



(12)发明专利申请

(10)申请公布号 CN 106651887 A

(43)申请公布日 2017.05.10

(21)申请号 201710025737.2

(22)申请日 2017.01.13

(71)申请人 深圳市唯特视科技有限公司

地址 518057 广东省深圳市高新技术产业
园区高新南一道009号中研发园新
产业孵化中心楼610室

(72)发明人 夏春秋

(51)Int.Cl.

G06T 7/11(2017.01)

G06N 3/08(2006.01)

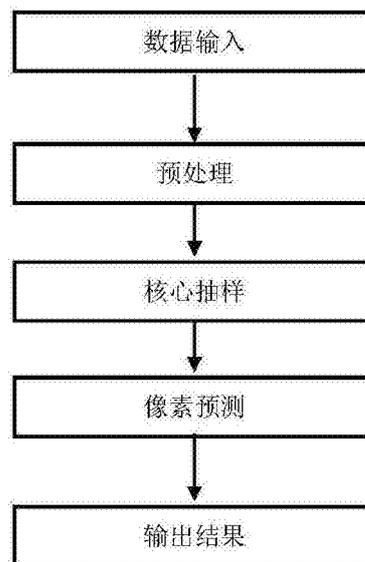
权利要求书2页 说明书4页 附图3页

(54)发明名称

一种基于卷积神经网络的图像像素分类方法

(57)摘要

本发明中提出的一种基于卷积神经网络的图像像素分类方法,其主要内容包括:数据输入、预处理、核心抽样、像素预测、输出结果,其过程为,输入图像首先被叠加在一起并调整为统一大小后被输入到预训练模型,获取对应每个图像的每一像素生成的中间映射,同一像素对应的多个中间映射形成皮层柱;接下来,从皮层柱的集合中随机抽取样本,并将得到的核心样本反馈到第二阶段的深度置信网络;然后网络对核心样本进行像素预测;最后,根据当前任务的不同输出不同的结果。本发明在图像分割方面,与现有的技术相比,它有助于产生细粒度分割;使用预训练的VGG-16模型,提高了整体性能,加快训练速度,另一个优点是避免使用巨大的数据集。



1. 一种基于卷积神经网络的图像像素分类方法,其特征在于,主要包括数据输入(一);预处理(二);核心抽样(三);像素预测(四);输出结果(五)。

2. 基于权利要求书1所述的一种基于卷积神经网络的图像像素分类方法,其特征在于,包括核心抽样框架,创建一个表示,结合来自测试数据的特征和从预训练网络的响应中获得的上下文知识,处理它并将其反馈到单独的深度置信网络;使用这种表示法从图像中在像素级上提取更多的信息,从而获得对整个图像的理解;使用这些映射响应,它从几个作为特征的层到另一个神经网络转移学习提供对输入图像的理解。

3. 基于权利要求书1所述的数据输入(一),其特征在于,采用由40个易于识别的动物对象类的图片组成的动物数据集,这40个对象类包含狗、猫、猫头鹰、海豹等;该数据集中的图像是单通道图像且大小不一致。

4. 基于权利要求书1所述的预处理(二),其特征在于,在地面真实图像集中较小的图片周围添加填充,使所有输入图像的大小都为 224×224 ;在提取映射响应之前,从该数据集中较大的图像中,通过使用 224×224 大小的滑动窗口和更小的步幅大小来创建同样大小的图片;通过改变对比度生成更多训练数据,单独地归一化用作特征的映射响应,并且在测试期间将相同的归一化参数用于相应的特征。

5. 基于权利要求书4所述的映射响应,其特征在于,将图像输入VGG-16模型后得到的来自卷积神经网络各层的响应,代表该图像各像素点的特征,可以被视为在图像不同位置处观看到图像的金字塔形状的不同水平;神经网络的第一层学习与Gabor特征或颜色块相似的特征,更深层的映射响应提取到越来越多的抽象特征,用于区分对象和对象的部分信息,同时失去空间和局部信息。

6. 基于权利要求书1所述的核心抽样(三),其特征在于,使用预训练模型提取映射时,首先从R,G和B值中减去常数来对图像进行归一化,接着从网络的每个层获取图像的映射响应,最后使用双线性插值法获得皮层柱;定义输入图像的每一个像素对应的皮层柱的集合为一个核心,从该核心随机抽取样本作为深度置信网络的输入。

7. 基于权利要求书6所述的皮层柱,其特征在于,在测试阶段期间,将图像叠加在一起并调整为统一大小后输入到VGG-16模型,对应每个像素的中间映射形成皮层柱;对应输入图像上的一个像素的皮层柱是一个有k列的向量,其中k是在VGG-16模型的中间映射的数量,向量的每一个成分都是一个映射;皮层柱不保留组成映射之间的任何空间相关性,是深度置信网络的训练数据。

8. 基于权利要求书1所述的像素预测(四),其特征在于,将核心抽样后得到的核心样本反馈到深度置信网络中,根据当前的任务,使用两种不同类型的输出层:执行线性回归并使用均方误差作为损失函数的回归层,以及可以使用负对数似然作为损失函数来对像素进行分类的logistic回归层。

9. 基于权利要求书8所述的线性回归层和logistic回归层,其特征在于,通常用于回归问题的损失函数是均方误差:

$$\frac{1}{n} \sum_{i=1}^n (\hat{y}^{(i)} - y^{(i)})^2 \quad (1)$$

其中在n个观测值中, \hat{y} 是预测值向量和y是实际值向量;大多数神经网络执行分类而不

是执行回归,任何两个标签之间的距离将是相同的;在这种情况下,似然(\mathcal{L})和损失(1)的功能是由公式(2)给出:

$$\begin{aligned}\mathcal{L}(\theta = \{W, b\}, \mathcal{D}) &= \sum_{i=0}^{|\mathcal{D}|} \log \left(P(Y = y^{(i)} | x^{(i)}, W, b) \right) \\ l(\theta = \{W, b\}, \mathcal{D}) &= -\mathcal{L}(\theta = \{W, b\}, \mathcal{D}),\end{aligned}\tag{2}$$

其中 W 、 b 分别是权重和偏差, \mathcal{D} 是数据集;给定一个输入,权重矩阵和偏差向量时,它输出该输入 x^i 属于某一类 y^i 的似然;由于这个方程是基于概率值而不是距离测量,它更适合于分类。

10. 基于权利要求书1所述的输出结果(五),其特征在于,综合每个像素的预测结果,并根据当前任务的不同输出不同的结果;执行分类任务得到图像分割后的结果,而执行回归任务则得到图像着色后的结果。

一种基于卷积神经网络的图像像素分类方法

技术领域

[0001] 本发明涉及图像分割领域,尤其是涉及了一种基于卷积神经网络的图像像素分类方法。

背景技术

[0002] 随着卷积网络架构的日益完善,卷积神经网络已经被广泛地用于图像分割,在没有任何人为干预的情况下,从图像提取信息并且能够以高精度识别图像中的对象。然而,这样的一个网络架构需要在一个巨大的数据集上训练,而且通常没有足够的训练数据用于特定任务,导致了在执行分割任务时前期训练时间漫长,整体性能差的问题。而如果采用基于卷积神经网络的图像像素分类方法,则可以通过使用预训练模型和转移学习解决上述问题,还可以应用于人脸识别、指纹识别、交通控制系统、行人检测、医学影像等领域。

[0003] 本发明提出了一种基于卷积神经网络的图像像素分类方法,采用核心抽样框架对图像进行自动分割,它包含两阶段:输入图像首先被叠加在一起并调整为统一大小后被输入到预训练的VGG-16模型,获取对应每个图像的每一像素生成的中间映射,同一像素对应的多个中间映射形成皮层柱;接下来,从皮层柱的集合中随机抽取样本,并将得到的核心样本反馈到第二阶段的深度置信网络;然后网络对核心样本进行像素预测;最后,根据当前任务的不同输出不同的结果,执行分类任务得到图像分割后的结果,而执行回归任务则得到图像着色后的结果。本发明在图像分割方面,与现有的技术相比,它有助于产生细粒度分割;使用预训练的VGG-16模型,提高了整体性能,加快训练速度,另一个优点是避免使用巨大的数据集;它还使用核心样本作为第二阶段深度置信网络的输入,实现了像素级别的图像分割,提高基于已学习知识的可分性。

发明内容

[0004] 针对图像分割过程耗时和分割效果不够精细的问题,本发明的目的在于提供一种基于卷积神经网络的图像像素分类方法。

[0005] 为解决上述问题,本发明提供一种基于卷积神经网络的图像像素分类方法,其主要内容包括:

[0006] (一)数据输入;

[0007] (二)预处理;

[0008] (三)核心抽样;

[0009] (四)像素预测;

[0010] (五)输出结果。

[0011] 其中,一种基于卷积神经网络的图像像素分类方法,包括核心抽样框架,创建一个表示,结合来自测试数据的特征和从预训练网络的响应中获得的上下文知识,处理它并将其反馈到单独的深度置信网络;使用这种表示法从图像中在像素级上提取更多的信息,从而获得对整个图像的理解;使用这些映射响应,它从几个作为特征的层到另一个神经网络

转移学习提供对输入图像的理解。

[0012] 其中,所述的数据输入,采用由40个易于识别的动物对象类的图片组成的动物数据集,这40个对象类包含狗、猫、猫头鹰、海豹等;该数据集中的图像是单通道图像且大小不一致。

[0013] 其中,所述的预处理,在地面真实图像集中较小的图片周围添加填充,使所有输入图像的大小都为 224×224 ;在提取映射响应之前,从该数据集较大的图像中,通过使用 224×224 大小的滑动窗口和更小的步幅大小来创建同样大小的图片;通过改变对比度生成更多训练数据,单独地归一化用作特征的映射响应,并且在测试期间将相同的归一化参数用于相应的特征。

[0014] 进一步地,所述的映射响应,将图像输入VGG-16模型后得到的来自卷积神经网络各层的响应,代表该图像各像素点的特征,可以被视为在图像不同位置处观看到图像的金字塔形状的不同水平;神经网络的第一层学习与Gabor特征或颜色块相似的特征,更深层的映射响应提取到越来越多的抽象特征,用于区分对象和对象的部分信息,同时失去空间和局部信息。

[0015] 其中,所述的核心抽样,使用预训练模型提取映射时,首先从R,G和B值中减去常数来对图像进行归一化,接着从网络的每个层获取图像的映射响应,最后使用双线性插值法获得皮层柱;定义输入图像的每一个像素对应的皮层柱的集合为一个核心,从该核心随机抽取样本作为深度置信网络的输入。

[0016] 进一步地,所述的皮层柱,在测试阶段期间,将图像叠加在一起并调整为统一大小后输入到VGG-16模型,对应每个像素的中间映射形成皮层柱;对应输入图像上的一个像素的皮层柱是一个有k列的向量,其中k是在VGG-16模型的中间映射的数量,向量的每一个成分都是一个映射;皮层柱不保留组成映射之间的任何空间相关性,是深度置信网络的训练数据。

[0017] 其中,所述的像素预测,将核心抽样后得到的核心样本反馈到深度置信网络中,根据当前的任务,使用两种不同类型的输出层:执行线性回归并使用均方误差作为损失函数的回归层,以及可以使用负对数似然作为损失函数来对像素进行分类的logistic回归层。

[0018] 进一步地,所述的线性回归层和logistic回归层,通常用于回归问题的损失函数是均方误差:

$$[0019] \quad \frac{1}{n} \sum_{i=1}^n (\hat{y}^{(i)} - y^{(i)})^2 \quad (1)$$

[0020] 其中在n个观测值中, \hat{y} 是预测值向量和y是实际值向量;大多数神经网络执行分类而不是执行回归,任何两个标签之间的距离将是相同的;在这种情况下,似然(L)和损失(1)的功能是由公式(2)给出:

$$[0021] \quad \mathcal{L}(\theta = \{W, b\}, \mathcal{D}) = \sum_{i=0}^{|\mathcal{D}|} \log \left(P(Y = y^{(i)} | x^{(i)}, W, b) \right) \quad (2)$$

$$l(\theta = \{W, b\}, \mathcal{D}) = -\mathcal{L}(\theta = \{W, b\}, \mathcal{D}),$$

[0022] 其中W、b分别是权重和偏差,D是数据集;给定一个输入,权重矩阵和偏差向量时,它输出该输入 x^i 属于某一类 y^i 的似然;由于这个方程是基于概率值而不是距离测量,它更适合于分类。

[0023] 其中,所述的输出结果,综合每个像素的预测结果,并根据当前任务的不同输出不同的结果;执行分类任务得到图像分割后的结果,而执行回归任务则得到图像着色后的结果。

附图说明

[0024] 图1是本发明一种基于卷积神经网络的图像像素分类方法的系统流程图。

[0025] 图2是本发明一种基于卷积神经网络的图像像素分类方法的核心抽样框架图。

[0026] 图3是本发明一种基于卷积神经网络的图像像素分类方法的VGG-16网络架构图。

[0027] 图4是本发明一种基于卷积神经网络的图像像素分类方法的映射响应示意图。

[0028] 图5是本发明一种基于卷积神经网络的图像像素分类方法的动物数据集。

具体实施方式

[0029] 需要说明的是,在不冲突的情况下,本申请中的实施例及实施例中的特征可以相互结合,下面结合附图和具体实施例对本发明作进一步详细说明。

[0030] 图1是本发明一种基于卷积神经网络的图像像素分类方法的系统流程图。主要包括数据输入;预处理;核心抽样;像素预测;输出结果。

[0031] 其中,所述的数据输入,采用由40个易于识别的动物对象类的图片组成的动物数据集,这40个对象类包含狗、猫、猫头鹰、海豹等;该数据集中的图像是单通道图像且大小不一致。

[0032] 其中,所述的预处理,在地面真实图像集中较小的图片周围添加填充,使所有输入图像的大小都为 224×224 ;在提取映射响应之前,从该数据集较大的图像中,通过使用 224×224 大小的滑动窗口和更小的步幅大小来创建同样大小的图片;通过改变对比度生成更多训练数据,单独地归一化用作特征的映射响应,并且在测试期间将相同的归一化参数用于相应的特征。

[0033] 进一步地,所述的映射响应,将图像输入VGG-16模型后得到的来自卷积神经网络各层的响应,代表该图像各像素点的特征,可以被视为在图像不同位置处观看到图像的金字塔形状的不同水平;神经网络的第一层学习与Gabor特征或颜色块相似的特征,更深层的映射响应提取到越来越多的抽象特征,用于区分对象和对象的部分信息,同时失去空间和局部信息。

[0034] 其中,所述的核心抽样,使用预训练模型提取映射时,首先从R,G和B值中减去常数来对图像进行归一化,接着从网络的每个层获取图像的映射响应,最后使用双线性插值法获得皮层柱;定义输入图像的每一个像素对应的皮层柱的集合为一个核心,从该核心随机抽取样本作为深度置信网络的输入。

[0035] 进一步地,所述的皮层柱,在测试阶段期间,将图像叠加在一起并调整为统一大小后输入到VGG-16模型,对应每个像素的中间映射形成皮层柱;对应输入图像上的一个像素的皮层柱是一个有k列的向量,其中k是在VGG-16模型的中间映射的数量,向量的每一个成分都是一个映射;皮层柱不保留组成映射之间的任何空间相关性,是深度置信网络的训练数据。

[0036] 其中,所述的像素预测,将核心抽样后得到的核心样本反馈到深度置信网络中,根

据当前的任务,使用两种不同类型的输出层:执行线性回归并使用均方误差作为损失函数的回归层,以及可以使用负对数似然作为损失函数来对像素进行分类的logistic回归层。

[0037] 进一步地,所述的线性回归层和logistic回归层,通常用于回归问题的损失函数是均方误差:

$$[0038] \quad \frac{1}{n} \sum_{i=1}^n (\hat{y}^{(i)} - y^{(i)})^2 \quad (1)$$

[0039] 其中在n个观测值中, \hat{y} 是预测值向量和y是实际值向量;大多数神经网络执行分类而不是执行回归,任何两个标签之间的距离将是相同的;在这种情况下,似然(\mathcal{L})和损失(1)的功能是由公式(2)给出:

$$[0040] \quad \begin{aligned} \mathcal{L}(\theta = \{W, b\}, \mathcal{D}) &= \sum_{i=0}^{|\mathcal{D}|} \log(P(Y = y^{(i)} | x^{(i)}, W, b)) \\ l(\theta = \{W, b\}, \mathcal{D}) &= -\mathcal{L}(\theta = \{W, b\}, \mathcal{D}), \end{aligned} \quad (2)$$

[0041] 其中W、b分别是权重和偏差,D是数据集;给定一个输入,权重矩阵和偏差向量时,它输出该输入 x^i 属于某一类 y^i 的似然;由于这个方程是基于概率值而不是距离测量,它更适合于分类。

[0042] 其中,所述的输出结果,综合每个像素的预测结果,并根据当前任务的不同输出不同的结果;执行分类任务得到图像分割后的结果,而执行回归任务则得到图像着色后的结果。

[0043] 图2是本发明一种基于卷积神经网络的图像像素分类方法的核心抽样框架图。映射响应由预训练网络通过输入图像生成,其被处理为核心样本并且被馈送到深度置信网络中。输出是对测试图像的每个像素的预测,并且可以根据手头的任务而不同。例如分割(使用分类),着色(使用回归)。

[0044] 图3是本发明一种基于卷积神经网络的图像像素分类方法的VGG-16网络架构图。该架构由多个卷积层组成,最大池层在卷基层和3个完全连接层之间,最后是softmax层。该网络在ImageNet数据集上进行训练,该数据集包含各种各样的对象。这使得它成为一个构建框架的完美模型,适用于各种数据集。

[0045] 图4是本发明一种基于卷积神经网络的图像像素分类方法的映射响应示意图。将图像输入VGG-16模型后得到的来自卷积神经网络各层的响应,代表该图像各像素点的特征,可以被视为在图像不同位置处观看到图像的金字塔形状的不同水平。较低的数字表示浅层映射,它检测图像的边缘,而较高的数字表示来自较深层的映射,提取到图像更多的抽象特征。

[0046] 图5是本发明一种基于卷积神经网络的图像像素分类方法的动物数据集。它是由40个易于识别的动物对象类的图片组成的动物数据集,这40个对象类包含狗、猫、猫头鹰、海豹等;该数据集中的图像是单通道图像且大小不一致。

[0047] 对于本领域技术人员,本发明不限制于上述实施例的细节,在不背离本发明的精神和范围的情况下,能够以其他具体形式实现本发明。此外,本领域的技术人员可以对本发明进行各种改动和变型而不脱离本发明的精神和范围,这些改进和变型也应视为本发明的保护范围。因此,所附权利要求意欲解释为包括优选实施例以及落入本发明范围的所有变更和修改。

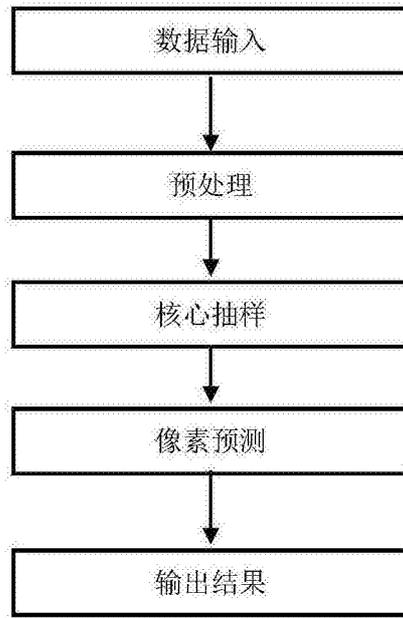


图1

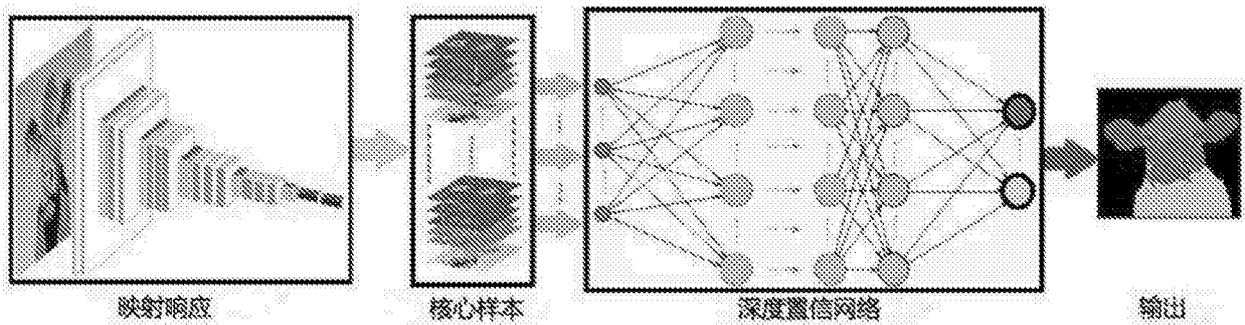


图2

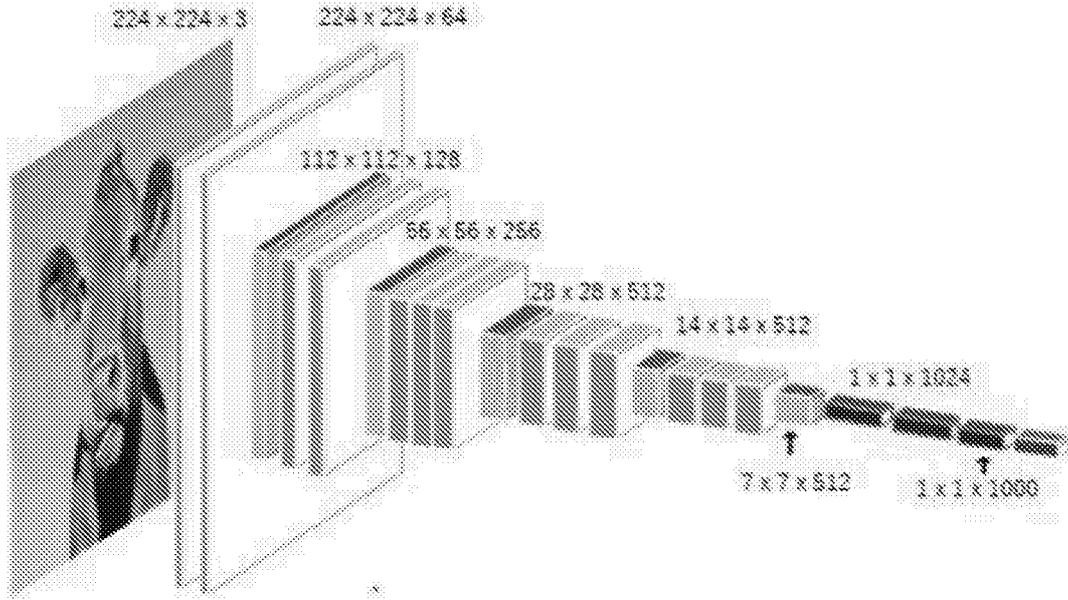


图3

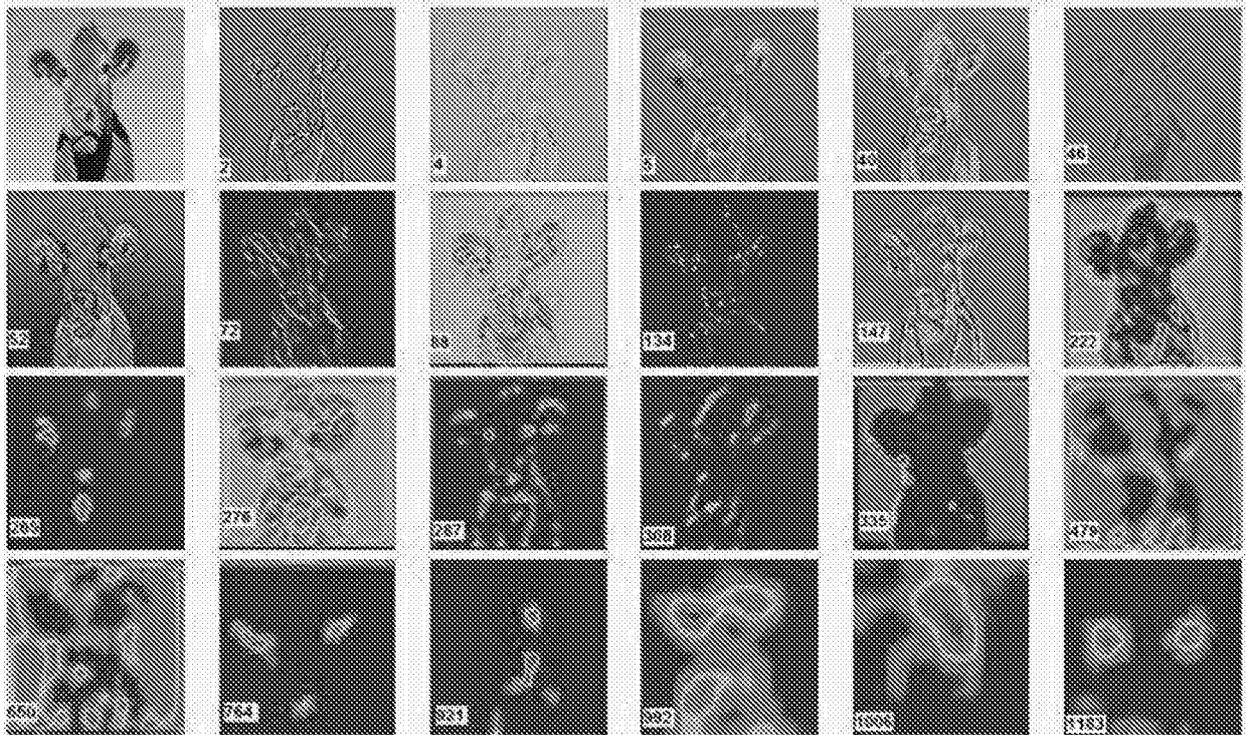


图4

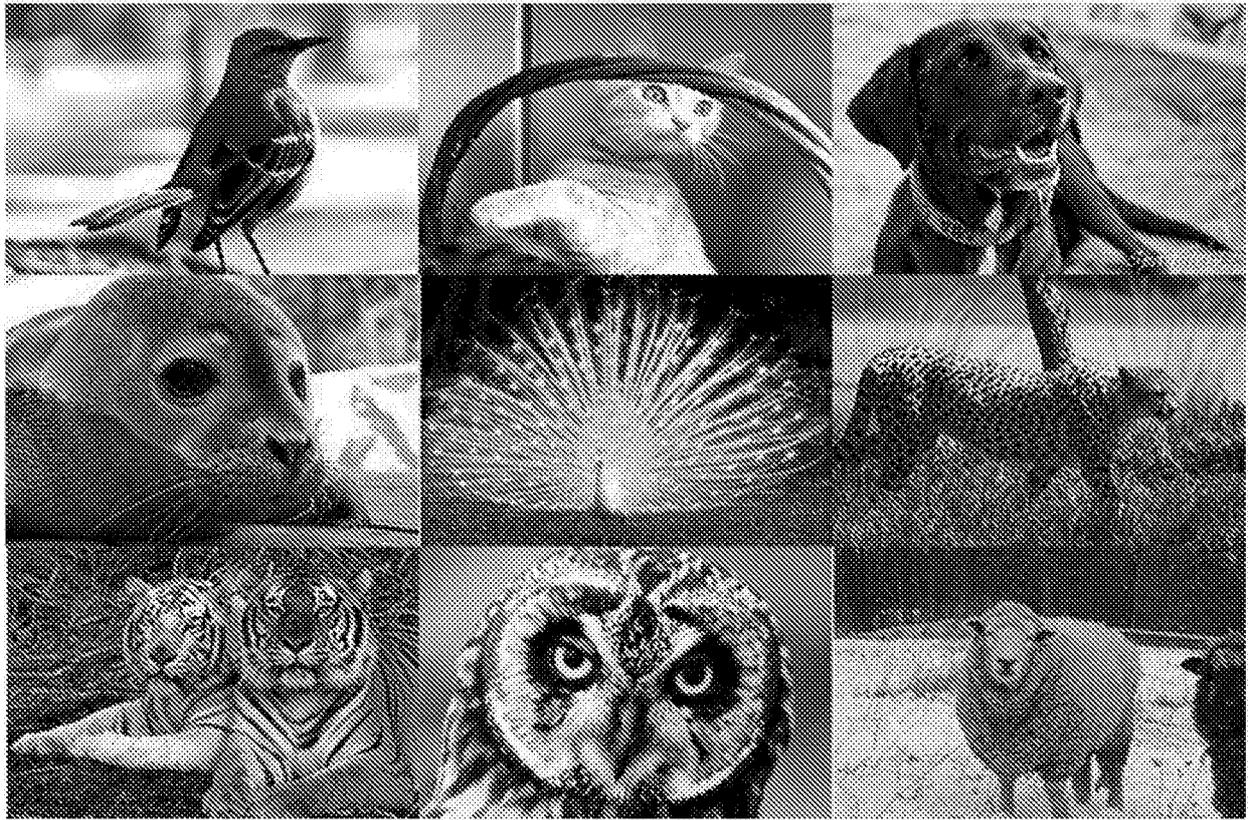


图5