吉林大学本科毕业设计（论文）文献综述

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 学 院 | 软件学院 | | | | 专业 | 软件工程 |
| 学生姓名 | 贾锋 | | | | 学号 | 55130218 |
| 指导教师 | 王欣 | | | | 职称 | 副教授 |
| 合作导师 |  | | 职称 |  | 单位 |  |
| 设计（论文）题目 | | 基于Qt的卷积神经网络辅助设计系统 | | | | |
| 摘要：  鉴于本项目的完成需要对深度学习有足够的了解，本文对深度学习算法进行了分类并对主流方法进行了简要的描述。首先介绍了深度学习的概念，然后根据深度学习算法的结构特征，概述了前馈深度网络、反馈深度网络和双向深度网络这三类主流深度学习算法的网络结构和训练方法，介绍了深度学习在图像处理领域的应用，最后，介绍了当前深度学习的研究热点，并对未来可能的研究方向做出了预测。可以看到：深度学习在图像处理领域有明显的优势，但仍存在需要进一步探索的问题，如网络模型规模与训练速度间的权衡、深度学习与传统方法的融合、特征理解和网络模型设计的方法等．  关键词：  深度学习；卷积神经网络；前馈神经网络；反卷积网络  **一、前言**  机器学习是一门致力于研究如何通过计算的手段，利用经验来改善系统自身性能的学科。在计算机系统中，“经验”通常是以“数据”的形式存在的，因此，机器学习研究的，是关于在计算机上从数据中产生“模型”的算法，即“学习算法”（learning algorithm）。有了学习算法，我们把经验数据提供给它，它就能基于这些数据产生模型，在面对新情况的时候，模型会给我们提供相应的判断。机器学习，研究的是关于“学习算法”的学问。  以“神经网络”为代表的机器学习方法，近年来在语音识别、计算机视觉等多类应用中取得了突破性进展[1-8]。为了能够在机器学习和图像处理领域进行相关研究工作，需要对该领域的研究现状和发展趋势有充分的认识。为此，本文筛选了从1982年到2016年的40篇论文，对其核心思想和方法进行综述和整理，并分析了该领域最新的研究进展和未来的发展趋势。  **二、综述**  **1、深度学习**  神经网络方法处理问题的思路主要来源于对生物的神经系统结构的研究。以图像数据为例，灵长类的视觉系统中对这类信号的处理依次为:首先检测边缘和初始形状[9]，然后再逐步形成更复杂的视觉。同样地，深度学习通过组合低层特征形成更加抽象的高层表示、属性类别或特征，给出数据的分层特征表示，进而通过特征对原始输入数据进行识别和分类。  深度学习之所以被称为“深度”，是相对支持向量机(support vector machine，SVM)、提升方法(boosting)、最大熵方法等“浅层学习”方法而言的，深度学习所学得的模型中，非线性操作的层级数［10］更多。浅层学习依靠人工经验抽取样本特征，网络模型通过学习所获得的是没有层次结构的单层特征［11］；而深度学习通过对原始信号进行逐层特征变换，将样本在原空间的特征表示变换到新的特征空间，自动地学习得到层次化的特征表示，从而更有利于分类或特征的可视化［12］。  深度学习的概念最早由多伦多大学的G．E．Hinton等于2006年提出［11］，指基于样本数据通过一定的训练方法得到包含多个层级的深度网络结构的机器学习过程［10］。传统的神经网络随机初始化网络中的权值，导致网络很容易收敛到局部最小值，为解决这一问题，Hinton提出使用无监督预训练优化网络权值的初值，再进行权值微调的方法，拉开了深度学习的序幕。  深度神经网络是由多个单层非线性网络叠加而成的［10，13］，常见的单层网络按照编码解码情况分为3类：只包含编码器部分、只包含解码器部分、既有编码器部分也有解码器部分。编码器提供从输入到隐含特征空间的自底向上的映射，解码器以重建结果尽可能接近原始输入为目标将隐含特征映射到输入空间［14］。深度神经网络分为以下3类(图1)。  1) 前馈深度网络(feed-forward deep networks，FFDN)，由多个编码器层叠加而成，如多层感知机(multi-layer perceptrons，MLP)［15］、卷积神经网络(convolutional neural networks，CNN)［16］等。  2) 反馈深度网络 (feed-back deep networks，FBDN)，由多个解码器层叠加而成，如反卷积网络(deconvolutional networks，DN)［14］、层次稀疏编码网络(hierarchical sparse coding，HSC)［17］等。  3) 双向深度网络(bi-directional deep networks，BDDN)，通过叠加多个编码器层和解码器层构成(每层可能是单独的编码过程或解码过程，也可能  既包含编码过程也包含解码过程)，如深度玻尔兹曼机(deep Boltzmann machines，DBM)［18］、深度信念网络(deep belief networks，DBN)［12］、栈式自编码器(stacked auto-encoders，SAE)［19］等。  **2、前馈神经网络**  前馈神经网络是最初的人工神经网络模型之一。在这种网络中，信息只沿一个方向流动，从输入层通过一个或多个隐层到达输出层，在网络中没有封闭环路。典型的前馈神经网络有多层感知机［15］和卷积神经网络［16］等。  1969年，M．Minsky等人证明单层感知机无法解决线性不可分问题(如异或操作)，这一结论将人工神经网络研究领域引入到一个低潮期，直到研究人员认识到多层感知机可解决线性不可分问题［15］，以及反向传播算法与神经网络结合的研究［20］使得神经网络的研究重新开始成为热点。但是由于传统的反向传播算法具有收敛速度慢、需要大量带标签的训练数据、容易陷入局部最优等缺点，多层感知机的效果并不是十分理想。  1984年，日本学者K．Fukushima等基于感受野概念提出的神经认知机可看作卷积神经网络的一种特例［21］，Y．Lecun等［16］提出的卷积神经网络是神经认知机的推广形式。卷积神经网络是由多个单层卷积神经网络组成的可训练的多层网络结构。每个单层卷积神经网络包括卷积、非线性变换和下采样3个阶段［22］，其中下采样阶段不是每层都必需的。每层的输入和输出为一组向量构成的特征图(feature map)（第一层的原始输入信号可以看作一个具有高稀疏度的高维特征图）。例如，输入部分是一张彩色图像，每个特征图对应的则是一个包含输入图像彩色通道的二维数组（对于音频输入，特征图对应的是一维向量;对于视频或立体影像，对应的是三维数组）；输出部分的每个特征图对应的是从输入图片的相应位置上提取的特定特征。  **2.1、单层卷积神经网络**  卷积阶段，通过一定数量的卷积核提取输入向量的特征，其定义源于D．H．Hubel等［44］基于对猫视觉皮层细胞研究提出的局部感受野概念。每个卷积核检测输入向量上所有位置上的特定特征，由于每次对输入向量执行卷积操作的时候，卷积核是不变的，即每次操作都是在整个输入向量中寻找特定的一种特征,这样的规则实现了权值共享［34］。为了提取输入特征图上不同的特征，使用不同的卷积核进行卷积操作。在网络结构层面上，卷积核以权值的形式存在。  非线性阶段（又称激活阶段），对卷积阶段得到的特征按照一定的原则进行筛选，筛选原则通常采用非线性变换的方式，以避免线性模型表达能力不够的问题。  激活阶段将卷积阶段提取的特征作为输入，进行非线性映射Ｒ=h(y)。传统卷积神经网络中激活函数往往采用sigmoid、tanh或softsign等饱和非线性(saturating nonlinearities)函数［23］，近几年的卷积神经网络中多采用不饱和非线性(non-saturating nonlinearity)函数ReLU(rectified linear units)［24-25］。ＲeLU比传统的饱和非线性函数有更快的收敛速度，因此在训练整个网络时，训练速度也比传统的方法快很多［1］。  下采样阶段，对每个特征图进行独立操作，通常采用平均池化（average pooling）或者最大池化（max pooling）的操作。平均池化操作的输入是进入池化窗口的像素，输出为窗口内像素值的均值Pa，要求窗口平移步长大于1且小于等于池化窗口的大小；最大池化则将均值Pa替换为最值Pm输出到下个阶段。池化操作后，输出特征图的分辨率降低，但能较好地保持高分辨率特征图所描述的特征。一些卷积神经网络不使用下采样阶段，而是通过在卷积阶段设置卷积窗口滑动步长大于1，以达到降低分辨率的目的［16，26］。  **2.2、卷积神经网络**  如图2所示，将单层的卷积神经网络进行多次堆叠，前一层的输出作为后一层的输入，便构成卷积神经网络。其中每2个节点间的连线，代表输入节点值经过卷积、非线性变换、下采样3个阶段传递到输出节点，一般最后一层后面接一个或数个全连接层和分类器。为了减少数据的过拟合，最近的一些卷积神经网络在全连接层引入“Dropout”［1］或“DropConnect”［27］的方法，即在训练过程中以一定概率P将隐含层节点的输出值(对于“DropConnect”为输入权值)清零，而用反向传播算法更新权值时，不再更新与该节点相连的权值。但是这2种方法都会降低训练速度［1，24，27］。 | | | | | | |

|  |
| --- |
| 在训练卷积神经网络时，最常用的方法是采用反向传播算法以及有监督的训练方式，算法流程如图3所示。网络中信号是前向传播的，即从输入特征向输出特征的方向传播，第1层的输入，经过多个卷积层，变成最后一层输出的特征向量，将输出特征向量与期望的标签进行比较，生成误差项，通过遍历网络的反向路径，将误差逐层传递到每个节点，根据权值更新公式，更新相应的权值。在训练过程中，网络中权值的初值可以随机初始化，也可以通过无监督的方式进行预训练［28］。网络误差随迭代次数的增加而减少，并且这一过程收敛于一个稳定的权值集合，额外的训练次数呈现出较小的影响。 |

|  |
| --- |
| **2.3、卷积神经网络的性质**  卷积神经网络的特点在于，采用原始信号（一般为图像）直接作为网络的输入，避免了传统识别算法中复杂的特征提取和图像重建过程；局部感受野方法获取的观测特征与平移、缩放和旋转无关。卷积阶段利用权值共享结构减少了权值的数量进而降低了网络模型的复杂度，这一点在输入数据是高分辨率图像时表现得更为明显。同时，下采样阶段利用图像局部相关性的原理对特征图进行抽样，在保留有用结构信息的同时有效地减少数据处理量。  **3、反馈深度网络**  与前馈网络不同，反馈网络并不是对输入信号进行编码，而是通过反卷积［14］或学习数据集的基［17，18］，对输入信号进行反解。前馈网络是对输入信号进行编码的过程，而反馈网络则是对输入信号解码的过程。典型的反馈深度网络有反卷积网络［14］、层次稀疏编码网络［17］等。  以反卷积网络为例，M．D．Zeiler等［14］ 提出的反卷积网络模型和Y．LeCun等［16］提出的卷积神经网络思想类似，但在实际的结构构建和实现方法上有所不同。卷积神经网络是一种自底向上的方法，该方法的每层输入信号经过卷积、非线性变换和下采样3个阶段处理，进而得到多层信息。相比之下，反卷积网络模型的每层信息是自顶向下的，组合通过滤波器组学习得到的卷积特征来重构输入信号。层次稀疏编码网络和反卷积网络非常相似，只是在反卷积网络中对图像的分解采用矩阵卷积的形式，而在稀疏编码网络中采用矩阵乘积的方式［17］。  **3.1、反卷积网络**  反卷积网络是通过先验学习，对信号进行稀疏分解和重构的正则化方法。图 4所示是一个单层反卷积网络模型，输入信号y由个特征通道组成，其中任意一个通道可看作个隐层特征图与滤波器组(个数为)的卷积。 |

|  |
| --- |
| **3.2、反卷积网络的性质**  反卷积网络的特点在于，通过求解最优化输入信号分解问题来计算特征，而不是利用编码器进行近似，这样能使隐层的特征更加精准，更有利于信号的分类或重建。  **4、双向深度网络**  双向深度网络由多个编码器层和解码器层叠加形成，每层可能是单独的编码过程或解码过程，也可能同时包含编码过程和解码过程。双向网络的结构结合了编码器和解码器2类单层网络结构，双向网络的学习则结合了前馈网络和反馈网络的训练方法，通常包括单层网络的预训练和逐层反向迭代误差2个部分，单层网络的预训练多采用贪心算法：每层使用输入信号与权值w计算生成信号传递到下一层，信号再与相同的权值w计算生成重构信号映射回输入层，通过不断缩小与间的误差，训练每层网络；网络结构中各层网络结构都经过预训练之后，再通过反向迭代误差对整个网络结构进行权值微调。其中单层网络的预训练是对输入信号编码和解码的重建过程，这与反馈网络训练方法类似；而基于反向迭代误差的权值微调与前馈网络训练方法类似。  典型的双向深度网络有深度玻尔兹曼机［18］、深度信念网络［12］、栈式自编码器［19］等。  **5、深度学习的应用**  深度学习目前在很多领域都优于过去的方法，下面重点介绍深度学习在图像处理方面的应用。  **5.1、深度学习在大规模图像数据集中的应用**  A．Krizhevsky等［1］首次将卷积神经网络应用于ImageNet大规模视觉识别挑战赛(ImageNet large scale visual recognition challenge，ILSVRC)［29］中，所训练的深度卷积神经网络［1］在ILSVRC-2012挑战赛中，取得了图像分类和目标定位任务的第一名。其中，图像分类任务中，前5选项错误率为15.3%，远低于第2名的26.2%的错误率；在目标定位任务中，前5选项错误率34%，也远低于第2名的50%。  在ILSVRC-2013比赛中，M．D．Zeiler等［30］采用卷积神经网络的方法，对文献［1］的方法进行了改进，并在每个卷积层上附加一个反卷积层用于中间层特征的可视化［14，30］，取得了图像分类任务的第一名。  在ILSVRC-2014比赛中，几乎所有的参赛队伍都采用了卷积神经网络及其变形方法。其中GoogleNet小组采用卷积神经网络结合Hebbian理论提出的多尺度的模型，以6.7%的分类错误，取得图形分类“指定数据”组的第一名。  从深度学习首次应用于ILSVRC挑战赛并取得突出的成绩，到2014年挑战赛中几乎所有参赛队伍都采用深度学习方法，并将分类错误率降低到6.7%，可看出深度学习方法在图像识别领域具有巨大优势。 |

|  |
| --- |
| **5.2、深度学习在人脸识别中的应用**  基于卷积神经网络的学习方法，香港中文大学的DeepID项目［3］以及Facebook的DeepFace项目［4］在户外人脸识别(labeled faces in the wild，LFW)数据库上的人脸识别正确率分别达97.45%和97.35%。  之后，采用基于卷积神经网络的学习方法，香港中文大学的DeepID2项目［31］将识别率提高到了99.15%，超过目前所有领先的深度学习［3-4］方法在LFW数据库上的识别率,甚至超过了人类手工识别在该数据库上的识别率［32］。DeepID2的深度结构包含4个卷积层，其中第3层采用2×2邻域的局部权值共享，第4层没有采用权值共享，且输出层与第3、4层都全连接。  **6、热点分析**  2016年，Google DeepMind的AlphaGo在围棋赛场上大显神威，让深度学习这一概念深入人心，深度学习的出现让很多人工智能相关技术取得了大幅度的进展，深度学习在语音识别、安防、机器人和自动驾驶等多个领域得到了应用。下面，本文将以图像处理和人工智能领域的最新论文为线索，对研究热点做出介绍和总结。  **6.1、创新的网络结构**  最近的很多研究，是围绕着表现更好的网络模型进行的。在这些工作中，有些是提出了新的网络框架，有些是对原有框架存在的问题提出了修补或解决方案，有些是将现有的深度学习方法和传统的计算机视觉方法相结合，为深度学习找到了新的应用点。  前微软研究员何凯明凭借着深度残差学习在ImageNet比赛的三个任务，以及COCO比赛的检测和分割任务上都获得了第一名，并获得了CVPR2016最佳论文[33]，深度残差网络使得设计上千层的超深网络成为可能。Redmon J等使用纯CNN，达到了实时检测的效果[34]。Shrivastava A等人的工作提供了F-RCNN框架下，在训练过程中进行样本筛选的方案[35]。 Tao Kong组的主要工作是提出了对CNN的多层特征进行融合的思路[36]。[37]提出了recursive-supervision方法和skip-connection方法来避免DRCN中会出现的梯度消失或爆炸问题。  **6.2、创新的研究方法**  以CNN及其衍生方案为代表的一些深度学习方法，存在着对最终效果缺乏理论解释的问题，拥有良好效果的网络模型往往是研究人员通过实验对比，不断的改良而产生的。围绕这个现象，不断有人提出新的研究方法，这些方法试图通过一些手段来帮助科研人员对网络模型进行挑选、设计和组合，他们的工作没有产生创新的网络结构，而是提供了创新的研究方法。  [38]的方法尝试帮助研究人员进行CNN的设计和选择，他们使用Bayesian和least squares SVM的方法来从预先训练的CNN模型中挑选适合的模型。面对各种CNN模型各有所长的问题，[39]使用了一个被称为“语义解释器”的前处理模块来动态选择CNN网络模型。Google DeepMind提出了用LSTM学习如何设计优化算法，即学习“如何学习”[40]。 |

|  |
| --- |
| **三、总结与展望**  深度学习算法在图像识别和语音识别中的应用，尤其是大规模数据集下的应用取得了突破性的进展，但仍有以下问题值得进一步研究:  1)无标记数据的特征学习。目前，标记数据的特征学习仍然占据主导地位，而真实世界存在着海量的无标记数据，将这些无标记数据逐一添加人工标签，显然是不现实的。所以，随着数据集体量和存储技术的发展，必将越来越重视对无标记数据的特征学习，以及为无标记数据进行自动添加标签的技术的研究。  2)模型规模与训练速度、训练精度之间的权衡。一般地，相同数据集下，模型规模越大，训练精度越高，训练速度会越慢。例如一些模型方法采用ReLU非线性变换、GPU运算，在保证精度的前提下，往往需要训练数天至一周的时间。对于模型优化，诸如模型规模调整、超参数设置、训练时调试等问题，训练时间会严重影响其效率。故而，如何在保证一定的训练精度的前提下，提高训练速度，依然是深度学习方向值得研究的课题之一。  3)与其他方法的融合。从上述应用实例中可发现，单一的深度学习方法，往往并不能带来最好的效果，很多具有实用意义的深度学习解决方案综合使用了其他的人工智能方法甚至是传统的图像处理算法，以获得更高的准确率。因此，深度学习方法与其他方法的融合，具有一定的研究意义。  4)缺乏充足的训练数据的情况下的训练方法。有些领域，例如医学图像处理，由于标注这些图片需要具备特殊的专业知识，很难获取足够的训练数据。在这种情况下，传统的训练方法往往表现欠佳，如何训练具有实用价值的深度学习模型值得研究。  5)创新的实验和研究方法。自从神经网络诞生以来，有很多人试图解释神经网络究竟学到了什么特征，然而效果却不尽人意。理解神经网络所学习到的特征，能够对设计新的网络架构起到指导作用，但是在直接对提出理论解释比较困难的情况下，设计和改进深度学习的研究方法，提出能够帮助科研人员进行模型筛选的工具，或者组合切换各种模型的方法，也是很有意义的，这也是当下的研究热点。 |

|  |
| --- |
| **四、参考文献**  [1]rizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[C] Advances in Neural Information Processing Systems. 2012: 1097-1105.  [2]Dahl G E, Yu D, Deng L, et al. Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2012, 20(1): 30-42.  [3]Sun Y, Wang X, Tang X. Deep learning face representation from predicting 10,000 classes[C] Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014: 1891-1898.  [4]Taigman Y, Yang M, Ranzato M A, et al. Deepface: Closing the gap to human-level performance in face verification[C] Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014: 1701-1708.  [5]Karpathy A, Toderici G, Shetty S, et al. Large-scale video classification with convolutional neural networks[C] Proceedings of theIEEE conference on Computer Vision and Pattern Recognition. 2014: 1725-1732.  [6]Ji S, Xu W, Yang M, et al. 3D convolutional neural networks for human action recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(1): 221-231.  [7]Dong C, Loy C C, He K, et al. Learning a deep convolutional network for image super-resolution[C] European Conference on Computer Vision. Springer International Publishing, 2014: 184-199.  [8] Roth H R, Lu L, Liu J, et al. Improving Computer-Aided Detection Using Convolutional Neural Networks and Random View Aggregation[J]. IEEE Transactions on Medical Imaging, 2016, 35(5): 1170-1181.  [9]Serre T, Kreiman G, Kouh M, et al. A quantitative theory of immediate visual recognition[J]. Progress in brain research, 2007, 165: 33-56.  [10]Bengio Y. Learning deep architectures for AI[J]. Foundations and trends in Machine Learning, 2009, 2(1): 1-127.  [11]Lowe D G. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004, 60(2): 91-110.  [12]Hinton G E, Osindero S, Teh Y W. A fast learning algorithm for deep belief nets[J]. Neural Computation, 2006, 18(7): 1527-1554.  [13]Bengio Y, Lamblin P, Popovici D, et al. Greedy layer-wise training of deep networks[J]. Advances in Neural Information Processing Systems, 2007, 19: 153.  [14]Zeiler M D, Krishnan D, Taylor G W, et al. Deconvolutional networks[C] IEEE Conference on Computer Vision and Pattern Recognition, 2010: 2528-2535.  [15]Hornik K, Stinchcombe M B, White H, et al. Multilayer feedforward networks are universal approximators[J]. Neural Networks, 1989, 2(5): 359-366.  [16]Cun Y L, Boser B E, Denker J S, et al. Handwritten digit recognition with a back-propagation network[C]. Conference and Workshop on Neural Information Processing Systems, 1990: 396-404.  [17]Yu K, Lin Y, Lafferty J, et al. Learning image representations from the pixel level via hierarchical sparse coding[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2011: 1713-1720.  [18]SALAKHUTDINOV，HINTON G E． Deep Boltzmann machines ［C］JML Workshop and Conference Proceedings Volume 5: AISTATS 2009． Brookline，MA:Microtome Publishing，2009: 448-455  [19]Vincent P, Larochelle H, Bengio Y, et al. Extracting and composing robust features with denoising autoencoders[C]. International Conference on Machine Learning, 2008: 1096-1103.  [20]Rumelhart D E, Hinton G E, Williams R J. Learning representations by back-propagating errors[J]. Cognitive Modeling, 1988, 5(3): 1.  [21]Fukushima K, Miyake S. Neocognitron: A new algorithm for pattern recognition tolerant of deformations and shifts in position[J]. Pattern Recognition, 1982, 15(6): 455-469.  [22] Lecun Y, Kavukcuoglu K, Farabet C, et al. Convolutional networks and applications in vision[C]. International Symposium on Circuits and Systems, 2010: 253-256.  [23]Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks.[J]. Journal of Machine Learning Research, 2010: 249-256  [24] Dahl G E, Sainath T N, Hinton G E, et al. Improving deep neural networks for LVCSR using rectified linear units and dropout[C]. International Conference on Acoustics, Speech, and Signal Processing, 2013: 8609-8613.  [25]Nair V, Hinton G E. Rectified linear units improve restricted boltzmann machines[C] Proceedings of the 27th International Conference on Machine Learning (ICML-10). 2010: 807-814.  [26] Simard P Y, Steinkraus D W, Platt J C, et al. Best practices for convolutional neural networks applied to visual document analysis[C]. International Conference on Document Analysis and Recognition, 2003: 958-963.  [27]Wan L, Zeiler M D, Zhang S, et al. Regularization of Neural Networks using DropConnect[C]. International Conference on Machine Learning, 2013: 1058-1066.  [28]Jarrett K, Kavukcuoglu K, Ranzato M, et al. What is the best multi-stage architecture for object recognition[C]. International Conference on Computer Vision, 2009: 2146-2153.  [29]Deng J, Dong W, Socher R, et al. Imagenet: A large-scale hierarchical image database[C] IEEE Conference on Computer Vision and Pattern Recognition, 2009: 248-255.  [30] Zeiler M D, Fergus R. Visualizing and understanding convolutional networks[C] European Conference on Computer Vision. Springer International Publishing, 2014: 818-833.  [31]Sun Y, Chen Y, Wang X, et al. Deep learning face representation by joint identification-verification[C]. Conference and Workshop on Neural Information Processing Systems, 2014: 1988-1996.  [32]Kumar N, Berg A C, Belhumeur P N, et al. Attribute and simile classifiers for face verification[C]. International Conference on Computer Vision, 2009: 365-372.  [33]He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C] Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 770-778.  [34]Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C] Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 779-788.  [35]Shrivastava A, Gupta A, Girshick R. Training region-based object detectors with online hard example mining[C] Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 761-769.  [36]Kong T, Yao A, Chen Y, et al. HyperNet: towards accurate region proposal generation and joint object detection[C] Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 845-853.  [37]Kim J, Kwon Lee J, Mu Lee K. Deeply-recursive convolutional network for image super-resolution[C] Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 1637-1645.  [38]Kim Y D, Jang T, Han B, et al. Learning to select pre-trained deep representations with bayesian evidence framework[C] Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 5318-5326.  [39]Andreas J, Rohrbach M, Darrell T, et al. Neural module networks[C] Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 39-48.  [40]Andrychowicz M, Denil M, Gomez S, et al. Learning to learn by gradient descent by gradient descent[C] Advances in Neural Information Processing Systems. 2016: 3981-3989. |