



(12) 发明专利

(10) 授权公告号 CN 110175615 B

(45) 授权公告日 2021.01.01

(21) 申请号 201910350741.5

CN 107122712 A, 2017.09.01

(22) 申请日 2019.04.28

CN 107967457 A, 2018.04.27

(65) 同一申请的已公布的文献号

CN 108647577 A, 2018.10.12

申请公布号 CN 110175615 A

US 2018181804 A1, 2018.06.28

WO 2018184195 A1, 2018.10.11

(43) 申请公布日 2019.08.27

Relja Arandjelovic 等. NetVLAD: CNN

(73) 专利权人 华中科技大学

architecture for weakly supervised place recognition. 《IEEE Transactions on Pattern Analysis and Machine Intelligence》. 2018, 第40卷(第6期),

地址 430074 湖北省武汉市洪山区珞喻路1037号

(72) 发明人 桑农 刘耀华 高常鑫

Yanghao Li 等. Adaptive Batch

(74) 专利代理机构 华中科技大学专利中心

Normalization for practical domain adaptation. 《Pattern Recognition》. 2018, 第80卷

42201

代理人 曹葆青 李智

(51) Int. Cl.

G06K 9/46 (2006.01)

G06K 9/62 (2006.01)

G06N 3/04 (2006.01)

仇晓松 等. 基于卷积神经网络的视觉位置识别方法. 《计算机工程与设计》. 2019, 第40卷(第1期),

王丽君 等. 基于卷积神经网络的位置识别. 《电子科技》. 2017, 第30卷(第1期),

(56) 对比文件

CN 109684977 A, 2019.04.26

CN 107767378 A, 2018.03.06

审查员 张露

权利要求书3页 说明书9页 附图2页

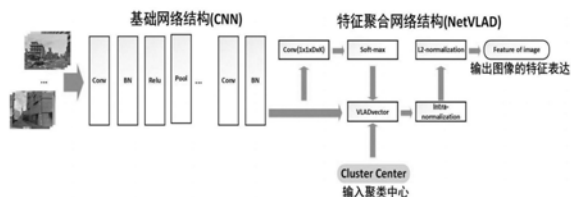
(54) 发明名称

模型训练方法、域自适应的视觉位置识别方法及装置

图像的特征向量。本发明能够提高视觉位置识别的鲁棒性。

(57) 摘要

本发明公开了一种模型训练方法、域自适应的视觉位置识别方法及装置,属于计算机视觉技术领域,包括:建立基于深度神经网络的图像特征提取模型;根据标准数据集构建训练集,训练集中每个训练样本包括目标图像及其正样本和s个负样本;利用训练集对图像特征提取模型进行训练;图像特征提取模型中,特征提取网络包括级联的多个第一网络;第一网络由一个或多个第二网络以及一个极大池化层依次连接而成,极大池化层用于特征选择;第二网络包括依次连接的卷积层,用于特征提取;批标准化层,用于进行零均值标准化处理;激活函数层,用于进行激活处理;局部特征聚合网络用于聚合局部特征以得到



CN 110175615 B

1. 一种图像特征提取模型训练方法,其特征在于,包括:

(1) 建立基于深度神经网络的图像特征提取模型,用于获取图像的特征向量;

所述图像特征提取模型包括特征提取网络和局部特征聚合网络;

所述特征提取网络包括级联的多个第一网络;所述第一网络由一个或多个第二网络以及一个极大池化层依次连接而成,所述极大池化层用于对其前的第二网络输出的图像进行特征选择;所述第二网络包括依次连接的卷积层、批标准化层以及激活函数层,所述卷积层用于对图像进行特征提取,所述批标准化层用于对所述卷积层输出的图像进行零均值标准化处理,所述激活函数层用于对所述批标准化层输出的图像进行激活处理;

所述局部特征聚合网络用于对所述特征提取网络输出的图像中所有的局部特征进行聚合,从而得到图像的特征向量;

(2) 在标准数据集中获得各目标图像的正样本和s个负样本,以由一张目标图像及其正样本和负样本构成一个训练样本,从而得到由所有训练样本构成的训练集;

目标图像的正样本为其临近图像中与其特征距离最近的图像,目标图像与其临近图像的位置距离 d_1 满足 $T_{NL} \leq d_1 \leq T_{NH}$;目标图像与其负样本的位置距离 d_2 满足 $d_2 \geq T_F$;

(3) 利用所述训练集对所述图像特征提取模型进行训练,从而得到各模型参数;

其中,所述标准数据集中各图像的位置信息已知,目标图像为所述标准数据集中预先筛选出的多张图像;图像间的特征距离为图像的特征向量之间的距离; T_{NL} 、 T_{NH} 和 T_F 均为预设的阈值, $0 < T_{NL} < T_{NH}$, $T_{NH} \leq T_F$; $s > 1$;

所述步骤(3)中,利用所述训练集对所述图像特征提取模型进行训练时,所采用的损失函数为:

$$loss = \sum_{k=1}^n \max \left(d_{q_k p_k} + m - \min_{i \in \{1, 2, \dots, s\}} (d_{q_k n_{ki}}), 0 \right);$$

其中,n为训练样本总数,k为训练样本序号,i为负样本序号, q_k 、 p_k 和 n_{ki} 分别表示第k个训练样本中的目标图像、正样本和第i个负样本, $d_{q_k p_k}$ 表示目标图像 q_k 与其正样本 p_k 之间的特征距离, $d_{q_k n_{ki}}$ 为目标图像 q_k 与其负样本 n_{ki} 之间的特征距离,m为预定义的超参数,max表示取最大值,min表示取最小值。

2. 如权利要求1所述的图像特征提取模型训练方法,其特征在于,所述局部特征聚合网络包括:降维卷积层、soft-max层、聚合层、内部归一化层以及整体归一化层;所述降维卷积层为一层卷积层,用于将待聚合图像的维度降维到与预设的聚类中心的个数相等,以使得所述待聚合图像的每个通道表示局部特征与每个聚类中心之差的权重;

所述soft-max层用于对局部特征与每个聚类中心之差的权重进行归一化;

所述聚合层用于根据局部特征、聚类中心以及归一化之后的权重聚合得到VLAD向量;

所述VLAD向量由N个D维度的向量组成,N为聚类中心个数,D为聚类中心的维度;

所述内部归一化层用于对所述VLAD向量中每个D维度的向量进行归一化,以使得每个D维度的向量的分布在同一数量级;

所述整体归一化层用于将经过所述内部归一化层处理后的D维度的向量串联为一个列向量后,对该列向量进行归一化,以使得所述待聚合图像的每个局部特征分布在同一数量

级；

其中，所述待聚合图像为所述特征提取网络输出的图像。

3. 一种基于权利要求1-2任一项所述的图像特征提取模型训练方法的域自适应的视觉位置识别方法，其特征在于，包括：

确定待识别图像所属的目标域，并获得所述目标域中不同位置处的多张图像，将所获取的图像与所述待识别图像均作为待检索图像；

以所述待检索图像为输入，利用所述图像特征提取模型获得各待检索图像的特征向量；获取图像特征向量时，对于每一个卷积层，统计所有待检索图像经过该卷积层后，所得到的特征图的均值和标准差，作为该卷积层之后的一层批标准化层的参数；所述图像特征提取模型中其余的模型参数为训练所得的模型参数；

利用所述图像特征提取模型获取测试数据集中各图像的特征向量；

根据所获取的特征向量获得所述测试数据集中与所述待识别图像的特征距离最近的图像，并将该图像的位置信息确定为所述待识别图像的位置信息，从而完成对所述待识别图像的视觉位置识别；

其中，所述测试数据集中各图像的位置信息已知，域为影响图像特征分布的因素集合。

4. 如权利要求3所述的域自适应的视觉位置识别方法，其特征在于，利用所述图像特征提取模型获取测试数据集中各图像的特征向量时，各模型参数的设置方式为：

利用训练所得的模型参数设置各模型参数；

或者，对于每一个卷积层，统计所述测试数据集中所有图像经过该卷积层后，所得到的特征图的均值和标准差，作为该卷积层之后的一层批标准化层的参数；所述图像特征提取模型中其余的模型参数为训练所得的模型参数。

5. 一种图像特征提取模型训练装置，其特征在于，包括：模型建立模块、训练集构造模块以及模型训练模块；

所述模型建立模块用于建立基于深度神经网络的图像特征提取模型，所述图像特征提取模型用于获取图像的特征向量；

所述训练集构造模块用于在标准数据集中，获得各目标图像的正样本和s个负样本，以由一张目标图像及其正样本和负样本构成一个训练样本，从而得到由所有训练样本构成的训练集；

所述模型训练模块用于利用所述训练集对所述图像特征提取模型进行训练，从而得到各模型参数；

其中，所述图像特征提取模型包括特征提取网络和局部特征聚合网络；

所述特征提取网络包括级联的多个第一网络；所述第一网络由一个或多个第二网络以及一个极大池化层依次连接而成，所述极大池化层用于对其前的第二网络输出的图像进行特征选择；所述第二网络包括依次连接的卷积层、批标准化层以及激活函数层，所述卷积层用于对图像进行特征提取，所述批标准化层用于对所述卷积层输出的图像进行零均值标准化处理，所述激活函数层用于对所述批标准化层输出的图像进行激活处理；

所述局部特征聚合网络用于对所述特征提取网络输出的图像中所有的局部特征进行聚合，从而得到图像的特征向量；

目标图像的正样本为其临近图像中与其特征距离最近的图像，目标图像与其临近图像

的位置距离 d_1 满足 $T_{NL} \leq d_1 \leq T_{NH}$; 目标图像与其负样本的位置距离 d_2 满足 $d_2 \geq T_F$;

所述标准数据集中各图像的位置信息已知, 目标图像为所述标准数据集中预先筛选出的多张图像; 图像间的特征距离为图像的特征向量之间的距离; T_{NL} 、 T_{NH} 和 T_F 均为预设的阈值, $0 < T_{NL} < T_{NH}$, $T_{NH} \leq T_F$; $s > 1$;

所述模型训练模块利用所述训练集对所述图像特征提取模型进行训练时, 所采用的损失函数为:

$$loss = \sum_{k=1}^n \max \left(d_{q_k p_k} + m - \min_{i \in \{1, 2, \dots, s\}} (d_{q_k n_{ki}}), 0 \right);$$

其中, n 为训练样本总数, k 为训练样本序号, i 为负样本序号, q_k 、 p_k 和 n_{ki} 分别表示第 k 个训练样本中的目标图像、正样本和第 i 个负样本, $d_{q_k p_k}$ 表示目标图像 q_k 与其正样本 p_k 之间的特征距离, $d_{q_k n_{ki}}$ 为目标图像 q_k 与其负样本 n_{ki} 之间的特征距离, m 为预定义的超参数, \max 表示取最大值, \min 表示取最小值。

6. 一种基于权利要求1-2任一项所述的图像特征提取模型训练方法的域自适应的视觉位置识别装置, 其特征在在于, 包括: 检索集获取模块、第一特征提取模块、第二特征提取模块以及识别模块;

所述检索集获取模块用于确定待识别图像所属的目标域, 并获得所述目标域中不同位置处的多张图像, 将所获取的图像与所述待识别图像均作为待检索图像;

所述第一特征提取模块用于以所述待检索图像为输入, 利用所述图像特征提取模型获得各待检索图像的特征向量; 获取图像特征向量时, 对于每一个卷积层, 统计所有待检索图像经过该卷积层后, 所得到的特征图的均值和标准差, 作为该卷积层之后的一层批标准化层的参数; 所述图像特征提取模型中其余的模型参数为训练所得的模型参数;

所述第二特征提取模块用于利用所述图像特征提取模型获取测试数据集中各图像的特征向量;

所述识别模块用于根据所述第一特征提取模块和所述第二特征提取模块所提取的特征向量, 获得所述测试数据集中与所述待识别图像的特征距离最近的图像, 并将该图像的位置信息确定为所述待识别图像的位置信息, 从而完成对所述待识别图像的视觉位置识别;

其中, 所述测试数据集中各图像的位置信息已知, 域为影响图像特征分布的因素集合。

模型训练方法、域自适应的视觉位置识别方法及装置

技术领域

[0001] 本发明属于计算机视觉技术领域,更具体地,涉及一种模型训练方法、域自适应的视觉位置识别方法及装置。

背景技术

[0002] 视觉位置识别,具体是指通过对图像进行特征提取,然后根据所提取的图像特征识别图像的地理位置。在现如今大力发展自动驾驶,自主导航移动机器人的需要越来越高以及虚拟现实和增强现实越来越普遍的情况下,视觉位置识别的研究在计算机视觉领域、机器人社区及其他相关领域中均引起了广泛的关注。

[0003] 在早期的计算机视觉研究时期,主要利用人工精心设计的提取图像特征点的方法提取图像特征,如尺度不变特征变换(SIFT)特征点。所提取特征的设计非常依赖经验,一些专家学者甚至历时数十年才能设计出一个比较好的特征,并且这些利用手工设计的提取图像特征点的算法在急剧的光照变化(如白天到夜间)和场景变化(场景中的行人和车辆发生编发)等情况下效果非常差,依赖这些特征的视觉位置识别方法,如视觉词袋模型(V-BOW)等,其性能也会急剧下降。近年来,随着深度学习的兴起,并被广泛应用于目标识别、目标检测、目标跟踪、语义分割等领域,一些基于深度学习的视觉位置识别方法被提了出来。例如,基于卷积神经网络的位置识别(Convolutional Neural Network-based Place Recognition),该方法利用深度卷积神经网络提取图像特征,由于深度卷积神经网络可以根据特定的任务进行端到端的训练,提取的图像特征更具有鲁棒性。又例如,用于弱监督位置识别的CNN架构NetVLAD(NetVLAD:CNN architecture for weakly supervised place recognition),该方法发挥了传统的局部特征聚合(VLAD)方法的优势,有效地聚合图像的局部特征,得到紧凑的图像表达特征向量,并该方法还使的利用深度神经网络提取的图像特征更具有鲁棒性。

[0004] 相比于传统的基于手工设计图像特征点的视觉位置识别方法,基于深度神经网络的视觉位置识别方法所提取的图像特征更具有鲁棒性,并且视觉位置识别较为准确。但是,深度神经网络在使用前需要进行训练,而由于视角、光照等因素的影响,用于训练的图像的特征分布与实际待识别图像的特征分布往往具有较大差异,在这种情况下,视觉位置识别的准确度得不到保证。总的来说,现有视觉位置识别方法的鲁棒性较低。

发明内容

[0005] 针对现有技术的缺陷和改进需求,本发明提供了一种模型训练方法、域自适应的视觉位置识别方法及装置,其目的在于,提高视觉位置识别的鲁棒性。

[0006] 为实现上述目的,按照本发明的第一方面,提供了一种图像特征提取模型训练方法,包括:

[0007] (1) 建立基于深度神经网络的图像特征提取模型,用于获取图像的特征向量;

[0008] 图像特征提取模型包括特征提取网络和局部特征聚合网络;

[0009] 特征提取网络包括级联的多个第一网络;第一网络由一个或多个第二网络以及一个极大池化层依次连接而成,极大池化层用于对其前的第二网络输出的图像进行特征选择;第二网络包括依次连接的卷积层、批标准化层以及激活函数层,卷积层用于对图像进行特征提取,批标准化层用于对卷积层输出的图像进行零均值标准化处理,激活函数层用于对批标准化层输出的图像进行激活处理;

[0010] 局部特征聚合网络用于对特征提取网络输出的图像中所有的局部特征进行聚合,从而得到图像的特征向量;

[0011] (2) 在标准数据集中,获得各目标图像的正样本和 s 个负样本,以由一张目标图像及其正样本和负样本构成一个训练样本,从而得到由所有训练样本构成的训练集;

[0012] 目标图像的正样本为其临近图像中与其特征距离最近的图像,目标图像与其临近图像的位置距离 d 满足 $T_{NL} \leq d < T_{NH}$;目标图像与其负样本的位置距离满足 $d \geq T_F$;

[0013] (3) 利用训练集对图像特征提取模型进行训练,从而得到各模型参数;

[0014] 其中,标准数据集中各图像的位置信息已知,目标图像为标准数据集中预先筛选出的多张图像;图像间的特征距离为图像的特征向量之间的距离; T_{NL} 、 T_{NH} 和 T_F 均为预设的阈值, $0 < T_{NL} < T_{NH}$, $T_{NH} \leq T_F$; $s \geq 1$ 。

[0015] 上述图像特征提取模型训练方法,所建立的图像特征提取模型中,每个用于特征提取的卷积之后都由一个批标准化层(Batch Normalization)对该卷积层所输出的图像进行零均值标准化处理,由此能够在加速模型训练的同时,使得经该图像特征提取模型所提取的图像特征都有相似分布,从而有效避免由于训练集中图像的特征分布差异较大而导致模型训练效果不佳,进而能够改善图像特征分布差异较大时,视觉位置识别的鲁棒性较低的问题。

[0016] 进一步地,局部特征聚合网络包括:降维卷积层、soft-max层、聚合层、内部归一化层以及整体归一化层;

[0017] 降维卷积层为一层卷积层,用于将待聚合图像的维度降维到与预设的聚类中心的个数相等,以使得待聚合图像的每个通道表示局部特征与每个聚类中心之差的权重;

[0018] soft-max层用于对局部特征与每个聚类中心之差的权重进行归一化;

[0019] 聚合层用于根据局部特征、聚类中心以及归一化之后的权重聚合得到VLAD(vector of locally aggregated descriptors)向量;VLAD向量由 N 个 D 维度的向量组成, N 为聚类中心个数, D 为聚类中心的维度;

[0020] 内部归一化层用于对VLAD向量中每个 D 维度的向量进行归一化,以使得每个 D 维度的向量的分布在同一数量级;

[0021] 整体归一化层用于将经过内部归一化层处理后的 D 维度的向量串联为一个列向量后,对该列向量进行归一化,以使得待聚合图像的每个局部特征分布在同一数量级;由此可以提高神经网络模型的收敛速度和网络模型的精度;

[0022] 其中,待聚合图像为特征提取网络输出的图像。

[0023] 进一步地, $s > 1$;通过选定多个负样本能够提高模型的训练精度,使得利用上述图像特征提取模型获取的图像特征向量进行视觉位置识别时,具有较高的鲁棒性。

[0024] 进一步地,步骤(3)中,利用训练集对图像特征提取模型进行训练时,所采用的损失函数为:

$$[0025] \quad loss = \sum_{k=1}^n \max \left(d_{q_k p_k} + m - \min_{i \in \{1, 2, \dots, s\}} (d_{q_k n_{ki}}), 0 \right);$$

[0026] 其中, n 为训练样本总数, k 为训练样本序号, i 为负样本序号, q_k 、 p_k 和 n_{ki} 分别表示第 k 个训练样本中的目标图像、正样本和第 i 个负样本, $d_{q_k p_k}$ 表示目标图像 q_k 与其正样本 p_k 之间的特征距离, $d_{q_k n_{ki}}$ 为目标图像 q_k 与其负样本 n_{ki} 之间的特征距离, m 为预定义的超参数, \max 表示取最大值, \min 表示取最小值;

[0027] 上述损失函数, 基于三元组损失的思想, 使得通过训练, 目标图像与正样本的特征距离最小化, 同时到负样本的特征距离最大化; 其中通过 $\min_{i \in \{1, 2, \dots, s\}} (d_{q_k n_{ki}})$ 这一项选择了损失最大的负样本, 由此能够基于难例挖掘的思想, 使得模型训练过程中更注意比较难于辨别的负样本, 进而可以在利用上述图像特征提取模型进行视觉位置识别时, 避免与待识别图像相似的负样本的干扰。

[0028] 按照本发明的第二方面, 还提供了一种基于本发明第一方面所提供的图像特征提取模型训练方法的域自适应的视觉位置识别方法, 包括:

[0029] 确定待识别图像所属的目标域, 并获得目标域中不同位置处的多张图像, 将所获取的图像与待识别图像均作为待检索图像;

[0030] 以待检索图像为输入, 利用图像特征提取模型获得各待检索图像的特征向量; 获取图像特征向量时, 对于每一个卷积层, 统计所有待检索图像经过该卷积层后, 所得到的特征图的均值和标准差, 作为该卷积层之后的一层批标准化层的参数; 图像特征提取模型中其余的模型参数为训练所得的模型参数;

[0031] 利用图像特征提取模型获取测试数据集中各图像的特征向量;

[0032] 根据所获取的特征向量获得测试数据集中与待识别图像的特征距离最近的图像, 并将该图像的位置信息确定为待识别图像的位置信息, 从而完成对待识别图像的视觉位置识别;

[0033] 其中, 测试数据集中各图像的位置信息已知, 域为影响图像特征分布的因素集合;

[0034] 根据实际应用中, 会根据光照、视角、季节等因素对图像特征分布的影响情况完成域的划定, 同一个域中, 图像的特征分布相似; 例如, 如果仅光照会对图像的特征分布产生较大的影响, 并且白天拍摄的图像具有相似的特征分布, 夜间拍摄的图像具有相似的特征分布, 则可根据光照条件划分得到两个域;

[0035] 上述域自适应的视觉位置识别方法, 在利用图像特征提取模型获取待识别图像的特征向量时, 模型中各批标准化层的参数不依赖于训练集, 而是利用多个与该待识别图像属于同一个域的图像获取相应的参数, 由于同一个域中的图像具有相似的特征分布, 因此, 本发明能够实现域自适应, 在训练集中图像与该待识别图像的特征分布差异较大时, 仍然能够准确完成视觉位置识别, 也即是说, 本发明能够提高视觉位置识别的鲁棒性。

[0036] 进一步地, 利用图像特征提取模型获取测试数据集中各图像的特征向量时, 各模型参数的设置方式为:

[0037] 利用训练所得的模型参数设置各模型参数;

[0038] 或者, 对于每一个卷积层, 统计测试数据集中所有图像经过该卷积层后, 所得到的

特征图的均值和标准差,作为该卷积层之后的一层批标准化层的参数;图像特征提取模型中其余的模型参数为训练所得的模型参数。

[0039] 按照本发明的第三方面,提供了一种图像特征提取模型训练装置,包括:模型建立模块、训练集构造模块以及模型训练模块;

[0040] 模型建立模块用于建立基于深度神经网络的图像特征提取模型,图像特征提取模型用于获取图像的特征向量;

[0041] 训练集构造模块用于在标准数据集中,获得各目标图像的正样本和s个负样本,由一张目标图像及其正样本和负样本构成一个训练样本,从而得到由所有训练样本构成的训练集;

[0042] 模型训练模块用于利用训练集对图像特征提取模型进行训练,从而得到各模型参数;

[0043] 其中,图像特征提取模型包括特征提取网络和局部特征聚合网络;

[0044] 特征提取网络包括级联的多个第一网络;第一网络由一个或多个第二网络以及一个极大池化层依次连接而成,极大池化层用于对其前的第二网络输出的图像进行特征选择;第二网络包括依次连接的卷积层、批标准化层以及激活函数层,卷积层用于对图像进行特征提取,批标准化层用于对卷积层输出的图像进行零均值标准化处理,激活函数层用于对批标准化层输出的图像进行激活处理;

[0045] 局部特征聚合网络用于对特征提取网络输出的图像中所有的局部特征进行聚合,从而得到图像的特征向量;

[0046] 目标图像的正样本为其临近图像中与其特征距离最近的图像,目标图像与其临近图像的位置距离d满足 $T_{NL} \leq d < T_{NH}$;目标图像与其负样本的位置距离d满足 $d \geq T_F$;

[0047] 标准数据集中各图像的位置信息已知,目标图像为标准数据集中预先筛选出的多张图像;图像间的特征距离为图像的特征向量之间的距离; T_{NL} 、 T_{NH} 和 T_F 均为预设的阈值, $0 < T_{NL} < T_{NH}$, $T_{NH} \leq T_F$; $s \geq 1$ 。

[0048] 按照本发明的第四方面,还提供了一种基于本发明第一方面所提供的图像特征提取模型训练方法的域自适应的视觉位置识别装置,包括:检索集获取模块、第一特征提取模块、第二特征提取模块以及识别模块;

[0049] 检索集获取模块用于确定待识别图像所属的目标域,并获得目标域中不同位置处的多张图像,将所获取的图像与待识别图像均作为待检索图像;

[0050] 第一特征提取模块用于以待检索图像为输入,利用图像特征提取模型获得各待检索图像的特征向量;获取图像特征向量时,对于每一个卷积层,统计所有待检索图像经过该卷积层后,所得到的特征图的均值和标准差,作为该卷积层之后的一层批标准化层的参数;图像特征提取模型中其余的模型参数为训练所得的模型参数;

[0051] 第二特征提取模块用于利用图像特征提取模型获取测试数据集中各图像的特征向量;

[0052] 识别模块用于根据第一特征提取模块和第二特征提取模块所提取的特征向量,获得测试数据集中与待识别图像的特征距离最近的图像,并将该图像的位置信息确定为待识别图像的位置信息,从而完成对待识别图像的视觉位置识别;

[0053] 其中,测试数据集中各图像的位置信息已知,域为影响图像特征分布的因素集合。

[0054] 总体而言,通过本发明所构思的以上技术方案,能够取得以下有益效果:

[0055] (1) 本发明所提供的图像特征提取模型训练方法,所建立的图像特征提取模型中,每个用于特征提取的卷积之后都由一个批标准化层对该卷积层所输出的图像进行零均值标准化处理,由此能够在加速模型训练的同时,使得经该图像特征提取模型所提取的图像特征都有相似的分布,从而有效避免由于训练集中图像的特征分布差异较大而导致模型训练效果不佳,进而能够改善图像特征分布差异较大时,视觉位置识别的鲁棒性较低的问题。

[0056] (2) 本发明所提供的图像特征提取模型训练方法,在其优选方案中,通过选定多个负样本完成训练样本的构建,能够提高模型的训练精度,使得利用上述图像特征提取模型获取的图像特征向量进行视觉位置识别时,具有较高的鲁棒性。

[0057] (3) 本发明所提供的图像特征提取模型训练方法,在其优选方案中,同时基于三元组损失和难例挖掘的思想构建损失函数,使得模型训练过程中更注意比较难于辨别的负样本,进而可以在利用上述图像特征提取模型进行视觉位置识别时,避免与待识别图像相似的负样本的干扰。

[0058] (4) 本发明所提供的域自适应的视觉位置识别方法,在利用图像特征提取模型获取待识别图像的特征向量时,模型中各批标准化层的参数不依赖于训练集,而是利用多个与该待识别图像属于同一个域的图像获取相应的参数,由此实现了域自适应,从而能够提高视觉位置识别的鲁棒性。

附图说明

[0059] 图1为本发明实施例提供的图像特征提取模型示意图;

[0060] 图2为本发明实施例提供的域自适应的视觉位置识别方法流程图;

[0061] 图3为本发明实施例提供的域自适应的视觉位置识别方法示意图。

具体实施方式

[0062] 为了使本发明的目的、技术方案及优点更加清楚明白,以下结合附图及实施例,对本发明进行进一步详细说明。应当理解,此处所描述的具体实施例仅仅用以解释本发明,并不用于限定本发明。此外,下面所描述的本发明各个实施方式中所涉及到的技术特征只要彼此之间未构成冲突就可以相互组合。

[0063] 本发明所提供的图像特征提取模型训练方法,包括:

[0064] (1) 建立基于深度神经网络的图像特征提取模型,用于获取图像的特征向量;

[0065] 如图1所示,图像特征提取模型包括特征提取网络和局部特征聚合网络;

[0066] 特征提取网络包括级联的多个第一网络;第一网络由一个或多个第二网络以及一个极大池化层(Pool)依次连接而成,极大池化层用于对其前的第二网络输出的图像进行特征选择;第二网络包括依次连接的卷积层(Conv)、批标准化层(BN)以及激活函数层(ReLU),卷积层用于对图像进行特征提取,批标准化层用于对卷积层输出的图像进行零均值标准化处理,激活函数层用于对批标准化层输出的图像进行激活处理;在本实施例中,各第二网络中卷积层的卷积核大小具体为3x3;各个第一网络所包含的第二网络的个数可以相同,也可以不同;

[0067] 局部特征聚合网络用于对特征提取网络输出的图像中所有的局部特征进行聚合,

从而得到图像的特征向量；

[0068] 在一个可选的实施方式中，如图1所示，局部特征聚合网络包括：降维卷积层(Conv)、soft-max层(Soft-max)、聚合层(VLAD)、内部归一化层(Intra-normalization)以及整体归一化层(L2-normalization)；

[0069] 降维卷积层为一层卷积层，其卷积和大小为 1×1 ，用于将待聚合图像的维度降维到与预设的聚类中心的个数相等，以使得待聚合图像的每个通道表示局部特征与每个聚类中心之差的权重；其中，待聚合图像为特征提取网络输出的图像；

[0070] soft-max层用于对局部特征与每个聚类中心之差的权重进行归一化；

[0071] 聚合层用于根据局部特征、聚类中心以及归一化之后的权重聚合得到VLAD向量；VLAD向量由 N 个 D 维度的向量组成， N 为聚类中心个数， D 为聚类中心的维度；

[0072] 设聚类中心有 N 个，用CluCenter表示， $CluCenter = [c_1, c_2, \dots, c_j, \dots, c_N]$ ，其中，每个聚类中心的维度为 D ， $c_j (j \in \{1, 2, \dots, N\})$ 表示第 j 个聚类中心；

[0073] 特征提取网络输出每张图像的局部特征为 n 个，用Features表示， $Features = [f_1, f_2, \dots, f_i, \dots, f_n]$ ，其中， $f_i (i \in \{1, 2, \dots, n\})$ 表示第 i 个局部特征；

[0074] 第 i 个局部特征与第 j 个聚类中心的差的权重用 a_{ij} 表示，可以得到VLAD向量中第 j 个 D 维度的向量VLADvector $_j$ (即VLAD向量的第 j 个元素)为：

$$[0075] \quad VLADvector_j = \sum_{c_j \in CluCenter} a_{ij}(f_i - c_j), \quad i \in (1, 2, \dots, n);$$

[0076] 内部归一化层用于对VLAD向量中每个 D 维度的向量进行归一化，以使得每个 D 维度的向量的分布在同一数量级；

[0077] 整体归一化层用于将经过内部归一化层处理后的 D 维度的向量串联为一个列向量后，对该列向量进行归一化，以使得待聚合图像的每个局部特征分布在同一数量级；由此可以提高神经网络模型的收敛速度和网络模型的精度；

[0078] 在本实施例中，内部归一化层和整体归一化层均通过L2范数归一化的方法完成归一化操作；

[0079] (2) 在标准数据集中，获得各目标图像的正样本和 s 个负样本，以由一张目标图像及其正样本和负样本构成一个训练样本，从而得到由所有训练样本构成的训练集；

[0080] 其中，标准数据集中各图像的位置信息已知，目标图像为标准数据集中预先筛选出的多张图像；

[0081] 目标图像的正样本为其临近图像中与其特征距离最近的图像，目标图像与其临近图像的位置距离 d 满足 $T_{NL} \leq d < T_{NH}$ ；目标图像与其负样本的位置距离满足 $d \geq T_F$ ；其中， T_{NL} 、 T_{NH} 和 T_F 均为预设的阈值， $0 < T_{NL} < T_{NH}$ ， $T_{NH} \leq T_F$ ； $s \geq 1$ ；图像间的特征距离为图像的特征向量之间的距离；

[0082] 在本实施例中，用于模型训练的标准数据集为TokyoTimeMachine谷歌街景数据集；该数据集包括采集自多个不同的位置、每个位置从12个角度方向采集的图像，总共大约47000张图像，每个图像带有地理坐标信息；在该数据集中，目标图像为随机选取的10000张图像，即训练样本总数为 $n = 10000$ ；在其他应用中，也可根据实际应用需求选用其他的数据集作为标准数据集；

[0083] 阈值 T_{NL} 、 T_{NH} 和 T_F 可根据所采用的标准数据集和实际的应用场景设定，一般情况下，

$T_{NH} \leq 25, 25 \leq T_F$; 在本实施例中, 阈值设置具体为 $T_{NL} = 1, T_{NH} = 10, T_F = 25$; 通过阈值 T_{NL} 和 T_{NH} 设置目标图像与其正样本间位置距离的上下限, 可以保证正样本与目标图像相似但又有所区别, 避免模型过拟合, 从而可以保证较好的模型训练效果;

[0084] 在本实施例中, 具体设置每个训练样本中, 负样本数量为 $s = 4$; 通过选定多个负样本能够提高模型的训练精度, 使得利用上述图像特征提取模型获取的图像特征向量进行视觉位置识别时, 具有较高的鲁棒性;

[0085] 本实施例所构建的训练集 $trainSet$ 具体可表示为:

$$[0086] \quad trainSet = \left[\begin{array}{l} S_1 [q_1, p_1, n_{11}, n_{12}, n_{13}, n_{14}], S_2 [q_2, p_2, n_{21}, n_{22}, n_{23}, n_{24}], \dots, S_k [q_k, p_k, n_{k1}, n_{k2}, n_{k3}, n_{k4}], \dots \\ S_{10000} [q_{10000}, p_{10000}, n_{100001}, n_{100002}, n_{100003}, n_{100004}] \end{array} \right]$$

[0087] 其中, 对于任意第 k 个训练样本 S_k , q_k , p_k 和 n_{ki} ($i \in \{1, 2, 3, 4\}$) 分别表示该训练样本中的目标图像、正样本和第 i 个负样本;

[0088] (3) 利用训练集对图像特征提取模型进行训练, 从而得到各模型参数。

[0089] 上述图像特征提取模型训练方法, 所建立的图像特征提取模型中, 每个用于特征提取的卷积之后都由一个批标准化层 (Batch Normalization) 对该卷积层所输出的图像进行零均值标准化处理, 由此能够在加速模型训练的同时, 使得经该图像特征提取模型所提取的图像特征都有相似的分布, 从而有效避免由于训练集中图像的特征分布差异较大而导致模型训练效果不佳, 进而能够改善图像特征分布差异较大时, 视觉位置识别的鲁棒性较低的问题。

[0090] 为了进一步提高视觉位置识别的鲁棒性, 上述图像特征提取模型训练方法的步骤 (3) 中, 利用训练集对图像特征提取模型进行训练时, 所采用的损失函数具体为:

$$[0091] \quad loss = \sum_{k=1}^{10000} \max \left(d_{q_k p_k} + m - \min_{i \in \{1, 2, 3, 4\}} (d_{q_k n_{ki}}), 0 \right);$$

[0092] 其中, $d_{q_k p_k}$ 表示目标图像 q_k 与其正样本 p_k 之间的特征距离, $d_{q_k n_{ki}}$ 为目标图像 q_k 与其负样本 n_{ki} 之间的特征距离, m 为预定义的超参数, \max 表示取最大值, \min 表示取最小值;

[0093] 上述损失函数, 基于三元组损失的思想, 使得通过训练, 目标图像与正样本的特征距离最小化, 同时到负样本的特征距离最大化; 其中通过 $\min_{i \in \{1, 2, \dots, s\}} (d_{q_k n_{ki}})$ 这一项选择了损失最大的负样本, 由此能够基于难例挖掘的思想, 使得模型训练过程中更注意比较难于辨别的负样本, 进而可以在利用上述图像特征提取模型进行视觉位置识别时, 避免与待识别图像相似的负样本的干扰。

[0094] 本发明还提供了一种基于上述图像特征提取模型训练方法的域自适应的视觉位置识别方法, 如图2所示, 包括:

[0095] 确定待识别图像所属的目标域, 并获得目标域中不同位置处的多张图像, 将所获取的图像与待识别图像均作为待检索图像;

[0096] 以待检索图像为输入, 利用图像特征提取模型获得各待检索图像的特征向量; 获取图像特征向量时, 对于每一个卷积层, 统计所有待检索图像经过该卷积层后, 所得到的特征图的均值和标准差, 作为该卷积层之后的一层批标准化层的参数; 图像特征提取模型中

其余的模型参数为训练所得的模型参数；

[0097] 利用图像特征提取模型获取测试数据集中各图像的特征向量；在本实施例中，用于视觉位置识别的测试数据集具体为tokyo247数据集，其中每张图像带有地理坐标信息；

[0098] 根据所获取的特征向量获得测试数据集中与待识别图像的特征距离最近的图像，并将该图像的位置信息确定为待识别图像的位置信息，从而完成对待识别图像的视觉位置识别；

[0099] 其中，测试数据集中各图像的位置信息已知，域为影响图像特征分布的因素集合；

[0100] 根据实际应用中，会根据光照、视角、季节等因素对图像特征分布的影响情况完成域的划定，同一个域中，图像的特征分布相似；例如，如果仅光照会对图像的特征分布产生较大的影响，并且白天拍摄的图像具有相似的特征分布，夜间拍摄的图像具有相似的特征分布，则可根据光照条件划分得到两个域；具体根据哪些因素完成域的划定，以及同一个域中图像特征分布的相似程度，可根据实际的应用需求确定，只要保证最终视觉位置识别的精度满足要求即可；

[0101] 上述域自适应的视觉位置识别方法，在利用图像特征提取模型获取待识别图像的特征向量时，模型中各批标准化层的参数不依赖于训练集，而是利用多个与该待识别图像属于同一个域的图像获取相应的参数，由于同一个域中的图像具有相似的特征分布，因此，本发明能够实现域自适应，在训练集中图像与该待识别图像的特征分布差异较大时，仍然能够准确完成视觉位置识别，也即是说，本发明能够提高视觉位置识别的鲁棒性。

[0102] 在上述视觉位置识别方法中，由于用于视觉位置识别的测试数据集tokyo247与用于模型训练的标准数据集TokyoTimeMachine中图像具有相似的特征分布，因此，在本实施例中，利用图像特征提取模型获取测试数据集中各图像的特征向量时，直接利用上述图像特征提取模型训练方法训练所得的模型参数设置各模型参数；

[0103] 在其他应用场景中，为了最大程度上避免对训练集的依赖，在利用图像特征提取模型获取测试数据集中各图像的特征向量时，还可以采用如下方式实现对模型参数的设置：对于每一个卷积层，统计测试数据集中所有图像经过该卷积层后，所得到的特征图的均值和标准差，作为该卷积层之后的一层批标准化层的参数；图像特征提取模型中其余的模型参数为训练所得的模型参数。

[0104] 图3所示为利用进行视觉位置识别的一个示例，其中，训练集图像表示用于模型训练的标准数据集，query图像为待检索图像，gallery图像为测试集数据库图像。

[0105] 本发明还提供了一种图像特征提取模型训练装置，用于实现上述图像特征提取模型训练方法，该装置包括：模型建立模块、训练集构造模块以及模型训练模块；

[0106] 模型建立模块用于建立基于深度神经网络的图像特征提取模型，图像特征提取模型用于获取图像的特征向量；

[0107] 训练集构造模块用于在标准数据集中，获得各目标图像的正样本和s个负样本，以由一张目标图像及其正样本和负样本构成一个训练样本，从而得到由所有训练样本构成的训练集；

[0108] 模型训练模块用于利用训练集对图像特征提取模型进行训练，从而得到各模型参数；

[0109] 其中，图像特征提取模型包括特征提取网络和局部特征聚合网络；

[0110] 特征提取网络包括级联的多个第一网络；第一网络由一个或多个第二网络以及一个极大池化层依次连接而成，极大池化层用于对其前的第二网络输出的图像进行特征选择；第二网络包括依次连接的卷积层、批标准化层以及激活函数层，卷积层用于对图像进行特征提取，批标准化层用于对卷积层输出的图像进行零均值标准化处理，激活函数层用于对批标准化层输出的图像进行激活处理；

[0111] 局部特征聚合网络用于对特征提取网络输出的图像中所有的局部特征进行聚合，从而得到图像的特征向量；

[0112] 目标图像的正样本为其临近图像中与其特征距离最近的图像，目标图像与其临近图像的位置距离 d 满足 $T_{NL} \leq d < T_{NH}$ ；目标图像与其负样本的位置距离 d 满足 $d \geq T_F$ ；

[0113] 标准数据集中各图像的位置信息已知，目标图像为标准数据集中预先筛选出的多张图像；图像间的特征距离为图像的特征向量之间的距离； T_{NL} 、 T_{NH} 和 T_F 均为预设的阈值， $0 < T_{NL} < T_{NH}$ ， $T_{NH} \leq T_F$ ； $s \geq 1$ ；

[0114] 在本实施例中，各模块的具体实施方式，可参考上述方法实施例中的描述，在此将不再复述。

[0115] 本发明还提供了一种域自适应的视觉位置识别装置，用于实现上述域自适应的视觉位置识别方法，该装置包括：检索集获取模块、第一特征提取模块、第二特征提取模块以及识别模块；

[0116] 检索集获取模块用于确定待识别图像所属的目标域，并获得目标域中不同位置处的多张图像，将所获取的图像与待识别图像均作为待检索图像；

[0117] 第一特征提取模块用于以待检索图像为输入，利用图像特征提取模型获得各待检索图像的特征向量；获取图像特征向量时，对于每一个卷积层，统计所有待检索图像经过该卷积层后，所得到的特征图的均值和标准差，作为该卷积层之后的一层批标准化层的参数；图像特征提取模型中其余的模型参数为训练所得的模型参数；

[0118] 第二特征提取模块用于利用图像特征提取模型获取测试数据集中各图像的特征向量；

[0119] 识别模块用于根据第一特征提取模块和第二特征提取模块所提取的特征向量，获得测试数据集中与待识别图像的特征距离最近的图像，并将该图像的位置信息确定为待识别图像的位置信息，从而完成对待识别图像的视觉位置识别；

[0120] 其中，测试数据集中各图像的位置信息已知，域为影响图像特征分布的因素集合；

[0121] 在本实施例中，各模块的具体实施方式，可参考上述方法实施例中的描述，在此将不再复述。

[0122] 本领域的技术人员容易理解，以上所述仅为本发明的较佳实施例而已，并不用以限制本发明，凡在本发明的精神和原则之内所作的任何修改、等同替换和改进等，均应包含在本发明的保护范围之内。

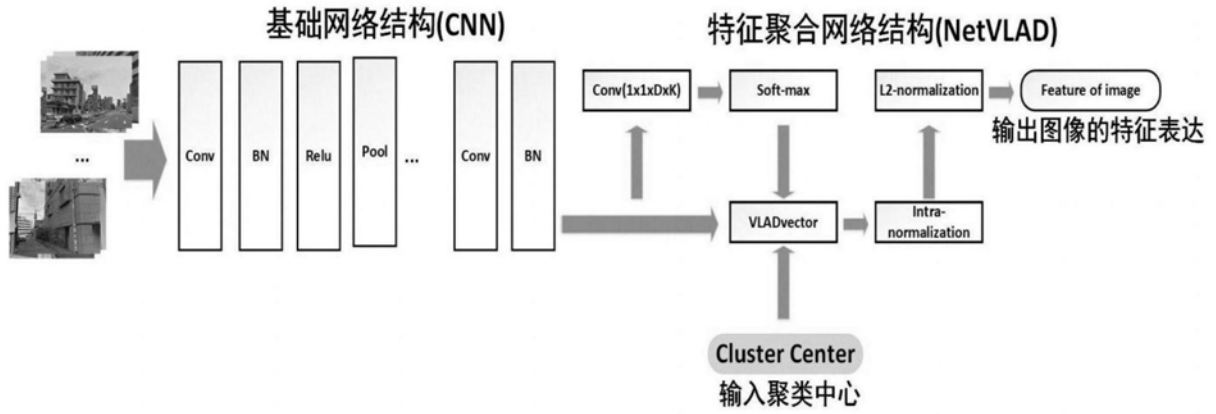


图1

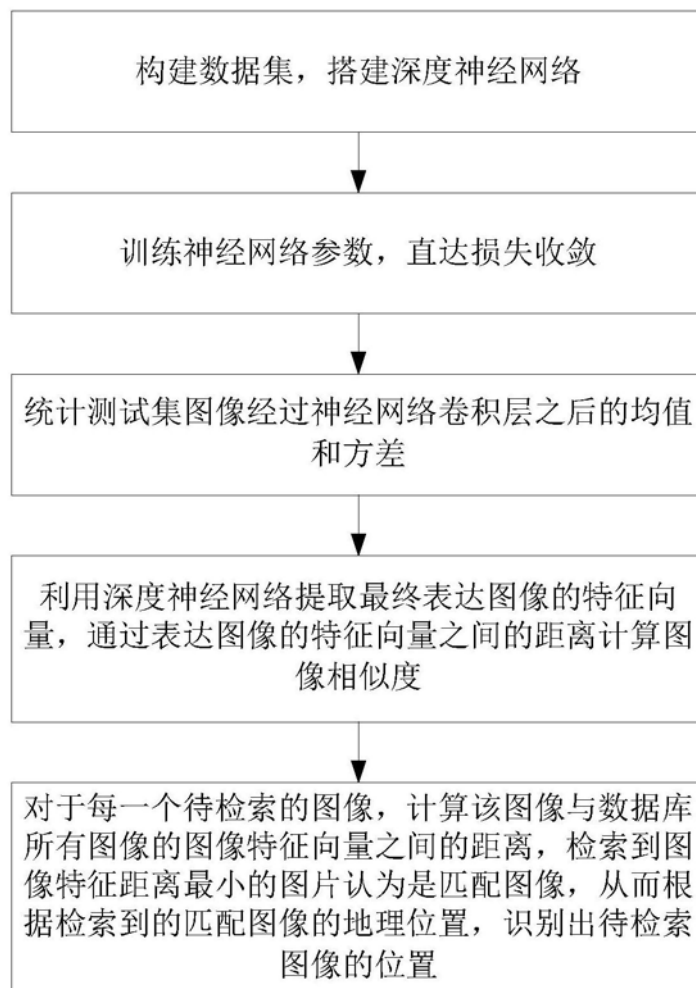


图2

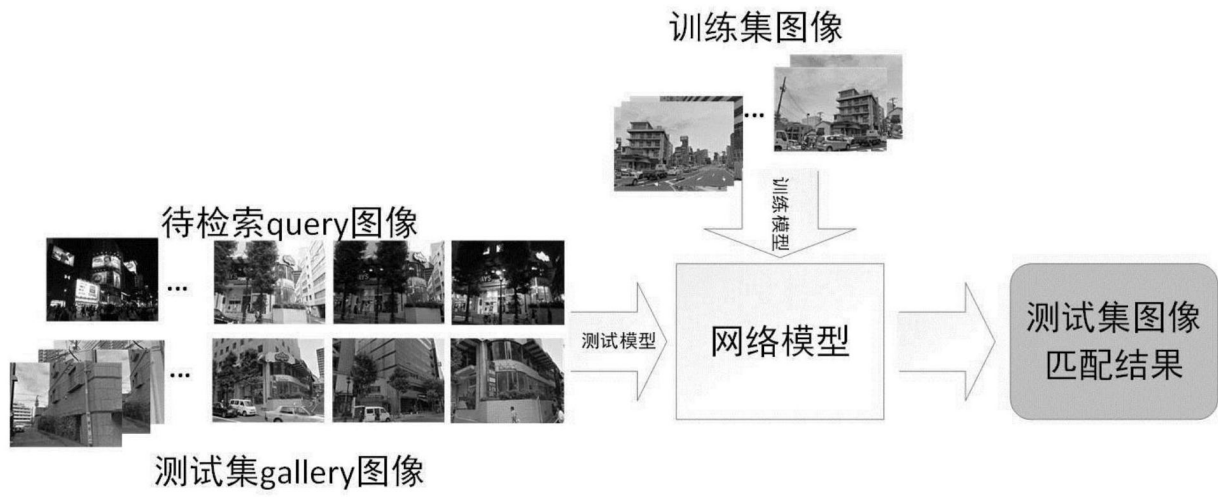


图3