这就会导致大量的冗余计算与低效。为了解决这个问题，我们提出了DELG，这是一种局部与全局特征统一的模型。

DELG模型应用了一个全卷积网络，这个全卷积网络包含两个头，一个处理全局特征，另一个处理局部特征。全局特征使用深度网络的局部池化特征图，高效的整合输入图像的显著特征，使得这个网络对于输入图像的改变更加鲁棒。在注意力机制的帮助下，局部特征分支使用交互特征层来检测图像的显著性区域，并且以一种可区分的方式产生xaiinggaun位置内容的描述符。



我们提出的DELG模型（左），全局模型在基于检索的系统中的第一个stage中使用可以高效选择最相似的图像（底部），局部特征在重新排列的结果中使用（上边，右边），提升系统的准确性。

由于系统能够提取全局与局部的特征，新型的设计允许高效的推理。首先这样一个统一的模型可以进行端到端的训练并且在实例级识别中得到最先进的结果。当与之前的全局特征对比时，我们的方法超出之前的模型7.5%MAP，并且对于局部特征的重排过程，基于DELG的方法可以超出之前7%。总之，DELG在GLDV2上实现了61.2%的AP，除了2019年挑战赛上的两种方法外，超出了所有的其他方法。所有的顶级的结果都使用了复杂的模型融合策略，我们只采用了单一模型。

Tensorflow 2 Open-Source Codebase（tensorflow2开源代码库）

为了促进研究的可复现性，我们发布了一个经过改进的[开源代码库](#)，其中包含了DELG和其他的实例级识别的技术，例如[DELG](#)和[检测到检索](#)。我们的代码应用最新的[Tensorflow2](#)，并且出了图像检索与匹配功能之外我们也开发了模型的训练与推理功能。我们邀请社区使用并且对代码[库作出贡献](#)以便在实际里识别领域能够发展出更强大的基础。

新的实例级分割挑战赛

聚焦于地表识别中，[谷歌地标数据集v2](#)是在实例级识别中可用的最大的shoji，其中包含500万张图像，包括了大概20万类。通过使用GLDV2训练地标检索模型，与之前数据集上训练模型的结果相比我们已经得到6%的MAP提升。最近，我们也已经发起了[新的浏览器接口](#)来可视化GLDV2数据集。

今年，在地标域上我们也发起了两个新的挑战赛，一个聚焦于[识别](#)另一个聚焦于[检索](#)。这些竞赛的特点是一种新收集的数据集与一种新的评估方法：参赛者不必上传一个计算好预测值的csv文件，而是提交一个模型和代码，这些模型和代码可以直接在Kaggle服务器上运行。来计算预测值与竞赛的排名。这种环境的计算限制强调高效性与实用性的解决方案。

这项挑战赛吸引了超过1200支队伍，同比增长3倍，并且参赛者基于我们的DELG实现了很大的提升。在识别任务中，最高的得分实现了大约43%的AP，在检索任务中，获胜的队伍实习拿了59%的MAP的提升。后者的实现通过多个高效模型的混合，池化方法，训练策略（细节问题可以参考：[Kaggle竞赛官网](#)）

除了地标识别与检索挑战赛，我们的学术和工业界的合作者也讨论他们在其他的域中的竞赛与基准上的成果。在艺术品识别中的大规模研究基准尚在构建中，这个数据集使用 [The Met's Open Access image collection](#)，并和一个由不同光度与集合变换的照片组成的新测试集。类似的一个新的商品检索竞赛将会包含个还总具有挑战性的方面，包含大量的商品，长尾类别分布，物体的外观以及全局环境的变化。更多的关于ILR研讨会的包括幻灯片和视频录像等信息可以在[官网](#)上查看

基于这些研究，开源代码，数据与挑战赛，我们希望能够促进实例级识别的成功，并且使得来自不同社区的研究者和机器学习爱好者能够开发能够泛化于不同域的方法。

致谢

这个项目主要的谷歌贡献者是**André Araujo, Cam Askew, Bingyi Cao, Jack Sim and Tobias Weyand**。我们也要感谢ILR研讨会的联合组织者Ondrej Chum, Torsten Sattler, Giorgos Tolias ([Czech Technical University](#)), Bohyung Han ([Seoul National University](#)), Guangxing Han ([Columbia University](#)), Xu Zhang ([Amazon](#)), collaborators on the artworks dataset Nanne van Noord, Sarah Ibrahimi ([University of Amsterdam](#)), Noa Garcia ([Osaka University](#)), 也要感谢来自the [Metropolitan Museum of Art](#)的合作者**Jennie Choi, Maria Kessler and Spencer Kiser**。对于开源代码库，我们要感谢下列开发者的帮助：Dan Anghel, Barbara Fusinska, Arun Mukundan, Yuewei Na and Jaeyoun Kim。我们也要感谢**Cukierski, Phil Culliton, Maggie Demkin对Kaggle竞赛的支持。此外我们也要感谢 Ralph Keller and Boris Bluntschli对我们收集数据的帮助。