(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 107636693 A
(43)申请公布日 2018.01.26

(21)申请号 201580080124.8
(22)申请日 2015.03.20
(85)PCT国际申请进入国家阶段日
2017.11.17
(86)PCT国际申请的申请数据
PCT/EP2015/056008 2015.03.20
(87)PCT国际申请的公布数据
WO2016/150472 EN 2016.09.29
(71)申请人 弗劳恩霍夫应用研究促进协会
地址 德国慕尼黑
申请人 柏林技术大学
(72)发明人 塞巴斯蒂安·巴赫
沃耶西·萨梅克
克劳斯-罗伯特·穆勒
亚历山大·宾德
格雷格里·蒙塔翁
(74)专利代理机构 中科专利商标代理有限责任公司 11021
代理人 周泉
(51)Int.Cl.
G06K 9/62(2006.01)
G06K 9/46(2006.01)
G06N 3/04(2006.01)
G06N 3/08(2006.01)

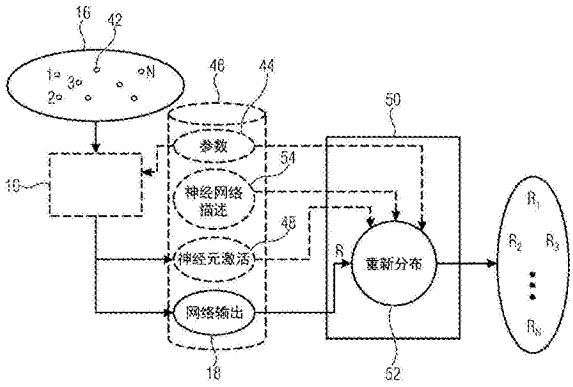
权利要求书5页 说明书39页 附图19页

(54)发明名称

针对人工神经网络的相关性分数指派

(57)摘要

通过穿过人工神经网络反向传播初始相关性分数，将从网络输出导出的初始相关性分数重新分布到项目集合上，来获得对应用人工神经网络的项目集合的相关性分数指派的任务，从而获得每个项目的相关性分数。具体地，这种反向传播适用于更广泛的人工神经网络集合，和/或实现更低的计算工作，这通过用以下方式同样地执行反向传播来实现：使得对于每个神经元，根据分布函数将相应神经元的下游邻居神经元集合的初步重新分布的相关性分数分布在相应神经元的上游邻居神经元集合上。



CN 107636693 A

1.一种用于将相关性分数指派给项目的集合的装置，所述相关性分数指示关于将由神经元(12)组成的人工神经网络(10)应用到项目(42)的集合(16)上以将项目(42)的集合(16)映射到网络输出(18)上的相关性，所述装置被配置为：

通过将从网络输出(18)导出的初始相关性分数(R)反向传播通过人工神经网络(10)来将所述初始相关性分数(R)重新分布到项目(42)的集合(16)上，以获得每个项目的相关性分数，

其中，所述装置被配置为通过以下方式执行所述反向传播：使得针对每个神经元，使用分布函数将相应神经元的下游邻居神经元集合的初步重新分布的相关性分数分布到该相应神经元的上游邻居神经元集合。

2.根据权利要求1所述的装置，其中，所述装置被配置为使得所述分布函数具有相关性守恒属性。

3.根据权利要求1或2所述的装置，其中，所述装置被配置为通过对所述人工神经网络的所有神经元均等地使用一个分布函数来执行所述反向传播。

4.根据权利要求1至3中任一项所述的装置，其中，所述装置被配置为使得所述分布函数是如下项的函数：

人工神经网络的权重，其确定所述相应神经元受所述相应神经元的上游邻居神经元集合影响的程度，

上游邻居神经元集合的神经元激活，其在将人工神经网络(10)应用于项目(42)的集合(16)上时表现自身；以及

所述相应神经元的下游邻居神经元集合的初步重新分布的相关性分数的和。

5.根据权利要求1至4中任一项所述的装置，其中，所述装置被配置为使得对于每个神经元j，给出多少相关性被重新分布为从相应神经元j到上游邻居神经元i的相关性消息R_{ij}的分布函数是

$R_{ij}=q(i)\cdot m(\{R_{jk},k\text{是}j\text{的下游邻居神经元}\})$

其中，$m(\mathbb{R}^K)$是其所有分量的单调递增函数并且给出相应神经元j的初步重新分布的相关性分数$R_j=m(\{R_{jk},k\text{是}j\text{的下游神经元}\})$，其中K是所述相应神经元j的下游邻居的数目，q(i)是取决于如下因素的函数：将上游邻居神经元i连接到所述相应神经元j的权重w_{ij}、由于将人工神经网络(10)应用于项目(42)的集合(16)上得到的所述相应神经元j的上游邻居神经元i的激活x_i、以及神经元j的可能的零值偏置项b_j。

6.根据权利要求5所述的装置，其中，$m(\{R_{jk},k\text{是}j\text{的下游神经元}\})=\sum_k R_{jk}$。

7.根据权利要求5或6所述的装置，其中，所述装置被配置为使得所述函数q(i)是通过函数s计算的加权激活$z_{ij}=s(x_i,w_{ij},b_j)$的函数p，从而$q(i)=p(\{z_{ij}|i\text{是}j\text{的上游邻居神经元}\})$。

8.根据权利要求7的装置，其中，函数s被选择成使得所述加权激活z_{ij}为：

$z_{ij}=x_iw_{ij}$，

或$z_{ij}=x_iw_{ij}+\frac{b_j}{I}$

其中，I是神经元j的上游邻居神经元i的数目。

9.根据权利要求5至8中任一项所述的装置，其中，所述装置被配置为使得对于满足$R_j>$

0的每个神经元j,所述函数q(i)满足排序属性,

在下列情况下满足所述排序属性:

a)如果$\Sigma_i z_{ij}>0$,则对于神经元j的上游邻居神经元中使得$z_{i_1 j} < z_{i_2 j}$的全部i_1和i_2,$q(i_1)\leq q(i_2)$成立;

b)或者,对于神经元j的上游邻居神经元中使得$z_{i_1 j}>0$且$z_{i_2 j}>0$且$z_{i_1 j} < z_{i_2 j}$的全部i_1和i_2,

$0\leq q(i_1)\leq q(i_2)$成立。

10.根据权利要求5至8中任一项所述的装置,其中,所述装置被配置为使得所述函数q(i)满足排序属性,在下述情况下满足所述排序属性:对于神经元j的上游邻居神经元中使得$g(z_{i_1 j}) < g(z_{i_2 j})$的全部$i_1$和$i_2$,

$|q(i_1)|\leq|q(i_2)|$成立,其中,函数g(•)的最小值为零并且在区间$(-\infty,0)$上单调递减,在区间$(0,+\infty)$上单调递增。

11.根据权利要求10所述的装置,其中,所述装置被配置为使得所述函数g(•)如下:

$g(z)=\alpha\max(0,z)-\beta\min(0,z)$,其中$\alpha>0,\beta\geq 0$

12.根据权利要求5至11中任一项所述的装置,其中,所述装置被配置为使得所述函数q(i)继承或与神经元的神经网络功能的泰勒分解成比例。

13.根据权利要求5至11中任一项所述的装置,其中,所述装置被配置为使得所述相关性消息R_{ij}与从数据中获得并将神经元j的上游邻居i的激活x_i映射到值m({R_{jk},k是j的下游神经元})的函数的所述泰勒分解成比例,至多存在近似误差。

14.根据权利要求1至13中任一项所述的装置,其中,所述装置被配置为使得所述分布函数是

$$R_{ij}=\frac{x_i w_{ij}+\frac{b_j}{n}}{h\left(\sum_r\left(x_r w_{rj}+\frac{b_j}{n}\right)\right)}\cdot m(\{R_{jk},k是j的下游邻居神经元\})$$

或者

$$R_{ij}=\frac{x_i w_{ij}}{h\left(\sum_r\left(x_r w_{rj}+\frac{b_j}{n}\right)\right)}\cdot m(\{R_{jk},k是j的下游邻居神经元\})$$

其中,n是相应神经元j的上游邻居神经元的数目,R_{ij}是从相应神经元j重新分布到上游邻居神经元i的相关性消息,R_{jk}是从下游邻居神经元k重新分布到相应神经元j的相关性消息,x_i是在将神经网络应用到项目(42)的集合(16)上期间上游邻居神经元i的激活,ω_{ij}是将上游邻居神经元i连接到相应神经元j的权重,w_{rj}也是将上游邻居神经元r连接到相应神经元j的权重,而b_j是相应神经元j的偏置项,以及h()是标量函数,并且其中,$m(\mathbb{R}^K)$是其所有分量的单调递增函数并且给出相应神经元j的初步重新分布的相关性分数$R_j=m(\{R_{jk},k是j的下游神经元\})$,其中K是所述相应神经元j的下游邻居的数目。

15.根据权利要求1至13中任一项所述的装置,其中,所述装置被配置为使得使用分布函数来执行到所述相应神经元j的上游邻居神经元i的集合上的分布,其中,所述分布函数

是

$$R_{ij}=\left(\alpha\frac{\left(x_i w_{ij}+\frac{b_j}{n}\right)_+}{h\left(\sum_r\left(x_r w_{rj}+\frac{b_j}{n}\right)_+\right)}-\beta\frac{\left(x_i w_{ij}+\frac{b_j}{n}\right)_-}{h\left(\sum_r\left(x_r w_{rj}+\frac{b_j}{n}\right)_-\right)}\right)$$

$$\cdot m(\{R_{jk},k是j的下游邻居神经元\})$$

或者

$$R_{ij}=\left(\alpha\frac{\left(x_i w_{ij}\right)_+}{h\left(\left(b_j\right)_+ +\sum_r\left(x_r w_{rj}\right)_+\right)}-\beta\frac{\left(x_i w_{ij}\right)_-}{h\left(\left(b_j\right)_- +\sum_r\left(x_r w_{rj}\right)_-\right)}\right)$$

$$\cdot m(\{R_{jk},k是j的下游邻居神经元\})$$

其中$(z)_+=\max(0,z)$，$(z)_-=\min(0,z)$，n是所述相应神经元的上游邻居神经元的数目，R_{ij}是从相应神经元j重新分布到所述上游邻居神经元i的相关性消息，而R_{jk}是从下游邻居神经元k重新分布到相应神经元j的相关性消息，x_i是在将神经网络应用到项目(42)的集合(16)上期间上游邻居神经元i的激活，ω_{ij}是将上游邻居神经元i连接到相应神经元j的权重，w_{rj}也是将上游邻居神经元r连接到相应神经元j的权重，而b_j是相应神经元j的偏置项，以及h()是标量函数，以及$\alpha>0$，$\beta\geq 0$，$\alpha-\beta=1$，并且其中，$m(\mathbb{R}^K)$是其所有分量的单调递增函数并且给出相应神经元j的初步重新分布的相关性分数$R_j=m(\{R_{jk},k是j的下游神经元\})$，其中K是相应神经元j的下游邻居的数目。

16.根据权利要求14或15所述的装置，其中，$m(\{R_{jk},k是j的下游神经元\})=\Sigma_k R_{jk}$。

17.根据权利要求14至16中任一项所述的装置，其中，h()是稳定函数$h(t)=t+\varepsilon\cdot\mathrm{sign}(t)$。

18.根据权利要求1至17所述的装置，其中，所述装置被配置为，针对每个项目i，通过对将所述相应项目作为上游邻居神经元的神经元的重新分布到所述相应项目的相关性消息进行求和，来计算相应项目i的相关性分数R_i。

19.根据权利要求1至18所述的装置，其中，所述人工神经网络直接应用于项目的集合上，使得项目(42)的集合(16)中的项目形成所述人工神经网络的人工神经元的子集的上游邻居，并且网络输出对应于在人工神经网络的下游端的神经元的神经元激活。

19.根据前述权利要求中任一项所述的装置，其中，所述网络输出(18)是标量，其中从所述标量导出的初始相关性分数等于所述标量的值，或者是通过将单调递增函数应用于所述标量的值而导出的，或者网络输出是向量，其中初始相关性值等于所述向量的一个或多个分量的值，或者是通过将单调递增函数应用于所述向量的一个或多个分量的值而导出的。

20.根据权利要求1至19中任一项所述的装置，其中，所述装置被配置为执行所述反向传播，使得$0.95\cdot R\leq f(\Sigma R_i)\leq 1.05\cdot R$，其中$\Sigma R_i$表示项目(42)中的集合(16)的所有项目i的相关性分数的和，并且f是仅取决于ΣR_i的单调函数。

21.根据权利要求20所述的装置，其中所述装置被配置为使得f是恒等函数。

22.根据权利要求1至21中任一项所述的装置，其中，所述装置被配置为使得对于每个神经元，通过所述分布函数分布到所述相应神经元的上游邻居神经元集合的相关性消息值的和等于$\xi(S_N)$，或者从其偏离不超过5％，其中S_N表示从所述相应神经元的下游邻居神经元集合到所述相应神经元的相关性消息的和，而ξ表示仅取决于S_N的单调函数。

23.根据权利要求22所述的装置，其中所述装置被配置为使得ξ是恒等函数。

24.根据权利要求1至23中任一项所述的装置，其中，所述人工神经网络被分层，使得每个神经元(12)属于层序列中的一层，并且所述装置被配置为通过对人工神经网络的所有神经元均等地使用一个分布函数来执行反向传播。

25.根据权利要求1至23中任一项所述的装置，其中所述人工神经网络被分层，使得每个神经元(12)属于层序列中的一层，并且所述装置被配置为执行反向传播，使得对于每层，分布到所述相应层的神经元的相关性消息值的和等于$\zeta(S_L)$或者从其偏离不超过5％，其中S_L表示位于所述相应层下游的层的神经元的初步重新分布的相关性分数的和，而ζ表示仅取决于S_L的单调函数。

26.根据前述权利要求中任一项所述的装置，其中，项目的集合(16)是以下各项或其组合：

图片，其中项目(42)的集合(16)中的每个项目(42)对应于所述图片的像素或子像素中的一个或多个，和/或

视频，其中项目(42)的集合(16)中的每个项目(42)对应于所述视频的图片的一个或多个像素或子像素、所述视频的图片或所述视频的图片序列，和/或

音频信号，其中项目(42)的集合(16)中的每个项目(42)对应于所述音频信号的一个或多个音频样本，和/或

局部特征的特征图或从图片、视频或音频信号局部或全局提取的变换，其中项目(42)的集合(16)中的所述项目(42)对应于局部特征，和/或

文本，其中项目(42)的集合(16)中的所述项目(42)对应于文本的词、句子或段落，和/或

诸如社交网络关系图的图形，其中项目(42)的集合(16)中的所述项目(42)对应于节点或边缘或节点集合或边缘集合或子图形。

27.一种用于数据处理的系统(100)，包括：

根据前述权利要求中任一项所述的用于将相关性分数指派给项目的集合的装置(50)，以及

用于处理项目的集合(16)或从项目集合导出的待处理的数据(106)的装置(102)，其中，根据相关性分数适配所述处理。

28.根据权利要求27所述的系统，其中所述处理是有损处理，并且用于处理的装置被配置为，与指派有较低相关性分数的项目相比，降低对指派有较高相关性分数的项目的有损处理的损失。

29.根据权利要求27所述的系统，其中所述处理是可视化，其中用于适配的所述装置被配置为根据所述相关性分数在所述可视化中执行突出。

30.根据权利要求27所述的系统，其中所述处理是通过从存储器读取或执行进一步测

量进行的数据补充，其中用于处理的装置(102)被配置为根据相关性分数关注数据补充。

31.一种用于突出感兴趣的区域的系统(110)，包括：

根据权利要求1至25中任一项所述的用于将相关性分数指派给项目的集合的装置(50)，以及

用于根据相关性分数生成相关性图形(114)的装置(112)。

32.一种用于优化神经网络的系统(120)，包括：

根据权利要求1至26中任一项所述的用于将相关性分数指派给项目的集合的装置(50)；

用于将用于指派的装置应用到项目的多个不同集合上的装置(122)；以及

用于在将用于指派的装置应用到项目的多个不同集合上期间，通过累积指派给网络的神经元的相关性来检测神经网络中相关性增加(128)的部分并且用于根据相关性增加的所述部分优化人工神经网络的装置(124)。

33.一种用于将相关性分数指派给项目的集合的方法，所述相关性分数指示关于将由神经元(12)组成的人工神经网络(10)应用到项目(42)的集合(16)上以将项目(42)的集合(16)映射到网络输出(18)上的相关性，所述装置被配置为：

通过将从网络输出(18)导出的初始相关性分数(R)反向传播通过人工神经网络(10)来将所述初始相关性分数(R)重新分布到项目(42)的集合(16)上，以获得每个项目的相关性分数，

其中，通过以下方式执行反向传播：使得针对每个神经元，使用分布函数将相应神经元的下游邻居神经元集合的初步重新分布的相关性分数分布到所述相应神经元的上游邻居神经元集合。

34.一种具有程序代码的计算机程序，用于当在计算机上运行时执行根据权利要求33所述的方法。

针对人工神经网络的相关性分数指派

技术领域

[0001] 本申请涉及人工神经网络的相关性分数指派。可以使用这种相关性分数指派，例如用于感兴趣区域(ROI)识别。

背景技术

[0002] 计算机程序能够成功地解决许多复杂的任务，例如图像和文本的自动分类或评估人的信誉。机器学习算法尤其成功，因为它们从数据中学习，即程序获得大的经过标记(或经过弱标记)的训练集，并且在某个训练阶段之后，它能够推广到新的未见过的示例。许多银行有对申请贷款的人的信誉进行分类(例如，基于年龄、地址、收入等)的系统。这种系统的主要缺点是可解释性，即系统通常不提供关于为什么以及如何作出决定的信息(例如为什么有人被归类为没有信誉)。确定分类决定的知识和关系是“隐含的”。

[0003] 理解和解释分类决定在许多应用中具有高价值，因为它使得能够验证系统的推理并向人类专家(例如，银行家、风险投资者或医生)提供额外的信息。在大多数情况下，机器学习方法的缺点是充当一个黑箱，而不提供关于什么使它们得出某一特定决定的任何信息。一般来说，复杂算法的性能比简单(线性)方法好得多(当有足够的训练数据可用时)，但是它们尤其缺乏可解释性。最近，一种类型的分类器，即神经网络，变得非常受欢迎，并产生了出色的结果。这类方法由一系列非线性映射组成，特别难以解释。

[0004] 在典型的图像分类任务中，例如，可以给出图像(例如，鲨鱼的图像)。参见图15。机器学习(ML)算法900将图像902归类为属于某个类904(例如“鲨鱼的图像”)。请注意，事先定义了类集合906(例如鲨鱼、人、夜生活、户外)。算法900是个黑箱，因为它不会告诉用户为什么它作出关于图像属于“鲨鱼的图像”类的决定。关于像素级别的这种分类决定的解释将是有趣的，例如，看到图像被归类为属于“鲨鱼图像”类，主要是因为鲨鱼鳍。这样的“相关性图”在908中示出。

[0005] 图像的分类已经成为许多计算机视觉应用(例如图像搜索[15]、机器人[10]、医学成像[50]、雷达图像中的对象检测[17]或面部检测[49])中的关键因素。神经网络[6]被广泛应用于这些任务，并且是关于诸如ImageNet[11]等的图像分类和排名的竞争中采用最多的方案。然而，像机器学习中的许多方法一样，这些模型往往缺乏对分类器预测的直接可解释性。换句话说，分类器充当黑箱，不提供关于为什么得出某种分类决定的详细信息。也就是说，图15中不具有解释可能性。

[0006] 可解释性的缺乏是由于处理原始图像像素到其特征表示以及从特征表示到最终分类器功能的各种映射的非线性。这在分类应用中是相当大的缺陷，因为它阻碍了人类专家仔细验证分类决定。一些简单的是或否答案在应用中有时价值有限，这些应用中，类似某事在哪里发生或它是如何结构化的等问题与仅仅存在或不存在某种结构的二进制或实值一维评估相比更为相关。

[0007] 一些工作已经致力于这一解释神经网络的课题。[54]致力于分析神经元上的分类器决定，这也适用于像素级别。它从卷积网络架构的输出层向输入像素执行层倒转[23]。这

项工作是特定于带修正线性激活函数的具有神经元层的卷积神经网络的架构。参见[42]，其将[54]中的工作的解释确立为对关于输入图像中的像素的偏导数的近似。在高层意义上，[54]中的工作使用解决优化问题的[55]中其自身的前导工作中的方法，以便重建图像输入，如何将响应向下投射到输入，[54]使用修正线性单元将信息从展开的地图投影到输入，目的是确保特征图非负。

[0008] 在[42]中给出了在输入点x的偏导数和围绕不同点x_0的完整泰勒级数之间的另一种方法。这项工作使用与输入点x不同的点x_0来计算导数和余数偏置，这两者都未进一步规定，而是避免不明确的理由使用泰勒级数的全线性加权项$x-x_0$。在特定领域(例如生态建模)中也研究了使用神经网络模型来量化输入变量的重要性，其中[16,34]调查了大量可能的分析，包括计算偏导数、扰动分析、权重分析和研究在训练时间包括和删除变量的影响。理解神经网络中的决定的不同途径是将更可解释的模型(例如决定树)适配到神经网络学习的功能[41]，并提取由该新模型学习的规则。

[0009] 然而，仍然需要一个强大的，易于实现和广泛应用的构思来实现人工神经网络的相关性分数指派任务。

发明内容

[0010] 因此，本发明的目的是提供一种用于将相关性分数指派给应用人工神经网络的项目集合的构思，该构思适用于更广泛的人工神经网络集合和/或降低计算工作。

[0011] 该目的是通过待审的独立权利要求的主题来实现的。

[0012] 本申请的基本发现是，可以通过以下操作来获得对应用人工神经网络的所述项目集合的相关性分数指派的任务：借助于人工神经网络反向传播初始相关性分数，将从网络输出导出的初始相关性值重新分布到项目的集合上，以获得针对每个项目的相关性分数。具体地，这种反向传播适用于更广泛的人工神经网络集合和/或需要较低的计算工作，通过以下方式执行反向传播来实现：使得对于每个神经元，根据分布函数将相应神经元的下游邻居神经元集合的初步重新分布的相关性分数分布在相应神经元的上游邻居神经元集合上。

附图说明

[0013] 根据各种实施例的本发明的优选实现方式和应用是从属权利要求的主题，下文将参照附图更详细地描述本申请的优选实施方式，其中

[0014] 图1a示出了使用人工神经网络的预测示例的示意图，在该人工神经网络上可以应用根据本发明的实施例的使用反向传播的相关性分数指派。

[0015] 图2a示出了说明根据本申请的实施例示例性地使用图1的人工神经网络作为基础来使用的反向传播过程的示意图。

[0016] 图1b和2b示出了图1a和2a的修改，根据该修改，网络和相关性指派在特征图而不是图像的像素上操作。

[0017] 图1c和2c示出了将图1a和2a应用于彩色图像的可能性。

[0018] 图1d和2d示出了图1a和2a的修改，根据该修改，网络和相关性指派在文本而不是图像上操作。

[0019] 图3示意性地示出了人工神经网络的中间神经元及其与上游和下游邻居神经元的连接，其中还示出了示例性的三个上游邻居神经元。

[0020] 图4示出了根据实施例的用于将相关性分数指派给项目集合的装置的框图。

[0021] 图5示出了在预测时间期间的神经网络形分类器；w_{ij}是连接权重，a_i是神经元i的激活。

[0022] 图6示出了在示出按层相关性计算时间期间图5的神经网络形分类器。$R_i^{(l)}$是要计算的神经元i的相关性。为了便于$R_i^{(l)}$的计算，引入消息$R_{i\leftarrow j}^{(l,l+1)}$。$R_{i\leftarrow j}^{(l,l+1)}$是需要计算使得等式(2)中的按层相关性守恒的消息。所述消息通过用于分类的连接从神经元i发送到其输入神经元j，例如，2是神经元4、5、6的输入神经元。神经元3是5、6的输入神经元。神经元4、5、6是神经元7的输入。

[0023] 图7示出了用于分类的示例性实值预测函数，其以虚黑线作为决定边界，该决定边界将-0.8区域的蓝点与0.6-0.9区域的绿点分开。前面的点被负面标记，后面的点被正面标记。在左侧，描绘了在预测点处的分类函数的局部梯度，并且在右侧示出了相对于决定边界上的根点的泰勒近似。

[0024] 图8示出了用描述神经元和权重连接的不同变量和索引来注释的多层神经网络的示例。左：正向通道；右：后向通道。

[0025] 图9示出了经过训练以从ImageNet数据集合中区分1000个类的神经网络的按像素分解。

[0026] 图10示出了一个实验，根据该实验，本申请实施例的构思被应用于包含0至9的数字图像的MNIST(混合国家标准与技术研究所)数据集合，从而示例性地示出了在右侧示例性示出围绕数字“3”和“4”的部分的热图，其具有高相关性，以便分别将这些数字识别为“3”并将相应数字与“9”区分开。

[0027] 图11示出了根据实施例的用于数据处理的系统的框图。

[0028] 图12示出了根据与图11不同的实施例的用于数据处理的系统的框图，所述不同之处在于对已从中导出该项目集合的数据执行处理；

[0029] 图13示出了根据实施例的ROI突出(highlight)系统的框图。

[0030] 图14示出了根据实施例的神经网络优化系统。

[0031] 图15示出了说明关于人工神经网络的相关性分数指派的任务以及与人工神经网络的通常的预测任务的关系的示意图。

具体实施方式

[0032] 在关于框图描述本申请的各个实施例之前，首先通过简要介绍人工神经网络并且然后解释实施例的构思底层的思想来描述这些实施例底层的构思。

[0033] 神经网络是互连的非线性处理单元的图形，其可以通过训练来逼近输入数据和输出数据之间的复杂映射。注意，输入数据例如是图像(像素集)，输出例如是分类决定(在最简单的情况下，+1/-1意味着“是的”图像中有鲨鱼或“不”图像中没有鲨鱼)。每个非线性处理单元(或神经元)由其应用了非线性激活函数的输入的加权线性组合组成。使用索引i来表示正进入具有索引j的神经元的各神经元，非线性激活函数定义为：

[0034] $$x_j = g\left(\sum_i x_i u_{ij} + b_j\right)$$

[0035] 其中g(•)是非线性单调增加的激活函数，w_{ij}是将神经元i连接到神经元j的权重，而b_j是偏置项。神经网络由其连接性结构、其非线性激活函数及其权重来定义。

[0036] 以下实施例使用一种可以是并且在随后的描述中称为相关性传播的构思。它将由输出神经元建模的数据中的特定结构的证据重新分布给输入神经元。因此，它试图根据输入变量(例如像素)产生其自己的预测的解释。注意，该构思对于每种类型的(无环路)神经网络都是有效的，与层数、激活函数的类型等无关。因此，它可以应用于许多流行模型，因为可以根据神经网络对许多算法进行描述。

[0037] 下面给出针对由卷积/次采样层后接完全连接层序列组成的网络的相关性传播过程的图示。

[0038] 具体地，图1a以简化的示例性方式示出了人工神经网络的示例。人工神经网络10由图1中被描绘为圆圈的神经元12组成。神经元12彼此互连或彼此交互。通常，每个神经元一方面连接到下游邻居(或后继(successor))神经元，另一方面连接到上游邻居(或前导(predecessor))神经元。术语“上游”、“前导”、“下游”和“后继”是指一般的传播方向，当将该一般的传播方向应用于项目集合16时，神经网络10沿着该传播方向14进行操作，以便将该项目集合16映射到网络输出18，即执行预测。

[0039] 如图1a所示，该项目集合16可以例如是通过如下方式形成图像的像素22集合：将每个像素与对应于场景颜色的像素值、或对应于相应像素在图像22的像素阵列中的位置的空间位置处的强度相关联。在这种情况下，集合16是项目的有序集合，即像素阵列。在这种情况下，项目将对应于各个像素值，即每个项目将对应于一个像素。之后，将会明确指出，本申请不限于图片领域。相反，项目集合16可以是在项目之间没有定义任何顺序的项目集合。它们之间的混合也可以是真实的。

[0040] 神经元12的第一或最低层24形成人工神经网络10的一种输入。也就是说，该最低层24的每个神经元12接收项目集合16的至少一个子集(即像素值的至少一个子集)作为其输入值。集合16中的项目子集(其值被输入到最低层24的某个神经元12中)的并集等于例如集合16，即在图1a的情况下等于整个图像22。换句话说，对于集合16的每个项目，其值被输入到最低层24的至少一个神经元12中。

[0041] 在神经网络10的相对侧，即在其下游/输出侧，网络10包括一个或多个输出神经元12’，其与神经元12的不同之处在于前者缺少下游邻居/后继神经元。在被应用到集合16之后以及在完成处理之后，存储在每个输出神经元12’中的值形成网络输出18。也就是说，网络输出可以例如是标量。在这种情况下，只有一个输出神经元12’将出现，并且其在网络10操作之后的值将形成网络输出。如图1所示，这样的网络输出可以例如是对于项目集合16(即图1a情况下的图像22)是否属于某个类的可能性的度量。然而，备选地，网络输出18可以是向量。在这种情况下，存在多于一个输出神经元12’，并且在网络10操作结束时获得的这些输出神经元12’中每一个的值形成网络输出向量的相应分量。例如，图1示出了网络输出18的每个分量是度量集合16属于与各个分量相关联的相应类(例如图像类“显示船”、“显示卡车、“显示轿车”)的可能性的度量。其他示例也是可以想象的，并在下文中给出。

[0042] 因此，总结上述内容，神经网络包括互连的神经元12，以便在正向传播或正常操作

中将项目集合16映射到神经输出。以类似于输出神经元12'的方式，其在网络操作结束时的值形成网络输出18，在图1a的示例性情况下，集合16的项目(即图像22的像素)可以被视为具有神经元12的网络10的输入神经元，并且由此形成的层分别是中间神经元或中间层。具体地，输入神经元因此可以被认为是中间神经元12的上游邻居或前导神经元，即层24的那些，正如输出神经元12'可以形成中间神经元12的下游邻居/后继神经元，形成例如网络10的最高中间层，或者在将一个或多个输出神经元12'解释为形成网络10的最高层的情况下形成网络10的次最高层。

[0043] 图1示出了神经网络10的简化示例，根据该示例，网络10的神经元12被严格排列在层26中，其中，层26形成层序列，特定神经元12的上游邻居/后继神经元全部是相应神经元12所属的层的紧邻较低层的成员，而所有下游邻居/后继神经元则全部是紧邻较高层的成员。然而，图1不应被解释为限制可以针对该问题应用下面进一步概述的本发明实施例的神经网络10的种类。相反，可以根据替代实施例来修改神经元12的这种严格分层布置，例如，上游邻居/前导神经元是多于一个先前层的神经元的子集，和/或下游邻居/后继神经元是多于一个更高层的神经元的子集。

[0044] 此外，尽管图1表明，在网络10的正向传播操作期间，每个神经元12将只被遍历一次，但是一个或多个神经元12可以被遍历两次或更多次。下面将讨论进一步的变化的可能性。

[0045] 如上所述，当将网络10应用于集合16(即在图1a的示例性情况下的图像22)时，网络10执行正向传播操作。在此操作期间，已经从其上游邻居/前导神经元接收其所有输入值的每个神经元12通过相应的神经元函数计算被称为其激活的输出值。在上述示例性等式中，这种被称为x_j的激活随后形成每个下游邻居/后继神经元的输入值。通过该度量，集合16的项目的值通过神经元12传播，以最终变为输出神经元12'。更精确地，集合16的项目的值形成网络10的最低层的神经元12的输入值，并且输出神经元12'接收其上游邻居/前导神经元12的激活作为输入值并通过各自的神经元函数计算它们的输出值，即网络输出18。与网络10的神经元12和12'相关联的神经元函数在所有神经元12和12'中可以是相等的，或者在其间可以是不同的，其中，“相等性”的意思是神经元函数是可参数化的，并且函数参数可以针对各个神经元不同，而不破坏所述相等性。在变化/不同的神经元函数的情况下，这些函数在网络10的相同层的神经元中可能相等，或者甚至可以在一层内的神经元之间不同。

[0046] 因此，网络10可以例如以在计算机上运行的计算机程序的形式即软件来实现，但是以诸如电路形式的硬连线形式的实现也是可行的。如上所述，每个神经元12使用神经元函数基于其输入值计算激活，所述神经元函数例如在上述显式示例中呈现为输入值的线性组合的非线性标量函数$g(\cdot)$。如上所述，与神经元12和12'相关联的神经元函数可以是参数化的函数。例如，在以下概述的具体示例之一中，可使用相应神经元的所有输入值i的偏置量b_j和权重w_{ij}对神经元j的神经元函数进行参数化。这些参数在图1a中用虚线框28示出。这些参数28可以通过训练网络10获得。为此，例如，网络10被重复地应用到正确的网络输出已知的项目集合16的训练集合，即在图1a的示例情况下的标记图像的训练集合。然而，其他可能性也可能存在。即使组合也可能是可行的。下面进一步描述的实施例不限于参数28的任何种类的起源或确定方式。图1a示出了例如网络10的上游部分21(其由从集合16(即，网络的输入)延伸到中间隐藏层的层26组成)已被人为地生成或学习，以便例如通过卷积滤波

器模拟对图像22的特征提取，使得(下游)拖尾层的每个神经元表示特征图20的特征值。每个特征图20例如与特定的特性或特征或脉冲响应等相关联。因此，例如，每个特征图20可以被认为是输入图像22的稀疏采样的滤波版本，其中，特征图20在相关联滤波器的相关联特征/特性/脉冲响应方面与另一特征图不同。如果例如集合16具有X·Y个项目，即像素，即X列和Y行像素，则每个神经元将对应于一个特征图20的一个特征值，该值将对应于与图像22的某一部分相关联的本地特征分数。例如，在具有P·Q特征分数样本的N个特征图的情况下，即P列和Q行特征值，部分21的下游拖尾层处的神经元的数量将是例如N·P·Q，其可以小于或大于X·Y。特征图20底层的特征描述符或滤波器的变换可以分别用于设置部分21内的神经元的神经元函数，或参数化神经元函数。然而，再次注意到，对于本申请及其实施例而言，网络的这种“变换的”而不是“学习的”部分21的存在不是强制性的，并且备选地，这种部分可以不存在。在任何情况下，在说明神经元12的神经元函数可以在所有神经元中相等或在一层神经元中相等等情况是可行的时，神经函数却可以是可参数化的，并且，尽管可参数化的神经函数可以在这些神经元中相等，但是该神经函数的函数参数也可在这些神经元之间变化。中间层的数量同样是自由的，并且可以等于1或大于1。

[0047] 综上所述，网络10在正常操作模式下的应用如下：输入图像22在其作为集合16的角色中受制于或耦合到网络10。也就是说，图像22的像素值形成第一层24的神经元12的输入值。如上所述，这些值沿着正向方向14通过网络10传播并导致网络输出18。在图1所示的输入图像22的情况下，例如，网络输出18例如将表示该输入图像22属于第三类，即显示轿车的图像类。更准确地说，虽然对应于“轿车”类的输出神经元结束于高值，但在此，示例性地对应于“卡车”和“船”的其他输出神经元将结束于(较)低值。

[0048] 然而，如本申请的说明书的介绍部分所述，关于图像22(即，集合16)是否显示轿车等的信息可能是不够的。相反，优选的是获得在像素的粒度级别的信息，所述信息指示对于网络10的决定，哪些像素(即集合16的哪些项目)是相关的，以及哪些像素不是相关的，例如，哪些像素显示轿车，哪些不显示轿车。该任务通过下面描述的实施例来处理。

[0049] 具体地，图2a以示例性方式示出了下面更详细描述的本发明的实施例如何操作、以便完成对集合16的项目的相关性分数指派的任务，其在图2a的示例性情况下是像素域。具体地，图2a示出了通过反向传播过程(后向或相关性传播)执行该相关性分数指派，根据该反向传播过程，相关性值R例如通过网络10向网络输入(即项目集合16)反向传播，从而针对图像的每个像素获得集合16的每个项目i的相关性分数R_i。对于包括例如X·Y个像素的图像，i可能在{1...X·Y}内，其中每个项目/像素i例如可以对应于像素位置(x_i,y_i)。在沿着与图1的正向传播方向14相反的反向传播方向32执行这种反向传播时，下文描述的实施例遵循某些约束，现在将更详细地解释这些约束，并且这些约束被称为相关性保护和相关性重新分布。

[0050] 简而言之，相关性分数指派开始于完成将人工神经网络10应用于集合16。如上所述，这一应用结束于网络输出18。从该网络输出18导出初始相关性值R。在下面描述的示例中，例如，将一个输出神经元12’的输出值用作该相关性值R。然而，使用例如应用于网络输出的单调函数，也可以不同地执行来自网络输出的导出。其他示例开始如下。

[0051] 在任何情况下，这个相关性值然后通过网络10在相反方向(即32)上传播，与正向传播方向14相比指向相反的方向，网络10沿该方向在被应用到集合16上时运行，从而导致

网络输出18。以如下方式进行反向传播，使得对于每个神经元12，将相应神经元的下游邻居神经元集合的初步重新分布的相关性值的总和分布在相应神经元的上游邻居神经元集合上，使得相关性“基本上守恒”。例如，可以选择分布函数，使得在完成反向传播之后，初始相关性值R等于集合16的项目i的相关性分数R_i的总和，确切地说即$R=\Sigma R_i$，或者通过单调函数f()，即$R=f(\Sigma R_i)$。在下文中，讨论有关分布函数的一些一般想法以及如何有利地选择它们。

[0052] 在反向传播期间，神经元12的神经元激活用于引导反向传播。也就是说，在将网络10应用于集合16以获得网络输出18期间，人工神经网络10的神经元激活被预先存储并重新使用，以便引导反向传播过程。如下面将更详细地描述的，可以使用泰勒近似来近似反向传播。因此，如图2a所示，反向传播的过程可以被认为是将初始相关性值R从输出神经元开始沿着反向传播方向32向网络10的输入侧分布。通过这种措施，相关性增加的相关性流动路径34从输出神经元36去向网络10的输入侧，即由项目集合16本身形成的输入神经元。图2示例性地示出了在通过网络10的过程中间歇地分支的路径。这些路径最终在项目集合16内相关性增加的热点中结束。在使用输入图像22的具体示例中，如图2a所示，相关性分数R_i表示在像素级别图像22内的相关性增加的区域，即图像22内在网络10中起主要作用、最终变为对应的网络输出18的区域。在下文中，使用上述非线性激活函数的示例作为网络10的神经元的神经元函数来更详细地讨论刚刚提及的相关性守恒和相关性重新分布属性。

[0053] 属性1：相关性守恒

[0054] 相关性传播模式的第一个基本属性强调，证据不能被创造也不能丢失。这适用于全局尺度(即从神经网络输出回到神经网络输入)和局部尺度(即在各个非线性处理单元的级别)。这种限制相当于将基尔霍夫电路定律应用于神经网络，并通过“语义证据”的概念取代“电流”的物理概念。具体参见图3。

[0055] 使用索引i和k来表示进入和去往具有索引j的神经元的神经元(进入的神经元在图3中示出为40，因此形成前导或上游邻居)，等式

[0056] $$\sum_i R_{ij} = \sum_k R_{jk}$$

[0057] 必须保持成立，其中R_{ij}表示从神经元j流向神经元i的相关性，而R_{jk}表示从神经元k流向神经元j的相关性。请注意，相关性守恒原理指出，“流入神经元”的相关性的总和必须与从该神经元流出的相关性的总和相同。相关性守恒确保输入神经元相关性(例如像素的相关性)的总和等于网络的输出值(例如，分类分数)。

[0058] 属性2：相关性重新分布

[0059] 相关性传播模型的第二个基本原理是，相关性的局部重新分布必须遵循一个固定的规则，该规则总是适用于网络中的所有神经元。可以为相关性重新分布定义许多不同的规则。一些规则是“有意义的”，其他的则不是。一个这样的有意义的规则是，例如，

[0060] $$R_{ij} = \frac{x_i w_{ij} + \frac{b_j}{n}}{\sum_{i'}\left(x_{i'} w_{i'j} + \frac{b_j}{n}\right)} \sum_k R_{jk}$$

[0061] 其中n是由i索引的神经元数。这种重新分布规则的解释在于：对神经元x_j的激活最有贡献的神经元x_i将被认为是大多数的进入相关性$\Sigma_k R_{jk}$。此外，通过将所有进入神经元

i上重新分布的相关性R_{ij}相加，应该清楚的是，满足属性1。

[0062] 然而，上面的确定性相关性传播规则有两个缺点：首先，当分母接近零时，它可能是数值不稳定的。第二，它可以给出R_{ij}的负值，它们具有未定义的含义。第一个问题通过重新定义如下规则来解决：

[0063] $$R_{ij}=\frac{x_i w_{ij}+\frac{b_j}{n}}{h\left(\sum_{i'}\left(x_{i'}w_{i'j}+\frac{b_j}{n}\right)\right)}\sum_k R_{jk}$$

[0064] 其中$h(t)=t+\varepsilon\,\mathrm{sign}(t)$是防止分母接近零的数值稳定器，其中ε被选择为非常小以符合属性1。第二个问题通过只考虑对神经元激活的积极贡献来解决，具体地，

[0065] $$R_{ij}=\frac{\max\left(0,x_i w_{ij}+\frac{b_j}{n}\right)}{\sum_{i'}\max\left(0,x_{i'}w_{i'j}+\frac{b_j}{n}\right)}\sum_k R_{jk}$$

[0066] 这里，注意到两个正数的比率必然是正的，因此相关性也是如此。这两个增强可以容易地组合以满足稳定性和正属性。

[0067] 请注意，相关性守恒说明了重新传播的作用(=将输出相关性分布给输入变量，同时保持总值(总和)不变)，而相关性重新分布说明了如何实现这一点(=“有意义的”重新分布应该确保对激活贡献最大的神经元(具有大的加权激活$x_i w_{ij}$)将被认为是大多数的进入相关性)

[0068] 在描述根据本申请的实施例的装置之前，应该扩展上述介绍，以便更清楚地呈现可能的替代方案。

[0069] 例如，虽然关于图1a和2a描述的实施例使用图像22作为项目集合16，并且可能以如下方式设计网络10：使得其一层的神经元的神经元激活代表图像的“局部特征”，即特征图20的样本，不过，图1b和2b的实施例使用特征图20作为项目集合16。也就是说，向网络10馈送特征图20的特征样本。可以通过使输入图像22经历特征提取器来从输入图像22获得特征图20，每个特征提取器从输入图像22提取相应的特征图20。使用箭头30在图1b中示出了该特征提取操作。例如，特征提取器可以将滤波器内核局部地应用到图像22上，以便通过使滤波器核心在图像上移动来每次应用导出一个特征样本，以便获得由例如以行和列排列的特征样本组成的对应特征图20。过滤器内核/模板针对相应的特征提取器和对应的特征图20可以分别是单独的。这里，图1b的网络10可以与图1a的网络10的剩余部分重合，即网络10在删除部分21之后的剩余。因此，在图1b的情况下，作为所谓的预测处理的一部分，特征样本值通过网络10沿着正向方向14传播，并且导致网络输出18。图2b示出了图1b的网络的相关性反向传播过程：反向传播过程通过网络10将相关性值R反向传播到网络输入，即项目集合16，从而获得每个项目的相关性分数R_i。因此，在图2b的情况下，每个特征样本获得一个相关性分数R_i。然而，由于特征图20通过特征图单独滤波器提取函数与图像内容相关，因此，即通过以固定的方式将集合16的项目的各个相关性分数分布到图像22的各个像素位置上，可以将每个相关性分数i变换到像素域中，即像素上。“固定方式”唯一地取决于与各个相关性分数的特征图相关联的特征提取器，并且表示特征提取30的一种反向函数38。因此，该反向函数38形成反向传播处理的一种扩展，以便缩小从像素的特征集合域到空间域的差距。

[0070] 此外，应注意，在图1a和图2a的情况下，已经预先假定图像22的每个像素(即16的每个项目)都携带标量。例如，这种解释可以适用于灰度图像22的情况，其中每个像素值对应于一个灰度值。然而，也存在其他可能性。例如，图像22可以是彩色图像。在这种情况下，集合16的每个项目可以对应于图像22的多个颜色平面或颜色分量之一的样本或像素值。在图1c和2c中示例性地示出了三个分量，其示出了图1a和2a向彩色图像22的延伸。因此，在对于X·Y个像素位置中的每个位置都针对三个颜色分量中每一个的颜色分量具有颜色分量值的情况下，图1c和2c中的项目集合16的数量为X·Y·3。然而，颜色分量的数量可能不仅仅是三个。此外，颜色分量的空间分辨率不必相同。图2c的反向传播结束于每个项目的相关性值，即颜色分量样本。在具有针对每个像素的所有分量的分量值的情况下，可以通过将针对各个像素的颜色分量获得的相关性值相加来获得最终相关性图。这在37中说明。

[0071] 尽管图1至图2c与图像和像素有关，但是本申请的实施例不限于这种数据。例如，文本及其词语可以用作基础。社交图分析应用可以如下所示：将相关性指派给图中的节点和连接，其中该图作为神经网络10的输入给出。在社交图分析的上下文中，节点可以表示用户，并且连接可以表示这些用户之间的关系。这种连接还可以被引导到组织内的模型信息流(例如引用网络)或责任链。例如，神经网络可以被训练来针对作为输入给出的图来预测该图的具体属性(例如与具体社交图相关联的生产率)

[0072] 在这种情况下，相关性传播和热图方法将设法在该图中识别用于解释预测属性(即，高生产率或低生产率)的子结构或节点。也可以对神经网络进行训练，以便在稍后时间点预测图的状态。在这种情况下，相关性传播过程将设法确定图中哪个子结构解释图的未来状态(例如，社交图中哪些子结构或节点在传播图中的信息或改变其状态的能力方面最有影响力)。因此，神经网络可以例如用于预测广告活动(回归任务)的成功(例如已售产品的数量)。相关性分数可用于确定对于成功有影响力的一些方面。仅仅通过专注于这些相关方面，公司就可以节省资金。相关性分数指派过程可以为广告活动的每个项目指派一个分数。然后，决定处理器可以接收该输入以及关于广告活动的每个项目的成本的信息，并且决定该活动的最佳策略。然而，相关性也可以用于如上所示的特征选择。

[0073] 相关性分数指派开始于导出初始相关性值R。如上所述，可以基于神经网络的输出神经元之一设置相同的值，以便通过反向传播获得该集合16的项目的相关性值，其指代该一个输出神经元的“含义”。然而，备选地，网络输出18可以是向量，并且输出神经元可以具有这样的含义：相同的可以被划分为重叠或非重叠子集。例如，对应于“卡车”和“轿车”含义(类别)的输出神经元可以组合以产生“汽车”含义的输出神经元的子集。因此，两个输出神经元的输出值可以用作反向传播中的起始点，由此导致项目16的相关性分数，即指示与该子集的含义的相关性的像素，即“汽车”。

[0074] 虽然上述内容表明该项目集合是与图片的一个像素对应的项目42的集合16的每个项目42的图片，但是这可以是不同的。例如，每个项目可以对应于像素集合或子像素(像素通常具有rgb值。子像素例如是像素的绿色分量)，如图2c所示的超像素。此外，备选地，项目集合16可以是视频，其中项目42的集合16的每个项目42对应于视频的图片的一个或多个像素、视频的图片或视频的图片序列。项目所指向的像素子集可以包含不同时间戳的图片的像素。此外，项目集合16可以是音频信号，其中项目42的集合16的每个项目42对应于诸如PCM样本的音频信号的一个或多个音频样本。集合16的各个项目可以是所述样本或音频记

录的任何其他部分。或者，该项目集合是频率和时间的乘积空间，并且每个项目是一个或多个频率时间间隔的集合，例如由例如重叠窗口序列的MDCT频谱组成的频谱图。此外，集合16可以是从图片、视频或音频信号局部提取的局部特征的特征图，其中项目42的集合16的项目42对应于所述局部特征，或从文本提取的局部特征的特征图，其中项目42的集合16的项目42对应于该文本中的词、句子或段落。

[0075] 为了完整起见，图1d和2d示出了一个变体，根据该变体，项目的数据集合16是文本而不是图像。对于这种情况，图1d示出了，通过将每个词w_i 43映射到具有相同长度(即具有相同的分量v_{ij} 47的数量J)的相应矢量v_i 45上，根据按词变换49将实际上是(例如I个)词43的序列41的文本变换成“抽象”或“可解释的”版本。每个组件可以与语义有关。可以使用的按词变换是例如Word2Vec或词指示符向量。向量v_i 45的分量v_{ij} 47代表集合16的项目并且受制于网络10，由此导致在网络的输出节点12处的预测结果18。图2d所示的反向传播导致每个项目，即对于每个向量分量v_{ij}(0<i<I;0<j<J)的的相关性值。总结53，对于每个词w_i，例如，与相应词w_i相关联的向量v_i的分量v_{ij}的相关性分数(其中0<j<J)导致每个词一个相关性和值(相关性分数)，因此，文本中的每个词可以按照其相关性分数和进行突出。突出选项的数量可以是两个或更多。也就是说，可以对词的相关性和值进行量化，以产生每个词的突出选项。突出选项可与不同的强度强度相关联，并且从相关性和值到突出选项的映射可导致相关性和值和突出强度之间的单调关联。再次，类似于神经网络涉及对图像的预测性能的示例，图1d和2d的网络10的输入侧部分可以具有一些可解释的含义。在所述图像的情况下，这是功能集合。在图1d和2d的情况下，网络10的输入部分可以表示从由集合16的分量组成的向量到最可能的较低维度向量的另一按向量映射，其中，所述较低维度向量的分量与由集合16的分量组成的向量的相当(rather)词族相关分量相比，可能具有相当的语义。

[0076] 图4示出了用于将相关性分数指派给项目集合的装置的示例。该装置例如以软件实现，即实现为编程的计算机。然而，其他实施可能性也是可以想象的。在任何情况下，装置50被配置为使用上述的反向传播过程，以便逐项目地将相关性分数指派给项目的集合16，其中相关性分数指示：对于每个项目而言，该项目在导出网络输出18所基于的网络10中具有什么样的相关性。因此，图4也示出了神经网络。网络10被示为不是装置50的一部分：相反，网络10定义要通过装置50将分数指派给项目集合16的“相关性”的含义来源。然而，备选地，装置50也可以包括网络10。

[0077] 图4示出了作为接收项目集合16的网络10，其中项目被示意性地表示为小圆圈42。图4还示出了网络10由神经元参数44控制的可能性，例如如上所述基于神经元的上游邻居/前导神经元来控制神经元激活计算的函数权重，即神经元函数的参数。这些参数44可以例如被存储在存储器或储存器46中。图4还示出了在使用参数44完成处理项目42的集合16之后网络10的输出，即网络输出18以及可选地由处理集合16产生的神经元12的神经元激活，该神经元激活由参考标记48表示。示例性地示出了神经元激活48、网络输出18和参数44将存储在存储器46中，但是它们也可以存储在单独的储存器或存储器中，或者可以不被存储。装置50可以访问网络输出18，并使用网络输出18和上述反向传播原理执行重新分布任务52，以获得集合16的每个项目i 52的相关性分数R_i。具体地，如上所述，装置50从网络输出中导出初始相关性值R，并且使用反向传播过程重新分布该相关性R，以便结束于针对项目i的单独的相关性分数R_i。图4中通过用附图标记42表示的小圆圈示出了集合16的各个项目。

如上所述，重新分布52可以由参数44和神经元激活48引导，因此，装置50也可以访问这些数据项。此外，如图4所示，实际神经网络10不需要在装置50内实现。相反，装置50可以访问(即了解)网络10的构造，例如神经元的数量、参数44所属的神经元函数和神经元互连，这些信息在图4中使用术语神经网络描述54示出，该术语神经网络描述54如图4所示也可以存储在存储器或储存器46中或其他地方。在替代实施例中，人工神经网络10也在装置50上实现，使得装置50可以除了包括执行重新分布任务52的重新分布处理器之外，还包括用于将神经网络10应用于集合16上的神经网络处理器。

[0078] 因此，上述实施例尤其能够缩小在计算机视觉中受欢迎的多层神经网络的分类和可解释性之间的差距。对于神经网络(例如[6,31])，将考虑基于广义p均值的具有任意连续神经元和汇集(pooling)函数的一般多层网络结构。

[0079] 下一节作为一般构思的按像素分解将解释分类器的按像素分解底层的基本途径。该按像素分解示于图1a和2c。多层网络的按像素分解将作为一般构思的按像素分解中解释的基于泰勒和按层相关性传播方法应用于神经网络架构。框架的实验评估将在实验中完成。

[0080] 作为一般构思的按像素分解

[0081] 按像素分解的总体思想是将图像x的单个像素的贡献理解为由分类器f在图像分类任务中进行的预测f(x)。针对每个图像x，分别找出哪些像素在多大程度上对正或负的分类结果有贡献。此外，通过度量来定量地表达这一程度。假设分类器具有阈值为零的实值输出。在这样的设置中，它是使得f(x)>0表示学习结构的存在的映射$f:\mathbb{R}^V \rightarrow \mathbb{R}^1$。两类分类器的概率输出可以通过减去0.5或对预测取对数然后加上2.0的对数来处理而不失一般性。有趣的是找到输入图像x的每个输入像素$x_{(d)}$对具体预测的贡献。特定于分类的重要约束在于找到相对于分类与最大不确定性状态相关的差分贡献，然后由根点的集合$f(x_0)=0$表示。一种可能的方式是将预测f(x)分解为相应像素单独的输入维数x_d的项的总和：

[0082] $$f(x) \approx \sum_{d=1}^{V} R_d \quad (1)$$

[0083] 定性解释是，$R_d<0$针对将要被分类的结构的存在贡献证据，而$R_d>0$为其存在贡献证据。对于随后的可视化，对于每个输入像素$x_{(d)}$，所得到的相关性R_d可以被映射到颜色空间并且以该方式可视化为传统的热图。在下面的工作中，一个基本的约束将是，R_d的符号应遵循上述定性解释，即正值应表示正贡献，负值表示负贡献。

[0084] 在下文中，为了实现如等式(1)中的按像素分解的目的，该构思被表示为按层相关性传播。还讨论了基于泰勒分解的方法，其产生了按层相关性传播的近似。将展示，对于广泛的非线性分类架构，可以进行按层相关性传播，而不需要通过泰勒扩展来使用近似。随后提出的方法不涉及细分。它们不需要像素训练作为训练阶段的学习设置或像素标记。这里使用的设置是图像分类，其中在训练期间针对整个图像提供一个标记，然而，该贡献不是关于分类器训练的。这些方法是建立在预训练分类器之上的。它们适用于已经预训练的图像分类器。

[0085] 按层相关性传播

[0086] 其一般形式的按层相关性传播假设可以将分类器分解成若干层计算。这些层可以是从图像进行特征提取的部分，或是计算特征上运行的分类算法的部分。如后所示，这对于神经网络是可能的。

[0087] 第一层可以是输入，图像的像素，最后一层是分类器f的实值预测输出。将第l层建模为具有维度V(l)的向量$z = (z_{(d,l)})_{d=1}^{V(l)}$。按层相关性传播假设获得了层l+1处的向量z的每个维度$z_{(d,l+1)}$的相关性分数$R_d^{(l+1)}$。这个想法是找到接近输入层的下一层l处向量z的每个维度的相关性分数，使得下面的等式成立。

[0088] $$f(x) = \cdots = \sum_{d\in l+1} R_d^{(l+1)} = \sum_{d\in l} R_d^{(l)} = \cdots = \sum_d R_d^{(1)} \quad (2)$$

[0089] 从作为分类器输出f(x)的最后一层向下到由图像像素构成的输入层x进行等式(2)的迭代，然后产生所需的等式(1)。输入层的相关性将作为等式(1)中的期望的和分解。如将展示的，这种分解本身既不是唯一的，也不能保证它对分类器预测产生有意义的解释。

[0090] 在这里给出一个简单的反例。假设存在一层。输入是$x \in \mathbb{R}^V$。使用线性分类器，其具有一些任意和维度特定的特征空间映射Φ_d以及偏置b

[0091] $f(x) = b + \sum_d \alpha_d \Phi_d(x_d)$ (3)

[0092] 将第二层的相关性简单地定义为$R_1^{(2)} = f(x)$。然后，一个可能的按层相关性传播公式是将输入x的相关性$R^{(1)}$定义为

[0093] $$R_d^{(1)} = \begin{cases} f(x) & \text{如果 } \sum_d |\alpha_d \phi_d(x_d)| \neq 0 \\ \frac{b}{V} & \text{如果 } \sum_d |\alpha_d \phi_d(x_d)| = 0 \end{cases} \quad (4)$$

[0094] 这显然满足等式(1)和(2)，然而所有输入维度的相关性$R^{(1)}(x_d)$具有与预测f(x)相同的符号。在按像素分解解释方面，如果f(x)>0则所有输入指向结构的存在，如果f(x)<0则指向结构的不存在。对于许多分类问题这不是一个现实的解释。

[0095] 讨论一种更有意义的定义按层相关性传播的方式。对于这个示例定义

[0096] $$R_d^{(1)} = \frac{b}{V} + \alpha_d \phi_d(x_d) \quad (5)$$

[0097] 然后，特征维度x_d的相关性取决于等式(5)中的项的符号。这对于许多分类问题来说是更加合理的解释。第二个示例表明按层相关性传播能够在某种程度上应对非线性，例如特征空间映射Φ_d，以及满足公式(2)的按层相关性传播的示例在实践中如何可能看起来可行。请注意，在这里，根本不需要对特征空间映射Φ_d进行规则性假设，它在Lebesgue度量下甚至可以是不连续或不可衡量的。底层公式(2)可以解释为特征处理的层之间的相关性R的守恒定律。

[0098] 上述示例进一步给出了关于什么是相关性R的直观情况，即对预测函数f(x)的局部贡献。在这个意义上，可以选择输出层的相关性作为预测本身f(x)。这第一示例显示了作为线性分解示例的预期结果。线性示例提供了第一个直观情况。

[0099] 给出更多图形和非线性的第二个示例。图5显示了具有神经元以及在神经元之间的连接上的权重w_{ij}的神经网络形分类器。每个神经元i具有来自激活函数的输出a_i。

[0100] 顶层由通过7索引的一个输出神经元组成。对于每个神经元i，计算相关性R_i。对于该示例，将丢弃层索引上标$R^{(1)}$，因为每当层索引明显时，所有神经元都具有显式的神经元索引。初始化顶层相关性$R_7^{(3)}$作为函数值，从而$R_7 = f(x)$。等式(2)中的按层相关性传播现

在需要保持以下等式成立

[0101] $$R_7^{(3)} = R_4^{(2)} + R_5^{(2)} + R_6^{(2)} \quad (6)$$

[0102] $$R_4^{(2)} + R_5^{(2)} + R_6^{(2)} = R_1^{(1)} + R_2^{(1)} + R_3^{(1)} \quad (7)$$

[0103] 将为这个示例做出两个假设。首先，在神经元i和j之间的消息中表达按层相关性，所述消息可以沿着每个连接发送。然而，如图6所示，与预测时间发生的情况相反，这些消息从神经元指向其输入神经元。其次，将除了神经元7之外的任何神经元的相关性定义为进入消息的总和：

[0104] $$R_i^{(l)} = \sum_{k:\ i\text{是针对神经元}k\text{的输入}} R_{i\leftarrow k}^{(l,l+1)} \quad (8)$$

[0105] 例如$R_3^{(1)} = R_{3\leftarrow 5}^{(1,2)} + R_{3\leftarrow 6}^{(1,2)}$。请注意，无论如何，神经元7还没有进入的消息。相反，将其相关性定义为$R_7^{(3)} = f(x)$。在等式(8)和以下文本中，术语输入和源的含义是在分类时间期间所定义的方向上作为对另一个神经元的输入，而不是在按层相关性传播的计算时间期间。例如，在图6中，神经元1和2是神经元4的输入和源，而神经元6是神经元2和3的宿。给定在等式(8)中编码的两个假设，通过等式(2)的按层相关性传播可以由以下充分条件来满足：

[0106] $$R_7^{(3)} = R_{4\leftarrow 7}^{(2,3)} + R_{5\leftarrow 7}^{(2,3)} + R_{6\leftarrow 7}^{(2,3)} \quad (9)$$

[0107] $$R_4^{(2)} = R_{1\leftarrow 4}^{(1,2)} + R_{2\leftarrow 4}^{(1,2)} \quad (10)$$

[0108] $$R_5^{(2)} = R_{1\leftarrow 5}^{(1,2)} + R_{2\leftarrow 5}^{(1,2)} + R_{3\leftarrow 5}^{(1,2)} \quad (11)$$

[0109] $$R_6^{(2)} = R_{2\leftarrow 6}^{(1,2)} + R_{3\leftarrow 6}^{(1,2)} \quad (12)$$

[0110] 一般来说，这个条件可以表示为：

[0111] $$R_k^{(l+1)} = \sum_{i:\ i\text{是针对神经元}k\text{的输入}} R_{i\leftarrow k}^{(l,l+1)} \quad (13)$$

[0112] 条件(13)和定义(8)之间的差异在于，在条件(13)中，所述和在针对层l+1处的固定神经元k的层l处的源上运行，而在定义(8)中，所述和在针对层l处的固定神经元i的层l+1处的宿上运行。这个条件是一个充分条件，而不是一个必要条件。这是定义(8)的结果。可以通过说明消息$R_{i\leftarrow k}^{(l,l+1)}$用于将神经元k的相关性$R_k^{(l+1)}$分布到层l处的输入神经元上来解释充分条件(13)。以下部分将基于这个概念，并且由定义(8)和充分条件(13)给出更严格的相关性守恒形式。

[0113] 现在，可以通过定义消息$R_{i\leftarrow k}^{(l,l+1)}$针对示例导出按层相关性传播的明确公式。按层相关性传播应反映在分类时间期间传递的消息。已知在分类时间期间，神经元i输入$a_i w_{ik}$到神经元k，只要i具有对k的正向连接即可。因此，可以通过如下表示等式(9)和(10)

[0114] $$R_7^{(3)} = R_7^{(3)} \frac{a_4 w_{47}}{\sum_{i=4,5,6} a_i w_{i7}} + R_7^{(3)} \frac{a_5 w_{57}}{\sum_{i=4,5,6} a_i w_{i7}} + R_7^{(3)} \frac{a_6 w_{67}}{\sum_{i=4,5,6} a_i w_{i7}} \quad (14)$$

[0115] $$R_4^{(2)} = R_4^{(2)} \frac{a_1 w_{14}}{\sum_{i=1,2} a_i w_{i4}} + R_4^{(2)} \frac{a_2 w_{24}}{\sum_{i=1,2} a_i w_{i4}} \quad (15)$$

[0116] 一般来说，这可以表示为

[0117] $$R_{i \leftarrow k}^{(l,l+1)} = R_k^{(l+1)} \frac{a_i w_{ik}}{\sum_h a_h w_{hk}} \quad (16)$$

[0118] 虽然该定义仍然需要被适配，使得当分母变为零时它是可用的，但是等式(16)中给出的示例给出了消息$R_{i \leftarrow k}^{(l,l+1)}$可能是什么的想法，即宿神经元$R_k^{(l+1)}$的相关性，其已经通过前一层l的神经元i的输入被成比例地加权计算得到。当使用不同的分类架构并且通过给定层的特征向量的维度来代替神经元的概念时，这个概念保持类似的方式。

[0119] 公式(16)具有第二属性：如果神经元$a_i w_{ik}$的贡献的符号与与来自所有输入神经元的贡献总和的符号不同，即如果神经元与该神经元从其继承了部分相关性的顶部神经元的整体趋势相反，则消息$R_{i \leftarrow k}^{(l,l+1)}$发送的相关性的符号将被交换。与等式(5)中的线性映射的示例相同，输入神经元可以根据其输入符号继承正或负相关性。

[0120] 此处还显示了另一个属性。相关性分布的公式适用于非线性甚至是非可微分(non-differentiable)或非连续神经元激活。算法将从已经计算的层l+1的相关性$R^{(l+1)}$开始。然后，将以保持等式(13)成立的方式，针对层l+1的所有元素k和前一层l的元素i来计算消息$R_{i \leftarrow k}^{(l,l+1)}$。然后，定义(8)将用于定义层l的所有元素的相关性$R^{(l)}$。

[0121] 泰勒型分解

[0122] 用于实现如(1)中对于一般可微分预测器f的分解的一种替代方法是一阶泰勒近似。

[0123] $$f(x) \approx f(x_0) + Df(x_0)[x - x_0]$$
$$= f(x_0) + \sum_{d=1}^{V} \frac{\partial f}{\partial x_{(d)}}(x_0)(x_{(d)} - x_{0(d)}) \quad (17)$$

[0124] 在此设置中，泰勒基点x_0的选择是自由参数。如上所述，在分类的情况下，有趣的是找出每个像素相对于预测的最大不确定性的状态的贡献，其由满足$f(x_0)=0$的点的集合给出，因为$f(x)>0$表示学习结构的存在，而$f(x)<0$表示学习结构的不存在。因此，x_0应选择为预测器f的根。为了预测的泰勒近似的精确性，x_0应该选择成接近欧氏标准的x，以便根据更高阶泰勒近似来最小化泰勒残差。在具有最小范数的多个现有根x_0的情况下，可以对它们进行平均或整合，以便获得所有这些解决方案的平均值。上述等式简化为

[0125] $$f(x) \approx \sum_{d=1}^{V} \frac{\partial f}{\partial x_{(d)}}(x_0)(x_{(d)} - x_{0(d)})$$从而$f(x_0)=0$ (18)

[0126] 由于需要找到接近根点x_0，因此按像素分解包含对超出泰勒级数的预测点x的非线性相关性。因此，整个按像素分解不是线性的，而是局部线性算法，因为根点x_0取决于预测点x。

[0127] 若干工作一直在使用灵敏度图[2,18,38]来可视化基于在预测点x使用偏导数的分类器预测。在基于预测点x的导数的灵敏度图和按像素分解方法之间存在两个本质区别。首先，在预测点x处的函数值f(x)与同一点x处的差分Df(x)之间没有直接的关系。其次，有

趣的是解释相对于由预测函数$f(x_0)=0$的根集合给出的某一状态的分类器预测。所述预测点的差分$Df(x)$不一定指向在欧几里得规范下接近的根。它指向可能仍然具有与预测$f(x)$相同符号的最近的局部最优值，因此对于解释与预测函数的根点的集合的差异是误导的。因此，在预测点x的导数对于实现目标是没有用的。图7示出了预测的局部梯度(向上箭头)和维度分解(向下箭头)之间的定性差异。具体地，该图直观地示出了预测点x处的梯度(在这里由平方表示)并不一定指向决定边界上的接近点。相反，它可能指向决定边界上的局部最优值或远点。在该示例中，来自预测点x的局部梯度的解释向量在不相关的方向上具有太大的贡献。另一类的最近邻居可以在非常不同的角度处找到。因此，预测点x处的局部梯度可能不能良好地解释单个维度对函数值$f(x)$的贡献。左侧图像中的预测点的局部梯度和右图像中的泰勒根点的局部梯度由黑色箭头表示。最近的根点x_0在决定边界上显示为三角形。右侧图像中的向下箭头通过围绕最近根点x_0的泰勒扩展使得$f(x)$的近似值可视化。给出该近似值作为表示$Df(x_0)$(右侧面板中的灰色箭头)和$x-x_0$(右侧面板中的虚线)之间的维度乘积的向量，其相当于$Df(x_0)$和$x-x_0$之间的外积的对角线。

[0128] 一个技术难题是找到根点x_0。对于连续分类器，可以使用未标记的测试数据或通过采样方法中从训练数据学习的生成模型产生的数据，并在预测点x和候选点集合$\{x'\}$之间执行线搜索，使得它们的预测具有相反的符号：$f(x)f(x')<0$.很明显，所述线$l(a)=ax+(1-a)x'$必须包含可以通过区间交集找到的f的根。因此，每个候选点x'产生一个根，并且可以选择使泰勒残差最小化的根点或使用具有低泰勒残余的根点子集上的平均值。

[0129] 注意，当应用于一层或多层的子集时，泰勒型分解可以被视为当函数为高度非线性时的相关性传播的近似方式。特别是当将其应用于输出函数f作为前一层的函数$f=f(Z_{i-1})$时满足这种情况，因为当输出层的相关性被初始化为预测函数$f(x)$的值时，等式(18)大致满足传播等式(2)。与泰勒近似不同，按层相关性传播不需要使用除输入点之外的第二点。在“多层网络的按像素分解”一节中的公式将证明可以针对广泛的架构实现按层相关性传播，而无需通过泰勒扩展进行近似。

[0130] 多层网络的按像素分解

[0131] 多层网络通常构建为以层方式组织的相互联系的神经元集合。它们在彼此组合时定义数学函数，将第一层神经元(输入)映射到最后一层神经元(输出)。用x_i表示每个神经元，其中i是神经元的索引。按照惯例，将网络的每层的不同索引进行关联。用“Σ_i”表示给定层的所有神经元的总和，并且用“Σ_j”表示另一层的所有神经元的总和。用$x_{(d)}$表示与像素激活对应的神经元(即，利用其来获得分类决定的分解)。从一层到下一层的公共映射由后跟非线性函数的线性投影组成：

[0132] $$z_{ij}=x_i w_{ij}, \quad (50)$$

[0133] $$z_j=\Sigma_i z_{ij}+b_j, \quad (51)$$

[0134] $$x_j=g(z_j), \quad (52)$$

[0135] 其中w_{ij}是将神经元x_i连接到神经元x_j的权重，b_j是偏置项，而g是非线性激活函数(参见图8，用于说明所使用的命名法)。多层网络将这些层中的若干层堆叠，每层都由大量的神经元组成。公共非线性函数是双曲正切$g(t)=\tanh(t)$或修正函数$g(t)=\max(0,t)$。当卷积与和汇集是线性运算时，神经网络的这种公式通常足以涵盖广泛的架构，例如简单的多层感知器[39]或卷积神经网络[25]。

[0136] 泰勒型分解

[0137] 通过用f:$\mathbb{R}^M \mapsto \mathbb{R}^N$表示实现网络的输入和输出之间的映射的向量值多变量函数,可以通过在决定函数f的近根点x_0处的泰勒扩展来获得分类决定$x \mapsto f(x)$的第一可能解释:

[0138] $$R_d^{(1)} = (x - x_0)_{(d)} \cdot \frac{\partial f}{\partial x_{(d)}}(x_0) \quad (53)$$

[0139] 通过使用反向传播算法[39]来重用网络拓扑,可以有效地计算按像素分解所需的导数$\partial f(x)/\partial x_{(d)}$。具体地,将导数反向传播到某一层j,可以使用链规则计算上一层i的导数:

[0140] $$\frac{\partial f}{\partial x_i} = \sum_j \frac{\partial f}{\partial x_j} \cdot \frac{\partial x_j}{\partial x_i} = \sum_j \frac{\partial f}{\partial x_j} \cdot w_{ij} \cdot g'(z_j) \quad (54)$$

[0141] 基于泰勒的分解的要求是找到支持对x的分类决定的局部解释的根x_0(即分类边界上的点)。这些根可以通过x附近的本地搜索来找到。然而,如[43]所述,这可能导致输入空间的点,其在感知上等同于原始样本x,并且其作为根的选择将产生非信息按像素分解。

[0142] 或者,可以通过在由x及具有不同类的其最近邻居定义的段上进行线搜索来找到根点。当数据流形被稀疏地填充时,该解决方案是有问题的,这和自然图像的情况一样。在这种情况下,可能遵循x及其最近邻居之间的直线将强烈偏离数据流形,并产生具有类似的差的按像素分解的根x_0。

[0143] 按层相关性反向传播

[0144] 作为泰勒型分解的替代方案,可以计算向后通路中的每层的相关性,即,将相关性$R_i^{(l)}$表达为上层相关性$R_j^{(l+1)}$的函数,并反向传播相关性直到达到输入(像素)。

[0145] 该方法的工作原理如下:知道某个神经元$R_j^{(l+1)}$对于分类决定f(x)的相关性,人们希望获得在发送到先前层的神经元的消息方面的这种相关性的分解。这些消息称为$R_{i\leftarrow j}$。具体地,如等式(8)和(13)所表示的那样,必须保持守恒属性:

[0146] $$\sum_i R_{i\leftarrow j}^{(l,l+1)} = R_j^{(l+1)} \quad (55)$$

[0147] 在线性神经元$x_j = \sum_i z_{ij}$的情况下,其中相关性$R_j = f(x)$,这种分解立即由$R_{i\leftarrow j} = z_{ij}$给出。然而,在一般情况下,神经元激活$x_j$是$z_j$的非线性函数。然而,对于双曲正切和修正函数(满足g(0)=0的两个简单单调递增函数),预激活z_{ij}仍然提供一个明智的方法来度量每个神经元x_i对R_j的相对贡献。相关性分解的第一个可能的选择基于局部和全局预激活的比,并且由下式给出:

[0148] $$R_{i\leftarrow j}^{(l,l+1)} = \frac{z_{ij}}{z_j} \cdot R_j^{(l+1)} \quad (56)$$

[0149] 这些相关性$R_{i\leftarrow j}$很容易显示为近似等式(2)的守恒属性,具体是:

[0150] $$\sum_i R_{i\leftarrow j}^{(l,l+1)} = R_j^{(l+1)} \cdot \left(1 - \frac{b_j}{z_j}\right) \quad (57)$$

[0151] 其中乘数考虑到偏置项吸收(或注入)的相关性。如果需要,残留偏置相关性可以

重新分布到每个神经元x_i上。

[0152] 等式(56)的传播规则的缺点在于:对于小值z_j,相关性$R_{i\leftarrow j}$可以取无限值。可以通过引入预定义的稳定器$\varepsilon \geq 0$来克服无限性。

[0153] $$R_{i\leftarrow j}^{(l,l+1)} = \begin{cases} \frac{z_{ij}}{z_j+\varepsilon} \cdot R_j^{(l+1)} & z_j \geq 0 \\ \frac{z_{ij}}{z_j-\varepsilon} \cdot R_j^{(l+1)} & z_j < 0 \end{cases} \quad (58)$$

[0154] 随后守恒定律变为

[0155] $$\sum_i R_{i\leftarrow j}^{(l,l+1)} = \begin{cases} R_j^{(l+1)} \cdot \left(1 - \frac{b_j+\varepsilon}{z_j+\varepsilon}\right) & z_j \geq 0 \\ R_j^{(l+1)} \cdot \left(1 - \frac{b_j-\varepsilon}{z_j-\varepsilon}\right) & z_j < 0 \end{cases} \quad (59)$$

[0156] 其中可以观察到稳定器吸收了一些进一步的相关性。具体地,如果稳定器ε变得非常大,则相关性被完全吸收。

[0157] 不泄漏相关性的替代稳定方法由分别处理负和正预激活构成。假设$z_j^+ = \sum_i z_{ij}^+ + b_j^+$和$z_j^- = \sum_i z_{ij}^- + b_j^-$,其中,"-"和"+"表示$z_{ij}$和$b_j$的负部和正部。相关性传播现在定义为

[0158] $$R_{i\leftarrow j}^{(l,l+1)} = R_j^{(l+1)} \cdot \left(\alpha \cdot \frac{z_{ij}^+}{z_j^+} + \beta \cdot \frac{z_{ij}^-}{z_j^-}\right) \quad (60)$$

[0159] 其中$\alpha>0$,$\beta<0$,$\alpha+\beta=1$。例如,对于$\alpha=2\beta=-1$,守恒定律变为:

[0160] $$\sum_i R_{i\leftarrow j}^{(l,l+1)} = R_j^{(l+1)} \cdot \left(1 - \frac{b_j^+}{2z_j^+} - \frac{b_j^-}{2z_j^-}\right) \quad (61)$$

[0161] 其与等式(57)具有相似的形式。这种替代传播方法还允许通过选择不同的因子α和β来手动控制正和负证据的重要性。

[0162] 在下文中,更一般地,针对从神经元j到作为神经元j的上游邻居的神经元i的相关性消息写入R_{ij}。在具有分层结构的神经网络的具体情况下,R_{ij}是$R_{i\leftarrow j}^{(l,l+1)}$的缩写方式,其中i和j分别是层I和I+1的神经元。类似地,可以删除神经元的相关性分数的层索引,写成R_j而不是$R_j^{(l+1)}$。

[0163] 除了上面的重新分布公式,可以定义替代的公式如下:

[0164] $$R_{ij} = \frac{x_i w_{ij} + \frac{b_j}{n}}{h\left(\sum_r \left(x_r\ w_{rj} + \frac{b_j}{n}\right)\right)} R_j \quad (等式A5)$$

[0165] 或者

[0166] $$R_{ij} = \frac{x_i w_{ij}}{h\left(\sum_r \left(x_r\ w_{rj} + \frac{b_j}{n}\right)\right)} R_j \quad (等式A6)$$

[0167] 其中,n是各个神经元的上游邻居神经元的数目,R_{ij}是从相应的神经元j重新分布给上游邻居神经元i的相关性值,R_j是作为神经元i的下游神经元的神经元j的相关性,x_i是

在神经网络的应用过程中上游邻居神经元i的激活，w_{ij}是将上游邻居神经元i连接到相应神经元j的权重，w_{rj}也是将上游邻居神经元r连接到相应神经元j的权重，并且b_j是相应神经元j的偏置项，以及h()是标量函数。通常情况下，h()是数值稳定项，它通过添加小的ε(例如h(x)＝x+ε•sign(x))来保持该值远离零。

[0168] 类似地，其他替代方案是：

[0169] $$R_{ij}=\left(\alpha\frac{\left(x_i\,w_{ij}+\frac{b_j}{n}\right)_+}{h\left(\sum_r\left(x_r\,w_{rj}+\frac{b_j}{n}\right)_+\right)}-\beta\frac{\left(x_i\,w_{ij}+\frac{b_j}{n}\right)_-}{h\left(\sum_r\left(x_r\,w_{rj}+\frac{b_j}{n}\right)_-\right)}\right)R_j$$

（等式A7）

[0170] 或者

[0171] $$R_{ij}=\left(\alpha\frac{(x_i\,w_{ij})_+}{h\left((b_j)_++\sum_r(x_r\,w_{rj})_+\right)}-\beta\frac{(x_i\,w_{ij})_-}{h\left((b_j)_-+\sum_r(x_r\,w_{rj})_-\right)}\right)R_j$$

（等式A8）

[0172] 一旦选择了相关性传播的规则，通过与等式(8)和(13)相一致地将来自所有上层神经元的相关性相加，来确定下层中每个神经元的总体相关性：

[0173] $$R_i^{(l)}=\sum_j R_{i\leftarrow j}^{(l,l+1)}\quad(62)$$

[0174] 相关性从一层反向传播到另一层，直到到达输入像素$x_{(d)}$，并且其中相关性$R_d^{(1)}$提供所需的决定f(x)的按像素分解。算法2中总结了神经网络的完整的按层相关性传播过程。

[0175] [0176]

算法2 神经网络的按像素分解

输入： $R^{(L)}=f(x)$

针对$l\in\{L-1,\ldots,1\}$执行

如等式（58）或（60）中计算$R_{i\leftarrow j}^{(l,l+1)}$

$R_i^{(l)}=\sum_j R_{i\leftarrow j}^{(l,l+1)}$

结束

输出： $\forall d: R_d^{(1)}$

[0177] 上述公式(58)和(60)可直接适用于满足某种结构的层。假设从某层获得神经元激活x_j，其被建模为来自前一层的激活x_i的输入函数。随后，按层相关性传播可直接适用于如下情况：存在函数g_j和函数h_{ij}，使得

[0178] $x_j=g_j(\sum_i h_{ij}(x_i))$ (63)

[0179] 在这种一般情况下，等式(50)的加权项$z_{ij}=x_iw_{ij}$必须由函数$h_{ij}(x_i)$相应地替换。再次重申，作为广义手段的限制，即使是最大汇集(pooling)也适用于这种结构，请参见例如等式(32)。对于具有较高非线性度的结构，例如局部重归一化[26,36]，可以再次使用应用于神经元激活x_j的泰勒近似来实现如等式(63)中给出的结构的近似。

[0180] 最后，从本节确立的公式可以看出，按层相关性传播与泰勒级数或偏导数不同。与泰勒级数不同，它不需要输入图像以外的第二点。泰勒级数的层应用可以解释为实现按层相关性传播的近似版本的通用方式。类似地，与依赖于导数的任何方法相反，神经元激活的微分或平滑属性不是能够定义满足按层相关性传播的公式的必要要求。在这个意义上，这是一个更一般的原则。

[0181] 泛化观点

[0182] 以上公式A5-A8可以进行泛化。

[0183] 假设已经获得在层I+1处的所有神经元k的相关性分数$R_k^{(l+1)}$。首先注意，基本思想是生成消息$R_{i\leftarrow k}^{(l,l+1)}$，使得满足等式(13)

[0184] $$R_k^{(l+1)} = \sum_{i:\ i\text{是针对神经元}k\text{的输入}} R_{i\leftarrow k}^{(l,l+1)}$$

[0185] 然后从这些消息计算层I处的所有神经元i的相关性$R_i^{(l)}$。如上所述，等式A5至A8是关于如何计算消息$R_{i\leftarrow k}^{(l,l+1)}$的示例。在上述方法中，使用等式(8)

[0186] $$R_i^{(l)} = \sum_{k:\ i\text{是针对神经元}k\text{的输入}} R_{i\leftarrow k}^{(l,l+1)}$$

[0187] 计算层I处的所有神经元i的相关性$R_i^{(l)}$。

[0188] 可以针对等式(8)进行第一次泛化：

[0189] 给定所有消息$R_{i\leftarrow k}^{(l,l+1)}$，，可以通过使用另一函数而不是相关性消息$R_{i\leftarrow k}^{(l,l+1)}$的和来计算层I处所有神经元i的相关性$R_i^{(l)}$，将其表示为m(•)，并且将消息$R_{i\leftarrow k}^{(l,l+1)}$作为输入：神经元i的相关性通过函数m(•)计算为：

[0190] $$R_i^{(l)} = \mathrm{m}\left(\left\{R_{i\leftarrow k}^{(l,l+1)} \middle|\ k:\ i\ \text{是针对神经元}k\text{的输入}\right\}\right)$$

[0191] 其在每个自变数中应该单调递增，并且可以看作是等式(8)中的和的泛化。当使用上游和下游神经元的术语时，可以写为：

[0192] $R_i = \mathrm{m}(\{R_{i\leftarrow k} | k\text{是}i\text{的下游神经元}\})$

[0193] 这种泛化的略失一般但可能经常使用的变体是：

[0194] $$R_i = \mathrm{m}_2\left(\sum_{k\text{是i的下游神经元}} m_3(R_{i\leftarrow k})\right)$$

[0195] 其中m_2和m_3是一个变量的单调递增函数。

[0196] 例如：

[0197] $$R_i = c\left(\sum_{k\text{是 i的下游神经元}} (R_{i\leftarrow k})^n\right)^{1/n}$$

[0198] 其中c是所选择的使得相关性守恒成立的常数。这个示例对于n的大值的大的项给予更多的权重。

[0199] 可以在如下情况下对等式(13)进行第二次泛化：当考虑公式A5至A8，其中$R_{i\leftarrow k}^{(l,l+1)}$始终是乘以$R_k^{(l+1)}$的项时：

[0200] $$R_{i\leftarrow k}^{(l,l+1)} = q(i)R_k^{(l+1)}$$

[0201] 其中q(i)是权重函数，使得

[0202] $$1 = \sum_{k:\ i\ \text{是针对神经元k的输入}} q(i) = \sum_{i\text{是 k的上游神经元}} q(i)$$

[0203] 这确保等式(13)仍然成立。

[0204] 由于先前已经从层I+2的神经元p的神经元相关性分数计算了层I+1的神经元k的神经元相关性分数，所以也可以将上面的公式重写为：

[0205] $$R_{i\leftarrow k}^{(l,l+1)} = q(i)\ \mathrm{m}\left(\{R_{k\leftarrow \mathrm{p}} | p\ \text{是k的下游神经元}\}\right)$$

[0206] 因此，达到了第一级泛化：

[0207] 泛化1

[0208] 给定神经元集合{k}的神经元相关性分数R_k的集合，计算与神经元集合{i}的相关性消息，神经元集合{i}是神经元集合{k}的上游神经元，使得存在消息加权函数q(•)

[0209] 使得$R_{i\leftarrow k}=q(i)R_k$

[0210] 给定相关性消息$R_{i\leftarrow k}$的集合，通过函数m(•)计算神经元i的相关性分数，该函数在其自变量中单调递增使得：

[0211] $R_i=m(\{R_{i\leftarrow k}|k是i的下游神经元\})$

[0212] 具体地，当仅使用相关性消息项，并且假定获得针对作为神经元i的下游神经元的所有神经元k的消息$\{R_{k\leftarrow p}|p是k的下游神经元\}$时，则可以计算：

[0213] $R_{i\leftarrow k}=q(i)m(\{R_{k\leftarrow p}|p是k的下游神经元\})$

[0214] 泛化1结束

[0215] 此外，可能要求相关性守恒属性得到满足。就是这种情况，例如，如果网络是分层的，则函数m(.)是元素之和，并且如果等式

[0216] $$1 = \sum_{k:\ \text{i是针对神经元k的输入}} q(i) = \sum_{\text{i是k的上游神经元}} q(i)$$

[0217] 成立。

[0218] 注意，数值稳定性的要求可能要求包括数字稳定项，使得相关性守恒属性仅大致满足，例如使得层的相关性的和等于5％的偏差。作为数值稳定器的一个示例，参见公式A5和A6中使用的函数$h(z)=z+\varepsilon \cdot sign(z)$。

[0219] 泛化2

[0220] 相关性守恒属性达到一定容差的要求可以通过以下条件来表达：

[0221] $$0.95 \sum_{k\text{是层 }l+1\text{的神经元}} R_k \leq \sum_{i\text{ 是层}l\text{的神经元}} R_i \leq 1.05 \sum_{k\text{是层 }l+1\text{的神经元}} R_k$$

[0222] 在上游和下游术语方面，这将是：

[0223] $$0.95 \sum_{k\text{ 是集合}I\text{的某神经元的下游神经元}} R_k \leq \sum_{i\text{ 是集合}I\text{的神经元}} R_i \leq 1.05 \sum_{k\text{是集合}I\text{的某神经元的下游神经元}} R_k$$

[0224] 这也可以用两种不同的观点进行重新设计。在第一种观点中，仅考虑来自输出的初始相关性R和针对输入项目集合中的每个项目的相关性R_i，所述输入项目作为神经网络的输入。那么可以在这些项中制定上面的要求，而不必规定神经网络的中间层的相关性的和：

[0225] $0.95R \leqslant \Sigma_{\text{项目中的}i} R_i \leqslant 1.05R$

[0226] 在第二种观点中，考虑进入和离开一个固定神经元的神经元之间的相关性消息，而不是神经元的相关性分数。

[0227] 要求从所有下游神经元进入特定神经元j的消息的和近似等于从神经元j发送到其上游神经元的消息的和，同样示例性地具有5％的容差：

[0228] $$0.95 \sum_{k\text{ 是神经元}j\text{的上游神经元}} R_{jk} \leq \sum_{i\text{ 是神经元}j\text{的下游神经元}} R_{ij} \leq 1.05 \sum_{k\text{是神经元}j\text{的上游神经元}} R_{jk}$$

[0229] 泛化2结束

[0230] 在如下情况下，这三种观点全都可以进一步泛化：当针对中间项考虑单调函数ζ、f或ξ，该单调函数仅取决于其输入时：

[0231] 泛化2B

[0232] 观点1：神经元的相关性分数R_k

[0233] $$0.95 \sum_{k\text{ 是集合}I\text{的某神经元的下游神经元}} R_k \leq \zeta\left(\sum_{i\text{是集合}I\text{的神经元}} R_i\right) \leq 1.05 \sum_{k\text{是集合}I\text{的某神经元的下游神经元}} R_k$$

[0234] 观点2：输出神经元的相关性分数R和输入项目集合中的项目的相关性分数

[0235] $0.95R \leqslant f(\Sigma_{\text{项目中的}i} R_i) \leqslant 1.05R$

[0236] 观点3：神经元j的上游和下游神经元邻居的相关性消息R_{jk}。

[0237] $$0.95\sum_{\substack{k\text{是神经元}j\\ \text{的上游神经元}}} R_{jk} \le \xi\left(\sum_{\substack{i\text{是神经元}j\\ \text{的下游神经元}}} R_{ij}\right) \le 1.05\sum_{\substack{k\text{是神经元}j\\ \text{的上游神经元}}} R_{jk}$$

[0238] 泛化2B结束

[0239] 现在考虑第三级泛化。

[0240] 检查等式A5至A8，可以确定上述级别泛化的一些额外要求。首先，等式A5至A8中的q(i)取决于加权激活z_{ij}。公式A5与A6和A7与A8之间的差异仅在于加权激活z_{ij}的定义。

[0241] 在A5和A7中，加权激活是$z_{ij}=x_iw_{ij}$。在A6和A8中，加权激活是$z_{ij}=x_iw_{ij}+\frac{b_j}{I}$，其中$b_j$是神经元j的偏置，而I是神经元j的上游神经元的数量。加权激活的定义差异来自偏置项的两种不同观点。在第一等式$z_{ij}=x_iw_{ij}$中，偏置项由单独的神经元建模，其输出具有等于b_j的值的恒定输出。由于偏置由单独的神经元产生，所以它不会进入加权激活的计算。

[0242] 在第二种观点中，偏置是一个附加项，它被加到神经元j的每个输入中，这解释了在加权激活的第二定义中的添加项$\frac{b_j}{I}$。

[0243] 因此实际上，仅仅从两个等式A5和A7得到两个基本公式，利用两种不同的方式来定义加权激活z_{ij}。

[0244] $$R_{ij}=\frac{z_{ij}}{h\left(b_j[\![A]\!]+\sum_r z_{ij}\right)}\sum_k R_{jk}$$ 等式（A5*）

[0245] 和

[0246]

$$R_{ij}=\left(\alpha\frac{(z_{ij})_+}{h\left((b_j)_+[\![A]\!]+\sum_r(z_{ij})_+\right)}-\beta\frac{(z_{ij})_-}{h\left((b_j)_-[\![A]\!]+\sum_r(z_{ij})_-\right)}\right)\sum_k R_{jk}$$ 等式（A7*）

[0247] 其中如果z_{ij}的定义不包含偏置，即如果z_{ij}被定义为$Z_{ij}=x_iw_{ij}$，则$[\![A]\!]$为1，否则为零。这里隐含地使用

[0248] $$R_j=\sum_{k\text{是}j\text{的下游神经元}} R_{jk}$$

[0249] 而不是通过单调递增函数m(•)对神经元相关性分数R_j的一般定义。在这些由等式A5*和A7*给出的特殊情况下，得到

[0250] $$q(i)=\frac{z_{ij}}{h\left(b_j[\![A]\!]+\sum_r z_{rj}\right)}$$

[0251] 和

[0252] $$q(i)=\left(\alpha\frac{(z_{ij})_+}{h\left((b_j)_+[\![A]\!]+\sum_r(z_{rj})_+\right)}-\beta\frac{(z_{ij})_-}{h\left((b_j)_-[\![A]\!]+\sum_r(z_{rj})_-\right)}\right)$$

[0253] 该检查导致第三级泛化：

[0254] 泛化3

[0255] 函数q(i)取决于加权激活z_{ij}，其中加权激活是神经元激活x_i，、连接权重w_{ij}和偏置项b_j的函数。

[0256] $z_{ij}=s(x_i,w_{ij},b_j)$，

[0257] 作为特殊情况

[0258] $z_{ij}=x_iw_{ij}$和$z_{ij}=x_iw_{ij}+\frac{b_j}{l}$

[0259] 泛化3结束

[0260] 最后还有第四级泛化。当检查等式A5*和A7*时，可以看到一个隐含属性，即q(i)与加权激活的排序的相关性。直观地，对于两个神经元i_1和i_2，如果加权激活中的一个大于另一个：$z_{i_1j}<z_{i_2j}$，那么神经元i_2也应该从神经元j接收比神经元i_1更大的相关性份额。然而，由于神经元相关性分数R_j、加权激活z_{ij}和权重q(i)可能具有不同的符号，导致得到的相关性消息$R_{i\leftarrow j}$中的符号翻转，因此必须注意定义这种直观的概念。这就是为什么不能简单地声称$z_{i_1j}<z_{i_2j}\Rightarrow q(i_1)<q(i_2)$。给出一个反例：在公式A5*中，如果$0<z_{i_1j}<z_{i_2j}$但$h\left(b_j[\![A]\!]+\sum_r z_{rj}\right)<0$，则遵从：$q(i_1)>q(i_2)>0$.但是在这种情况下成立的是：$|q(i_1)|<|q(i_2)|$，因为项$h\left(b_j[\![A]\!]+\sum_r z_{rj}\right)$对于$q(i_1)$和$(i_2)$是相同的。

[0261] 检查公式A5*和A7*，可以得出这些公式满足的一组排序属性。定义排序属性的一种方法是考虑加权激活z_{ij}的绝对值和消息加权函数(•)的绝对值的泛化。

[0262] 对于公式A5*，以下排序属性成立：

[0263] $$|z_{i_1j}|<|z_{i_2j}|\Rightarrow|q(i_1)|<|q(i_2)|$$

[0264] 对于公式A7*，略有不同的排序属性成立。考虑

[0265] $$\alpha^*=\frac{\alpha}{h\left((b_j)_+[\![A]\!]+\sum_r(z_{rj})_+\right)}$$

[0266] 和

[0267] $$\beta^*=\frac{\beta}{h\left((b_j)_-[\![A]\!]+\sum_r(z_{rj})_-\right)}$$

[0268] 那么针对函数

[0269] $g(z)=\alpha^*(z)_+-\beta^*(z)_-=\alpha^*\max(0,z)-\beta^*\min(0,z)$

[0270] 以下排序属性成立：

[0271] $$g(z_{i_1j})<(z_{i_2j})\Rightarrow|q(i_1)|<|q(i_2)|$$

[0272] 这里请注意，$|z|=\alpha(z)_+-\alpha(z)_-$，其中$\alpha=1$，$\beta=1$，使得函数g(.)还包括公式A5*的

排序属性，其中α，β具有不同值。

[0273] 以上给出的函数g(•)的进一步泛化导致如下函数：其最小值为零，并且在区间(-∞，0)上单调递减，在区间(0，+∞)上单调递增。

[0274] 所以到达

[0275] 泛化4

[0276] 消息函数q(•)需要满足排序函数：对于作为神经元j的上游邻居神经元的所有i_1和i_2，其中

[0277] $$g(z_{i_1 j}) < g(z_{i_2 j})$$

[0278] 对于最小值为零、并且在区间(-∞，0)上单调递减、在区间(0，+∞)上单调递增的函数g(•)而言，$|q(i_1)| \leq |q(i_2)|$成立。

[0279] 特别对于函数g(•)的一个选择是$g(z) = \alpha \max(0,z) - \beta \min(0,z)$，其中$\alpha \geq 0$，。$\beta \geq 0$

[0280] 泛化4结束

[0281] 定义排序属性的另一种方法是将自身限制为当$R_j>0$时的情况。这在人们对传播负神经元相关性不感兴趣时是合理的。为了理解这一点，应该考虑：当神经网络作出的预测确定了结构的存在时，人们通常有兴趣对项目集合中的单个项目进行预测，这意味着神经的输出在作为输入的项目集合上具有高的正分数。如果神经的输出具有高的正分数，则可以预期大多数神经元相关性也是正的，仅因为大多数神经元都支持神经网络的高的正预测，因此在实践中可以忽略具有负相关性的小部分神经元。

[0282] 为了推断另一排序属性，请注意，如果$\sum_i z_{ij}>0$，那么也得到：对于$h(t) = t+\varepsilon \operatorname{sign}(t)$如果$h(\sum_i z_{ij})>0$。

[0283] 具体地，当考虑公式A5*时，那么以下排序属性成立：

[0284] 如果$\sum_i z_{ij}>0$，那么对于作为神经元j的上游神经元的所有i_1和i_2，得到：

$$z_{i_1 j} < z_{i_2 j} \Rightarrow q(i_1) < q(i_2)$$

[0285] 如果$\sum_i z_{ij}<0$，那么对于作为神经元j的上游神经元的所有i_1和i_2，得到：

$$z_{i_1 j} < z_{i_2 j} \Rightarrow q(i_1) > q(i_2)$$

[0286] 此属性对于公式A7*不成立。

[0287] 泛化5

[0288] 消息函数q(•)需要满足排序属性：如果$R_j>0$ and $\sum_i z_{ij}>0$，那么对于作为神经元j的上游神经元的所有i_1和i_2，得到：$z_{i_1 j} < z_{i_2 j} \Rightarrow q(i_1) < q(i_2)$

[0289] 泛化5结束

[0290] 对于$R_j>0$情况下可能有用的另一排序属性将是：

[0291] $$0 < z_{i_1 j} < z_{i_2 j} \Rightarrow 0 < q(i_1) < q(i_2)$$

[0292] 这对于公式A7*是成立的。

[0293] 存在对于这两个公式A5*和A7*也成立的另一排序属性，即，如果只比较具有相同符号的加权激活：

[0294] 消息函数q(•)需要满足排序属性，如果

[0295] $|z_{i_1 j}| < |z_{i_2 j}|$和$(z_{i_1 j}) = sign(z_{i_2 j})$，则$|q(i_1)| \leq |q(i_2)|$成立。

[0296] 这是用绝对值来替换函数g(•)的方式。

[0297] 注意，公式A5*满足更窄的排序属性，即

[0298] $$|z_{i_1 j}| < |z_{i_2 j}| \Rightarrow |q(i_1)| < |q(i_2)|$$

[0299] 当插入$z_{ij}=x_i w_{ij}$或$z_{ij} = x_i w_{ij} + \frac{b_j}{I}$时，所有这些公式都成立，以便可以根据使用的加权激活z_{ij}的定义，从上述每个排序属性创建两个版本。

[0300] 请注意，还可以定义排序属性。

[0301] 例如，以下八个条件也产生有意义的排序属性，这些属性用相关性消息表达：

[0302] $$|x_i\, w_{ij}| < |x_k\, w_{kj}| \Rightarrow |R_{ij}| < |R_{kj}|$$

[0303] 或者

[0304] $$\left|x_i\, w_{ij} + \frac{b_j}{n}\right| < \left|x_k\, w_{kj} + \frac{b_j}{n}\right| \Rightarrow |R_{ij}| < |R_{kj}|$$

[0305] 或者

[0306] $$|x_i\, w_{ij}| < |x_k\, w_{kj}|\ and\ sign(x_i\, w_{ij}) = sign(x_k\, w_{kj}) \Rightarrow |R_{ij}| < |R_{kj}|$$

[0307] 或者

[0308] $$\left|x_i\, w_{ij} + \frac{b_j}{n}\right| < \left|x_k\, w_{kj} + \frac{b_j}{n}\right|\ and\ sign\left(x_i\, w_{ij} + \frac{b_j}{n}\right) = sign(x_k\, w_{kj} + \frac{b_j}{n}) \Rightarrow |R_{ij}| < |R_{kj}|$$

[0309] 或者

[0310] $$\left(R_j > 0 \ \wedge \sum_i x_i w_{ij} > 0\right) \Rightarrow (\forall\, i,k:\ x_i w_{ij} < x_k w_{kj} \Rightarrow R_{ij}$$

[0311] 或者

[0312] $$\left(R_j > 0 \ \wedge \sum_i x_i w_{ij} > -b_j\right) \Rightarrow (\forall\, i,k:\ x_i w_{ij} < x_k w_{kj} \Rightarrow R_{ij} \le R_{kj})$$

[0313] 或者

[0314] $$\forall\, i,k:\ (R_j > 0 \ \wedge\ (0 < x_i w_{ij} < x_k w_{kj})) \Rightarrow (0 \le R_{ij} \le R_{kj})$$

[0315] 或者

[0316] $$\forall\, i,k:\ \left(R_j > 0 \ \wedge\ \left(-\frac{b_j}{n} < x_i w_{ij} < x_k w_{kj}\right)\right) \Rightarrow (0 \le R_{ij} \le R_{kj})$$

[0317] 作为根据网络输入将泰勒扩展应用到网络输出函数上，泰勒扩展也可以用于将单个神经元的相关性分数重新分布到其上游邻居上。这允许将上述针对一组神经元所提出的策略与针对另一组神经元的符合泰勒分布的相关性分布相结合。泰勒扩展可以以下列方式

使用：假设$x_j\ (x_{i_1},\ldots,x_{i_n})$是神经元j的神经元激活函数，作为上游邻居神经元$i_1,\ldots,i_n$的输入$x_{i_k}$的函数。然后使$\sum_k \frac{\partial x_j}{\partial x_{i_k}}\Big|_{\{(\tilde{x}_{i_1},\ldots,\tilde{x}_{i_n})\}}\cdot\left(x_{i_k}-\tilde{x}_{i_k}\right)$为围绕点$\left(\tilde{x}_{i_1},\ldots,\tilde{x}_{i_n}\right)$的输入$(x_{i_1},\ldots,x_{i_n})$的泰勒扩展。那么可以通过设置下式来使用泰勒扩展和上述公式：$z_{i_k j}=\frac{\partial x_j}{\partial x_{i_k}}\Big|_{\{(\tilde{x}_{i_1},\ldots,\tilde{x}_{i_n})\}}\cdot\left(x_{i_k}-\tilde{x}_{i_k}\right)$。

[0318] 各种附加说明

[0319] 因此，最先进的分类器如深度神经网络(DNN)按如下方式工作。

[0320] 1)网络结构(例如层数、单元等)由人指定。

[0321] 2)使用潜在的数百万标记(和未标记)数据样本(例如图像)来对网络参数(权重)进行训练/优化。请注意，网络上可获得一些预先训练的网络。

[0322] 3)网络可以应用于新的图像，并且能够例如将图像归类为属于特定类，例如“包含鲨鱼的图像”、“作为新闻文章的文本文档”或“缺乏信誉的人”类。

[0323] 4)由于网络高度非线性且非常复杂，因此很难理解为什么这种特定图像被归类为“鲨鱼”。因此，网络充当黑箱(见图4)。

[0324] 5)所提出的实施例能够解释为什么分类器到达其决定，即，能够可视化重要信息所在的位置(例如，像素)。抽象地说，能够将大规模(例如整个图像、整个文本文档)计算的分类决定分解成更小的尺度(例如，单个像素、单个词)。

[0325] 6)由于DNN不仅可以对图像进行训练，而且已被应用于几乎每种类型的数据，例如，时间序列、词、物理度量等，所描述的实施例的原理可应用于许多不同的场景。

[0326] 以下关于图5至10提到的描述将用来提供对图4的相关性分数指派装置的描述的一些额外的说明。上面已经描述过，装置50可以仅被配置为执行重新分布52。然而，另外，装置50还可以被配置为执行人工神经网络10在集合16上的实际应用。因此，对于该替代方案，装置50可以被认为包括可以重用参考标记10的神经网络处理器，以及可以重用参考标记52的重新分布处理器。在任一情况下，装置50可以例如包括储存器或存储器46。然而，有趣的是，应当注意到，在诸如预测过程(比如分类过程)开始涉及网络10的层与反向传播过程52反向遍历网络10所至的层之间可能存在差距。在图1a-c和2a-c的情况下，例如，已经示出了预测过程中涉及的正向传播14跨越或包含与反向传播过程32相同的网络10的层。也就是说，正向传播过程14或网络10被直接应用到集合16上，反向传播32直接结束于集合16的相关性分数。在图1b和图2b的情况下，例如，在预测过程中，通过特征提取过程30预先填充该集合16，并且为了突出相关性分数增加的相关部分，例如，在与原始图像22重叠的方式下，已经使用该特征提取的反转(即38)来扩展反向传播过程并执行空间(像素)域中的相关部分的突出。然而，上面提及的描述也揭示，使用人工神经网络的一个或多个附加层，即正向传播方向14上网络10的实际(训练)部分之前的神经元层，即层或部分21，可以备选地转换或描述特征提取过程30。在相关性指派过程中的反向传播中实际上不需要遍历仅仅镜像特征提取30的任务的这些层。然而，较高级侧的部分21的这些附加(转换)层可以在预测过程期间在正向传播过程中遍历，即在遍历网络10的实际(训练)部分之前开始的端处。因此，将获得特征样本而不是像素的相关性分数R_i。换而言之，相关性不仅可以关于输入变量(例

如，在文本的情况下与每个词相关联的矢量的分量或图像情况下的每个像素的红、绿和蓝分量)进行分解，还可以关于这些项目(例如网络某一层的神经元)的非线性变换进行分解。因此，可能希望停止某个中间层的相关性反投影。自然地，一方面正向传播的起点与另一方面反向传播32的终点之间的这种差距的示例也可以应用于其他种类的数据，即应用于图像之外的数据例如，音频信号、文本等。

[0327] 对于网络输出18和集合16的项目42的种类，附加说明似乎是值得的。关于网络输出18，上面还概述了相同的可以是标量或向量，例如标量或向量的分量是实值。从其导出的相关性值R可以分别是从标量或向量的分量之一导出的实值。对于“项目”42，上述示例应该已经使之足够清楚，同样可以是标量或向量。一方面图1a和2a与另一方面图1c和2c的并置使得这一点变得清楚。在如图1c和2c所示的彩色图片的像素情况下，像素值是向量，即这里的示例性地对应于三(或更多)个标量颜色分量(诸如RGB、CMYK等)的三个或更多个分量的向量。集合16的项目42是像素的标量分量。将相关性值重新分布到项目集合，得到每个项目(即每个像素的每个分量)的相关性值R_i。为了导出每个像素的一个标量相关性值，可以对相应像素的所有分量的相关性值进行求和，以获得该像素的这样的公共相关性值。这已经在图2c中的37处示出。在文本的情况下也可能出现类似的措施。因此，关于输入变量的相关性分解可以以允许容易地可视化和解释相关性分解的方式来重新分组。例如，为了将相关性可视化为像素域中的热图，可以针对每个像素对与其红、绿和蓝分量相关联的相关性进行求和，如关于图2c所解释的。类似地，对于文本分析，为了将文档的相关性分解可视化为热图文本，可以针对每个词对与对应向量的每个分量相关联的相关性进行求和。

[0328] 还可以评估其他示例。然而，由稳定函数h(•)(见等式A5*和A7*)应用的情况可导致相关性“泄漏”，使得例如可能不满足每个项目集合16的利用前述来自泛化2B的函数f、ξ和ζ描述的相关性属性。例如，对于项目集合，只能满足导致网络输出达到最大网络输出的75％。举例来讲，假设人工神经网络所执行的预测是某一图片是否示出“猫”，然后，当经受反向传播时，针对图像(针对该图像，网络输出处的预测导致关于图片示出猫的高于75％的值)的预测可能导致像素的相关性分数满足关于f的条件(对于所有或至少大于99％)，而其它图片可能不满足或者肯定不满足。

[0329] 从另一观点来看，应该有利地选择分布函数，从而导致“有意义的”反向传播的相关性分数。为此，除了或替代相关性守恒属性之外，分布函数还可以遵守一些“排序”属性。换句话说，即使不遵守上述讨论的相关性守恒属性，该分布函数也可能导致有意义的反向传播相关性分数。具体地，对于每个神经元j，产生多少相关性R_{ij}从相应的神经元j重新分布到上游邻居神经元i的分布函数可以是

[0330] $R_{ij}=q(i)\cdot m(\{R_{jk},k\text{是}j\text{的下游神经元}\})$

[0331] 其中$m(\mathbb{R}^K)$是其所有分量的单调递增函数(K是相应神经元j的下游邻居的数目)，并产生相应神经元j的初步重新分布的相关性值。

[0332] q(i)是满足取决于相应神经元j的上游邻居神经元i的激活x_i的排序属性的函数，其中I是上游邻居神经元i的数目，并且权重w_{ij}将上游邻居神经元i连接到相应的神经元j，并且，如果存在，相应神经元j的偏置项b_j则假定为零，如果不存在，其中排序属性是泛化4和泛化5中以及泛化4和泛化5之类给出的其中之一。

[0333] 还应当注意，图4同时揭示了相关性分数指派过程的图以及其中所示的元素(例如

10和52)表示在这种方法/过程期间执行的处理步骤，其中诸如30和38的步骤表示在该过程中附加执行的可选步骤或任务。或者，装置50可以被配置为附加地执行任务30和38或30。例如，所有这些任务可以表示实现所述过程或装置50所基于的计算机程序的代码的不同部分。

[0334] 此外，在下文中将使用一些不同的术语来描述以上描述，以避免对于本申请的范围的误解。

[0335] 具体地，以上描述揭示了对样本进行的预测分析，其中“样本”是项目的集合16。该预测是基于项目集合16导出网络输出的过程，并且通过将样本作为输入的映射来执行。对样本作为整体进行预测，并产生向量值或实值输出或可以被变换为向量值或实值的输出，即网络输出18。预测映射涉及通过神经网络的正向传播14。它可以按以下方式分解：它由元素12组成，元素12获取输入并通过对输入应用函数(即神经函数)来计算输出。至少一个元素12具有样本(即集合16)的一个项目作为输入。该模型是在不失一般性的情况下进行的，因此每个元素最多只取样本的一个项目作为输入。至少一个元素12将其它元件的输出作为输入。如上所述，这些可以通过乘以取决于元素12及其输入的值来加权。至少一个权重不为零。至少一个元素的输出用于进行样本的预测。在模型中存在从样本项目到预测的连接。

[0336] 换句话说，上述概述(分层)反向传播是在已经执行对该项目集合的预测的假设下执行的。该过程开始于所有这些元素的相关性的初始化，这些元素是通过预测(即基于网络输出)直接计算的。如果该输出是实值，则相关性R形成输出神经元，该输出神经元计算相应预测网络输出，通过使用模型的预测值来初始化。如果输出为向量值，则可以针对所有输出神经元设置相关性R，可以通过使用针对一个输出神经元情况的实值输出的情况描述的初始化，并且针对剩余输出神经元将相关性设置为零，来初始化相关性R。初始化后，以下两个公式以交替方式进行计算。

[0337] 具体地，对于已经计算相关性R_k的每个元素(神经元)k，计算用于向元素k提供输入的所有元素的消息$R_{i\leftarrow k}$，使得

[0338] $$R_k = \sum_{i:\ i\text{向}k\text{提供输入}} R_{i\leftarrow k} \qquad \text{(等式A1)}$$

[0339] $$R_i = \sum_{k:\ k\text{从}i\text{接收输入}} R_{i\leftarrow k} \qquad \text{(等式A2)}$$

[0340] 或者，可以仅使用等式A2并且仅隐含地计算消息$R_{i\leftarrow k}$，使得它们满足等式A1。

[0341] 在神经网络包含循环的情况下，即神经网络是复现的并且具有时间依赖状态，其结构可以在时间上展开，从而产生前馈映射，对于该前馈映射可以应用与上述相同的过程。通过在时间上展开，意味着得到在每个时间步长中对网络状态进行建模的一层。

[0342] 在计算输入元素i的相关性R_i之前，至少一个消息$R_{i\leftarrow k}$可以被随机值替代(即使可以计算消息$R_{i\leftarrow k}$，因为在某些步骤中已经计算了其计算所需的相关性R_k)。

[0343] 在计算输入元素i的相关性R_i之前，至少一个消息$R_{i\leftarrow k}$可以被恒定值替代(即使可以计算消息$R_{i\leftarrow k}$，因为在某些步骤中已经计算了其计算所需的相关性R_k)。

[0344] 在下文中，提供了关于按层相关性传播原理的更多技术观点。每层应指派一个索引。第一层索引为1，最后一层的索引最高。集合16中的每个项目的分数可以以下列方式计

算：

[0345] 假设已经对排序的项目集合进行了预测。

[0346] • 首先，如下所述初始化作为输出层的最后一层的相关性：

[0347] 如果输出是实值，则将最后一层中的单个元素的相关性初始化为模型的预测值。

[0348] 如果输出为向量值，则通过使用针对输出层中至少一个元素的实值输出的情况所描述的初始化，以及通过将剩余元素的相关性设置为零，来初始化最后一层中所有元素的相关性。

[0349] 其次，在从一层索引到上游层的层上执行迭代。

[0350] 迭代完成如下：

[0351] 给定当前层(索引为I+1)中所有元素的相关性$R_k^{(l+1)}$，计算从当前层(索引I+1)中的每个元素到上游层(索引I)中的所有元素的消息项$R_{i\leftarrow k}^{(l,l+1)}$，使得

[0352] $$R_i^{(l)} = \sum_{\text{(元素i的输入)}} R_{i\leftarrow k}^{(l,l+1)} \qquad \text{（等式A3）}$$

[0353] 持有近似误差。

[0354] 给定从层到其上游层的所有消息$R_{i\leftarrow j}^{(l,l+1)}$，通过下式计算上游层的相关性：

[0355] $$R_i^{(l)} = \sum_{\text{(元素i的输入)}} R_{i\leftarrow k}^{(l,l+1)} \qquad \text{（等式A4）}$$

[0356] 从这里，将针对下一个上游层I-1进行迭代，因为已经计算了层I处的所有相关性$R_i^{(l)}$。

[0357] 穿过所有层向下到达层1的迭代结果是在第一层中所有元素的相关性分数$R_d^{(1)}$，它们是排序集合中的项目的分数。

[0358] 该方法的结果是每个项目一个分数，其表示对于排序的项目集合进行的预测的项目的相关性，或者该结果是与以下的至少一个相结合的分数：

[0359] 这些分数到颜色上的映射，使得每个分数区间被映射到一个颜色上

[0360] 根据由每个项目的分数确定的顺序的项目的排序列表

[0361] 可以是

[0362] -如果函数在层I，那么将用字母i索引的元素的输出值表示为$x_i^{(l)}$

[0363] -从索引为i的一个元素到索引为j的另一个元素的连接可以具有w_{ij}，该权重与先前元素的输出相乘。因此，从层I中索引为i的元素对索引为j的元素的输入可以写为

[0364] $$z_{ij} = x_i^{(l)} w_{ij}$$

[0365] 偏置项可由不需要输入并提供恒定输出的元素表示。

[0366] 具体地，通过将以下一组公式中的至少一个应用到模型中的至少一个元素和该元素的输入集合来计算消息项$R_{i\leftarrow j}^{(l,l+1)}$：

[0367] 等式A5或A6或A7或A8(上文给出)

[0368] 可以通过将上述等式A1-A26中的至少一个应用到模型中的至少一个元素和该元素的输入集合来计算消息项$R_{i\leftarrow j}^{(l,l+1)}$。

[0369] 所述样本可以是排序的项目集合。下面将列出排序的项目集合的若干可能示例。

[0370] 排序的项目集合可以是图像，并且每个项目可以是其一个或多个像素的集合。

[0371] 排序的项目集合可以是文本，并且每个项目可以是其一个或多个词的集合。

[0372] 排序的项目集合可以是文本，并且每个项目可以是其一个或多个句子的集合。

[0373] 排序的项目集合可以是文本，并且每个项目可以是其一个或多个段落的集合。

[0374] 排序的项目集合可以是键值对的列表，并且每个项目可以是其一个或多个键值对的集合。

[0375] 排序的项目集合可以是财务数据或公司相关数据的键值对的列表，并且每个项目可以是一个或多个键值对的集合。

[0376] 排序的项目集合可以是视频，并且每个项目可以是具有时间戳的一对或多对像素的集合。

[0377] 排序的项目集合可以是视频，并且每个项目可以是一个或多个帧的集合。

[0378] 排序的项目集合可以是视频，并且每个项目可以是一个或多个像素的集合。

[0379] 可学习神经网络的技术规范

[0380] 以下段落描述了一种神经网络，其方式是使其大多数层在训练阶段被学习，这与其他类型的浅层学习算法是不同的。它可以具有以下属性

[0381] -如果模型在测试时间是两层的，则使用一组训练数据和取决于训练数据的子集的误差度量来优化第一层权重。

[0382] -如果模型在测试时间是三层或四层的，则使用一组训练数据和取决于训练数据的子集的误差度量来优化至少第一层或第二层权重。

[0383] -如果模型在测试时间是五层或更多层的，则使用一组训练数据和取决于训练数据的子集的误差度量来优化至少从第一层到最后的第三层中一层的权重。(这允许最后的层也被优化)

[0384] 层中的至少一个元素可以是修正的线性激活单元。

[0385] 层中的至少一个元素可以是Heaviside激活单元。

[0386] 层中的至少一个元素可以是tanh激活单元。

[0387] 层中的至少一个元素可以是物流激活单元。

[0388] 层中的至少一个元素可以是S形激活单元。

[0389] 实验

[0390] 在两个数据集合上显示结果，MNIST上的两组结果很容易解释，第二组实验依靠作为Caffe开源软件包[20]的一部分提供的15层已训练网络，其预测ILSVRC挑战的1000个类别。一方面，通过MNIST数字的实验，旨在表明能够发现具体到训练阶段的细节。另一方面，来自Caffe工具箱的预训练网络的结果表明，该方法与箱外的深度神经网络协同工作，并且在训练阶段期间不依赖可能的技巧。

[0391] 使用预训练的网络将参考分数指派应用于其他逼真图像。以相关性分数形式对分类决定的解释突出了类的有意义的特征，例如，“鲨鱼”的鲨鱼翅、“杯子”的圆形、“火山”的

山形等。请注意，相关性分数指派不突出图像中的所有梯度，而是突出区别性特征。例如，图9示出了将上述相关性分数指派应用于经过训练以从ImageNet数据集合中区分1000个类的神经网络：上部图像示出了对网络的输入，即集合16，并且下部图像示出了根据上述实施例的指示指派给像素的相关性分数的热图，每个输入图像一个热图。如上所述，热图可以覆盖在输入图像上。可以看出，在蛇(左侧图像)的情况下，表示壳的像素接收大部分初始相关性分数，即被确定为导致网络预测将图像归类为显示蛇的主要原因，在鲨鱼(左侧图像的第二个)的情况下，表示鳍的像素接收大部分的初始相关性分数，在山(从右侧图像数的第二个)的情况下，表示峰的像素接收大部分初始相关性分数，并且在火柴(左侧图像)的情况下，表示火柴和火的像素接收大部分初始相关性分数。

[0392] 还在MNIST数据集合上训练了神经网络。该数据集合包含从0到9的数字图像。在训练后，网络能够对新的不可见的图像进行归类。通过反向传播相关性分数指派，可以问问为什么网络将3的图像归类为“3”类，换句话说，是什么使3与其他数字不同。在图10的热图中可以看到，3(相对于其他数字)的最重要的特征是中间水平笔画和左侧没有垂直连接(对于数字8有)。也可以问问例如为什么4的图像不被归类为“9”，换句话说，当看到4的图像时，不会认为是9。可以看出，反对是“9”的证据是4的顶部的差距。请注意，使用箭头62表示的红色代表某一类的证据，60表示的蓝色表示反对该类的证据。总之，已经表明，该方法对分类决定提供了有意义的解释。

[0393] 应用

[0394] 到目前为止，描述集中于相关性分数指派过程。在下文中，将简要描述指派给集合16的项目的相关性分数可以用于什么。

[0395] 一般应用将是使用这里提出的相关性分数指派(RS指派)作为更大、更复杂的算法(CA)的一部分。可以想到应用算法CA是非常昂贵的情况，所以RS指派可以定义可以应用算法CA的一些兴趣区域。例如，

[0396] -医生的时间是宝贵的。RS指派可以在筛选癌症时识别图像中的重要区域。

[0397] -在视频编码中，通道带宽是宝贵的。RS指派可以通知算法CA关于视频的哪些部分比其他部分更重要，以例如确定更好的编码策略(例如，针对重要部分使用更多比特)或更好的传输调度(例如首先传输重要信息)。

[0398] -热图可用于计算某些预测任务的附加特征。例如，可以使用经过训练的网络，将其应用于某些图像，并从更重要的区域中提取更多的特征。这可能导致计算时间或信息传输的减少。或者，从其提取的区域或附加信息可以用于重新训练和改进经过训练的网络。

[0399] -在用户或公司想要知道哪些区域或特征对某个任务很重要的情况下，RS指派可以用作调查工具。

[0400] 此外，在图像应用领域，

[0401] -RS指派可用于医疗应用，例如，帮助医生识别病理图像中的肿瘤或识别MRI图像中的观察值。更具体的示例包括：

[0402] -检测生物组织图像中的炎症征象

[0403] -检测生物组织图像中的癌症征象

[0404] -检测生物组织图像的病理变化

[0405] -RS指派可以应用于一般图像。例如，社交网站平台或搜索引擎具有许多图像，并

且可能感兴趣的是什么使图像“好笑”、“不寻常”、“有趣”或什么使人、房屋的图像或房屋的内部有吸引力/美观或少吸引力/不太美观。

[0406] -RS指派可用于监视应用中，以检测图像的哪个部分触发系统检测异常事件。

[0407] -检测卫星、飞机拍摄的图像或遥测数据中的土地利用变化。

[0408] 在视频应用领域，

[0409] -热图可用于设置编码的压缩强度，例如，对于包含重要信息的区域使用更多比特，对于其他区域使用较少比特。

[0410] -RS指派可用于视频摘要，即以识别视频中的“相关”帧。这将允许智能视频浏览。

[0411] -动画电影有时看起来不太现实。不清楚“丢失”什么，使电影看起来更逼真。在这种情况下可以使用热图来突出视频的不太逼真的部分。

[0412] 在文本应用的情况下，

[0413] -将文本文档按类别归类可以由DNN或BoW模型执行。RS指派可以可视化文档被归类到特定类中的原因。可以突出或选择主题文本的相关性以便进一步处理。RS指派可以突出重要的词，从而提供长文本的摘要。这样的系统可用于例如专利律师快速浏览许多文本文档。

[0414] 在财务数据应用的情况下，

[0415] 银行使用诸如(深度)神经网络的分类器来确定某人是否获得信用贷款(例如德国Schufa系统)。这些算法如何工作是不透明的，例如一些没有得到贷款的人不知道原因。RS指派可以准确地显示为什么有人不能获得贷款。

[0416] 在营销/销售领域，

[0417] -RS指派可用于确定什么使特定产品描述图像/文本销售产品(例如，公寓租赁、ebay产品)。

[0418] -RS指派可用于确定什么使某个在线视频/博客文章被高度评价或喜欢

[0419] -公司可能普遍感兴趣的是什么“特征”使例如他们的网站或产品有吸引力

[0420] -公司感兴趣的是某些用户购买某产品而其它用户不购买的原因。RS指派可用于识别用户不购买产品的原因，并相应地改进广告策略。

[0421] 在语言学/教育领域

[0422] -RS指派可用于确定哪一部分文本将特定语言(诸如英语、法语、西班牙语或德语)的母语与非母语讲话者区分开来。

[0423] -Rs指派可用于在文本中查找文档是否由特定人员撰写的证明元素。

[0424] 在上面的描述中，已经提供了不同的实施例用于将相关性分数指派给项目集合。例如，已经提供了关于图片的示例。关于后面的示例，已经提供了关于相关性分数的使用的实施例，即为了使用相关性分数(即通过使用可以与原始图片重叠的热图)突出图片中的相关部分。在下文中，呈现使用或利用相关性分数的实施例，即使用上述相关性分数指派作为基础的实施例。

[0425] 图11示出了一种用于处理项目集合的系统。通常使用附图标记100来指示该系统。除了装置50之外，该系统还包括处理装置102。两者都在集合16上操作。处理装置102被配置为处理该项目集合，即集合16，以获得处理结果104。这样做，处理装置102被配置为根据由相关性分数指派器50指派给集合16的项目的相关性分数R_i来适配其处理。装置50和装置

102可以使用在一个或多个计算机上运行的软件来实现。它们可以在单独的计算机程序或一个通用计算机程序上实现。关于集合16，上述所有示例都是有效的。例如，假设处理装置102执行诸如数据压缩的有损处理。例如，由装置102执行的数据压缩可以包括无关性降低。例如，集合16可以表示诸如图像或视频的图像数据，并且由装置102执行的处理可以是有损性质的压缩，即该装置可以是编码器。在这种情况下，例如，与指派有较低相关性分数的项目相比，装置102可以被配置为降低对指派有较高相关性分数的项目的处理的损失。例如，可以通过量化步长或通过改变编码器的速率控制的可用比特率来改变损失。例如使用较高的比特率、使用较低的量化步长等，相关性分数高的样本区域的编码损失可以较小。因此，例如，相关性分数指派关于视频场景中对嫌疑人的检测/预测执行其相关性分数指派。在这种情况下，处理装置102能够针对感兴趣的场景(即由于在其中已“检测到”嫌疑人而感兴趣的时空部分)而在有损压缩视频(根据该示例该视频表示集合16)中花费更多的数据速率。或者处理装置102使用相同的数据速率，但是由于相关性分数实现的加权，对于具有高相关性分数的样本的项目的压缩较低，而对于具有低相关性分数的样本的项目的压缩较高。在这种情况下，处理结果104是有损压缩数据或数据流，即视频16的压缩版本。然而，如前所述，集合16不限于视频数据。它可以是图片或音频流等。

[0426] 为了完整起见，图12示出了图11的系统的变型。这里，相关性分数指派50对集合16进行操作，以便导出集合16的项目的相关性分数R_i，但是处理装置102对待处理的不等于集合16的数据106进行操作。相反，集合16已经从数据106导出。例如，图12示出了图1的示例性情况，根据该情况，通过特征提取过程30已经从数据106导出集合16。因此，集16“描述”数据106。以上述方式，相关性值R_i可以经由反向映射过程38与原始数据106相关联，反向映射过程38表示针对特征提取过程30的反转或反向映射。因此，处理装置102对数据106进行操作，并根据相关性分数R_i适配或精简其处理。

[0427] 由图11和图12中的处理装置102执行的处理不限于有损处理，诸如有损压缩等。例如，在集合16或数据106的上述许多示例中，集合16的项目形成以1、2或更多维度排列的项目的排序集合。例如，像素以至少二维排序，即x和y是两个横向维度，并且在包含时间轴时是三维的。在音频信号的情况下，诸如时域(例如PCM)样本或MDCT系数的样本沿着时间轴排序。然而，集合16的项目也可以在频域中排序。也就是说，集合16的项目可以表示例如图片、视频或音频信号的频谱分解的系数。在这种情况下，过程30和反向过程38可以分别表示正向变换或反向变换的谱分解。在所有这些情况下，由相关性分数指派器50获得的相关性分数R_i也被排序，即它们形成相关性分数的排序集合，或者换句话说，形成“相关性图”，其可以与集合16重叠，或者经由处理38与数据106重叠。因此，处理装置102可以例如使用集合16的项目中的顺序或数据106的样本的顺序来执行数据106的集合16的可视化，并且使用相关性图来突出可视化的相关部分。例如，处理结果104将是屏幕上图像的呈现，并且使用相关性图装置102利用例如闪烁、颜色反转等突出屏幕上的某些部分，以便分别指示集合16或数据106中增加了相关性的一部分。例如，这样的系统100可以用于视频监视的目的，以便例如将保安人员的注意力吸引到由数据106或集合16(即例如视频或图片)表示的场景的某一部分上。

[0428] 或者，由装置102执行的处理可以表示数据补充。例如，数据补充可以指代从存储器读取。作为另一选择，数据补充可以涉及另外的度量。假设，例如，该集合16再次是排序的

集合，即是属于图片106的特征图，是图片本身或视频。在这种情况下，处理装置102可以相关性分数R_i导出ROI(即感兴趣区域)的信息，并且可以将数据补充集中到该ROI上，以避免对集合16所指代的完整场景执行数据补充。例如，可以由装置50对低分辨率显微镜图片执行第一相关性分数指派，然后装置102可以对低分辨率显微镜图片中、相关性分数指示高相关性的局部部分进行另一显微镜测量。因此，处理结果104将是数据补充，即以高分辨率显微镜图片形式的另一测量。

[0429] 因此，在使用图11或12的系统来控制数据速率支出的情况下，系统100产生了有效的压缩构思。在使用系统100进行可视化处理的情况下，系统100能够增加观看者意识到某些感兴趣区域的可能性。在使用系统100来精简数据补充的情况下，系统100能够通过避免针对不感兴趣的区域执行数据补充来避免数据补充的量。

[0430] 图13示出了用于突出项目集合的感兴趣区域的系统110。也就是说，在图13的情况下，再次将该项目集合假定为诸如特征图、图片、视频、音频信号等的排序集合。除了图形生成器112之外，系统110还包括相关性分数指派器50，所述图形生成器根据由相关性分数指派器50提供的相关性分数R_i生成相关性图形。如上所述，相关性图形114可以是使用颜色以便“度量”相关性R_i的热图。如上所述，相关性分数R_i是标量，或可以通过将共同归属的映射相关性分数(例如属于图像的一个彩色像素的不同颜色分量的子像素的相关性分数)相加而成为标量。然后，例如，使用例如单个像素的一维标量相关性分数作为CCT值，可以将标量相关性分数R_i映射到灰度或颜色上。然而，可以使用从一维到三维颜色空间(如RGB)的任何映射来生成彩色图。例如，将分数映射到色调区间，修复饱和度和值维度，然后将HSV表示形式变换为RGB表示形式。

[0431] 然而，相关性图形114可以备选地以直方图等形式来表示。图形生成器112可以包括用于显示相关性图形114的显示器。除此之外，图形生成器112可以使用诸如计算机程序之类的软件来实现，该计算机程序可以与实现相关性分数指派器50的计算机程序分离或包括在其中。

[0432] 作为具体的示例，假设项目的集合16是图像。根据指派器获得的每个像素的像素相关性分数可以被离散化/量化成一组值，并且离散化/量化索引可以被映射到一组颜色上。映射可以在图形生成器112中完成。在颜色的一些CCT(颜色温度)度量之后的相关性-颜色映射的情况下，将像素最终指派给颜色(例如“热图”)可以被保存为数据库中或存储介质上的图像文件或者由生成器112呈现给观看者。

[0433] 或者，将像素指派给颜色可以与原始图像重叠。在这种情况下，图11和12的处理器102可以用作图形生成器。所得到的重叠图像可以作为图像文件保存在介质上或呈现给观看者。“重叠”可以例如通过将原始图像变换成灰度图像来完成，并且用于将像素相关性分数映射到颜色值以映射到色调值。可以由处理器102通过使用色相饱和度值表示形式来创建重叠图像，即所述值(然而，由于几乎黑色的像素没有清晰可见的颜色而具有太小的值的上限，并且还可能从原始图像获得饱和度)从原始图像的灰度版本的相应样本的灰度值获得，并且从颜色图中获取色调值。处理器102可以对如刚才概述生成的图像(例如，颜色图或重叠或相关性分数的排序集合(可以表示为图像，但不是要求))进行分段。这种分段图像中对应于具有非常高的分数的区域、或具有绝对值大的分数的区域的那些分段可以被提取、存储在数据库或存储介质中，并且用作分类器训练程序的附加训练数据。如果项目的集合

16是文本，则相关性指派的结果可以是如上所述的每个词或每个句子的相关性分数。然后可以将相关性分数离散化为一组值并映射到一组颜色上。然后，处理器102可以用颜色来标记词，所得到的颜色突出的文本可以被保存在数据库中或存储介质上或呈现给人。备选地或附加地突出这些词，处理器102仅仅选择文本的词、句子部分或句子的子集，即具有最高分数或最高的绝对值分数(例如，通过对分数或其绝对值设置阈值)的那些部分，并将此选择保存在数据库中或存储介质上，或将其呈现给人。如果将相关性指派应用于数据集合16，使得样本由存储在数据库中的表中的一组键值对(例如关于公司的财务数据)组成，则每个样本的结果将是每个键值对的相关性分数。对于给定的样本，随后可以选择具有最高分数或最高绝对值分数(例如通过对分数或其绝对值设置阈值)的键值对的子集，并且可以将该选择保存在数据库中或存储介质上或将其呈现给人。这可以由处理器102或生成器112完成。

[0434] 上面已经关于图12概述了数据集合16可以是图像或视频。然后可以使用像素相关性分数来找到具有高分数的区域。为此，可以示例性地使用上述分段或视频分段。在视频的情况下，高分数区域将是时空子集或视频的部分。对于每个区域，例如通过计算区域像素的像素分数的分位数或p均值$M_p(x_1,\ldots,x_N)=\left(\frac{1}{N}\sum_{i=1}^{N}x_i^p\right)^{\frac{1}{p}}$，可以计算每个区域的分数。数据集合(例如视频)然后将受到处理器102的压缩算法的影响，根据所计算的分数，可对该区域调整压缩率。可以使用区域分数与压缩率的单调(下降或上升)映射。然后根据区域分数与压缩率的映射对每个区域进行编码。

[0435] 此外，在图像作为集合16的情况下，处理器102可以如下操作：刚刚概述的分段可以应用于所有像素或重叠图像或颜色图的分数集合，并且对应于具有非常高的分数的区域或具有大绝对值的分数的区域可以提取分段。然后，处理器可以将原始图像16的这些共同定位的分段呈现给人或其他算法，用于检查内容是否存在显眼或异常内容的可能性。这可以用于例如安全防护应用。同样，集合16可以是视频。整个视频又由一组帧组成。如上所述，项目集合16中的项目可以是帧或帧的子集或来自帧的子集的一组区域。时空视频分段可以应用于项目的相关性分数指派，以便找到具有高平均分数的项目或高平均绝对值分数的项目的时空区域。如上所述，指派给区域内的项目的平均分数可以例如使用p均值或分位数估计器来度量。可以由处理器102(例如通过图像或视频分段)提取具有最高分数(诸如高于某个阈值的分数)的时空区域，并呈现给人或其他算法以便检查内容是否存在显眼或异常内容的可能性。用于检查的算法可以被包括在处理器102中，或者可以在其外部，对于上面提及的检查(最)高分数区域的情况也是如此。

[0436] 根据实施例，使用刚刚提到的具有这种最高分数的时空区域来实现改善对视频进行预测的训练目的。如上所述，项目集合16是可以由一组帧表示的整个视频。该项目集合中的项目是帧或帧的子集或来自帧的子集的一组区域。然后应用视频分段来找到具有高平均分数项目或高平均绝对值分数项目的时空区域。处理器102可以选择连接到其他神经元的神经网络的神经元，使得经由间接连接，以上区域成为所选神经元的输入的一部分。处理器102可以以下列方式优化神经网络：给定输入图像和如上所选择的神经元(例如通过获得来自具有高相关性分数或其高绝对值的区域的直接或间接输入)，处理器102试图通过改变所选择的神经元的输入的权重、以及作为所选神经元的直接或间接的上游邻居的那些神经元

的权重，来增加网络输出或网络输出的平方，或减少网络输出。这样的改变可以例如通过相对于要改变的权重计算给定图像的神经元输出的梯度来完成。然后通过梯度乘以步长常数来更新权重。不用说，时空区域也可以通过像素分数的分段(即通过使用像素作为集合16的项目)来获得，然后执行上面概述的优化。

[0437] 甚至备选地，相关性指派可以应用于由节点、以及具有或不具有权重的有向或无向边缘组成的图形数据；例如，集合16的项目将因此是子图。将针对每个子图计算元素相关性分数。例如，如果通过将节点及其边缘以整数权重进行编码、同时用预留为停止标志的整数分隔语义单元、来将子图作为整体来编码的情况下，可以将子图作为神经网络的输入。或者，用于计算每个项目的相关性分数的集合16的项目可以是节点。然后计算项目相关性分数。之后，可以通过图形分段来找到具有高平均分数的子图集合(该平均分数可以通过节点上分数的分位数或p均值$M_p(x_1,\ldots,x_N)=\left(\frac{1}{N}\sum_{i=1}^{N}x_i^p\right)^{\frac{1}{p}}$来计算)。将每个节点的分数离散化为一组值，并将离散化索引映射到一组颜色上。节点和子图对颜色的最终指派和/或提取的子图可以作为文件保存在数据库中或存储介质上或呈现给观看者。

[0438] 图14示出了用于优化神经网络的系统。该系统通常使用附图标记120表示，并且包括相关性分数指派器50、应用装置122以及检测和优化装置124。应用装置122被配置为将装置50应用到多个不同的项目集合16上。因此，对于每个应用，装置50确定集合16的项目的相关性分数。然而，这一次，装置50还在反向传播期间输出指派给神经网络10的各个中间神经元12的相关性值，从而获得用于每个应用的上述相关性路径34。换句话说，针对装置50在各个集合16上的每个应用，检测和优化装置124获得神经网络10的相关传播图126。在装置50在不同集合16上应用期间，装置124通过累积130或叠加指派给网络10的中间神经元12的相关性，来检测神经网络10内相关性增加的部分128。换句话说，装置124通过叠加来叠加或累积不同的相关性传播图126，以获得神经网络10的部分128，其包括在装置50的反向传播过程中对集合16的项目传播高百分比的相关性的那些神经元。然后可以由装置124使用该信息，以优化132人工神经网络10。具体地，例如，人工神经网络10的神经元12的一些互连可以被忽略，以便使人工神经网络10更小而不损害其预测能力。然而，还有其他可能性。

[0439] 此外，可以是相关性分数指派过程给出热图，并且针对例如平滑性和其它特性对热图进行分析。根据分析，可以触发一些动作。例如，可以停止神经网络的训练，因为它根据热图分析捕获到概念“足够好”。此外，应当注意，热图分析结果可以与神经网络预测结果(即所述预测)一起使用来做某事。具体地，依赖于热图和预测结果两者可能优于仅依赖于预测结果，这仅仅是因为例如，热图可以告知关于预测的确定性的一些事。可以通过分析热图来潜在地评估神经网络的质量。

[0440] 最后强调的是，所提出的相关性传播主要关于在分类任务上训练的网络在上面进行了说明，但是在不失一般性的情况下，上述实施例可以应用于指派认为是输出类别的分数的任何网络。可以使用其他技术(如回归或排名)来学习这些分数。

[0441] 因此，在上述描述中，已经提出了体现可以被称为按层相关性传播从而允许了解神经网络预测器的方法的实施例。证明了这种新颖原理的不同应用。对于图像，已经示出了像素贡献可以被可视化为热图，并且可以被提供给不仅可以直观地验证分类决定的有效性、而且还可以将进一步的分析集中在潜在兴趣的区域上的人类专家。该原理可以应用于

各种任务、分类器和数据类型，即不限于图像，如上所述。

[0442] 虽然已经在装置的上下文中描述了一些方面，但是将清楚的是，这些方面还表示对应方法的描述，其中，块或设备对应于方法步骤或方法步骤的特征。类似地，在方法步骤的上下文中描述的方面也表示对相应块或项或者相应装置的特征的描述。可以由(或使用)硬件装置(诸如，微处理器、可编程计算机或电子电路)来执行一些或全部方法步骤。在一些实施例中，可以由这种装置来执行最重要方法步骤中的某一个或多个方法步骤。

[0443] 取决于某些实现要求，可以在硬件中或在软件中实现本发明的实施例。可以使用其上存储有电子可读控制信号的数字存储介质(例如，软盘、DVD、蓝光、CD、ROM、PROM、EPROM、EEPROM或闪存)来执行实现，该电子可读控制信号与可编程计算机系统协作(或者能够与之协作)从而执行相应方法。因此，数字存储介质可以是计算机可读的。

[0444] 根据本发明的一些实施例包括具有电子可读控制信号的数据载体，该电子可读控制信号能够与可编程计算机系统协作从而执行本文所述的方法之一。

[0445] 通常，本发明的实施例可以实现为具有程序代码的计算机程序产品，程序代码可操作以在计算机程序产品在计算机上运行时执行方法之一。程序代码可以例如存储在机器可读载体上。

[0446] 其他实施例包括存储在机器可读载体上的计算机程序，该计算机程序用于执行本文所述的方法之一。

[0447] 换言之，本发明方法的实施例因此是具有程序代码的计算机程序，该程序代码用于在计算机程序在计算机上运行时执行本文所述的方法之一。

[0448] 因此，本发明方法的另一实施例是包括其上记录有计算机程序的数据载体(或者数字存储介质或计算机可读介质)，该计算机程序用于执行本文所述的方法之一。数据载体、数字存储介质或记录介质通常是有形的和/或非瞬时性的。

[0449] 因此，本发明方法的另一实施例是表示计算机程序的数据流或信号序列，所述计算机程序用于执行本文所述的方法之一。数据流或信号序列可以例如被配置为经由数据通信连接(例如，经由互联网)传送。

[0450] 另一实施例包括处理装置，例如，配置为或适用于执行本文所述的方法之一的计算机或可编程逻辑器件。

[0451] 另一实施例包括其上安装有计算机程序的计算机，该计算机程序用于执行本文所述的方法之一。

[0452] 根据本发明的另一实施例包括被配置为向接收机(例如，以电子方式或以光学方式)传输计算机程序的装置或系统，该计算机程序用于执行本文所述的方法之一。接收机可以是例如计算机、移动设备、存储设备等。装置或系统可以例如包括用于向接收机传送计算机程序的文件服务器。

[0453] 在一些实施例中，可编程逻辑器件(例如，现场可编程门阵列)可以用于执行本文所述的方法的功能中的一些或全部。在一些实施例中，现场可编程门阵列可以与微处理器协作以执行本文所述的方法之一。通常，方法优选地由任意硬件装置来执行。

[0454] 本文描述的装置可以使用硬件装置，或使用计算机，或者使用硬件装置和计算机的组合来实现。

[0455] 本文描述的方法可以使用硬件装置，或使用计算机，或者使用硬件装置和计算机

的组合来执行。

[0456] 上述实施例对于本发明的原理仅是说明性的。应当理解的是：本文所述的布置和细节的修改和变形对于本领域其他技术人员将是显而易见的。因此，旨在仅由所附专利权利要求的范围来限制而不是由借助对本文的实施例的描述和解释所给出的具体细节来限制。

[0457] 参考列表

[0458] [6]Christopher M Bishop et al.Pattern recognition and machine learning,volume 1.springer New York,2006.

[0459] [10]Hendrik Dahlkamp,Adrian Kaehler,David Stavens,Sebastian Thrun,and Gary R.Bradski.Self-supervised monocular road detection in desert terrain.In Robotics:Science and Systems,2006.

[0460] [11]Jia Deng,Alex Berg,Sanjeev Satheesh,Hao Su,Aditya Khosla,and Fei-Fei Li.The ImageNet Large Scale Visual Recognition Challenge 2012 (ILSVRC2012).

[0461] http://www.image-net.org/challenges/LSVRC/2012/.

[0462] [12]Dumitru Erhan,Yoshua Bengio,Aaron Courville,and Pascal Vincent.Visualizing higher-layer features of a deep network.Technical Report 1341,University of Montreal,June 2009.

[0463] [15]L.Fei-Fei and P.Perona.Abayesian hierarchical model for learning natural scene categories.In Computer Vision and Pattern Recognition,2005.CVPR 2005.IEEE Computer Society Conference on,volume 2,pages 524-531vol.2,2005.

[0464] [16]Muriel Gevrey,Ioannis Dimopoulos,and Sovan Lek.Review and comparison of methods to study the contribution of variables in artificial neural network models.Ecological Modelling,160(3):249-264,2003.

[0465] [17]Ronny Hänsch and Olaf Hellwich.Object recognition from polarimetric SAR images.In Uwe Soergel,editor,Radar Remote Sensing of Urban Areas,volume 15of Remote Sensing and Digital Image Processing,pages 109-131.Springer Netherlands,2010.

[0466] [20]Yangqing Jia.Caffe:An open source convolutional architecture for fast feature embedding.http://caffe.berkeleyvision.org/,2013.

[0467] [23]Alex Krizhevsky,Ilya Sutskever,and Geoffrey E.Hinton.Imagenet classification with deep convolutional neural networks.In Peter L.Bartlett,Fernando C.N.Pereira,Christopher J.C.Burges,Léon Bottou,and Kilian Q.Weinberger,editors,NIPS,pages 1106-1114,2012.

[0468] [25]Yann LeCun and Corinna Cortes.The MNIST database of handwritten digits.http://yann.lecun.com/exdb/mnist/,1998.

[0469] [26]Yann LeCun,Koray Kavukcuoglu,and Clément Farabet.Convolutional networks and applications in vision.In ISCAS,pages 253-256.IEEE,2010.

[0470] [27]Quoc V.Le.Building high-level features using large scale

unsupervised learning.In ICASSP,pages 8595-8598,2013.

[0471] [31]Grégoire Montavon,Geneviève B.Orr,and Klaus-Robert Müller, editors.Neural Networks:Tricks of the Trade,Reloaded,volume 7700 of Lecture Notes in Computer Science(LNCS).Springer,2nd edn edition,2012.

[0472] [34]Julian D Olden,Michael K Joy,and Russell G Death.An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data.Ecological Modelling,178(3-4):389-397, 2004.

[0473] [36]Nicolas Pinto,David D Cox,and James J DiCarlo.Why is real-world visual object recognition hard?PLoS Comput Biol,4(1):27,1 2008.

[0474] [39]David E.Rumelhart,Geoffrey E.Hinton,and Ronald J.Williams.Learning representations by back-propagating errors.Nature,323: 533-536,Oct 1986.

[0475] [41]Rudy Setiono and Huan Liu.Understanding neural networks via rule extraction.In IJCAI,pages 480-487.Morgan Kaufmann,1995.

[0476] [42]Karen Simonyan,Andrea Vedaldi,and Andrew Zisserman.Deep inside convolutional networks:Visualising image classification models and saliency maps.CoRR,abs/1312.6034,2013.

[0477] [43]Christian Szegedy,Wojciech Zaremba,Ilya Sutskever,Joan Bruna, Dumitru Erhan,Ian J.Goodfellow,and Rob Fergus.Intriguing properties of neural networks.CoRR,abs/1312.6199,2013.

[0478] [49]Paul A.Viola and Michael J.Jones.Rapid object detection using a boosted cascade of simple features.In CVPR(1),pages 511-518,2001.

[0479] [50]Ross Walker,Paul Jackway,Brian Lovell,and Dennis Longstaff.Classification of cervical cell nuclei using morphological segmentation and textural feature extraction.In Australian New Zealand Conference on Intelligent Information Systems,1994.

[0480] [54]Matthew D.Zeiler and Rob Fergus.Visualizing and understanding convolutional networks.CoRR,abs/1311.2901,2013.

[0481] [55]Matthew D.Zeiler,Graham W.Taylor,and Rob Fergus.Adaptive deconvolutional networks for mid and high level feature learning.In ICCV, pages 2018-2025,2011.

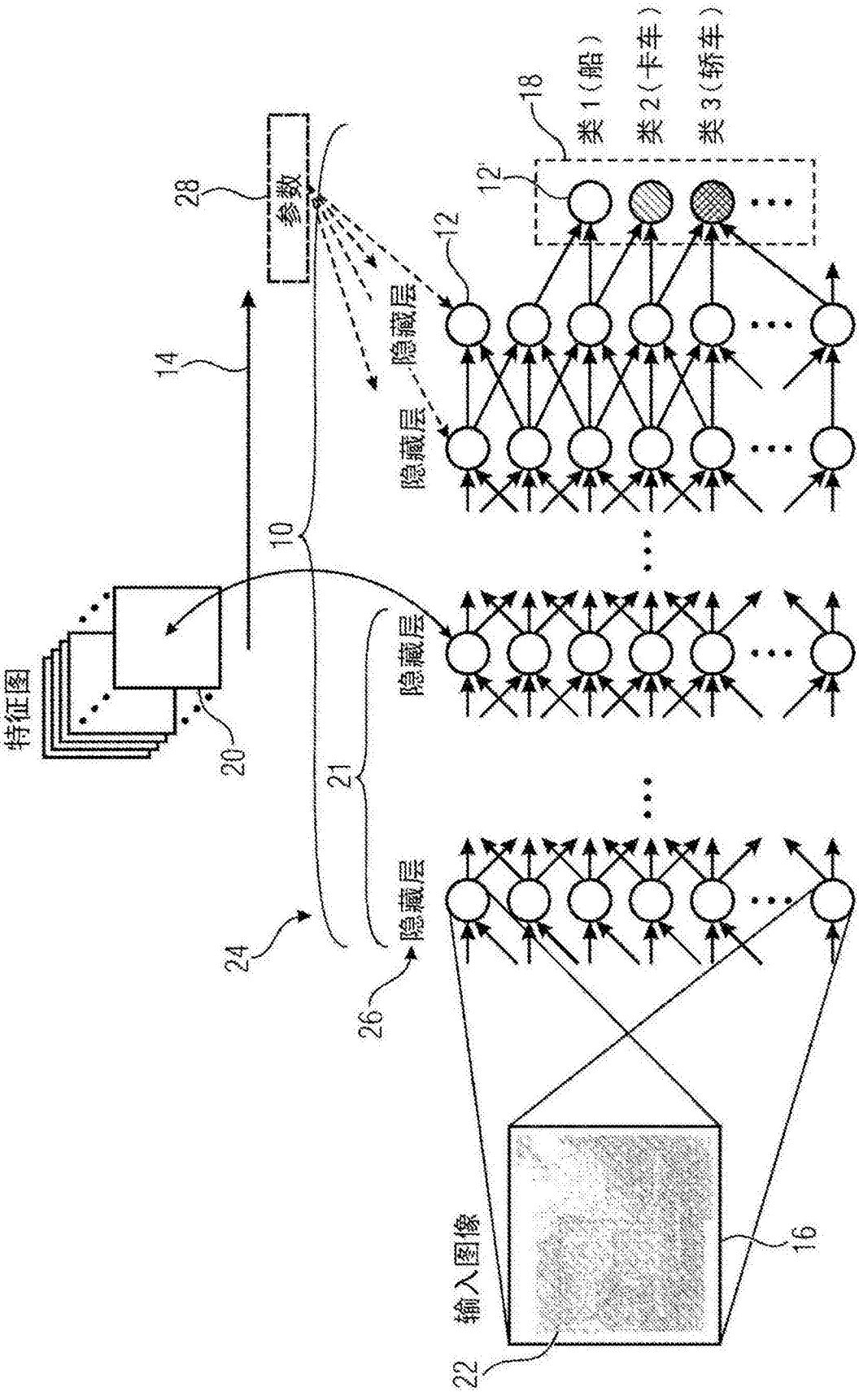

类 1（船）
类 2（卡车）
类 3（轿车）
18
12'
12
28
参数
隐藏层
隐藏层
14
10
特征图
20
隐藏层
21
24
26
隐藏层
输入图像
16
22

图1A

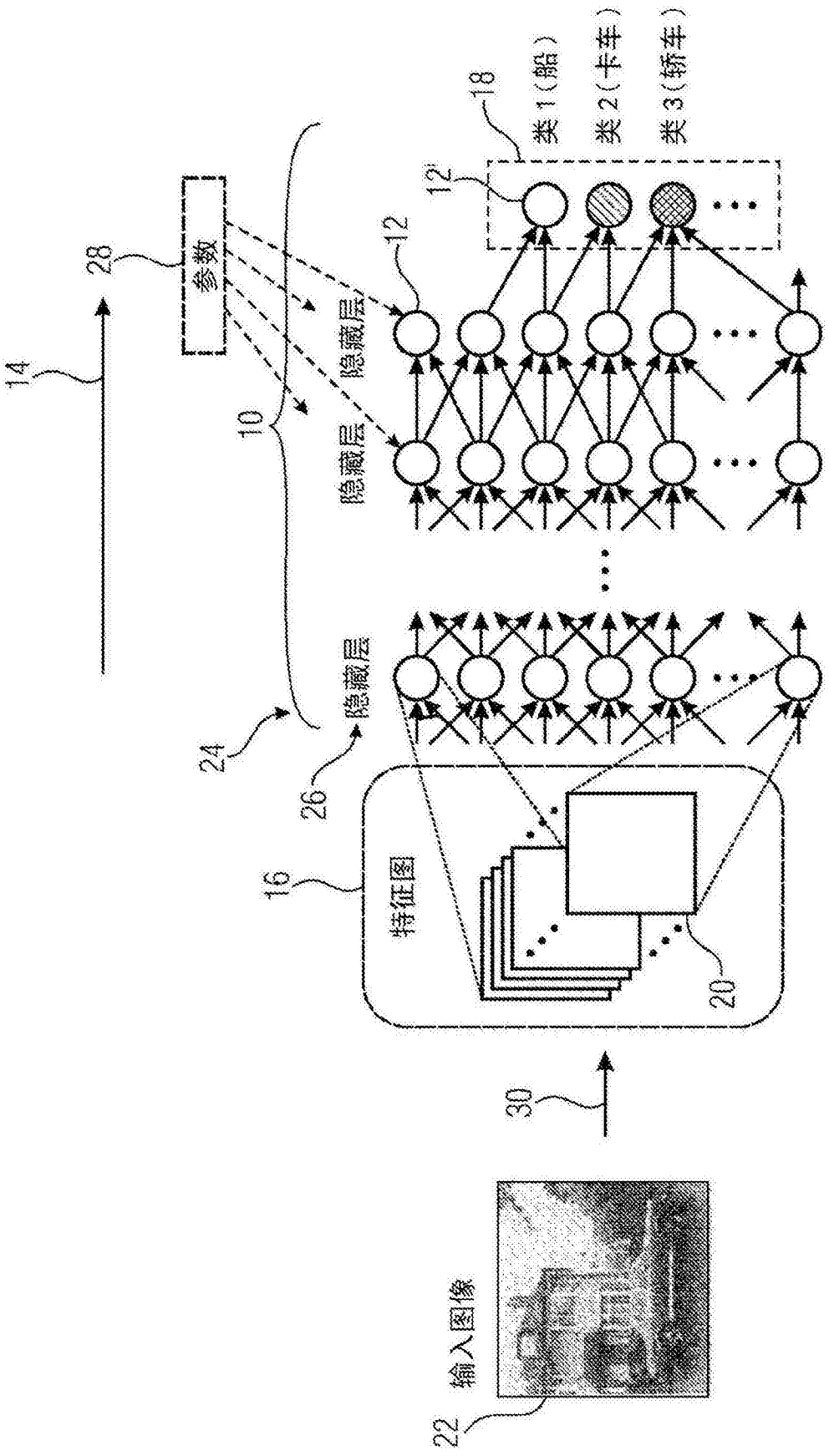

类1（船）
类2（卡车）
类3（轿车）
18
12'
12
28
参数
14
10
隐藏层
隐藏层
隐藏层
24
26
16
特征图
20
30
输入图像
22

图1B

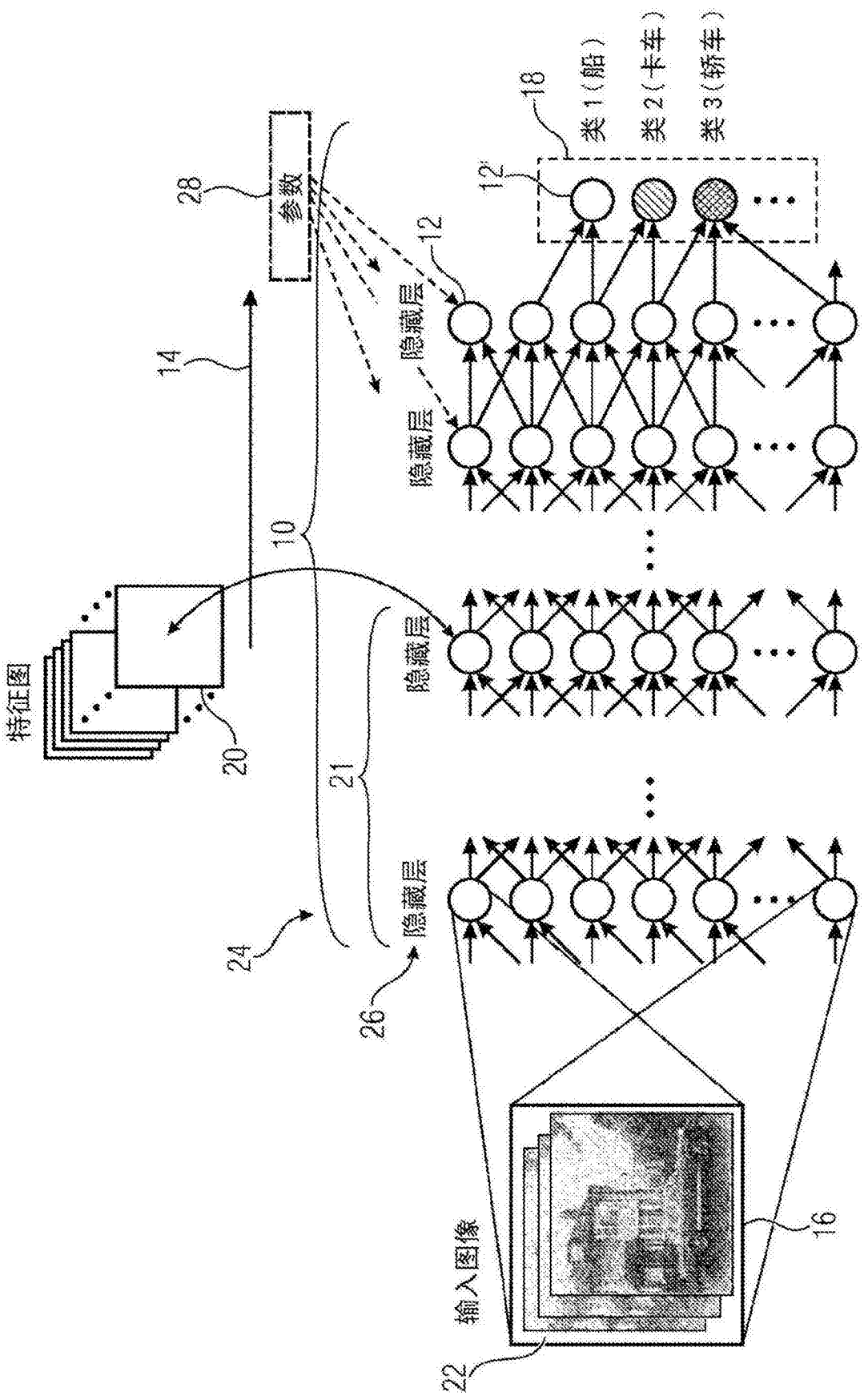

类1（船）
类2（卡车）
类3（轿车）
18
12'
12
28
参数
隐藏层
隐藏层
14
10
特征图
20
隐藏层
21
隐藏层
24
26
输入图像
22
16

图1C

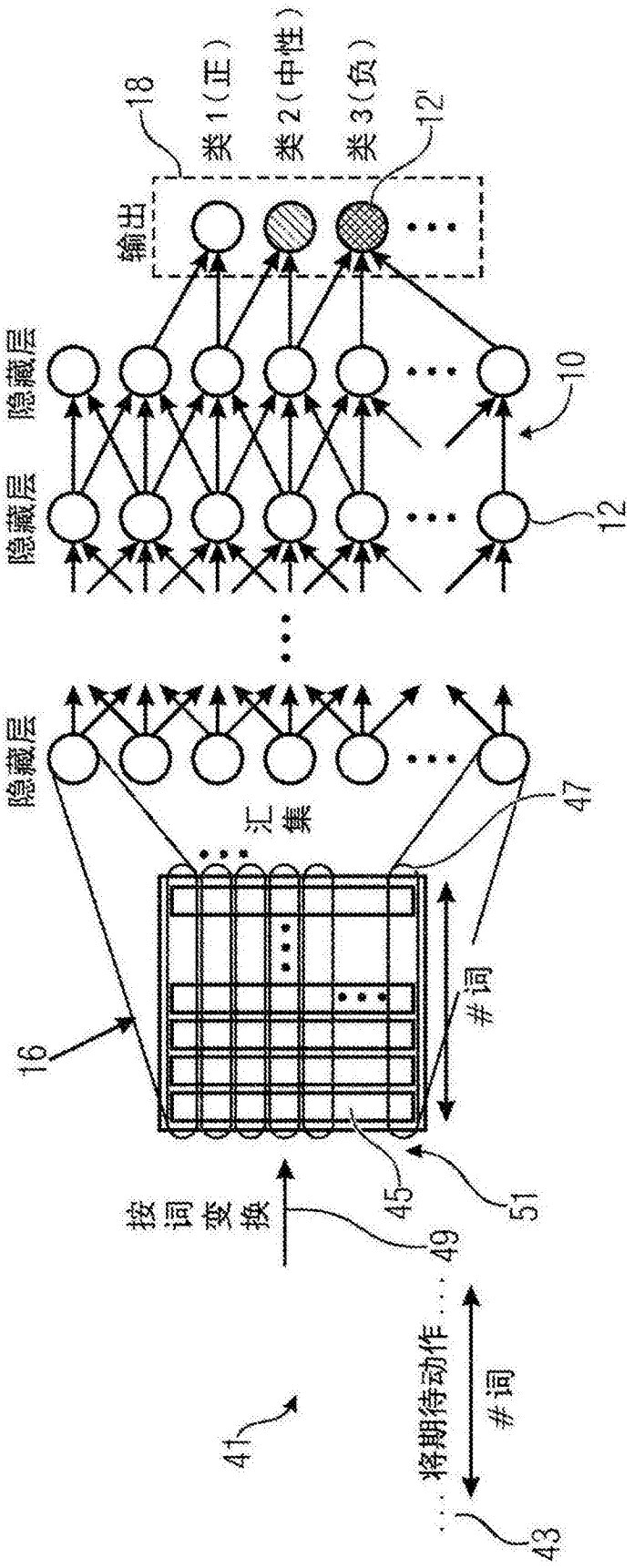

18
类1(正)
类2(中性)
类3(负)
12'
输出
隐藏层
隐藏层
10
12
隐藏层
汇集
47
16
#词
45
51
按词取样
49
将期待动作
#词
41
43

图1D

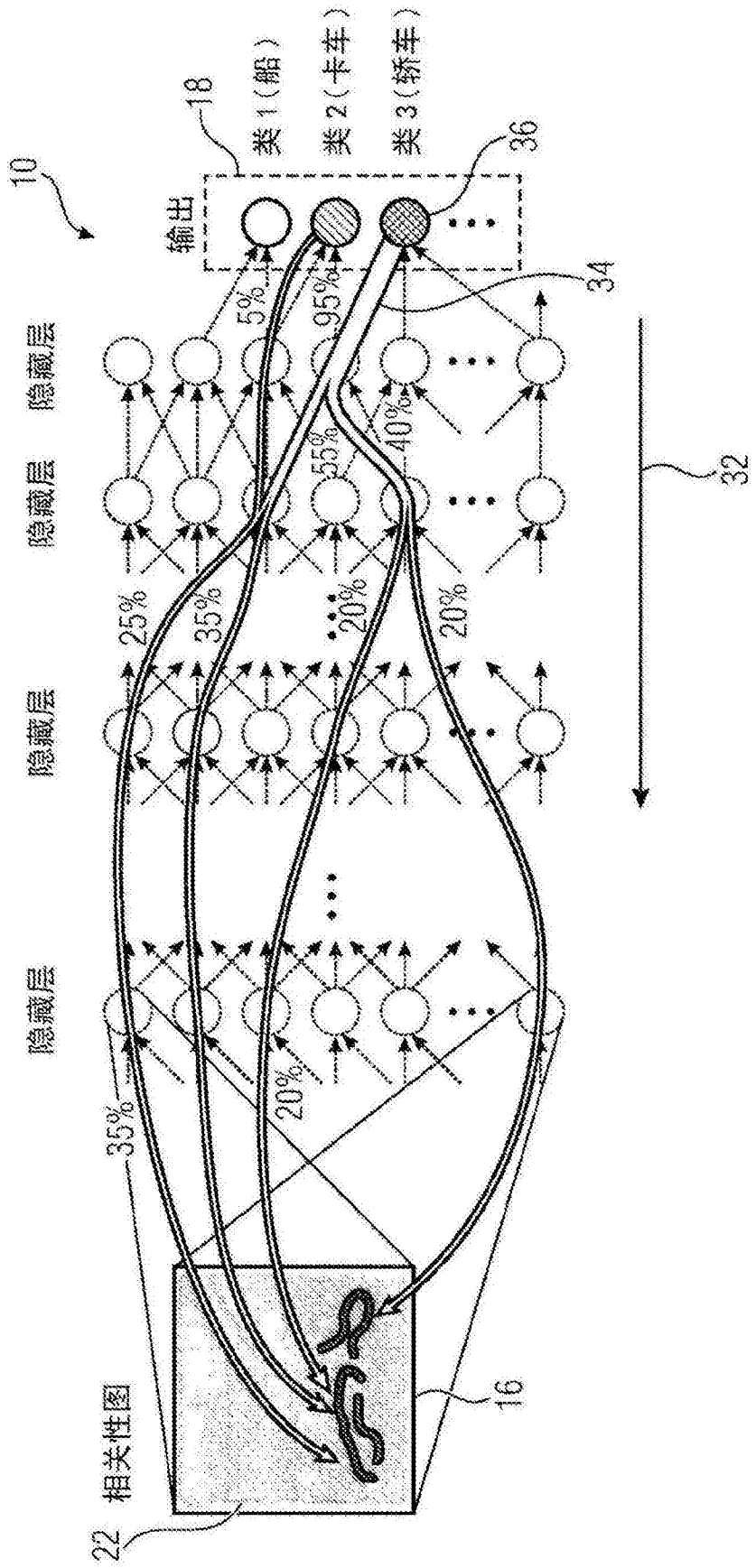

10
18
类1（船）
类2（卡车）
类3（轿车）
36
输出
34
5%
95%
隐藏层
隐藏层
隐藏层
隐藏层
32
25%
35%
20%
55%
40%
20%
35%
20%
相关性图
16
22

图2A

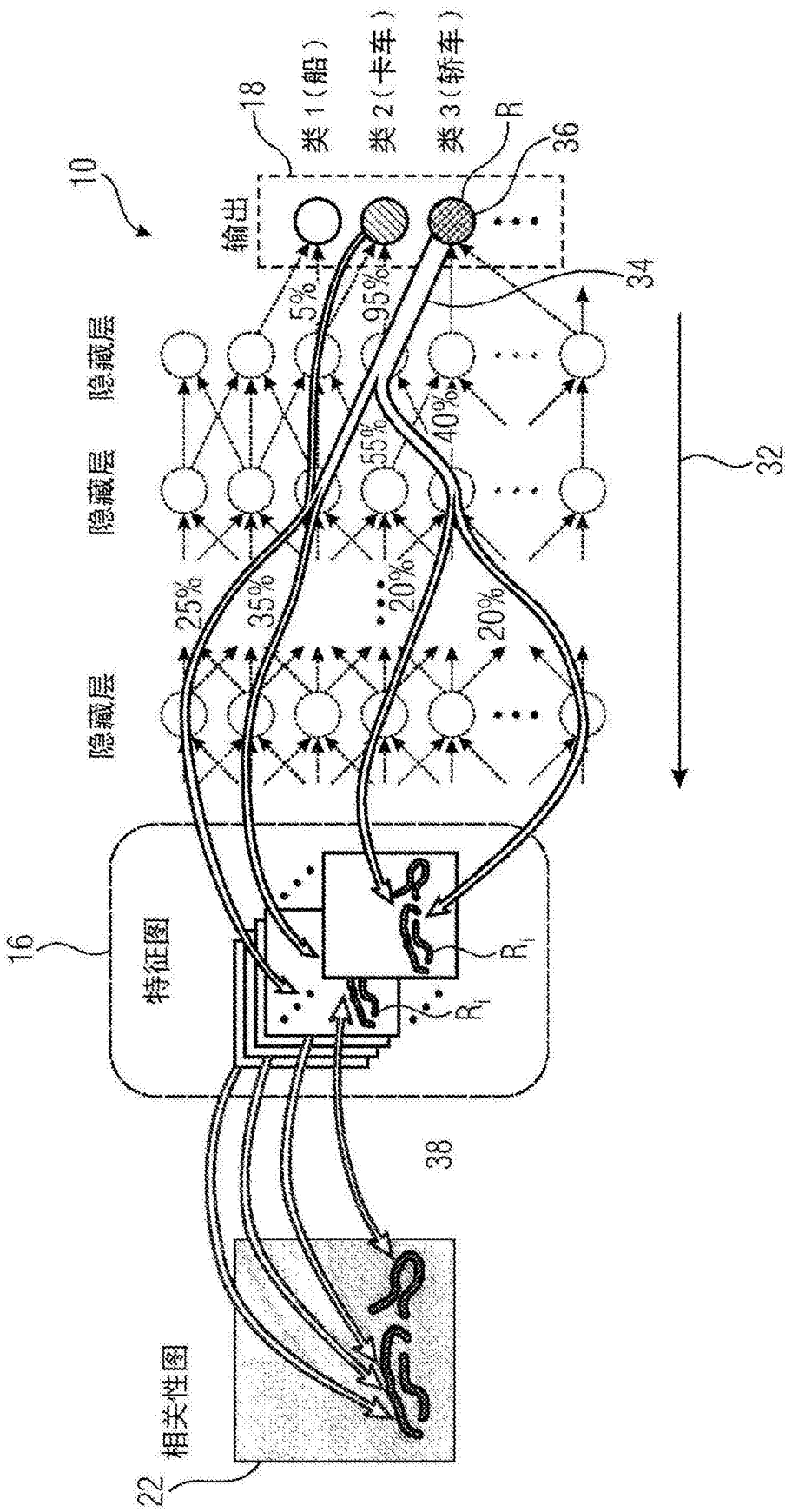

10
18
类1（船）
类2（卡车）
类3（轿车）
R
36
输出
34
5%
95%
隐藏层
隐藏层
隐藏层
55%
40%
32
25%
35%
20%
20%
16
特征图
R_i
R_i
38
相关性图
22

图2B

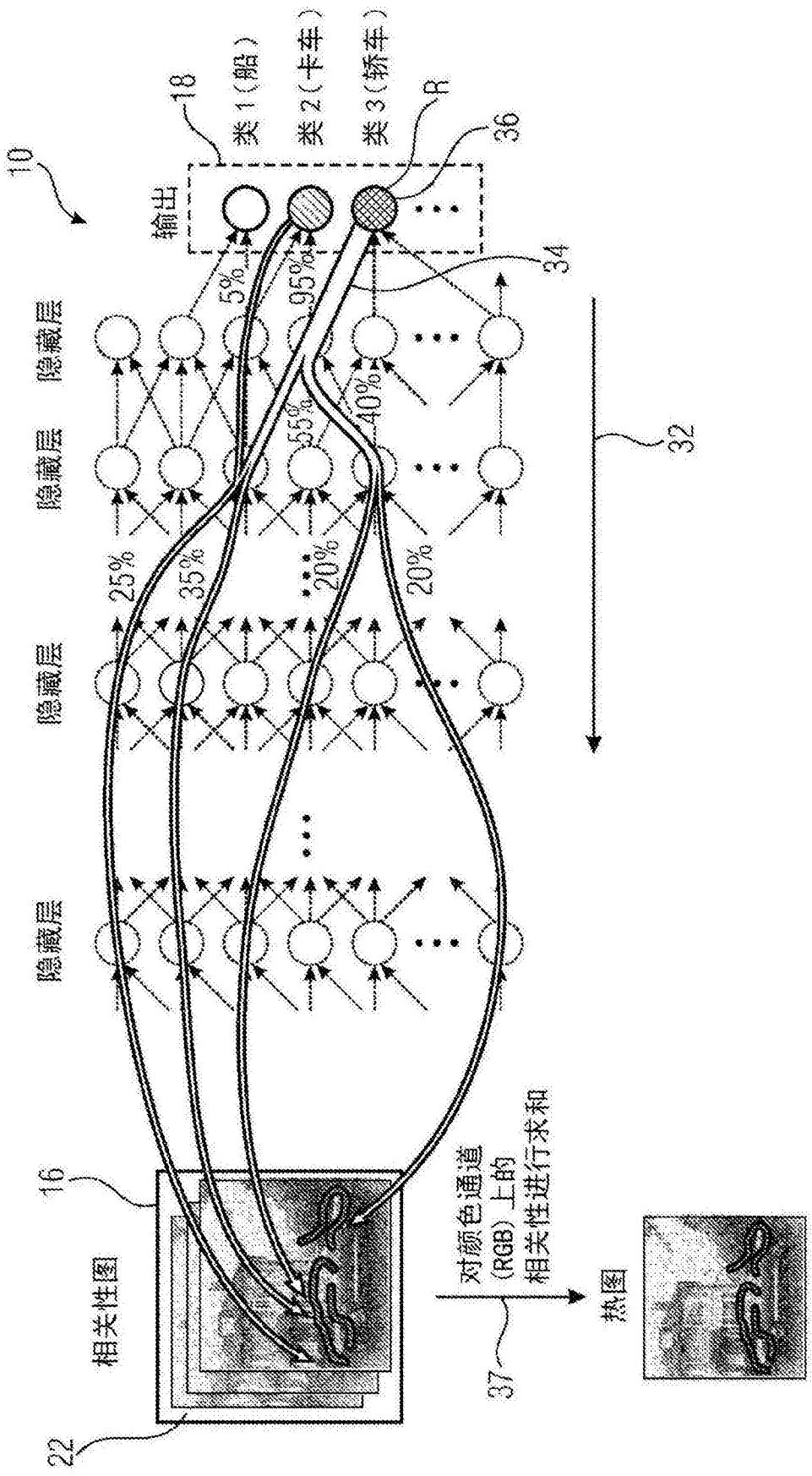

10
18
类1（船）
类2（卡车）
类3（轿车）
R
36
输出
34
5%
95%
隐藏层
隐藏层
隐藏层
隐藏层
25%
35%
20%
55%
40%
20%
32
16
相关性图
22
对颜色通道
(RGB)上的
相关性进行求和
37
热图

图2C

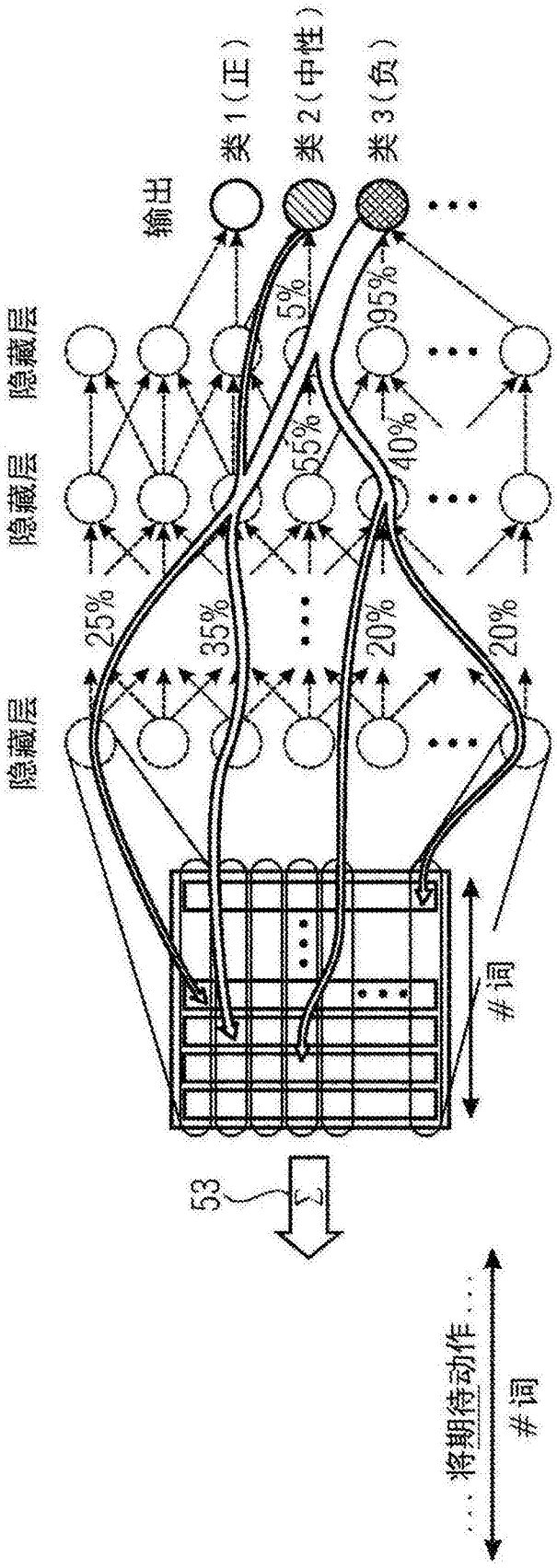


图2D

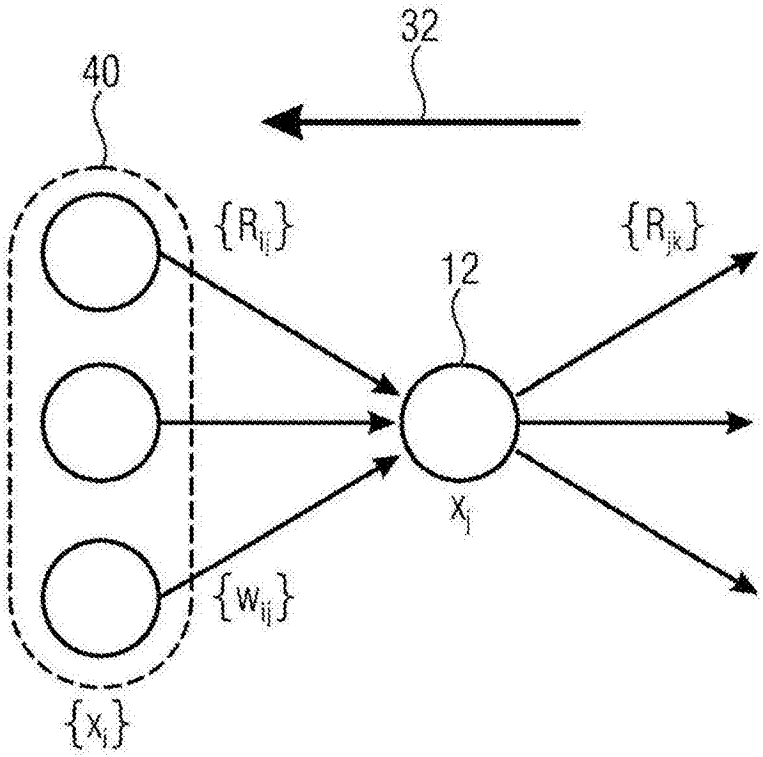


图3

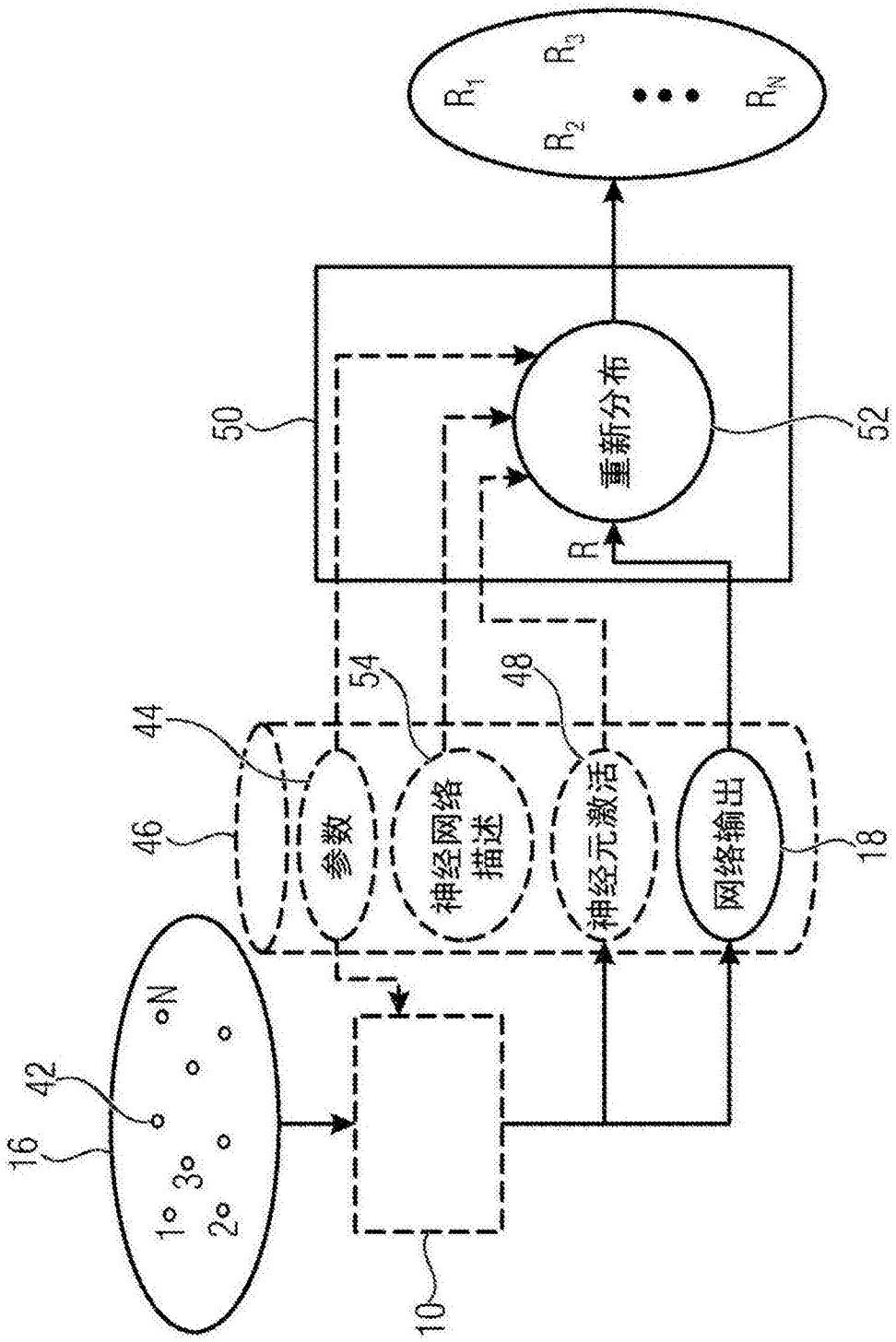

R1
R3
R2
RN
50
重新分布
52
R
54
48
44
46
参数
神经网络描述
神经元激活
网络输出
18
N
42
16
1
3
2
10

图4

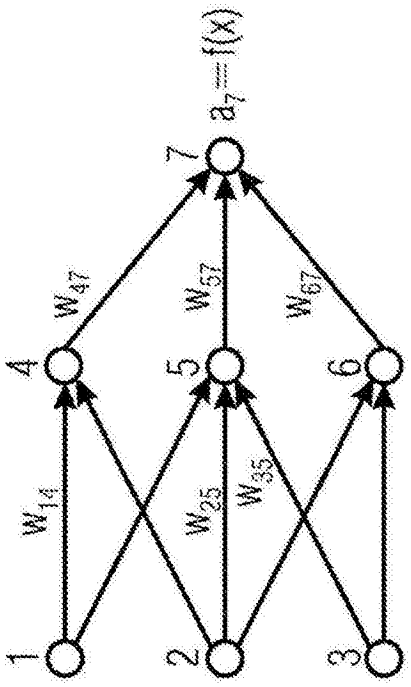

a7=f(x)
7
W47
W57
W67
4
5
6
W14
W25
W35
1
2
3

图5

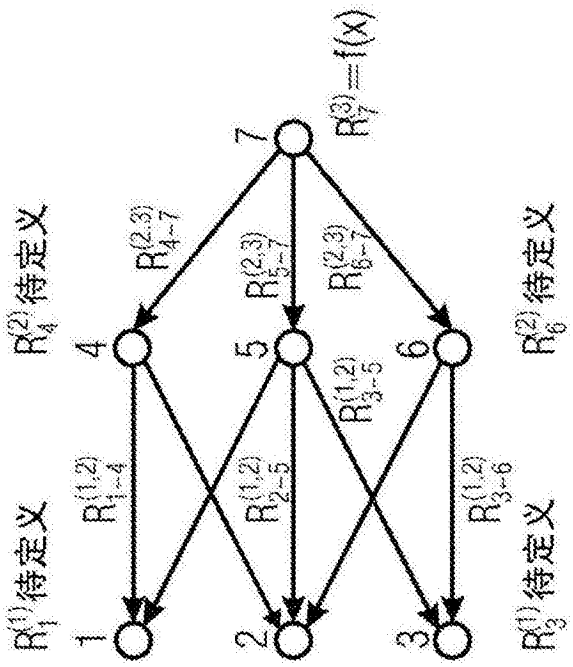

R7(3)=f(x)
7
R4(2)待定义
R4-7(2,3)
R5-7(2,3)
R6-7(2,3)
R6(2)待定义
4
5
6
R3-5(1,2)
R1-4(1,2)
R2-5(1,2)
R3-6(1,2)
R1(1)待定义
R3(1)待定义
1
2
3

图6

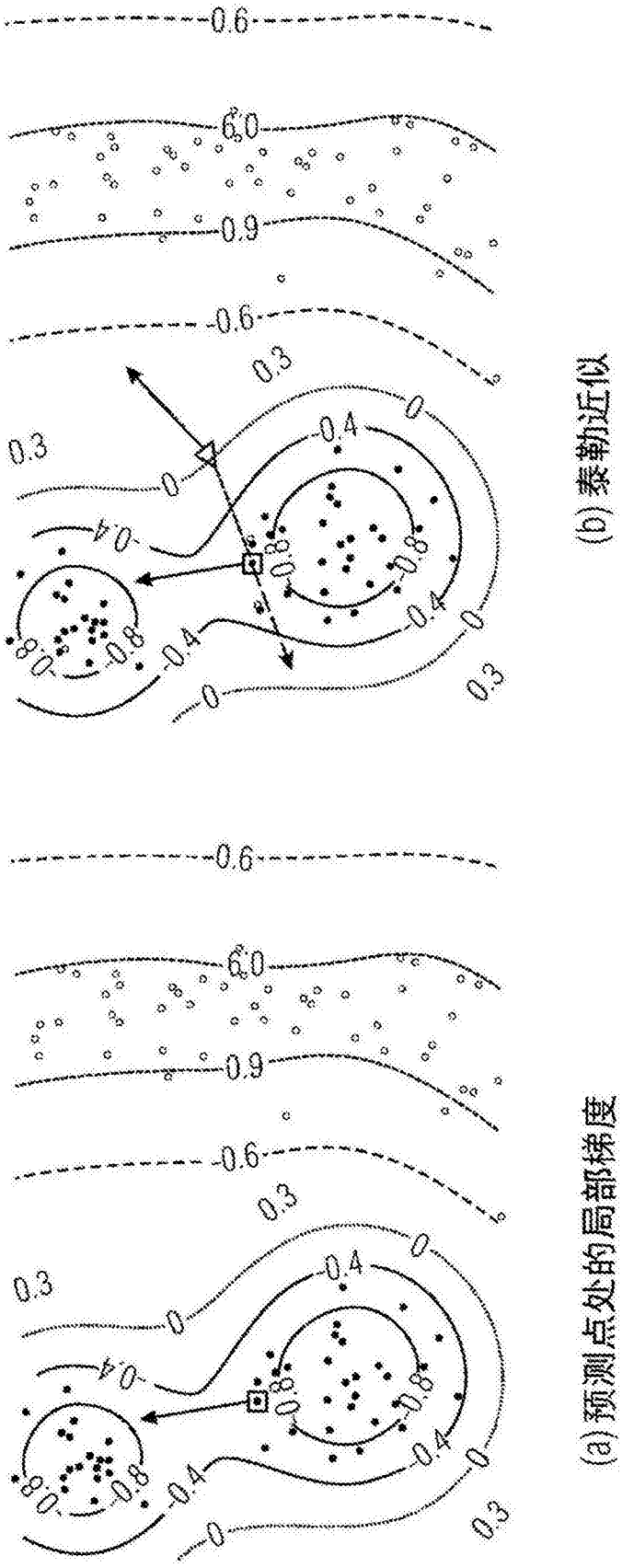
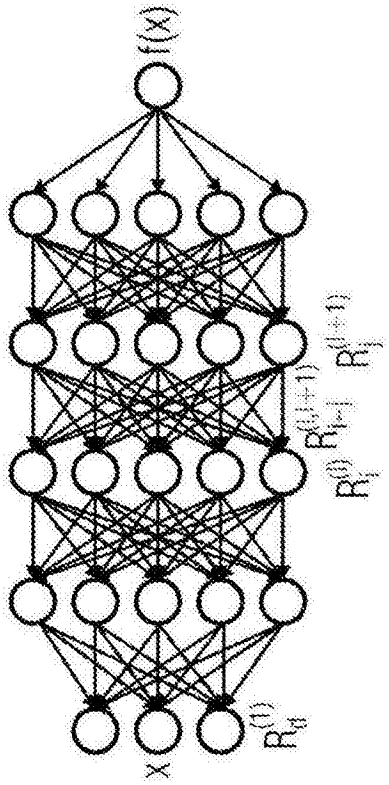

0.6
6.0
0.9
0.6
0.3
0
0.4
0.3
0.8
0.4
0
0.3
(b) 泰勒近似
0.6
6.0
0.9
0.6
0.3
0
0.4
0.3
0.8
0.4
0
0.3
(a) 预测点处的局部梯度

图7



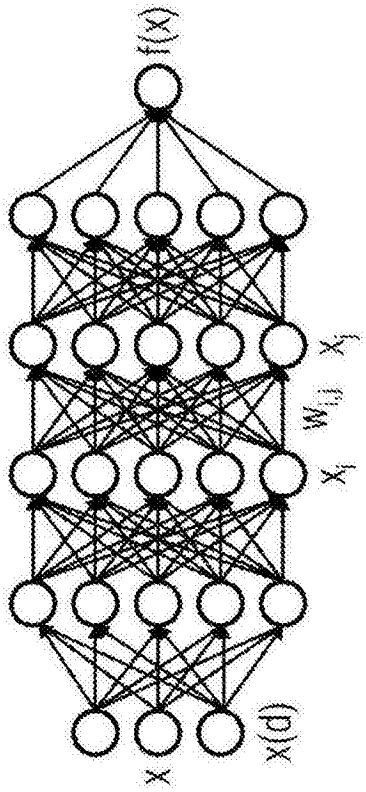


图8

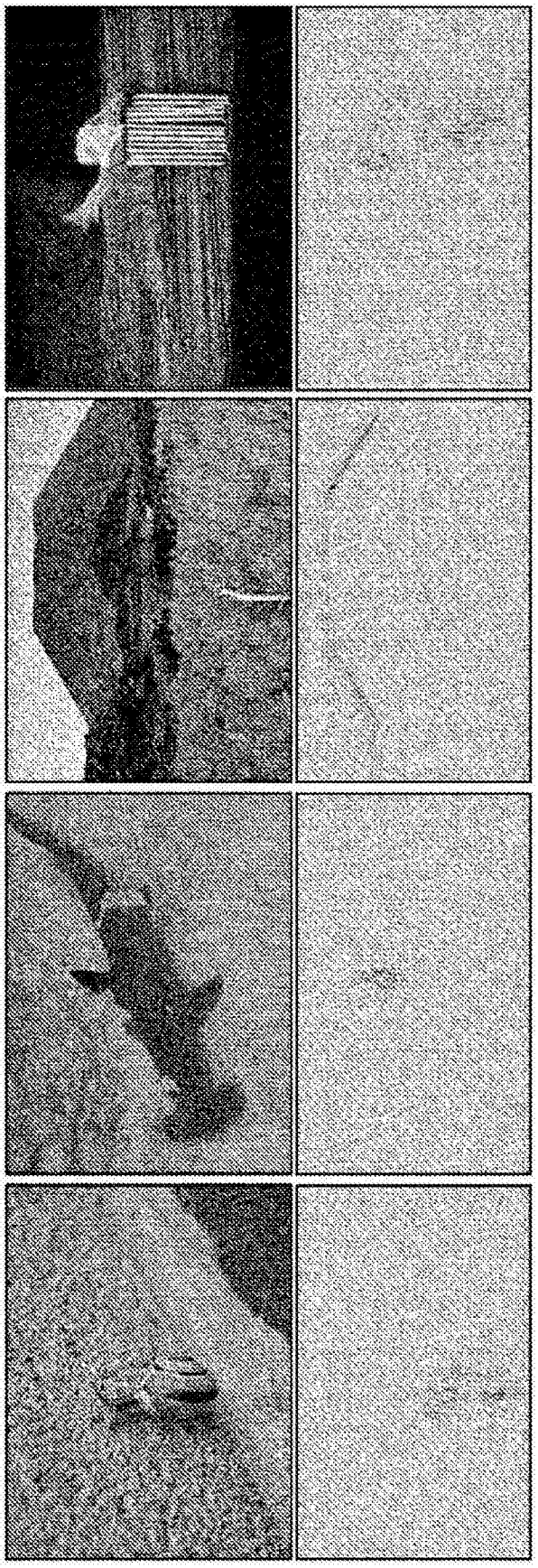


图9

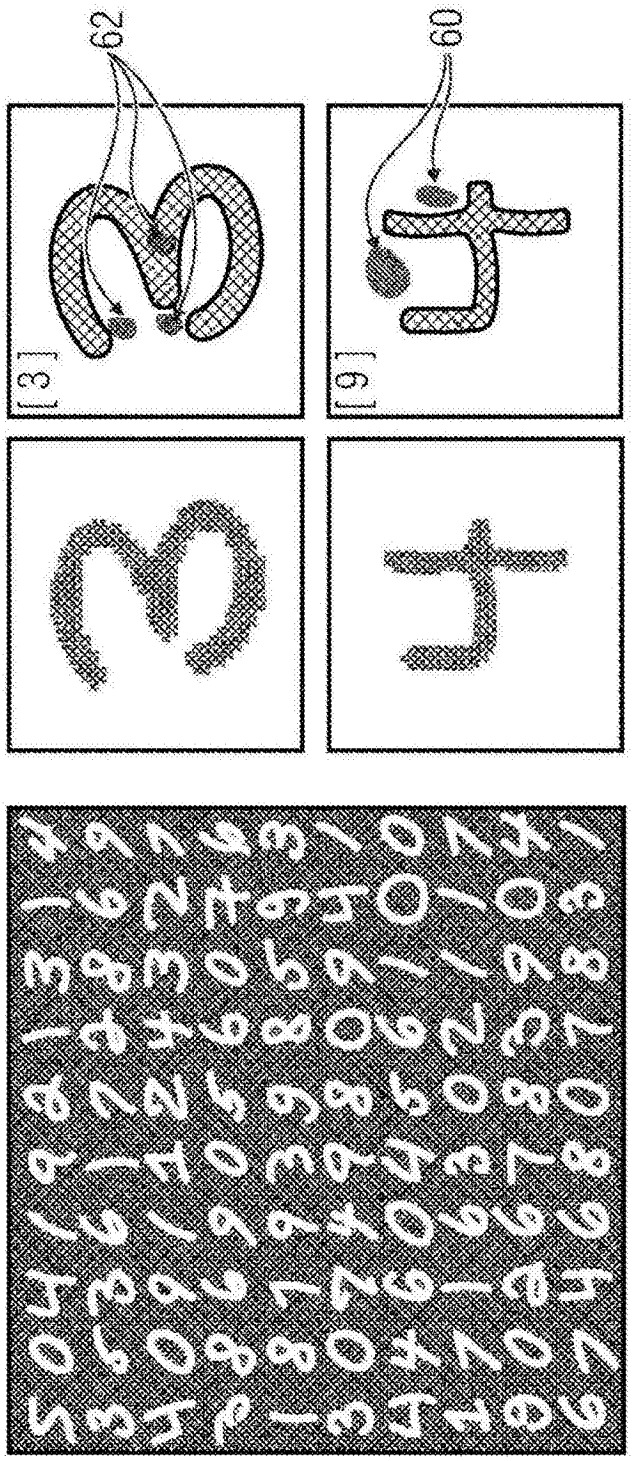


图10

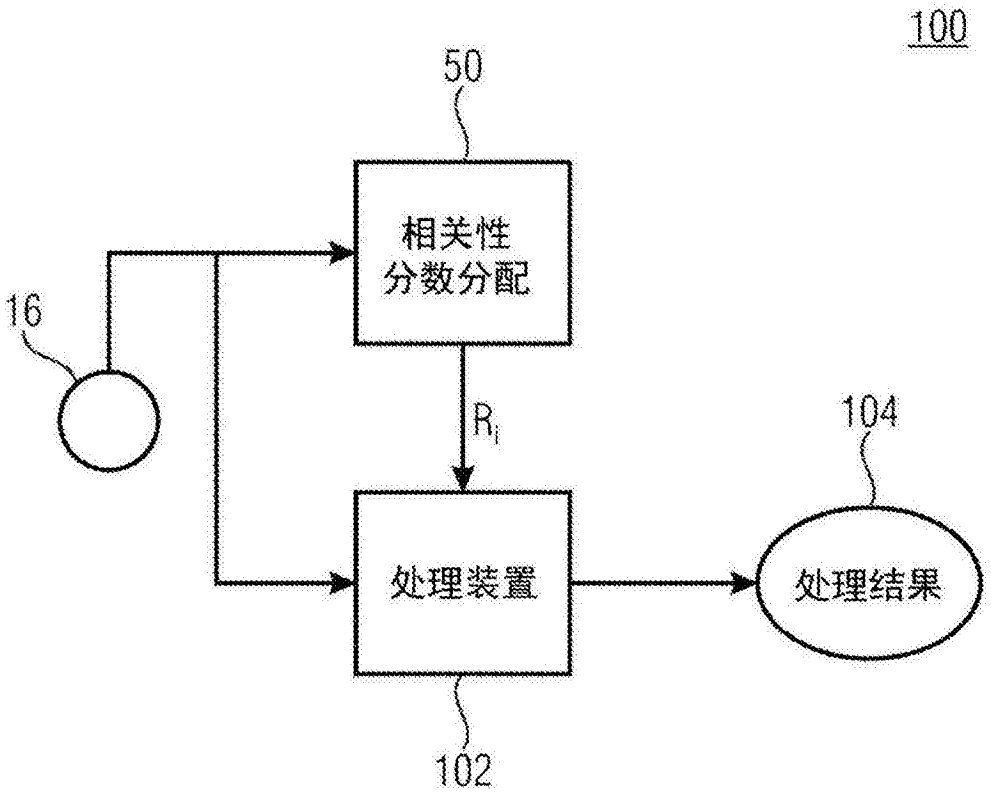

100
50
相关性
分数分配
16
R_i
104
处理装置
处理结果
102

图11

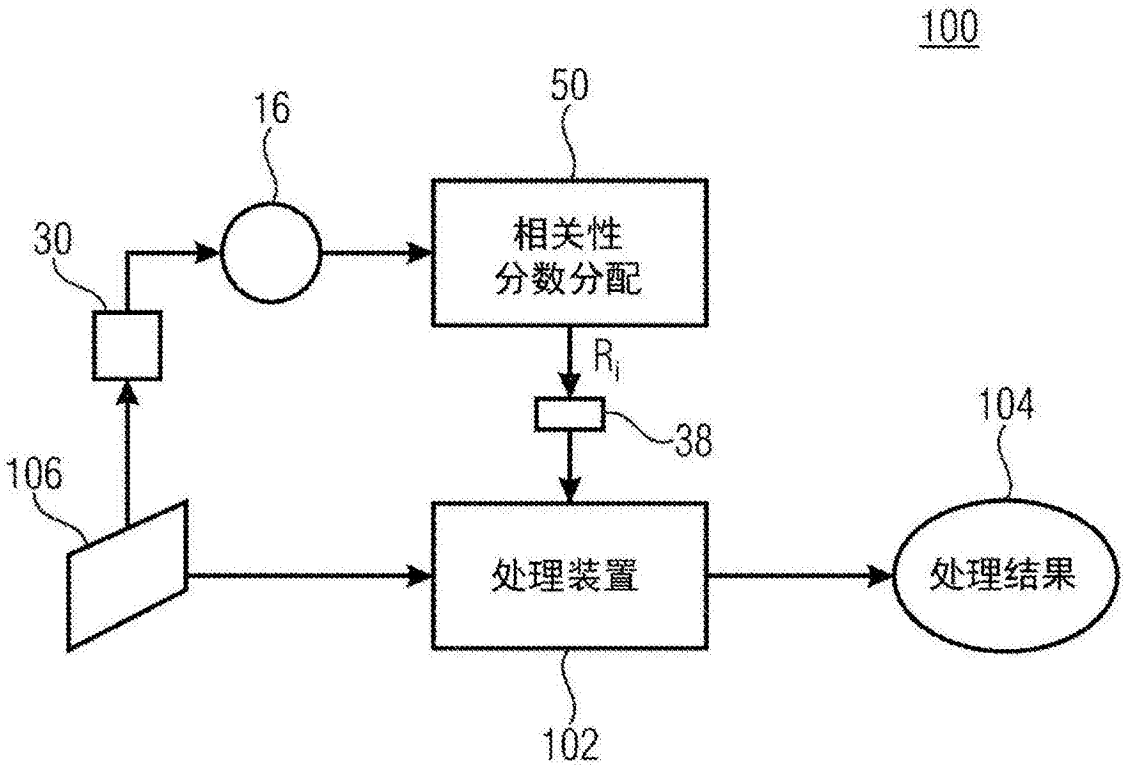

100
16
50
30
相关性
分数分配
R_i
38
104
106
处理装置
处理结果
102

图12

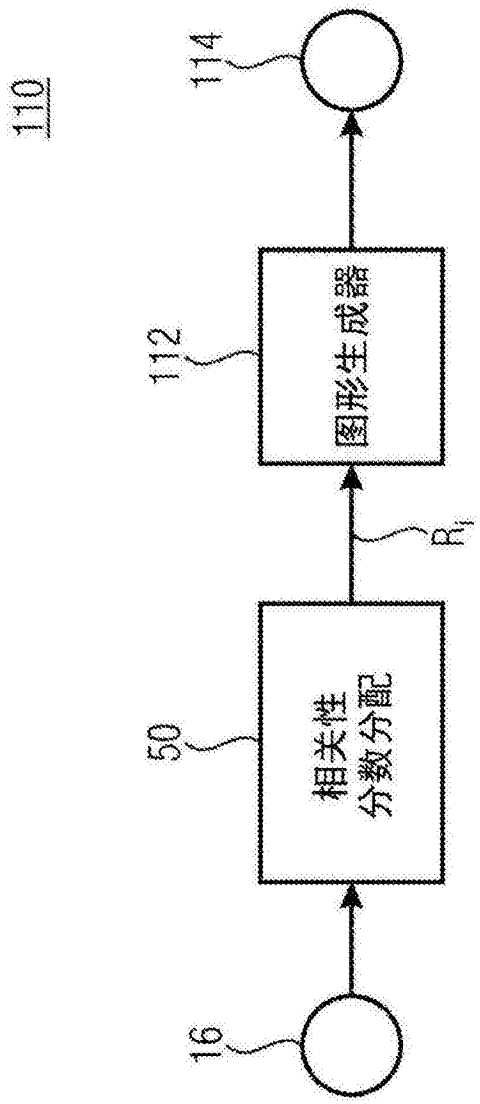


图13

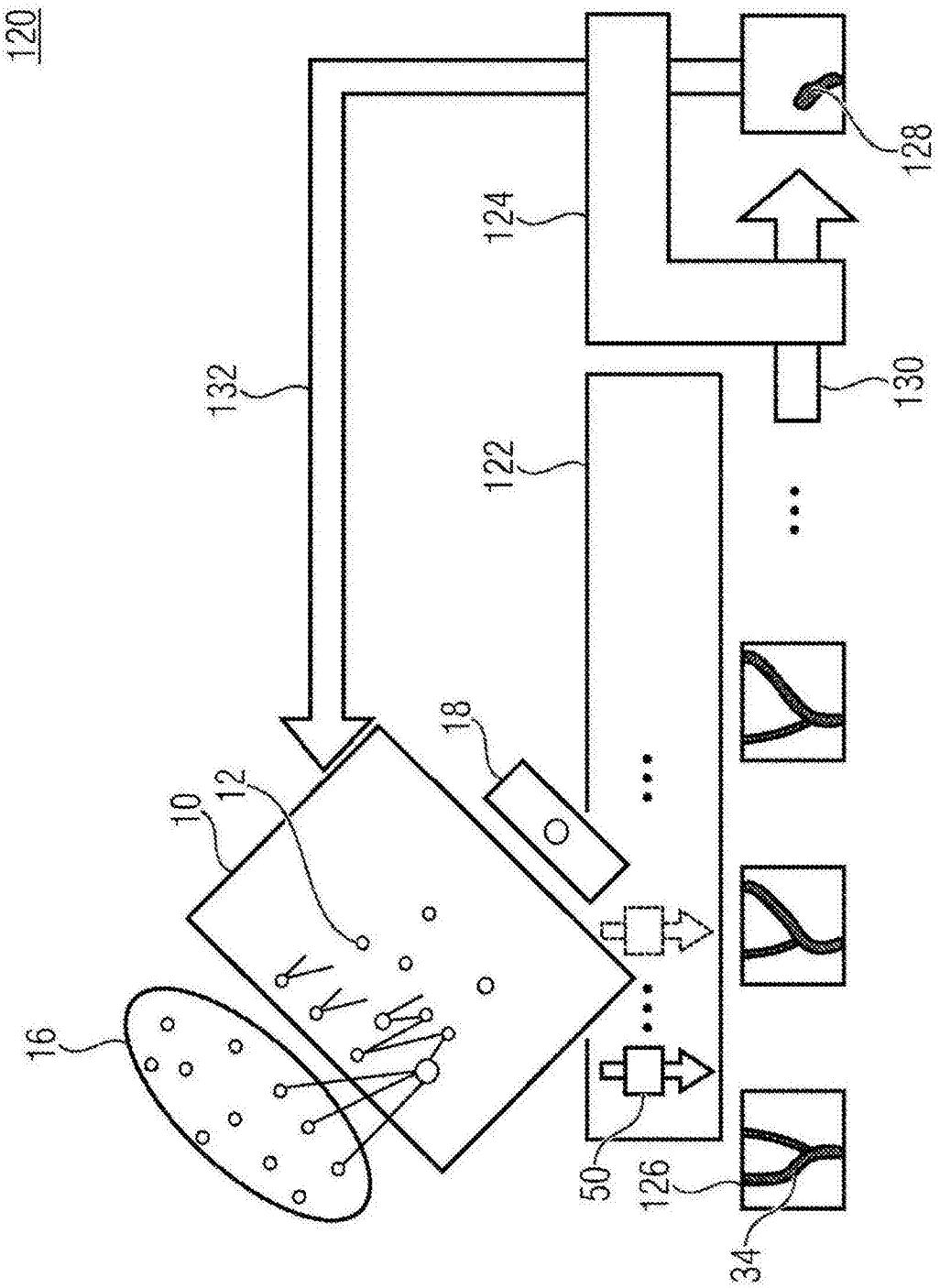


图14

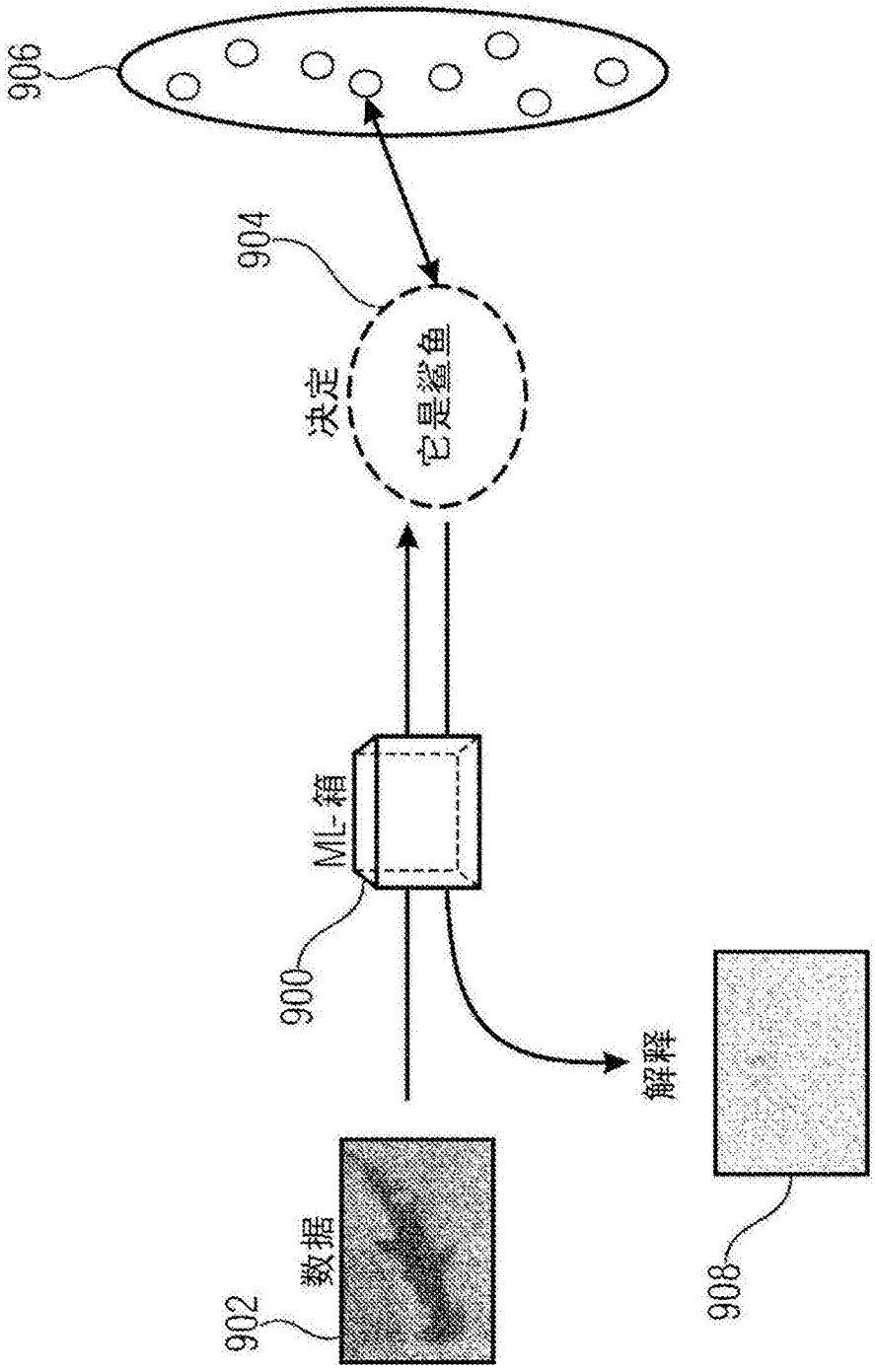


图15