

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2018 年 8 月 2 日 (02.08.2018)



(10) 国际公布号
WO 2018/137358 A1

(51) 国际专利分类号:
G06F 17/30 (2006.01)

(21) 国际申请号: PCT/CN2017/104397

(22) 国际申请日: 2017 年 9 月 29 日 (29.09.2017)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201710060334.1 2017 年 1 月 24 日 (24.01.2017) CN

(71) 申请人: 北京大学 (PEKING UNIVERSITY) [CN/CN]; 中国北京市海淀区颐和园路 5 号段凌宇, Beijing 100871 (CN)。

(72) 发明人: 段凌宇 (DUAN, Lingyu); 中国北京市海淀区颐和园路 5 号, Beijing 100871 (CN)。 白燕 (BAI,

Yan); 中国北京市海淀区颐和园路 5 号段凌宇, Beijing 100871 (CN)。 楼燚航 (LOU, Yihang); 中国北京市海淀区颐和园路 5 号段凌宇, Beijing 100871 (CN)。 高峰 (GAO, Feng); 中国北京市海淀区颐和园路 5 号段凌宇, Beijing 100871 (CN)。

(74) 代理人: 北京辰权知识产权代理有限公司 (BEIJING CHEN QUAN INTELLECTUAL PROPERTY LAW FIRM); 中国北京市海淀区中关村东路 66 号世纪科贸大厦 B 座 1501 室冯楠, Beijing 100190 (CN)。

(81) 指定国 (除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK,

(54) Title: DEEP METRIC LEARNING-BASED ACCURATE TARGET RETRIEVAL METHOD

(54) 发明名称: 基于深度度量学习的目标精确检索方法

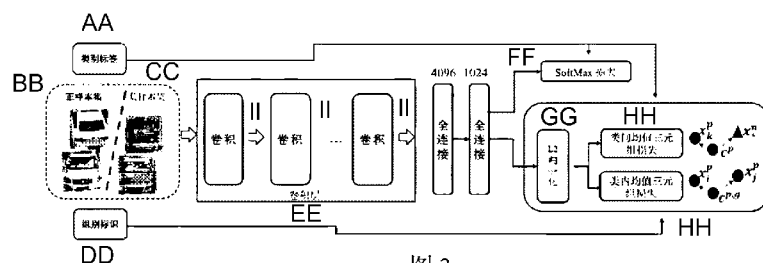


图 3

- AA Class label
- BB Positive sample set
- CC Negative sample set
- DD Group identification
- EE Convolution layer
- FF SoftMax loss
- GG L2 normalization
- HH Inter-class mean triad loss
- II Convolution
- 1024, 4096 Full-connection

(57) Abstract: Deep metric learning-based accurate target retrieval method, the method comprising: in the iterative training of a deep neural network structure, enabling the same class of target objects to be close to each other and enabling different classes of target objects to be far away from each other in a process wherein characteristics of multiple extracted pictures of the same class of target object are processed, wherein intra-class individuals having a similar attribute are close to each other in the characteristic distribution of individuals of the same class, and the distance between intra-class individuals having different attributes is greater than a preset distance,

LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX,
MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL,
PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL,
SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG,
US, UZ, VC, VN, ZA, ZM, ZW。

(84) 指定国(除另有指明, 要求每一种可提供的地区
保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ,
NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM,
AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG,
CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU,
IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT,
RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI,
CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

根据细则4.17的声明:

- 关于发明人身份(细则4.17(i))
- 关于申请人有权申请并被授予专利(细则
4.17(ii))
- 关于申请人有权要求在先申请的优先权(细则
4.17(iii))
- 发明人资格(细则4.17(iv))

本国际公布:

- 包括国际检索报告(条约第21条(3))。

so as to obtain a trained deep neural network structure; using a trained deep neural network structure to extract respective characteristics from pictures to be queried and a preset reference picture respectively; and obtaining Euclidean distances of characteristics between the query pictures and the reference picture, and sorting the distances from small to large so as to obtain an accurate retrieval target. With said method, the problem of accurate retrieval in a vertical field is solved.

(57) 摘要: 一种基于深度度量学习的目标精确检索方法, 方法包括: 深度神经网络结构的迭代训练中, 对提取的同类目标对象的多张图片的特征进行处理过程中, 使得相同类别的目标对象相互靠近, 不同类别的目标对象相互远离, 且属于同一类别的个体的特征分布中, 具有相似属性的类内个体之间距离相互靠近, 具有不同属性的类内个体之间大于预设距离, 以获得训练后的深度神经网络结构; 采用训练后的深度神经网络结构对待查询图片与预设的参考图片分别提取各自的特征, 并获取查询图片与参考图片之间特征的欧式距离, 对该距离进行从小到大的排序, 获得精确检索的目标。所述方法解决了垂直领域的精确检索问题。

基于深度度量学习的目标精确检索方法

技术领域

本发明涉及计算机视觉技术，具体涉及一种基于深度度量学习的目标精确检索方法。

背景技术

目标精确检索一直是计算机领域中的至关重要的问题，同时也是对象跟踪、行为分析等应用分析的基础。精确检索（同时也称细粒度识别），旨在精细地区分不同种视觉上相似的物体类别。例如，细粒度车辆识别可以识别出图片内特定的汽车模型，比如“奥迪 A6 2015 款”。近年来，随着计算机大规模并行计算能力的突飞猛进以及深度卷积神经网络的成功应用，人们在大量细粒度图像分类的垂直领域投入了更多研究，比如识别不同品种的动物、植物、车辆、衣服、行人的身份等。

目标精确检索的目的在于将某个特定查询目标从一系列具有相似外表的参考对象数据集中识别出来。识别的过程是一个相似度比对的过程，依据与查询图像的相似度得分从高到低排序。相似度得分的产生过程通常包含三个步骤：提取特征、做相应变换、相似度度量。

对于查询目标和参考对象，常见的提取特征有颜色、纹理、以及尺度不变特征等，这些特征再经过某种数学模型进行变换或融合，再选取适当的距离度量生成特征之间的相似度得分并排序。

细粒度图像识别的难点在于类间的相似性和类内的多样性。一方面，不同的细粒度类别对象本身属于某一个大类别（如：奥迪 A6 和奥迪 A8 都属于奥迪车），其具有类间相似性，使得提取出的特征

也非常相似导致难以区分。另一方面，同一细粒度类别的对象由于不同的视角，姿态和光照条件会呈现出极大的差异性，称为类内多样性。

为了减轻在细粒度物体识别过程中类间相似性和类内差异性带来的负作用，为此，通过基于局部部件的方法被用来捕捉细微的局部特征，区分不同类别物体，减小由于视角和姿势变化导致的特征上的差异。例如，细粒度鸟类识别中学习鸟头和身体的局部特征模型，在局部部件之间加入几何约束。

但是，基于局部部件的方法依赖于精确的部件定位，当缺少大量视角变化时性能低下。

最近业内人士提出了基于度量学习的方法，这种方法旨在最大化类间距离同时最小化类内距离，以降低类间相似性和类内多样性的影响。基于度量学习方法中使用的目标检索算法大多依赖手工设计的特征，如：HOG、LBP、SIFT等，然后主要研究两幅图像之间的距离度量方法。

发明内容

鉴于上述问题，本发明提出了克服上述问题或者至少部分地解决上述问题的一种基于深度度量学习的目标精确检索方法。

为此目的，第一方面，本发明提出一种基于深度度量学习的目标精确检索方法，包括：

深度神经网络结构的迭代训练中，对提取的同类目标对象的多张图片的特征进行处理过程中，使得相同类别的目标对象的特征距离减小，不同类别的目标对象的特征距离增加，且属于同一类别的个体的特征分布中，具有相似属性的类内个体之间的特征距离减小，具有不同属性的类内个体之间的特征距离大于预设距离，以获得训练后的深度神经网络结构；

采用训练后的深度神经网络结构对待查询图片与预设的参考图片分别提取各自的特征，并获取查询图片与参考图片之间特征的欧式距离，对该距离进行从小到大的排序，获得精确检索的目标。

可选地，所述深度神经网络结构为根据选择对象的属性信息能够区分不同对象类别的网络结构；

和/或，

迭代训练之前的深度神经网络结构为通用的深度卷积神经网络结构。

可选地，深度神经网络结构的迭代训练中，还用于执行前向传播计算损失，并通过反向传播算法调整深度神经网络结构中的权重值，使得训练后的深度神经网络结构中损失收敛。

可选地，所述深度卷积神经网络的最后一层全连接的输出有两个分支，一个分支连接Softmax归一化指数损失函数，另一分支连接融入类内多样性的均值三元组损失函数；

迭代训练中执行前向传播计算的损失为：Softmax损失函数和均值三元组损失函数根据各自的权重值确定的损失。

可选地，在深度卷积神经网络迭代训练过程中，融入类内多样性的均值三元组损失函数的损失确定方式如下：

采用KmeansK均值聚类算法对类内的所有特征点进行聚类，得到类内的若干分组；

将一个类别中所有样本均值作为该类内均值点；以及

根据聚类得到的类内若干分组，在一个类内计算组内均值，作为组内均值点；根据所有样本的类别标签和每一类内的组别标识，获取所有正样本、负样本分别到类内均值点和各自分组后的组内均值点的距离，并计算融入类内多样性的均值三元组损失函数。

可选地，迭代训练中执行前向传播计算的损失为：Softmax损失函数和融入类内多样性的均值三元组损失函数根据各自的权重值确

定的损失，包括：

根据 $L_{GS-TRS} = \omega L_{softmax} + (1 - \omega) L_{ICV-triplet}$ 确定前向传播计算的损失

L_{GS-TRS} ；

其中， ω 是加权重， $L_{softmax}$ 为Softmax损失函数的损失， $L_{ICV-triplet}$ 为均值三元组损失函数的损失。

可选地，采用Kmeans聚类算法对类内的所有特征点进行聚类，得到类内的若干分组，包括：

使用 $\argmin \sum_{g=1}^G \sum_{x=1}^{N^{p,g}} \|f(x) - \mu_g\|^2$ 对类内的所有特征点进行聚类，得到类内的若干分组；

其中， $f(x)$ 是提取的图像的特征， G 是组数，代表聚类族的数目， $N^{p,g}$ 是在组 $S_{c,g}$ 内的样本数量， μ_g 是聚类中心。

可选地，将一个类别中所有样本均值作为该类内均值点，包括：

若一个包含 N^p 个类内样本集 X^p 为 $X^p = \{x_1^p, x_2^p, \dots, x_{N^p}^p\}$ ， $1 \leq i \leq N^p$ ，则该类内样本集的类内均值参考点 c^p 为： $c^p = \frac{1}{N^p} \sum_i^{N^p} f(x_i^p)$ 。

可选地，根据聚类得到的类内若干分组，在一个类内计算组内均值，作为组内均值点，包括：

若一个包含 N^p 个组内样本集 X^p 为 $X^p = \{x_1^p, x_2^p, \dots, x_{N^p}^p\}$ ， $1 \leq i \leq N^p$ ，则组内均值参考点 c^p 为： $c^p = \frac{1}{N^p} \sum_i^{N^p} f(x_i^p)$ 。

可选地，根据所有样本的类别标签和每一类内的组别标识，获取所有正样本、负样本分别到类内均值点和各自分组后的组内均值点的距离，并计算融入类内多样性的均值三元组损失函数，包括：

根据公式一获取融入类内多样性的三元组损失函数的损失值；

公式一： $L_{ICV-triplet} = L_{inter}(c^p, x_k^p, x_*^n) + \sum_{g=1}^G L_{intra}(c^{p,g}, x_i^p, x_j^p)$

$$= \sum_{k=1}^{N^p} \frac{1}{2} \max \left\{ \|c^p - f(x_k^p)\|_2^2 + \alpha_1 - \|c^p - f(x_*^n)\|_2^2, 0 \right\} \\ + \sum_{g=1}^G \sum_{i=1}^{N^{p,g}} \frac{1}{2} \max \left\{ \|c^{p,g} - f(x_i^p)\|_2^2 + \alpha_2 - \|c^{p,g} - f(x_j^p)\|_2^2, 0 \right\}$$

其中， c^p 代表类别 c 中所有样本的均值中心， $c^{p,g}$ 代表类别 c 的组 g 中所有样本的均值中心；每一个类别 c 有一个类中心 c^p 和 G 个组中心 $c^{p,g}$ ，

x_i^p, x_j^p 代表两个取自不同组的同类样本。对于类间关系， $x_k^p \in c$ 是正样本， $x_*^n \notin c$ 是负样本；

α_1 是类中心与不同类别样本距离的间隔， α_2 是类内组中心与不同组别样本距离的间隔；

反向传播的偏导数：

正样本 x_i^p 对损失的偏导数为：

$$\frac{\partial L}{\partial f(x_i^p)} = f(x_i^p) - c^p + \frac{1}{N^p} (f(x_*^n) - f(x_i^p))$$

负样本 x_*^n 对损失的的偏导数为：

$$\frac{\partial L}{\partial f(x_*^n)} = c^p - f(x_*^n)。$$

由上述技术方案可知，本发明的基于深度度量学习的目标精确检索方法，能实现类内个体具有相似属性的距离更接近，相比传统的三元组损失函数训练的模型，检索的准确率得到了极大的提高；通过多损失函数联合优化的模型产生的特征更加鲁棒，检索性能比单一损失函数优化的模型更高。

附图说明

图 1 为本发明一实施例提供的同类样本在模型优化过程中相应

的特征分布的示意图；

图 2 为现有技术中基于聚类方法产生的车辆图像的类内分组示意图；

图 3 为本发明一实施例提供的以 VGGM 为基本网络结构的训练阶段示意图。

具体实施方式

为使本发明实施例的目的、技术方案和优点更加清楚，下面将结合本发明实施例中的附图，对本发明实施例中的技术方案进行清楚地描述，显然，所描述的实施例是本发明一部分实施例，而不是全部的实施例。

结合图 1 所示，本发明实施例的基于深度度量学习的目标精确检索方法，包括：

步骤A01、深度神经网络结构的迭代训练中，对提取的同类目标对象的多张图片的特征进行处理过程中，使得相同类别的目标对象的特征距离减小，不同类别的目标对象的特征距离增加，且属于同一类别的个体的特征分布中，具有相似属性的类内个体之间的特征距离减小，具有不同属性的类内个体之间的特征距离大于预设距离（如图1中的（a）、（b）所示），以获得训练后的深度神经网络结构。

应说明的是，结合图1中的（a）、（b），该步骤中，训练后的深度神经网络结构可使得相同类别的目标对象相互靠近，不同类别的目标对象相互远离；相应地，属于同一类别的个体的特征分布中，具有相似属性的类内个体之间的特征距离相互靠近，具有不同属性的类内个体之间的特征距离大于预设距离。

步骤A02、采用训练后的深度神经网络结构对待查询图片与预设的参考图片分别提取各自的特征，并获取查询图片与参考图片之间特征的欧式距离，对该距离进行从小到大的排序，获得精确检索的目标。

在具体应用中，上述的深度神经网络结构可为根据选择对象的属

性信息能够区分不同对象类别的网络结构。

在本实施例中，迭代训练之前的深度神经网络结构优选为通用的深度卷积神经网络结构。

上述实施例的目标精确检索方法属于计算机视觉领域，可，适用于基于视频处理的智能交通和智能监控技术。由此，可解决垂直领域的精确检索问题，即给定一个查询图片，需要在一个大型的垂直领域的数据库（如：车辆的数据集）中查找相同的对象目标或类别。

需要说明的是，本实施例中，深度神经网络结构的迭代训练中，还用于执行前向传播计算损失，并通过反向传播算法计算到各个层对应的损失，再根据梯度下降算法更新深度神经网络结构中的权重值，使得训练后的深度神经网络结构中损失收敛。

也就是说，深度卷积神经网络的最后一层全连接的输出有两个分支，一个分支连接Softmax归一化指数损失函数，另一分支连接融入类内多样性的均值三元组损失函数；

迭代训练中执行前向传播计算的损失 L_{GS-TRS} 为：Softmax损失函数和均值三元组损失函数根据各自的权重值确定的损失。

例如，可根据 $L_{GS-TRS} = \omega L_{softmax} + (1 - \omega) L_{ICV-triplet}$ 确定前向传播计算的损失 L_{GS-TRS} ；

其中， ω 是加权权重， $L_{softmax}$ 为Softmax损失函数的损失， $L_{ICV-triplet}$ 为均值三元组损失函数的损失。

也就是说，三元组损失网络使用相似度距离学习而不是选择超平面，这种网络对特征的判别能力可以通过在学习目标中加入分类损失（Softmax损失）来显著提升性能。

因此，上述方法通过多任务学习的方法联合优化融入类内多样性的均值三元组损失和Softmax损失。在网络的前向传播计算损失的过

程中,使用线性加权的方法实现这两种损失的结合。本发明实施例中,优化这个多损失函数可以实现有力的细粒度分类性能和提取有分辨性的细粒度检索特征。

上述实施例中目标精确检索的方法,在距离度量的优化过程创新性的中融入了类内多样性的表达。利用该方法训练的模型提取出的图片特征表达,可以较好的保留细粒度类别类内的多样性,在精确检索时拥有相同或相似的类内属性的图片可以在检索结果中得到较高的排名。不同对象图片的相似度时表现为两幅图像生成的特征之间的欧式距离,距离越近越相似。如图1所示,为同类样本在模型优化过程中相应的特征分布。

图1(a)示出的是现有传统的三元组损失示意图,图1(b)示出的是本发明的加入类内多样性(ICV)的均值三元组损失的特征分布示意图。在图1(a)和图1(b)中可以看出,本发明的损失函数在模型优化过程中,同类对象会聚集在一起,同时同类对象内部具有相似属性的对象会在类内聚集成更小的团。

另外,在深度网络训练阶段,输入是正负样本集合,正样本集合包含同一个对象或类别但是拍摄自不同摄像头的多张图片,负样本集合包含不同对象或类别的多张图片,图片的数目没有特殊限制。每一个对象除了本身的类别信息,还拥有在一个类别内的组别信息。前者是根据类别定义界定的不同类别标签,后者是在一个细粒度类别内进行划分的组别,划分的依据是通过Kmeans算法聚类得到。所有的输入图片经过相同的深度神经网络被映射到一个高维欧式空间,优化目标是在欧式空间中相同类别的对象尽可能靠近,不同类别的对象尽可能远离,同时在类内的特征分布上,拥有相似属性的同一个组别的对象尽可能接近,不同组别的对象保持一定的距离间隔。

为此,在深度卷积神经网络迭代训练过程中,融入类内多样性的均值三元组损失函数的损失确定方式如下步骤B01至步骤B04:

B01、采用Kmeans K均值聚类算法对类内的所有特征点进行聚类，得到类内的若干分组；

举例来说，使用 $\operatorname{argmin} \sum_{g=1}^G \sum_{x=1}^{N^{p,g}} \|f(x) - \mu_g\|^2$ 对类内的所有特征点进行聚类，得到类内的若干分组；

其中， $f(x)$ 是提取的图像的特征， G 是组数，代表聚类族的数目， $N^{p,g}$ 是在组 $S_{c,g}$ 内的样本数量， μ_g 是聚类中心。

B02、将一个类别中所有样本均值作为该类内均值点。

可以理解的是，如果一个包含 N^p 个类内样本集 X^p 为 $X^p = \{x_1^p, x_2^p, \dots, x_{N^p}^p\}$ ， $1 \leq i \leq N^p$ ，则该类内样本集的内均值参考点 c^p 为： $c^p = \frac{1}{N^p} \sum_i^{N^p} f(x_i^p)$ 。

B03、根据聚类得到的类内若干分组，在一个类内计算组内均值，作为组内均值点；

若一个包含 N^p 个组内样本集 X^p 为 $X^p = \{x_1^p, x_2^p, \dots, x_{N^p}^p\}$ ， $1 \leq i \leq N^p$ ，则组内均值参考点 c^p 为： $c^p = \frac{1}{N^p} \sum_i^{N^p} f(x_i^p)$ 。

B04、根据所有样本的类别标签和每一类内的组别标识，获取所有正样本、负样本分别到类内均值点和各自分组后的组内均值点的距离，并计算融入类内多样性的均值三元组损失函数。

即根据公式一获取融入类内多样性的三元组损失函数的损失值；

公式一： $L_{ICV-triplet} = L_{inter}(c^p, x_k^p, x_*^n) + \sum_{g=1}^G L_{intra}(c^{p,g}, x_i^p, x_j^p)$

$$= \sum_{k=1}^{N^p} \frac{1}{2} \max \left\{ \|c^p - f(x_k^p)\|_2^2 + \alpha_1 - \|c^p - f(x_*^n)\|_2^2, 0 \right\} \\ + \sum_{g=1}^G \sum_{i=1}^{N^{p,g}} \frac{1}{2} \max \left\{ \|c^{p,g} - f(x_i^p)\|_2^2 + \alpha_2 - \|c^{p,g} - f(x_j^p)\|_2^2, 0 \right\}$$

其中， c^p 代表类别 c 中所有样本的均值中心， $c^{p,g}$ 代表类别 c 的组 g

中所有样本的均值中心；每一个类别 c 有一个类中心 c^p 和 G 个组中心 $c^{p,g}$,

x_i^p, x_j^p 代表两个取自不同组的同类样本。对于类间关系, $x_k^p \in c$ 是正样本, $x_*^n \notin c$ 是负样本;

α_1 是类中心与不同类别样本距离的间隔, α_2 是类内组中心与不同组别样本距离的间隔;

反向传播的偏导数:

正样本 x_i^p 对损失的偏导数为:

$$\frac{\partial L}{\partial f(x_i^p)} = f(x_i^p) - c^p + \frac{1}{N^p} (f(x_*^n) - f(x_i^p))$$

负样本 x_*^n 对损失的的偏导数为:

$$\frac{\partial L}{\partial f(x_*^n)} = c^p - f(x_*^n).$$

下面将对具体对用于目标精确检索的方法进行详细描述:

在深度网络训练阶段, 输入是正负样本集合, 正样本集合包含同一个对象或类别但是拍摄自不同摄像头的多张图片, 负样本集合包含不同对象或类别的多张图片, 图片的数目没有特殊限制。每一个对象除了本身的类别信息, 还拥有类别内的组别信息。

1.类内组别划分

为了表现类内多样性, 需要将相同类别的样本进行分组。不同于类别标签, 类内数据的内在属性特征非常抽象难以精确表述。

为此, 本发明中使用无监督的方法给每个类别中的对象进行分组。

首先, 用一个在ImageNet数据集上做过1000类分类任务的VGG_CNN_M_1024 (VGGM) 深度神经网络作为特征提取器;

接着, 对所有的训练图像提取网络最后的全相连层的输出作为特

征，并用主成分分析（PCA）进行特征降维；

最后，使用K-means方法聚类：

$$\operatorname{argmin} \sum_{g=1}^G \sum_{x=1}^{N^{p,g}} \|f(x) - \mu_g\|^2$$

这里 $f(x)$ 是提取的图像的特征， G 是组数，代表聚类族的数目， $N^{p,g}$ 是在组 $S_{c,g}$ 内的样本数量， μ_g 为聚类中心。每个图片实例在聚类后都有一个组标签，并且划分在一个组内的对象成员往往有相似的性质如颜色、拍摄视角，背景等。

2.均值三元组损失

当前现有技术中正样本中，三元组的参考点通常是随机选取的。为了降低不恰当参考点选择带来的负影响，本发明实施例中使用通过计算所有正样本平均值的方法选择参考点，并建立均值三元组损失。

例如，给定一个包含 N^p 个样本的正样本集 $X^p = \{x_1^p, x_2^p, \dots, x_{N^p}^p\}$ 和包含 N^n 个从其他类选取的样本的负样本集 $X^n = \{x_1^n, x_2^n, \dots, x_{N^n}^n\}$ 。这样，均值参考点可以被表述为：

$$c^p = \frac{1}{N^p} \sum_i^{N^p} f(x_i^p)$$

其中 $1 \leq i \leq N^p$ ， $1 \leq j \leq N^n$ 。区别于使用随机选取的参考点，均值三元组损失表述为：

$$L(c^p, X^p, X^n) = \sum_i^{N^p} \frac{1}{2} \max \left\{ \|f(x_i^p) - c^p\|_2^2 + \alpha - \|f(x_*^n) - c^p\|_2^2, 0 \right\}$$

这里 x_*^n 是最接近均值参考点 c^p 的样本。虽然损失函数只用到了作为正样本均值的参考点，但在反向传播过程中涉及到所有正样本。

如果用所有正样本计算得到的均值参考点不满足约束 $\|f(x_i^p) - c^p\|_2^2 + \alpha \leq \|f(x_*^n) - c^p\|_2^2$ ，所有的正样本都需要反向传播，正样本 x_i^p 对损失的偏导数为：

$$\frac{\partial L}{\partial f(x_i^p)} = f(x_i^p) - c^p + \frac{1}{N^p} (f(x_*^n) - f(x_i^p))$$

关于正样本 x_i^p ($k! = i$)对损失的偏导数为:

$$\frac{\partial L}{\partial f(x_k^p)} = \frac{1}{N^p} (f(x_*^n) - f(x_i^p))$$

关于负样本(x_*^n)对损失的的偏导数为:

$$\frac{\partial L}{\partial f(x_*^n)} = c^p - f(x_*^n)$$

可理解的是, 下面的融入类内多样性的均值三元组损失是在上述三元组损失的基础上进一步优化的。本实施例可以选择上述 1 和 2 实现, 也可以选择 1 和 3 实现。本发明不对其进行限定。

3. 融入类内多样性的均值三元组损失

为了保存与类内属性相似的样本相对距离更近, 本发明在提出的均值三元组损失中加入了类内多样性。

令 c^p 代表类别 c 中所有样本的均值中心, $c^{p,g}$ 代表类别 c 的组 g 中所有样本的均值中心。每一个类别 c 有一个类中心 c^p 和 G 个组中心 $c^{p,g}$ 。

对于类内多样性, 令 x_i^p , x_j^p 代表两个取自不同组的同类样本。

对于类间关系, $x_k^p \in c$ 是正样本, $x_*^n \notin c$ 是负样本。

将类内方差加入三元组中, 约束为:

$$\begin{aligned} \|c^p - f(x_i^p)\|_2^2 + \alpha_1 &\leq \|c^p - f(x_*^n)\|_2^2 \\ \|c^{p,g} - f(x_i^p)\|_2^2 + \alpha_2 &\leq \|c^{p,g} - f(x_j^p)\|_2^2 \end{aligned}$$

这里 α_1 是类中心与不同类别样本距离的间隔, α_2 是类内组中心与不同组别样本距离的间隔。因此, 加入类内多样性的均值三元组损失可表示为:

$$\begin{aligned}
L_{ICV-triplet} &= L_{inter}(c^p, x_k^p, x_*^n) + \sum_{g=1}^G L_{intra}(c^{p,g}, x_i^p, x_j^p) \\
&= \sum_{k=1}^{N^p} \frac{1}{2} \max \left\{ \|c^p - f(x_k^p)\|_2^2 + \alpha_1 - \|c^p - f(x_*^n)\|_2^2, 0 \right\} \\
&\quad + \sum_{g=1}^G \sum_{i=1}^{N^{p,g}} \frac{1}{2} \max \left\{ \|c^{p,g} - f(x_i^p)\|_2^2 + \alpha_2 - \|c^{p,g} - f(x_j^p)\|_2^2, 0 \right\}
\end{aligned}$$

以下以监控场景下的车辆精确检索为例进行说明。

本实施例着重介绍如何将本发明应用在车辆的精确检索问题上。需要注意的是，本实施例针对的问题是针对于车牌识别失效的场景如无牌、套牌、遮挡牌照等情况，利用车辆本身的细节特性进行卡口车辆精确检索。车辆本身的细节包括车辆上的喷绘，年检表、纸巾盒，装饰性图案、喷漆等。

现有技术中在模型训练阶段，对训练图像进行聚类获得类内的组别信息，如图 2 所示，现有技术中的针对同一款捷豹车的聚类效果图。图 2 为现有基于聚类方法产生的车辆图像类内分组示意图。对车辆数据集进行组内划分，同一分组下的图像体现了相似的属性，如：角度，颜色等，其中聚类中心 $K=5$ 。

图 3 为本实施例的以 VGGM 为基本网络结构的训练阶段示意图。该网络结构的损失函数由两部分组成，分别是 Softmax 损失函数和融入类内多样性的均值三元组损失函数。

本实施例的深度神经网络结构图如图 3 所示，此处以 VGG_M 神经网络为例。网络的训练过程是一个多损失函数的优化过程包括 Softmax 损失函数与本发明中提出的融入类内多样性的均值三元组损失函数。两个损失函数均连接在网络的最后一层全连接层 FC7 之

后，在 VGG_M 网络中，FC7 层的网络输出是 1024 维。在特征进入融入类内多样性的均值三元组损失之前需经过 L2 归一化层。

关于标签，每个训练样本均标注一个类别标签和组别 ID。对于 Softmax 损失函数，需要提供类别标签，对于融入类内多样性的均值三元组损失函数需要提供类别标签与类内的组别 ID。

在实际的测试阶段，可以将训练的网络中的损失函数层都除去，测试图片送入网络前向传播并提取 FC7 层的输出向量作为图片的特征表达。图片之间的相似度只需计算特征向量之间的欧式距离。在检索过程中，依次比较查询图片与数据库中的参考图片之间的特征之间的欧氏距离，按从小到大排序寻找与查询图像最像的目标对象。

由此，本发明中提出的融入类内多样性的均值三元组损失函数，能实现类内个体具有相似属性的距离更接近，相比传统的三元组损失函数训练的模型，检索的准确率得到了极大的提高；通过多损失函数联合优化的模型产生的特征更加鲁棒，检索性能比单一损失函数优化的模型更高。

本领域技术人员可以理解，实施例中的各步骤可以以硬件实现，或者以在一个或者多个处理器上运行的软件模块实现，或者以它们的组合实现。本领域的技术人员应当理解，可以在实践中使用微处理器或者数字信号处理器（DSP）来实现根据本发明实施例的一些或者全部部件的一些或者全部功能。本发明还可以实现为用于执行这里所描述的方法的一部分或者全部的设备或者装置程序（例如，计算机程序和计算机程序产品）。

虽然结合附图描述了本发明的实施方式，但是本领域技术人员可以在不脱离本发明的精神和范围的情况下做出各种修改和变型，

这样的修改和变型均落入由所附权利要求所限定的范围之内。

权利要求书

1、一种基于深度度量学习的目标精确检索方法，其特征在于，包括：

深度神经网络结构的迭代训练中，对提取的同类目标对象的多张图片的特征进行处理过程中，使得相同类别的目标对象的特征距离减小，不同类别的目标对象的特征距离增加，且属于同一类别的个体的特征分布中，具有相似属性的类内个体之间的特征距离减小，具有不同属性的类内个体之间的特征距离大于预设距离，以获得训练后的深度神经网络结构；

采用训练后的深度神经网络结构对待查询图片与预设的参考图片分别提取各自的特征，并获取查询图片与参考图片之间特征的欧式距离，对该距离进行从小到大的排序，获得精确检索的目标。

2、根据权利要求1所述的方法，其特征在于，所述深度神经网络结构为根据选择对象的属性信息能够区分不同对象类别的网络结构；和/或，

迭代训练之前的深度神经网络结构为通用的深度卷积神经网络结构。

3、根据权利要求2所述的方法，其特征在于，深度神经网络结构的迭代训练中，还用于执行前向传播计算损失，并通过反向传播算法调整深度神经网络结构中的权重值，使得训练后的深度神经网络结构中损失收敛。

4、根据权利要求3所述的方法，其特征在于，所述深度卷积神经网络的最后一层全连接的输出有两个分支，一个分支连接Softmax归一化指数损失函数，另一分支连接融入类内多样性的均值三元组损失函数；

迭代训练中执行前向传播计算的损失为：Softmax损失函数和均值三元组损失函数根据各自的权重值确定的损失。

5、根据权利要求4所述的方法，其特征在于，在深度卷积神经网络迭代训练过程中，融入类内多样性的均值三元组损失函数的损失确定方式如下：

采用KmeansK均值聚类算法对类内的所有特征点进行聚类，得到类内的若干分组；

将一个类别中所有样本均值作为该类内均值点；以及

根据聚类得到的类内若干分组，在一个类内计算组内均值，作为组内均值点；根据所有样本的类别标签和每一类内的组别标识，获取所有正样本、负样本分别到类内均值点和各自分组后的组内均值点的距离，并计算融入类内多样性的均值三元组损失函数。

6、根据权利要求5所述的方法，其特征在于，迭代训练中执行前向传播计算的损失为：Softmax损失函数和融入类内多样性的均值三元组损失函数根据各自的权重值确定的损失，包括：

根据 $L_{GS-TRS} = \omega L_{softmax} + (1 - \omega) L_{ICV-triplet}$ 确定前向传播计算的损失 L_{GS-TRS} ；

其中， ω 是加权权重， $L_{softmax}$ 为Softmax损失函数的损失， $L_{ICV-triplet}$ 为均值三元组损失函数的损失。

7、根据权利要求5所述的方法，其特征在于，采用Kmeans聚类算法对类内的所有特征点进行聚类，得到类内的若干分组，包括：

使用 $\argmin \sum_{g=1}^G \sum_{x=1}^{N^{p,g}} \|f(x) - \mu_g\|^2$ 对类内的所有特征点进行聚类，得到类内的若干分组；

其中， $f(x)$ 是提取的图像的特征， G 是组数，代表聚类族的数目， $N^{p,g}$ 是在组 $S_{c,g}$ 内的样本数量， μ_g 是聚类中心。

8、根据权利要求5所述的方法，其特征在于，将一个类别中所

有样本均值作为该类内均值点，包括：

若一个包含 N^p 个类内样本集 X^p 为 $X^p = \{x_1^p, x_2^p, \dots, x_{N^p}^p\}$, $1 \leq i \leq N^p$ ，则该类内样本集的内均值参考点 c^p 为： $c^p = \frac{1}{N^p} \sum_i^{N^p} f(x_i^p)$ 。

9、根据权利要求5所述的方法，其特征在于，根据聚类得到的类内若干分组，在一个类内计算组内均值，作为组内均值点，包括：

若一个包含 N^p 个组内样本集 X^p 为 $X^p = \{x_1^p, x_2^p, \dots, x_{N^p}^p\}$, $1 \leq i \leq N^p$ ，则组内均值参考点 c^p 为： $c^p = \frac{1}{N^p} \sum_i^{N^p} f(x_i^p)$ 。

10、根据权利要求5所述的方法，其特征在于，根据所有样本的类别标签和每一类内的组别标识，获取所有正样本、负样本分别到类内均值点和各自分组后的组内均值点的距离，并计算融入类内多样性的均值三元组损失函数，包括：

根据公式一获取融入类内多样性的三元组损失函数的损失值；

公式一： $L_{ICV-triplet} = L_{inter}(c^p, x_k^p, x_*^n) + \sum_{g=1}^G L_{intra}(c^{p,g}, x_i^p, x_j^p)$

$$= \sum_{k=1}^{N^p} \frac{1}{2} \max \left\{ \|c^p - f(x_k^p)\|_2^2 + \alpha_1 - \|c^p - f(x_*^n)\|_2^2, 0 \right\} \\ + \sum_{g=1}^G \sum_{i=1}^{N^{p,g}} \frac{1}{2} \max \left\{ \|c^{p,g} - f(x_i^p)\|_2^2 + \alpha_2 - \|c^{p,g} - f(x_j^p)\|_2^2, 0 \right\}$$

其中， c^p 代表类别 c 中所有样本的均值中心， $c^{p,g}$ 代表类别 c 的组 g 中所有样本的均值中心；每一个类别 c 有一个类中心 c^p 和 G 个组中心 $c^{p,g}$ ，

x_i^p, x_j^p 代表两个取自不同组的同类样本。对于类间关系， $x_k^p \in c$ 是正样本， $x_*^n \notin c$ 是负样本；

α_1 是类中心与不同类别样本距离的间隔， α_2 是类内组中心与不同

组别样本距离的间隔；

反向传播的偏导数：

正样本 x_i^p 对损失的偏导数为：

$$\frac{\partial L}{\partial f(x_i^p)} = f(x_i^p) - c^p + \frac{1}{N^p} (f(x_*^n) - f(x_i^p))$$

负样本 x_*^n 对损失的的偏导数为：

$$\frac{\partial L}{\partial f(x_*^n)} = c^p - f(x_*^n)。$$

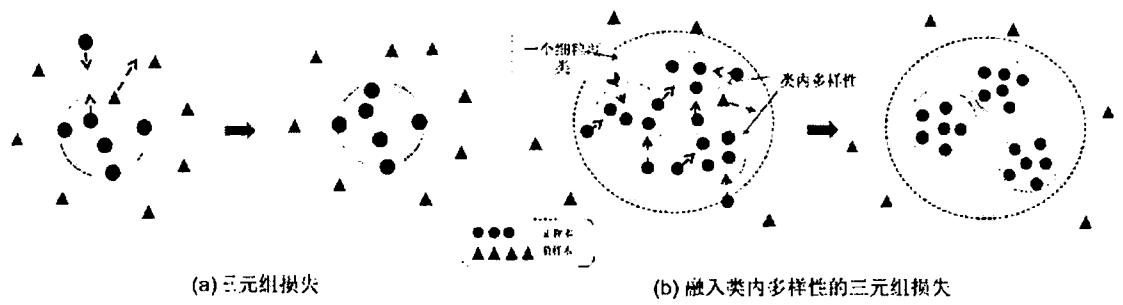


图 1

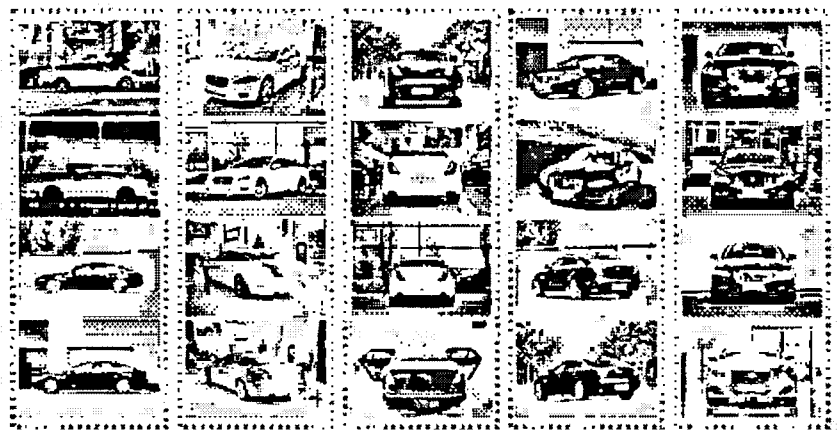


图 2

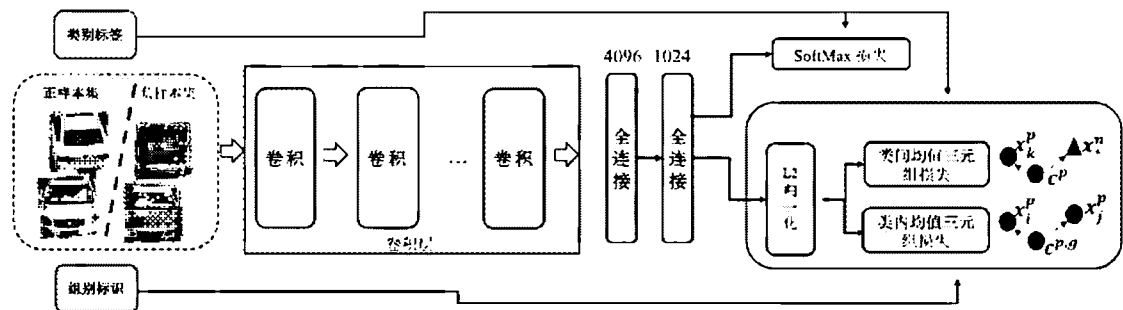


图 3

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2017/104397

A. CLASSIFICATION OF SUBJECT MATTER

G06F 17/30 (2006.01) i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F; G06K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNPAT, WPI, EPODOC, CNKI, IEEE, GOOGLE: 深度度量学习, 搜索, 检索, 神经网络, 图像, 图片, 特征, 距离, 相似, 传播, depth, measuremen, search, neural network, picture, image, feature, distance, similarity, propagation

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 106897390 A (PEKING UNIVERSITY), 27 June 2017 (27.06.2017), claims 1-10	1-10
X	CN 105808732 A (PEKING UNIVERSITY), 27 July 2016 (27.07.2016), description, paragraphs [0027]-[0054] and [0067]-[0086]	1-10
X	CN 106203242 A (ALIBABA GROUP HOLDING LIMITED), 07 December 2016 (07.12.2016), description, paragraphs [0127]-[0157]	1-10
A	CN 106022226 A (TONGJI UNIVERSITY), 12 October 2016 (12.10.2016), entire document	1-10
A	US 2007297675 A1 (LEE, S.J.J.), 27 December 2007 (27.12.2007), entire document	1-10
A	US 2014307958 A1 (THE PENN STATE RESEARCH FOUNDATION), 16 October 2014 (16.10.2014), entire document	1-10

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents:	“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
“A” document defining the general state of the art which is not considered to be of particular relevance	“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
“E” earlier application or patent but published on or after the international filing date	“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	“&” document member of the same patent family
“O” document referring to an oral disclosure, use, exhibition or other means	
“P” document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search 14 November 2017	Date of mailing of the international search report 09 January 2018
Name and mailing address of the ISA State Intellectual Property Office of the P. R. China No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088, China Facsimile No. (86-10) 62019451	Authorized officer KONG, Xin Telephone No. (86-10) 62413668

International application No.
PCT/CN2017/104397

Form PCT/ISA/210 (patent family annex) (July 2009)

国际检索报告

国际申请号

PCT/CN2017/104397

A. 主题的分类

G06F 17/30(2006.01)i

按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

G06F; G06K

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用))

CNPAT, WPI, EPDOC, CNKI, IEEE, GOOGLE: 深度度量学习, 搜索, 检索, 神经网络, 图像, 图片, 特征, 距离, 相似, 传播, depth, measurement, search, neural network, picture, image, feature, distance, similarity, propagation

C. 相关文件

类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
PX	CN 106897390 A (北京大学) 2017年 6月 27日 (2017 - 06 - 27) 权利要求1-10	1-10
X	CN 105808732 A (北京大学) 2016年 7月 27日 (2016 - 07 - 27) 说明书第[0027]-[0054], [0067]-[0086]段	1-10
X	CN 106203242 A (阿里巴巴集团控股有限公司) 2016年 12月 7日 (2016 - 12 - 07) 说明书第[0127]-[0157] 段	1-10
A	CN 106022226 A (同济大学) 2016年 10月 12日 (2016 - 10 - 12) 全文	1-10
A	US 2007297675 A1 (LEE, SHIH-JONG J.) 2007年 12月 27日 (2007 - 12 - 27) 全文	1-10
A	US 2014307958 A1 (THE PENN STATE RESEARCH FOUNDATION) 2014年 10月 16日 (2014 - 10 - 16) 全文	1-10

☐ 其余文件在C栏的续页中列出。☒ 见同族专利附件。

* 引用文件的具体类型:

“A” 认为不特别相关的表示了现有技术一般状态的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的)

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件

“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性

“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性

“&” 同族专利的文件

国际检索实际完成的日期

2017年 11月 14日

国际检索报告邮寄日期

2018年 1月 9日

ISA/CN的名称和邮寄地址

中华人民共和国国家知识产权局(ISA/CN)
中国北京市海淀区蓟门桥西土城路6号 100088

传真号 (86-10)62019451

授权官员

孔昕

电话号码 (86-10)62413668

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2017/104397

检索报告引用的专利文件			公布日 (年/月/日)	同族专利	公布日 (年/月/日)
CN	106897390	A	2017年 6月 27日	无	
CN	105808732	A	2016年 7月 27日	无	
CN	106203242	A	2016年 12月 7日	WO 2016177259 A1	2016年 11月 10日
CN	106022226	A	2016年 10月 12日	无	
US	2007297675	A1	2007年 12月 27日	无	
US	2014307958	A1	2014年 10月 16日	无	

表 PCT/ISA/210 (同族专利附件) (2009年7月)