

模型训练和场景识别方法、装置、设备及介质

本申请要求在 2021 年 10 月 09 日提交中国专利局、申请号为 202111174534.2 的中国专利申请的优先权，以上申请的全部内容通过引用结合在本申请中。

技术领域

本申请涉及图像处理技术领域，例如涉及一种模型训练和场景识别方法、装置、设备及介质。

背景技术

机器审核技术（简称机审）在大规模短视频/图片审核中的应用越来越广泛，机审确定出的违规图片再推送至工作人员审核（简称人审），最终确定图片是否违规。机审的出现大大提高了图片审核的效率。但是机审倾向于依靠图像视觉上的共性来作出违规判决，从而忽略了因大环境变化而导致的审核结果的变化。例如枪支违规的审核，机审在识别到图像上出现枪支时，一般会认为图片违规，但是这样的机审结果准确性较差。这是因为例如是动漫或游戏场景下的枪支，则图片并非是违规图片。因此场景识别对于机审的审核结果的准确性影响较大。目前亟需一种场景识别的方案。

发明内容

本申请实施例提供了一种模型训练和场景识别方法、装置、设备及介质。

本申请实施例提供了一种场景识别模型训练方法，所述场景识别模型包括核心特征提取层以及与所述核心特征提取层连接的全局信息特征提取层、至少一个层级的带注意力机制的局部监督学习 LCS 模块、全连接决策层，所述方法包括：

通过样本图像的第一场景标签和标准交叉熵损失，训练得到所述核心特征提取层和所述全局信息特征提取层的参数；

根据每个层级的 LCS 模块输出的特征图和所述样本图像的第一场景标签逐像素计算得到的损失值，训练每个层级的 LCS 模块的权重参数；

通过所述样本图像的第一场景标签和标准交叉熵损失，训练得到所述全连接决策层的参数。

另一方面，本申请实施例提供了一种基于上述所述的方法训练得到的场景识别模型的场景识别方法，所述方法包括：

获取待识别的图像；

将所述待识别的图像输入预先训练完成的场景识别模型，基于所述场景识别模型确定所述待识别的图像对应的场景信息。

另一方面，本申请实施例提供了一种场景识别模型训练装置，所述装置包括：

第一训练单元，设置为通过样本图像的第一场景标签和标准交叉熵损失，训练得到所述核心特征提取层和所述全局信息特征提取层的参数。

第二训练单元，设置为根据每个层级的 LCS 模块输出的特征图和所述样本图像的第一场景标签逐像素计算得到的损失值，训练每个层级的 LCS 模块的权重参数。

第三训练单元，设置为通过所述样本图像的第一场景标签和标准交叉熵损失，训练得到所述全连接决策层的参数。

另一方面，本申请实施例提供了一种基于上述所述的装置训练得到的场景识别模型的场景识别装置，所述装置包括：

获取模块，设置为获取待识别的图像；

识别模块，设置为将所述待识别的图像输入预先训练完成的场景识别模型，基于所述场景识别模型确定所述待识别的图像对应的场景信息。

再一方面，本申请实施例提供了一种电子设备，所述电子设备包括处理器，所述处理器用于执行存储器中存储的计算机程序时实现上述所述场景识别模型训练方法的步骤，或实现上述所述场景识别方法的步骤。

再一方面，本申请实施例提供了一种计算机可读存储介质，所述计算机可读存储介质存储有计算机程序，所述计算机程序被处理器执行时实现如上述所述场景识别模型训练方法的步骤，或实现如上述所述场景识别方法的步骤。

本申请实施例提供了一种模型训练和场景识别方法、装置、设备及介质，所述场景识别模型包括核心特征提取层以及与所述核心特征提取层连接的全局信息特征提取层、各层级的 LCS 模块、全连接决策层，所述方法包括：

通过样本图像的第一场景标签和标准交叉熵损失，训练得到所述核心特征提取层和所述全局信息特征提取层的参数；

根据所述各层级的 LCS 模块输出的特征图和所述样本图像的第一场景标签逐像素计算得到的损失值，训练所述各层级的 LCS 模块的权重参数；

通过所述样本图像的第一场景标签和标准交叉熵损失，训练得到所述全连接决策层的参数。

附图说明

为了更清楚地说明本申请实施例，下面将对实施例描述中所需要使用的附图作简要介绍，下面描述中的附图仅仅是本申请的一些实施例，对于本领域的普通技术人员来讲，在不付出创造性劳动的前提下，还可以根据这些附图获得其他的附图。

图1为本申请实施例提供的场景识别模型训练过程示意图；

图2为本申请实施例提供的场景识别方法应用示意图；

图3为本申请实施例提供的模型主体训练阶段流程图；

图4为本申请实施例提供的模型分支扩展阶段流程图；

图5为本申请实施例提供的场景识别模型核心特征提取部分的结构示意图；

图6为本申请实施例提供的全局信息特征提取层的结构与执行原理示意图；

图7为本申请实施例提供的局部监督学习模块原理详解示意图；

图8为本申请实施例提供的场景识别模型扩展分支网络的结构和第一轮训练原理示意图；

图9为本申请实施例提供的场景识别模型分支扩展阶段结构与训练示意图；

图10为本申请实施例提供的场景识别过程示意图；

图11为本申请实施例提供的场景识别模型训练装置结构示意图；

图12为本申请实施例提供的场景识别装置结构示意图；

图13为本申请实施例提供的电子设备结构示意图；

图14为本申请实施例提供的另一电子设备结构示意图。

具体实施方式

下面将结合附图对本申请进行描述，所描述的实施例仅仅是本申请一部分实施例，而不是全部的实施例。基于本申请中的实施例，本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其它实施例，都属于本申请保护的范围。

本申请实施例涉及的专用缩略语或自定义名词，解释如下：

卷积神经网络：一种用于提取图像或视频特征，并根据所提取的特征完成分类、检测等视觉任务的端到端复杂映射，通常由多个基础卷积模块堆叠而成。

卷积层：一个利用带特定感受野的核对图像进行加权求和特征提取的操作层，一般该层还会结合非线性激活函数提高映射能力。

池化：一种归总操作，比如对特定范围或者特定维度的像素值进行归总，通常包括最大化、最小化和取平均等等。

分组卷积：按通道把特征图组分成若干个小组，每个小组的特征图执行相同或不同的卷积操作，可以用来降低计算开销。

特征金字塔：一种多尺度特征的提取方法，通常是从网络的不同层级取出特征图，然后通过某种上采样方案对齐特征图，并融合这些特征图来产出多尺度特征。

残差块：一种带跨层连接旁路的由多个卷积层组成的模块，使用该模块可以搭建较深的卷积神经网络，并且避免梯度消失现象，加速网络的训练。

热力图：一种可以反映图像局部重要性的特征图，一般重要性越高，则局部的热力值就越大，或者相反。

局部监督学习：对模型的某些部分或特征图的某个局部使用直接相连的标签和损失，进行参数或提取能力的学习。

注意力机制：一种通过拟合不同部分重要性程度来迫使网络关注重要区域，并基于重要区域特征作出决策的机制。

Sigmoid：一种不考虑类别互斥关系的激活函数，通常被激活以后的输出值都会落在 $[0, 1]$ 区间，以完成标准化。

可变形卷积：一种卷积核不是规范几何形状的卷积操作，不规范的几何形状通常由原形状加上偏移生成。

标准交叉熵：一种常规的用于简单分类问题的损失评估函数，常用于训练分类网络，包括单标签分类和多标签分类。

Focal Loss：一种针对类别不平衡问题的损失函数，可以让数据量较少的类别获得较大的惩罚，防止模型完全倾向于数据量较多的类别。

需要说明的是，本申请实施例并非直接应用于机审环节，直接产生推审结果；而是以场景信号的方式输出具体机审模型所需要的场景信息，通过合适的策略与机审模型共同产生推送结果。最终的推送结果认为违规的视频或图片将被推送至人审环节进行多轮审核，得到处罚结果；而最终的推送结果认为正常的视频或图片，也会按照一定的采样率在不同区域抽样巡查，或根据举报结果推送至人审环节复审，避免漏掉严重违规的视频/图片。

实施例 1：

图 1 为本申请实施例提供的场景识别模型训练过程示意图，该过程包括以下步骤：

S101: 通过样本图像的第一场景标签和标准交叉熵损失, 训练得到所述核心特征提取层和所述全局信息特征提取层的参数。

S102: 根据所述各层级的 LCS 模块输出的特征图和所述样本图像的第一场景标签逐像素计算得到的损失值, 训练所述各层级的 LCS 模块的权重参数。

S103: 通过所述样本图像的第一场景标签和标准交叉熵损失, 训练得到所述全连接决策层的参数。

其中, 场景识别模型包括核心特征提取层以及与所述核心特征提取层连接的全局信息特征提取层、各层级的 LCS 模块、全连接决策层。

本申请实施例提供的场景识别方法应用于电子设备, 该电子设备可以是个人电脑 (personal computer, PC)、平板电脑等智能设备, 也可以是服务器。

为了适应不同细粒度场景识别的要求, 本申请实施例中, 所述场景识别模型还包括分支扩展结构; 所述分支扩展结构包括卷积层和局部对象关联关系模块;

根据所述分支扩展结构的卷积层输出的特征图和所述样本图像的第二场景标签逐像素计算得到的损失值, 训练所述分支扩展结构各层级的卷积层的权重参数; 通过带场景置信正则项的损失函数, 训练得到局部对象关联关系模块的参数; 其中, 所述第一场景标签和所述第二场景标签的粒度不同。

一般第一场景标签为粗粒度场景标签, 第二场景标签为细粒度场景标签。

如图 2 所示, 本申请实施例提出的多层次细粒度场景识别模型会以同步的方式运行, 即针对某个视频或图片集中的待识别图像, 它会优先于机审流程中的其它现有的机审模型先执行, 把执行得到的场景信息存储在缓存队列中; 然后通过同步信号激活一系列的机审模型 (图 2 中的机审模型 a、机审模型 b 和机审模型 c 等), 以同样的视频或图片集中的待识别图像作为输入, 运行机审模型, 得到初步的推审结果。其中, 配置了场景策略的机审模型/标签, 将从缓存队列中取出相应的场景信号, 与机审初步推审结果共同计算, 得到最终的推审结果, 并决定是否推送至人审环节; 而没有配置场景策略的机审模型/标签则直接根据机审模型所给的结果决定是否推送。

若确定出的待识别图像对应的场景信息属于违规场景信息, 且机审的审核结果为所述待识别图像为违规图像, 则确定所述待识别图像为违规图像。并决定推送至人审环节。若确定出的待识别图像对应的场景信息不属于违规场景信息, 或机审的审核结果为所述待识别图像不是违规图像, 则确定所述待识别图像不是违规图像, 此时不推送至人审环节, 或者按照一定的采样率在不同区域

抽样巡查，或根据举报结果推送至人审环节复审。

其中，电子设备中可以预先保存哪些场景信息属于违规场景信息，在确定出的待识别图像对应的场景信息之后，便能够判断待识别图像对应的场景信息是否属于违规场景信息。机审模型审核待识别图像是否为违规图像的过程可以采用相关技术，在此不再进行赘述。

图3为本申请实施例提供的场景识别模型的主体结构训练过程示意图，图4为本申请实施例提供的场景识别模型的分支扩展结构训练过程示意图。

图3和图4展示了本申请实施例提出的多层次细粒度场景识别模型的整体训练过程，由两个阶段组成，分别为“模型主体训练阶段”和“模型分支扩展阶段”，完成训练后将生成一个如图2左侧所示结构的场景识别模型，用以配合提高机审模型的精度。对于场景识别模型，主体结构部分的特征提取能力非常重要，因为主体结构部分的各个组成部件，一般都会作为分支扩展结构时的前置部件，影响具体细粒度分支的特征提取能力。为了大幅提高模型主体结构部分的特征提取能力，使其去挖掘高丰富度的特征；如图3所示，在模型主体结构训练阶段，使用了三轮具有不同针对目标的训练策略对主体结构部分的各个模块进行训练。其中，第一轮主要通过第一场景标签（此时一般为图片级的场景语义标签）和标准交叉熵损失，经多次迭代，专门优化了场景识别模型的核心特征提取层和多尺度全局信息特征提取层，局部监督学习模块的权重参数先不优化。第一轮优化的主要是为了让模型获得抽象语义特征和多尺度上全局特征的提取能力。然后，固定住第一轮优化的参数（权重参数），给每个池化层后的卷积特征图组接上一个“带注意力机制的局部监督学习模块”(Local Supervised 模块, LCS 模块)，每个层级均以池化后的卷积特征图组作为输入，经LCS模块后，输出一张“聚焦化”特征图，通过该特征图和样本图像的第一场景标签（此时一般为框图级的标签），以逐像素的二值sigmoid损失对各层级的LCS模块权重参数进行优化。第二轮优化完成后，模型将对局部对象的特征敏感，可通过LCS模块提取局部对象特征，并通过Fisher卷积特征编码，在减少特征冗余的同时，尽量降低有决策影响的细微特征的缺失。第三轮优化则集中在针对融合特征的决策层权重参数，此时第一轮优化使用的全连接输出层将被移除，换上对应三种特征融合的全连接决策层，同样通过第一场景标签（此时一般为图片级的场景语义标签）和标准交叉熵损失进行训练优化，决策层以外的权重参数固定。

在模型主体训练阶段结束后，模型可以根据后续细粒度场景需求在主体

结构上进行分支扩展。图 4 举了扩展一个分支的例子。一般分支会从主体结构部分的某个卷积层输出开始，接入若干的卷积池化操作层，训练时会固定住主体结构部分的权重参数，经过两轮优化完成训练。第一轮优化使用了逐像素的二值 sigmoid 损失直接对关联的卷积层进行优化，不再使用 LCS 模块作为中继。第一轮优化的主要也是为了让扩展分支对局部对象具备特征学习能力。而第二轮优化则会在扩展分支中嵌入一些由“变形卷积”组成的“局部对象关联关系学习模块”，在分支已具备局部对象特征提取能力的基础上，通过带场景置信正则项的 focus 损失，学习局部对象关联关系的挖掘能力。使用类似的方法，可以在主体网络上扩展多个不同的分支，处理不同的任务需求。

本申请实施例提供了一种基于场景识别模型进行图像场景识别的方案，在训练场景识别模型时，首先通过样本图像的第一场景标签和标准交叉熵损失，训练得到核心特征提取层和全局信息特征提取层的参数，然后根据各层级的 LCS 模块输出的特征图和所述样本图像的第一场景标签逐像素计算得到的损失值，训练各层级的 LCS 模块的权重参数，最后训练得到场景识别模型的全连接决策层的参数。使得场景识别模型具备高丰富度特征的提取能力，基于场景识别模型进行场景识别，大幅提高场景识别的准确性。并且，场景识别模型还包括分支扩展结构，从而适应不同细粒度场景识别的要求。

由图 4 可以看出，场景识别模型中的核心特征提取层分别与全局信息特征提取层、各层级的 LCS 模块、全连接决策层（图 4 中的全连接层（Fully Connected Layer, FC 层））以及分支扩展结构连接。在基于场景识别模型对待识别的图像进行场景识别时，首先由核心特征提取层对待识别的图像进行特征提取，然后在将结果分别输出至全局信息特征提取层、各层级的 LCS 模块、全连接决策层以及分支扩展结构。全局信息特征提取层、各层级的 LCS 模块、全连接决策层以及分支扩展结构根据收到的特征图再并行进行处理，得到最终的场景识别结果。

图 5~图 9 展示了场景识别模型以及模型训练的细节。

实施例 2:

所述核心特征提取层包括第一类分组多感受野残差卷积模块和第二类分组多感受野残差卷积模块；

所述第一类分组多感受野残差卷积模块包括第一分组、第二分组和第三分组，所述第一分组、第二分组和第三分组的卷积尺寸不同，所述第一分组、第二分组和第三分组包括残差计算旁路结构；每个分组通过卷积操作和残差计算

输出特征图，每个分组输出的特征图在通道维度拼接并进行通道混洗，卷积融合后输出到下一模块；

所述第二类分组多感受野残差卷积模块包括第四分组、第五分组和第六分组，所述第四分组、第五分组和第六分组的卷积尺寸不同，所述第五分组和第六分组分别包括 1×1 卷积旁路结构和残差计算旁路结构；每个分组输出的特征图在通道维度拼接并进行通道混洗，卷积融合后输出到下一模块。

本申请实施例提出的场景识别模型，其核心特征提取层结构如图 5 右侧所示，由一系列的“分组多感受野残差卷积模块”组成。图 5 左侧描述了两类分组多感受野残差卷积模块，它们由三个不同感受野的卷积分支组成，为了节省计算量，上一模块计算得到的特征图组将会分成三组分别传送至不同的卷积分支进行卷积操作，进一步提取特征。第一类分组多感受野残差卷积模块为图 5 中的“GM-Resblock”，为了覆盖不同的感受野，第一分组、第二分组和第三分组这三个分支的卷积核分别使用了 1×1 ， 3×3 ， 5×5 三种不同的尺寸，其中 5×5 卷积操作使用两层 3×3 卷积操作进行替换，这样可以在保持相同感受野的同时，增加非线性映射的个数，提高拟合能力。每个分支还加入了用来计算残差的旁路结构，以在扩大模型深度的同时避免梯度消失。本申请实施例采用多感受野卷积作为模块，主要是因为场景识别属于视觉复杂问题，不同尺度下的局部特征都有可能对场景的判别造成影响，而多感受野机制可以尽量捕获更多的促进决策的因素。为了保证通道数的规整性，GM-Resblock 三个卷积分支输出的结果会在通道维度拼接并进行通道混洗，最后使用 1×1 卷积融合以后传到下一个模块。需要说明的是，第二类分组多感受野残差卷积模块为图 5 中的“GM 投影 block”，GM 投影 block 包括第四分组、第五分组和第六分组，第四分组、第五分组和第六分组三个分支的卷积核分别使用了 1×1 ， 3×3 ， 5×5 三种不同的尺寸，其中 5×5 卷积操作使用两层 3×3 卷积操作进行替换，GM 投影 block 还会被用于对特征图进行下采样，所以它的结构稍微做了修改，比如取消了 1×1 卷积分支的旁路，在 3×3 和 5×5 卷积分支的旁路加入 1×1 卷积，用来维持特征图尺寸和通道数的一致性。为了保证通道数的规整性，GM 投影 block 三个卷积分支输出的结果会在通道维度拼接并进行通道混洗，最后使用 1×1 卷积融合以后传到下一个模块。

实施例 3：

所述通过样本图像的第一场景标签和标准交叉熵损失，训练得到所述核心特征提取层和全局信息特征提取层的参数包括：

对所述核心特征提取层中不同层级的特征图使用膨胀因子不同的反卷积操作进行上采样，在通道维度使用双线性插值算法对齐通道数，各个层级的特征图逐通道进行相加合并，合并后的特征图组进行卷积融合，并通过逐通道的全局平均池化得到全局信息特征向量，将所述全局信息特征向量和全连接层 FC 特征向量拼接，并通过标准交叉熵损失训练得到所述核心特征提取层和全局信息特征提取层的参数。

在模型主体训练阶段的第一轮，全局信息特征提取模块伴随着模型的核心特征提取层一起被训练。图 6 简单展示了其细节和原理。为了提取图像的全局信息，本申请从多尺度的特征图出发，融合不同尺度的全局信息，获取高质量的全局信息特征向量。与单一尺度的全局信息特征相比，多尺度的特征一方面可以减少信息丢失，另一方面还可以使模型对全局空间层面的重要区域更敏感。本申请实施例借鉴特征金字塔的思想，对模型核心特征提取层中不同层级的特征图组使用膨胀因子不同的反卷积操作进行上采样，确保特征图的尺寸一致。此处使用反卷积而非普通的填充上采样，主要是为了缓解上采样操作导致的图像失真问题。完成上采样操作后，不同层级的特征图的通道数仍然不一致，此处简单地在通道维度循环使用双线性插值算法补充不足的通道，对齐通道数。然后各个层级的特征图执行逐通道相加操作完成合并，合并以后的特征图组使用 1×1 卷积进行融合，并通过逐通道的全局平均池化获得全局信息特征向量。根据图 3，该向量将和记录抽象特征的 FC 特征向量拼接，然后接入标准交叉熵损失进行优化。

实施例 4:

所述根据各层级的 LCS 模块输出的特征图和所述样本图像的第一场景标签逐像素计算得到的损失值，训练所述各层级的 LCS 模块的权重参数包括：

通过通道维度的注意力机制使用激活函数获取每个通道的重要性权重，根据所述每个通道的重要性权重对每个通道的特征图进行加权求和，得到汇总热力图；

根据所述汇总热力图、对象场景关联重要度和对象的面积逐像素计算损失值，根据所述损失值训练所述各层级的 LCS 模块的权重参数。

在模型主体结构部分，另一个重要步骤是令其对局部对象特征也具备较好的提取能力。如图 7 所示，在模型主体训练阶段第一轮完成以后，同样是在多个层级上，本申请提出使用“带注意力机制的局部监督学习模块”和“局部对象监督损失”来增强模型该部分的提取能力。其中，带注意力机制的局部监督

学习模块(LCS 模块)的结构如图 7 左下所示。首先,取出各层级的特征图组,先经过 3×3 卷积进行映射并保持通道数不变。考虑到同一个像素位置上,不同通道的特征图重要性并不相同,所以本申请实施例在尺寸下采样的同时,还通过通道维度的注意力机制对不同通道的特征图重要性进行控制,用来获取一张更为准确地表明各个位置像素信息的汇总热力图,更好地指引 LCS 模块对局部对象特征的学习。另外,LCS 模块使用普通 3×3 卷积来完成下采样,而非池化操作,这样做是为了在降冗余的时候避免局部对象激活偏移过大。通道维度的注意力机制使用 Sigmoid 激活获取重要性权重,因为通道间的重要性不是互斥关系;最终会通过全连接层输出一个重要性权重向量,然后重要性值与对应通道相乘,完成通道特征图的加权。

LCS 模块输出经注意力增强的特征图组以后,会专门接入“局部对象监督损失”来有监督地引导该模块学习局部对象特征的提取能力。例如,先对经注意力增强的特征图组跨通道逐像素求和,得到一张反映不同像素位置激活情况的热力图。然后使用该热力图和基于框图对象与“对象-场景关联重要度”标记的 mask 图求损失,并进行反向传播。mask 图是在图像级的场景语义标签的基础上,根据场景图像中对象对场景判决的影响程度得到的标签,其中,图像中的对象会根据其所占据的框图范围给出 mask,对场景判决影响大的对象标为“重要”(mask 值给 1.0),对场景判决影响较小且在多个场景中出现的公共对象标为“不重要”(mask 值给 0.5),背景 mask 值给 0.0。为了达到“局部监督学习”的效果,损失使用逐像素的二值 sigmoid 损失,根据“重要对象”的面积与“不重要对象”的面积之比来选定惩罚权重,当“重要对象”的面积远小于“不重要对象”时,“重要对象”的惩罚权重与“不重要对象”的惩罚权重的相对差距将扩大,让 LCS 模块在“重要对象”为小目标时,加大对其的学习力度,避免偏向于“不重要对象”或“背景”的学习。需要注意的是,因为 LCS 模块的目标是提取局部对象特征,所以“背景”的惩罚权重在两种情况下都会取一个较小的值。具体的损失表达式如下:

$$LCS_Loss = \sum_{i=1}^H \sum_{j=1}^W l_{bsigmoid}(p_{i,j}),$$

$$l_{bsigmoid} = \begin{cases} \lambda_{im} \left(\text{sigmoid}(p_{i,j}) - \text{mask}_{i,j} \right)^2, & \text{if } \text{mask}_{i,j} = 1.0 \text{ and } \frac{\text{mask_area}_{im}}{\text{mask_area}_{unim}} > T_{area}, \\ \lambda_{unim} \left(\text{sigmoid}(p_{i,j}) - \text{mask}_{i,j} \right)^2, & \text{if } \text{mask}_{i,j} = 0.5 \text{ and } \frac{\text{mask_area}_{im}}{\text{mask_area}_{unim}} > T_{area}, \\ \lambda_{back} \left(\text{sigmoid}(p_{i,j}) - \text{mask}_{i,j} \right)^2, & \text{if } \text{mask}_{i,j} = 0.0 \text{ and } \frac{\text{mask_area}_{im}}{\text{mask_area}_{unim}} > T_{area} \\ \lambda'_{im} \left(\text{sigmoid}(p_{i,j}) - \text{mask}_{i,j} \right)^2, & \text{if } \text{mask}_{i,j} = 1.0 \text{ and } \frac{\text{mask_area}_{im}}{\text{mask_area}_{unim}} < T_{area}, \\ \lambda'_{unim} \left(\text{sigmoid}(p_{i,j}) - \text{mask}_{i,j} \right)^2, & \text{if } \text{mask}_{i,j} = 0.5 \text{ and } \frac{\text{mask_area}_{im}}{\text{mask_area}_{unim}} < T_{area}, \\ \lambda_{back} \left(\text{sigmoid}(p_{i,j}) - \text{mask}_{i,j} \right)^2, & \text{if } \text{mask}_{i,j} = 0.0 \text{ and } \frac{\text{mask_area}_{im}}{\text{mask_area}_{unim}} < T_{area} \end{cases}$$

其中, $p_{i,j}$ 代表热力图上像素的激活值, $\text{mask}_{i,j}$ 代表像素级的标签, area 代表面积, 本申请 $\lambda_{im}, \lambda_{unim}, \lambda'_{im}, \lambda'_{unim}, \lambda_{back}$ 分别取值 0.8, 0.6, 1.0, 0.5, 0.3。需要注意的是, 本申请在训练 LCS 模块时, 各个层级的模块都是与损失直连, 独自反向传播, mask 图会根据需要进行相应的下采样。

其中, H, W 代表图片的高和宽, i, j 代表像素的行号和列号, $l_{bsigmoid}$ 代表每个像素对应损失值的计算方式, T_{area} 代表触发不同计算损失计算方式的阈值, mask_area_{im} 代表重要对象的 mask 区域面积, mask_area_{unim} 代表普通对象的 mask 区域面积, mask_area_{im} 与 mask_area_{unim} 代表可人为标出。

LCS 模块完成训练后, 其直接提取的特征仍然是特征图组, 尺寸为 $H \times W \times C$, 直接作为特征冗余度仍然会过大, 而使用非线性全连接层来提取特征向量又会导致一些细微的决定性特征丢失。因此, 本申请实施例利用 Fisher 卷积编码方法降低特征图的维度, 使用 Fisher 卷积特征编码技术来提取局部对象特征向量, 在减少细微的决定性特征丢失的同时避免冗余特征带来的几何变换影响。Fisher 卷积特征编码的流程比较简单, 主要利用了多种通用高斯分布来对不同像素上的向量进行混合, 减少尺寸维度的特征数, 步骤如下:

在尺寸维度展平特征图, 使其表示为 $H \times W$ 个 C 维向量。

利用 PCA 把每个 C 维向量降维为 M 维。

在 $H \times W$ 个 M 维向量上利用 K 个高斯分布计算 K 个高斯混合参数值。

把 $H \times W$ 个 M 维向量演化为 K 个 M 维高斯向量。

计算所有高斯向量的均值向量和方差向量，将其拼接并 L2 正则化，最终输出长度为 $2MK$ 的局部对象特征向量，每个层级输出一个向量。

此处和全局信息特征提取不同在于，为了获取一些细微的局部对象特性，不同层级的特征不再融合，而是分别输出。如图 3 步骤 3 所示，获取了局部对象特征、全局信息特征以后，这些特征还会和 FC 抽象特征结合起来，重新构建一个面向高丰富度特征的主体决策层，利用这些特征完成高精度决策。

实施例 5：

所述分支扩展结构使用深度分离卷积残差块 DW 构造，在残差块的主路中，中间层使用 DW 卷积层，DW 卷积前后使用 1×1 卷积层。

所述局部对象关联关系学习模块包括可变形卷积层、卷积层和平均池化层；

可变形卷积层获取当前像素位置的卷积核偏移值，卷积核参数的当前位置加上偏移值作为其真实生效的位置，获取真实生效的位置的特征图像素值，经过卷积操作和平均池化操作输出特征图。

完成模型主体的训练后，进入分支扩展阶段。通常会根据新的细粒度场景需求来扩展分支，可以根据需求采用合适的网络结构去设计新分支。本申请实施例考虑到分支的多重扩展性，为了控制每个分支的开销，使用深度分离卷积残差块 (Depth-Wise, DW) 来构造分支，如图 8 所示。在残差块的主路中，中间层使用 DW 卷积来替换普通卷积层，约减少三分之二的计算开销，DW 卷积前后使用 1×1 卷积实现逆通道收缩操作，同时输出采用线性激活，这样做主要是为了避免 Relu 在负值激活时抛弃过多的特征。本申请最终使用三个模块（分支模块组成部分 a、分支模块组成部分 b 和分支模块组成部分 c）串接起来组成细粒度分支。由于分支是由场景模型主体的核心特征提取层扩展得到，该部分并未特别针对局部对象特征学习能力进行优化，所以扩展分支网络的对应层级会直接接入前面提出的 LCS 损失进行预训练优化，此处和主体部分的训练不同，并未额外增加 LCS 模块，而是使用和扩展分支网络共享卷积层参数。这样做一方面是为了降低开销，另一方面是为了在分支扩展阶段的第二轮训练时，可以在局部对象特征提取的基础上，同时学会局部对象的关联关系，结合局部对象特征和局部对象的全局空间关联实现细粒度复杂场景的识别。

为了在局部对象特征提取能力的基础上获得关联关系的学习能力，本申请在分支扩展阶段的第二轮，在每个分支模块的组成部分间嵌入一个“关联关系

学习模块”，这些模块和原分支网络的组成部分一起训练。如图9下侧所示，关联关系学习模块由一个可变形卷积层、一个 1x1 卷积层和平均池化层组成。其中可变形卷积层是其核心，它在进行卷积操作时使用了变形的卷积核，这主要是因为局部对象的全局空间关联一般不会是规则的几何形状，其关联逻辑通过变形的卷积核能更准确地进行建模。可变形卷积的执行过程很简单，它在执行卷积操作前需要先通过一个分支去获取当前像素位置的卷积核偏移，该偏移包括 X 偏移和 Y 偏移(因为卷积核参数通常只需要关注尺寸维度)，然后卷积核参数的当前位置加上偏移值作为其真实生效的位置，考虑到该位置的坐标可能为浮点数，所以卷积核参数对应位置的特征图像素值可以使用双线性插值获得。完成可变形卷积操作后，还会执行一次 1x1 卷积操作和平均池化操作(非全局平均池化，不改变尺寸)，这主要是用于平滑输出结果。需要注意的是，关联关系学习模块仅仅是分支扩展网络连结位置的旁路，原模块间仍然会直接相连。在本轮训练中，由于其关注的一般是细粒度的场景，这就比较容易出现数据类别不平衡以及跨类别特征重叠的现象。因此本轮训练采用了 focal loss 作为损失函数的主要部分，这种 loss 会对数量较少的类别给予更多的训练关注，同时它也很适合作为多标签训练损失。另外，本申请还用了主体部分各个场景的置信度来作为正则项，提高本轮训练的效率。损失函数的格式如下：

$$L_{branch} = L_{focus} + \sum_{i=1}^{Num_{class}} C_{score}^i \mathbb{R}$$

其中 L_{focus} 代表标准 focus loss, C_{score}^i 代表图像在主体部分针对某一场景类别 i 的置信分, $\mathbb{R}$ 为正则项, 本申请使用 L2 正则项作为扩增的处罚项。分支扩展可在主体识别特征提取层的任意层级进行, 以树状方式展开。

其中 Num_{class} 代表类别数量。

本申请的实施例带来了以下的有益效果：

本申请使用三阶段的训练方案从抽象特征、全局信息特征、局部对象特征三个角度对模型主体特征提取部分进行训练，使模型具备高丰富度特征的提取能力，并根据它们作场景判别，大幅提高场景识别精度。

本申请结合特征金字塔的思想，从多个尺度去挖掘全局信息特征，避免了因过多的下采样和非线性变换而导致的全局空间关联信息丢失，提供了高质量的全局信息特征，提高了背景类场景的识别能力。

本申请在多个层级上分别通过局部监督学习的方式，为不同层级提供局部

对象特征的提取能力，与单一层级的局部对象特征提取相比，减少了细微的场景决策信息的丢失，丰富了局部对象特征。

本申请通过注意力机制增强了局部监督学习模块对不同通道的关注程度，加强了重要局部对象特征的激活，给后续的 Fisher 编码指明方向。

本申请首次提出基于归总热力图，结合框图级的局部对象重要性，使用新的逐像素二值 Sigmoid 损失进行优化，迫使局部监督学习模块可以聚焦于“重要局部对象”的学习，并且减少“不重要局部对象”和“背景”对决策的干扰。

本申请使用 Fisher 卷积编码从特征图提取特征向量，减少冗余的同时避免过度抽象丢失信息。

在主体训练阶段，为了增加特征的丰富度，本申请使用了多分支残差卷积作为基本模块，保证了特征提取能力；而在模型分支扩展阶段，本申请则使用了深度分离卷积、共享局部学习模块等策略减少开销。

本申请首次提出使用可变形卷积来搭建关联关系学习模块，利用可变形卷积的几何灵活性对局部对象的关联关系准确建模。

本申请还使用主体部分的场景置信度作为正则项，结合 focus loss，很好地对类别不平衡的细粒度场景识别进行优化。

模型主体训练阶段的第一轮也可以使用 focal loss 只针对核心特征提取层进行充分训练，然后再单独训练全局信息特征提取模块。

全局信息特征提取模块可以单纯使用两层反卷积同时完成尺寸上采样和通道扩充，但这样会减慢收敛速度。

全局信息特征提取模块也可使用通道级的注意力机制和全连接层完成特征融合。

局部监督学习模块可使用图像级的语义标签结合辅助损失与全连接层一起训练。

细粒度分支扩展网络也可在已有的分支扩展网络上进行扩展，而不必都以主体网络作为扩展的起点。

模型主体部分也可以使用基于深度分离卷积的基础模块来降低开销，同时 $n \times n$ 卷积也可以转化为等价的 $1 \times n$ 和 $n \times 1$ 卷积减少开销。

关联关系学习可以设计专门的损失函数，多层次独自训练，而无需混合在分支扩展网络一起训练。

实施例 6:

图 10 为本申请实施例提供的场景识别过程示意图，该过程包括:

S201: 获取待识别的图像。

S202: 将所述待识别的图像输入预先训练完成的场景识别模型, 基于所述场景识别模型确定所述待识别的图像对应的场景信息。

在本申请实施例提供的场景识别方法应用于电子设备, 该电子设备可以为PC、平板电脑等智能设备, 也可以为服务器。该进行场景识别的电子设备和上述实施例中进行模型训练的电子设备可以相同, 也可以不同。

由于模型训练的过程一般是离线的, 进行模型训练的电子设备通过上述实施例中的方法训练模型, 可以直接将训练完成的场景识别模型保存在进行场景识别的电子设备中, 以便后续进行场景识别的电子设备, 直接通过该训练完成的场景识别模型进行相应的处理。

在本申请实施例中, 将输入到的场景识别模型进行处理的图像作为待识别的图像。获取到该待识别的图像之后, 将该待识别的图像输入预先训练完成的场景识别模型, 基于场景识别模型确定待识别的图像对应的场景信息。

实施例 7:

图 11 为本申请实施例提供的场景识别模型训练装置结构示意图, 该装置包括:

第一训练单元 11, 设置为通过样本图像的第一场景标签和标准交叉熵损失, 训练得到所述核心特征提取层和所述全局信息特征提取层的参数。

第二训练单元 12, 设置为根据所述各层级的 LCS 模块输出的特征图和所述样本图像的第一场景标签逐像素计算得到的损失值, 训练所述各层级的 LCS 模块的权重参数。

第三训练单元 13, 设置为通过所述样本图像的第一场景标签和标准交叉熵损失, 训练得到所述全连接决策层的参数。

所述装置还包括:

第四训练单元 14, 设置为根据所述分支扩展结构的卷积层输出的特征图和所述样本图像的第二场景标签逐像素计算得到的损失值, 训练所述分支扩展结构各层级的卷积层的权重参数; 通过带场景置信正则项的损失函数, 训练得到局部对象关联关系学习模块的参数; 其中, 所述第一场景标签和所述第二场景标签的粒度不同。

所述第一训练单元 11, 还设置为对所述核心特征提取层中不同层级的特征图使用膨胀因子不同的反卷积操作进行上采样, 在通道维度使用双线性插值算法对齐通道数, 各个层级的特征图逐通道进行相加合并, 合并后的特征图组进

行卷积融合，并通过逐通道的全局平均池化得到全局信息特征向量，将所述全局信息特征向量和全连接层 FC 特征向量拼接，并通过标准交叉熵损失训练得到所述核心特征提取层和全局信息特征提取层的参数。

所述第二训练单元 12，还设置为通过通道维度的注意力机制使用激活函数获取每个通道的重要性权重，根据所述每个通道的重要性权重对每个通道的特征图进行加权求和，得到汇总热力图；根据所述汇总热力图、对象场景关联重要度和对象的面积逐像素计算损失值，根据所述损失值训练所述各层级的 LCS 模块的权重参数。

实施例 8：

图 12 为本申请实施例提供的场景识别装置结构示意图，该装置包括：

获取单元 21，设置为获取待识别的图像；

识别单元 22，设置为将所述待识别的图像输入预先训练完成的场景识别模型，基于所述场景识别模型确定所述待识别的图像对应的场景信息。

所述装置还包括：

确定单元 23，设置为响应于确定出的所述待识别图像对应的场景信息属于违规场景信息且机审的审核结果为所述待识别图像为违规图像，确定所述待识别图像为违规图像。

实施例 9：

在上述各实施例的基础上，本申请实施例中还提供了一种电子设备，如图 13 所示，包括：处理器 301、通信接口 302、存储器 303 和通信总线 304，其中，处理器 301，通信接口 302，存储器 303 通过通信总线 304 完成相互间的通信；

所述存储器 303 中存储有计算机程序，当所述程序被所述处理器 301 执行时，使得所述处理器 301 执行如下步骤：

通过样本图像的第一场景标签和标准交叉熵损失，训练得到所述核心特征提取层和所述全局信息特征提取层的参数；

根据所述各层级的 LCS 模块输出的特征图和所述样本图像的第一场景标签逐像素计算得到的损失值，训练所述各层级的 LCS 模块的权重参数；

通过所述样本图像的第一场景标签和标准交叉熵损失，训练得到所述全连接决策层的参数。

基于同一发明构思，本申请实施例中还提供了一种电子设备，由于上述电子设备解决问题的原理与场景识别模型训练方法相似，因此上述电子设备的实施可以参见方法的实施，重复之处不再赘述。

实施例 10:

在上述各实施例的基础上，本申请实施例中还提供了一种电子设备，如图 14 所示，包括：处理器 401、通信接口 402、存储器 403 和通信总线 404，其中，处理器 401，通信接口 402，存储器 403 通过通信总线 404 完成相互间的通信；

所述存储器 403 中存储有计算机程序，当所述程序被所述处理器 401 执行时，使得所述处理器 401 执行如下步骤：

获取待识别的图像；

将所述待识别的图像输入预先训练完成的场景识别模型，基于所述场景识别模型确定所述待识别的图像对应的场景信息。

基于同一发明构思，本申请实施例中还提供了一种电子设备，由于上述电子设备解决问题的原理与场景识别方法相似，因此上述电子设备的实施可以参见方法的实施，重复之处不再赘述。

实施例 11:

在上述各实施例的基础上，本申请实施例还提供了一种计算机存储可读存储介质，所述计算机可读存储介质内存储有可由电子设备执行的计算机程序，当所述程序在所述电子设备上运行时，使得所述电子设备执行时实现如下步骤：

通过样本图像的第一场景标签和标准交叉熵损失，训练得到所述核心特征提取层和所述全局信息特征提取层的参数；

根据所述各层级的 LCS 模块输出的特征图和所述样本图像的第一场景标签逐像素计算得到的损失值，训练所述各层级的 LCS 模块的权重参数；

通过所述样本图像的第一场景标签和标准交叉熵损失，训练得到所述全连接决策层的参数。

基于同一发明构思，本申请实施例中还提供了一种计算机可读存储介质，由于处理器在执行上述计算机可读存储介质上存储的计算机程序时解决问题的原理与场景识别模型训练方法相似，因此处理器在执行上述计算机可读存储介质存储的计算机程序的实施可以参见方法的实施，重复之处不再赘述。

实施例 12:

在上述各实施例的基础上，本申请实施例还提供了一种计算机存储可读存储介质，所述计算机可读存储介质内存储有可由电子设备执行的计算机程序，当所述程序在所述电子设备上运行时，使得所述电子设备执行时实现如下步骤：

获取待识别的图像；

将所述待识别的图像输入预先训练完成的场景识别模型，基于所述场景识

别模型确定所述待识别的图像对应的场景信息。

基于同一发明构思，本申请实施例中还提供了一种计算机可读存储介质，由于处理器在执行上述计算机可读存储介质上存储的计算机程序时解决问题的原理与场景识别方法相似，因此处理器在执行上述计算机可读存储介质存储的计算机程序的实施可以参见方法的实施，重复之处不再赘述。

本申请实施例提供了一种模型训练和场景识别方法、装置、设备及介质，用以提供一种准确性较高的场景识别方案。

本申请是参照根据本申请实施例的方法、设备（系统）、和计算机程序产品的流程图和 / 或方框图来描述的。应理解可由计算机程序指令实现流程图和 / 或方框图中的每一流程和 / 或方框、以及流程图和 / 或方框图中的流程和 / 或方框的结合。可提供这些计算机程序指令到通用计算机、专用计算机、嵌入式处理机或其他可编程数据处理设备的处理器以产生一个机器，使得通过计算机或其他可编程数据处理设备的处理器执行的指令产生用于实现在流程图一个流程或多个流程和 / 或方框图一个方框或多个方框中指定的功能的装置。

这些计算机程序指令也可存储在能引导计算机或其他可编程数据处理设备以特定方式工作的计算机可读存储器中，使得存储在该计算机可读存储器中的指令产生包括指令装置的制造品，该指令装置实现在流程图一个流程或多个流程和 / 或方框图一个方框或多个方框中指定的功能。

这些计算机程序指令也可装载到计算机或其他可编程数据处理设备上，使得在计算机或其他可编程设备上执行一系列操作步骤以产生计算机实现的处理，从而在计算机或其他可编程设备上执行的指令提供用于实现在流程图一个流程或多个流程和 / 或方框图一个方框或多个方框中指定的功能的步骤。

尽管已描述了本申请的一些实施例，但本领域内的技术人员一旦得知了基本创造性概念，则可对这些实施例作出另外的变更和修改。所以，所附权利要求意欲解释为包括一些实施例以及落入本申请范围的所有变更和修改。

本领域的技术人员可以对本申请进行各种改动和变型而不脱离本申请的精神和范围。这样，倘若本申请的这些修改和变型属于本申请权利要求及其等同技术的范围之内，则本申请也意图包含这些改动和变型在内。

权利要求书

1、一种场景识别模型训练方法，所述场景识别模型包括核心特征提取层以及与所述核心特征提取层连接的全局信息特征提取层、至少一个层级的带注意力机制的局部监督学习 LCS 模块、全连接决策层，所述方法包括：

通过样本图像的第一场景标签和标准交叉熵损失，训练得到所述核心特征提取层和所述全局信息特征提取层的参数；

根据每个层级的 LCS 模块输出的特征图和所述样本图像的第一场景标签逐像素计算得到的损失值，训练每个层级的 LCS 模块的权重参数；

通过所述样本图像的第一场景标签和标准交叉熵损失，训练得到所述全连接决策层的参数。

2、如权利要求 1 所述的方法，其中，所述场景识别模型还包括分支扩展结构；所述分支扩展结构包括卷积层和局部对象关联关系学习模块；

所述方法还包括：

根据所述分支扩展结构的卷积层输出的特征图和所述样本图像的第二场景标签逐像素计算得到的损失值，训练所述分支扩展结构至少一个层级的卷积层的权重参数；通过带场景置信正则项的损失函数，训练得到局部对象关联关系学习模块的参数；其中，所述第一场景标签和所述第二场景标签的粒度不同。

3、如权利要求 1 所述的方法，其中，所述核心特征提取层包括第一类分组多感受野残差卷积模块和第二类分组多感受野残差卷积模块；

所述第一类分组多感受野残差卷积模块包括第一分组、第二分组和第三分组，所述第一分组、第二分组和第三分组的卷积尺寸不同，所述第一分组、第二分组和第三分组包括残差计算旁路结构；每个分组通过卷积操作和残差计算输出特征图，每个分组输出的特征图在通道维度拼接并进行通道混洗，卷积融合后输出到下一模块；

所述第二类分组多感受野残差卷积模块包括第四分组、第五分组和第六分组，所述第四分组、第五分组和第六分组的卷积尺寸不同，所述第五分组和第六分组分别包括 1×1 卷积旁路结构和残差计算旁路结构；每个分组输出的特征图在通道维度拼接并进行通道混洗，卷积融合后输出到下一模块。

4、如权利要求 1 所述的方法，其中，所述通过样本图像的第一场景标签和标准交叉熵损失，训练得到所述核心特征提取层和所述全局信息特征提取层的参数包括：

对所述核心特征提取层中不同层级的特征图使用膨胀因子不同的反卷积操作进行上采样，在通道维度使用双线性插值算法对齐通道数，至少一个层级的

特征图逐通道进行相加合并，合并后的特征图组进行卷积融合，并通过逐通道的全局平均池化得到全局信息特征向量，将所述全局信息特征向量和全连接层 FC 特征向量拼接，并通过标准交叉熵损失训练得到所述核心特征提取层和全局信息特征提取层的参数。

5、如权利要求 1 所述的方法，其中，所述根据每个层级的 LCS 模块输出的特征图和所述样本图像的第一场景标签逐像素计算得到的损失值，训练每个层级的 LCS 模块的权重参数包括：

通过通道维度的注意力机制使用激活函数获取每个通道的重要性权重，根据所述每个通道的重要性权重对每个通道的特征图进行加权求和，得到汇总热力图；

根据所述汇总热力图、对象场景关联重要度和对象的面积逐像素计算损失值，根据所述损失值训练每个层级的 LCS 模块的权重参数。

6、如权利要求 2 所述的方法，其中，所述分支扩展结构使用深度分离卷积残差块 DW 构造，在残差块的主路中，中间层使用 DW 卷积层，DW 卷积层前后使用 1x1 卷积层。

7、如权利要求 2 所述的方法，其中，所述局部对象关联关系学习模块包括可变形卷积层、卷积层和平均池化层；

可变形卷积层获取当前像素位置的卷积核偏移值，卷积核参数的当前位置加上偏移值作为所述卷积核参数的真实生效的位置，获取真实生效的位置的特征图像素值，经过卷积操作和平均池化操作输出特征图。

8、一种基于如权利要求 1-7 任一项所述的方法训练得到的场景识别模型的场景识别方法，所述方法包括：

获取待识别的图像；

将所述待识别的图像输入预先训练完成的场景识别模型，基于所述场景识别模型确定所述待识别的图像对应的场景信息。

9、如权利要求 8 所述的方法，还包括：

响应于确定出的所述待识别图像对应的场景信息属于违规场景信息且机审的审核结果为所述待识别图像为违规图像，确定所述待识别图像为违规图像。

10、一种场景识别模型训练装置，包括：

第一训练单元，设置为通过样本图像的第一场景标签和标准交叉熵损失，训练得到核心特征提取层和全局信息特征提取层的参数；

第二训练单元，设置为根据每个层级的 LCS 模块输出的特征图和所述样本

图像的第一场景标签逐像素计算得到的损失值，训练每个层级的 LCS 模块的权重参数；

第三训练单元，设置为通过所述样本图像的第一场景标签和标准交叉熵损失，训练得到全连接决策层的参数。

11、一种基于如权利要求 10 所述的装置训练得到的场景识别模型的场景识别装置，所述装置包括：

获取单元，设置为获取待识别的图像；

识别单元，设置为将所述待识别的图像输入预先训练完成的场景识别模型，基于所述场景识别模型确定所述待识别的图像对应的场景信息。

12、一种电子设备，所述电子设备包括处理器，所述处理器用于执行存储器中存储的计算机程序时实现如权利要求 1-7 中任一所述场景识别模型训练方法的步骤，或实现如权利要求 8-9 所述场景识别方法的步骤。

13、一种计算机可读存储介质，所述计算机可读存储介质存储有计算机程序，所述计算机程序被处理器执行时实现如权利要求 1-7 中任一所述场景识别模型训练方法的步骤，或实现如权利要求 8-9 所述场景识别方法的步骤。