

Summary of Application of Convolution Neural Network on Image Classification

Zeming Yang, Jun Liu, Cheng Xue, Zihong Yu

Beijing Wuzi University, Beijing
Email: 83770561@qq.com, liujun006@aliyun.com

Received: Feb. 2nd, 2018; accepted: Feb. 21st, 2018; published: Feb. 28th, 2018

Abstract

In recent years, image classification occupies an important position in pattern recognition. The rapid growth of image data for image information analysis and processing put forward new requirements. Convolution neural network came into being, with its powerful image recognition classification ability is widely used in a variety of image classification system, and achieved very significant results. Firstly, this paper reviews the development of image classification technology, and introduces the classification methods of feature space and feature space and the image feature extraction methods. Secondly, the research of convolution neural network on image classification is introduced. Finally, the problems existing in the image classification of convolution neural networks are summarized, and the future development direction is summarized.

Keywords

Image Classification, Convolution Neural Network, Image Feature Extraction Method

卷积神经网络在图像分类上的应用综述

杨泽明, 刘 军, 薛 程, 于子红

北京物资学院, 北京
Email: 83770561@qq.com, liujun006@aliyun.com

收稿日期: 2018年2月2日; 录用日期: 2018年2月21日; 发布日期: 2018年2月28日

摘 要

近些年来, 图像分类在模式识别领域占据着重要的地位。迅速增长的图像数据对于图像信息的分析与处理提出了新的要求。卷积神经网络应运而生, 以其强大的图像识别分类能力被广泛的应用于各种图像分

类系统, 并取得了十分显著的效果。本文首先, 回顾了图像分类技术的发展历程, 介绍了图样空间与特征空间的分类方法以及图像特征提取方法。其次, 介绍了卷积神经网络在图像分类问题上的研究情况。最后, 总结卷积神经网络在图像分类问题上存在的问题, 以及未来的发展方向。

关键词

图像分类, 卷积神经网络(CNN), 图像特征提取方法

Copyright © 2018 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

由于科技与社会的不断发展, 人类传达信息的数量与方式发生了翻天覆地的变化。尤其是在近些年来, 信息数据爆炸式增长, 传统的文字已无法满足人们对于信息获取的需求, 而图像以其包含信息量大, 信息表达生动直接的优势逐渐成为主流信息传播方式之一。图像数据量也迅猛不断的增长。对于这些图像数据, 人们需要一种可以快速高效且合理的手段对它们进行处理、分析、解读, 从而在海量的图像数据中高效精确的提取到所需的信息。由于采用人工识别判断的方式不仅效率低下, 而且带有很大的主观性, 浪费大量的人力及时间资源。所以, 亟需快速提升计算机对于图像数据的处理能力, 开发出可以快速、高效、稳定的图像识别及分类算法。

图像分类[1]是指利用人工智能技术特别是机器学习方法, 使得计算机能够对图像进行识别和分类的过程。其在模式识别领域扮演者重要的角色, 涉及到手写字识别、人脸识别、车辆识别等方面。图像分类技术的深入研究对于计算机模式识别领域有着重要的意义。

2. 图像分类方法

图像分类[2]是近年来才发展起来的一门新兴科学技术, 它的主要研究内容为图像的分类与描述。一个图像分类系统主要由图像信息的获取、信息加工和处理、特征抽取、判断或分类等部分组成。图像分类的方法目前主要可分为两大类; 基于图像空间的分类方法和基于特征空间的分类方法。

基于图像空间的分类方法主要是利用图像的颜色、灰度、纹理、形状、位置等底层特征来对图像进行分类。基于颜色特征, 由于物体表面的颜色分布不同, 所以可以根据颜色特征将图像分类 Swain [3]提出的颜色直方图法是最早将颜色特征应用于图像分类的方法, 该方法通过每种颜色在图像空间中所占比例的不同来对图像进行分类, 但这种方法无法识别区分图像所描述的信息。基于纹理特征, 描述图像灰度空间的分布特征来对图像进行分类。纹理无处不在, 每个物体的表面都存在着不同的纹理。图像的纹理经计算机处理、数字化之后便可以得到对图像进行分析和处理的信号。上世纪 70 年代, Haralick [4]提出了灰度共生矩阵表示方法, 该方法利用两个位置的灰度像素的联合概率密度定义了灰度共生矩阵, 反映出图像灰度的方向、相邻间隔、变化幅度等信息, 从而得到纹理特征向量。然而, 由于纹理图像的多样性和分析算法的复杂性, 使得同一种纹理特征很难在不同的领域中应用。基于形状特征[5], 它描述的是封闭轮廓曲线所包围的区域, 形状特征一般情况下与图像中所描绘的目标对象有关。目前, 基于形状特征的分类方法基本上是通过轮廓特征、区域特征建立图像索引来对图像进行分类的。不过, 通过基于图像空间的分类方法对图像进行分类, 所需要计算数据量大, 计算过程十分复杂、却分类效果一般。

基于特征空间的分类方法[6]通过如 K-L 变换、小波变换等变换方法将原图像映射到高维空间, 然后提取其高层特征来实现图像的分类。该方法在有效的降低计算数据的维度和计算过程的复杂程度。但是, 分类的结果很大程度上取决于特征提取方法的适应性。

3. 图像特征提取方法

之前介绍的两类图像分类方法都是通过提取图像特征来进行图像分类的, 不同的是一个提取底层特征, 一个提取高层特征。从而我们可以看出图像特征的提取是图像分类的基础。目前, 特征提取的主要方法分为两个大类方法: 线性特征提取方法和非线性特征提取方法。

线性特征提取方法是通过线性映射方法来提取特征的, 其中比较有代表性的方法有主成分分析法(Principle Component Analysis, PCA)、基于 Fisher 准则的线性鉴别分析法(Linear Discriminate Analysis, LDA)和投影寻踪(Projection Pursuit, PP)等。PCA [7]方法采用了 K-L 变换中的协方差变化矩阵, 通过一系列的线性变换之后, 找到一组最优单位投影轴, 并利用该投影轴重建原样本。这种方法能够对多维数据进行降维, 用尽可能少的数据重建原样本。不过, PCA 方法在数据压缩时容易造成高阶统计特征的丢失, 使得特征描述不充分。LDA 方法[8]是通过找到一组最优线性变换, 使得投影后的样本类内方差最小, 类间方差最大, 从而达到图像分类的目的。PP 方法[9] [10]通过将样本中的高维观测数据映射到相对维度较低的子空间上, 从中找到可以反映其数据结构的投影, 以此来研究高维数据。但是, 该方法对于非线性问题的处理能力较弱以及需要庞大的计算量。以上线性特征方法都是通过一定的线性映射提取到样本的线性特征, 这使得这些方法对于非线性分布结构问题并没有很好地效果。

非线性特征提取方法是利用非线性映射方法来得到特征, 其中主要的方法有支持向量机(Support Vector Machine, SVM)、核主成分分析(Kernel Principal Component Analysis, KPCA)等。SVM [11]方法是通过一个非线性映射将样本空间映射到一个高维空间, 使得在样本空间中的线性不可分问题转化为高维空间中的线性可分问题, 以此来进行图像分类。KPCA [12] [13]方法是将输出空间通过一个非线性映射映射到特征空间上并进行 PCA。上述两种方法都是通过将原样本映射到高维空间中, 再进行线性分割, 这样容易引发“维度灾难”。

由于图像数据多种多类, 而传统的图像分类方法大多是针对某一种具体的图像进行分类。当图像种类发生变化时, 之前在某种图像的分类任务上取得不错效果的方法, 在另一种图像的分类中效果并不令人满意。因此找到一种通用且高效的图像分类方法十分必要。1998 年, 由 LeCun [14]提出的卷积神经网络(Convolution Neural Networks, CNN)经过多年以来人们不懈的努力, 在这方面上有着十分突出的表现。

4. 卷积神经网络在图像分类问题上的研究

4.1. 卷积神经网络研究现状

20 世纪 60 年代, Hubel 和 Wiesel [15]研究了猫的视网膜以及视觉皮层中枢神经细胞的信息处理机制, 发现视网膜所接收的视觉信息传递到大脑的过程中是由多个层次的感受野(receptive field)激发完成的。1980 年, Fukushima [16]提出基于感受野概念的神经认知机(neocognitron), 其模式识别不受位置变化, 较小的形状变化以及图像尺寸大小的影响。

20 世纪 90 年代, 基于卷积神经网络的应用大量出现, 其最初应用于语音识别及文档阅读。1995 年, LeCun [14]提出了 LeNet-5 模型, 该模型通过一系列交替相连的卷积层与池进行分类层将输入图像转化为权值共享的特征图, 再通过全连接的方式对图像的特征表达进行分类, 并采用梯度 BP 算法对网络进行监督训练。其中, 卷积层的卷积核完成了感受野功能, 将底层局部区域信息激发到高层。其结构如图 1 所示。

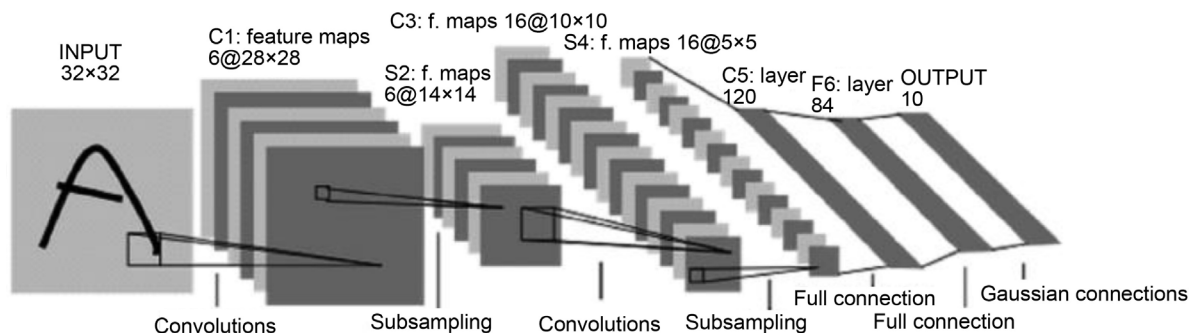


Figure 1. LeNet-5 model [14]

图 1. LeNet-5 模型[14]

由于模型深度不断加深,网络结构及训练参数变得复杂,使得计算量迅速增长,同时,网络得训练容易陷入局部最优,以及在高位数据的处理中出现过拟合现象。加之当时硬件条件的限制,使其仅应用于小尺度图形上。直到 2006 年, Hinton [17]教授提出两个重要观点:第一,在数据的特征表达上,具有多个隐含层的深度神经网络其学习能力明显优于浅层学习时,其所得特征更加接近于数据的本质;第二,可以通过逐层初始化的方法大幅降低深度神经网络的训练难度。至此,上述情况才得以改善。

4.2. 卷积神经网络在图像分类问题上的进展

2010 年至今,每年举办的 ImageNet ILSVRC (Large Scale Visual Recognition Challenge)图像分类比赛是具有相当知名度的国际赛事,取得了很多优异的成果。该比赛选用 ImageNet [18]数据集的子集,其中包含有上百万张图像,这些图像被划分为 1000 多个类别。2010 年与 2011 年,在该项比赛中取得桂冠的团队使用的都是传统图像分类算法,他们主要使用 SIFT, LBP [19]等算法提取图像特征,再使用 SVM 等分类器对其进行分类,取得的最好成绩为错误率 28.2% [20]。ILSVRC2012 比赛是图像分类研究领域的一个重要转折点。在此次比赛中, Alex Krizhevsk 等提出的 AlexNet [21]第一次在大规模图像分类任务中采用卷积神经网络模型,成功的将错误率降低至 16.4%,相比第二名降低了约 10%的错误率。如图 2 所示, AlexNet 是一个前五层为卷积层、后三层为全连接层的八层卷积神经网络,其全连接层的最后一层采用 softmax 分类方法,并且采用 ReLU (Rectified linear units)函数作为非线性激活函数。此外,该模型提出了 Dropout 方法来降低过拟合现象的发生。

ILSVRC2013 的获胜队伍 Clarifai [22]提出了一种将反卷积网络用于 AlexNet 的每一卷积层,借此可以对每一卷积层所学到的特征进行可视化的分析,该方法加深了人们对于卷积神经网络对于图像分类原理的理解,而且将错误率降低至 11.7%。

ILSVRC2014 比赛中图像分类结果取得了重大突破, Google 团队提出的 GoogleNet [23]以 6.7% 错误率的好成绩摘得桂冠,先比之前的最佳成绩将错误率价低至一半。该网络应用了赫布学习规则,同时使用多尺度处理方法对卷积神经网络进行优化。该团队受到 Network in network [24]思想的启发提出了 Inception 模块。Inception 模块的结构如图 3 所示,此模块是用稠密组件近似地替代图像中的最优局部系数结构,以实现有效的降维,从而达到拓宽网络的深度与宽度,减少训练参数的目的。同时,有效的降低了过拟合现象的发生。

ILSVRC2014, 微软亚洲研究院团队提出了 SPP-NET [25]模型,以 8.3% 的错误率取得了当年的季军。该模型采用了一种名为空间金字塔池化的新池化方法,如图 4 所示。在此之前,由于在卷积神经网络中全连接层的参数数量固定,所以要保障全连接层输入维数固定。全连接层的输入维数是卷基层的输出维

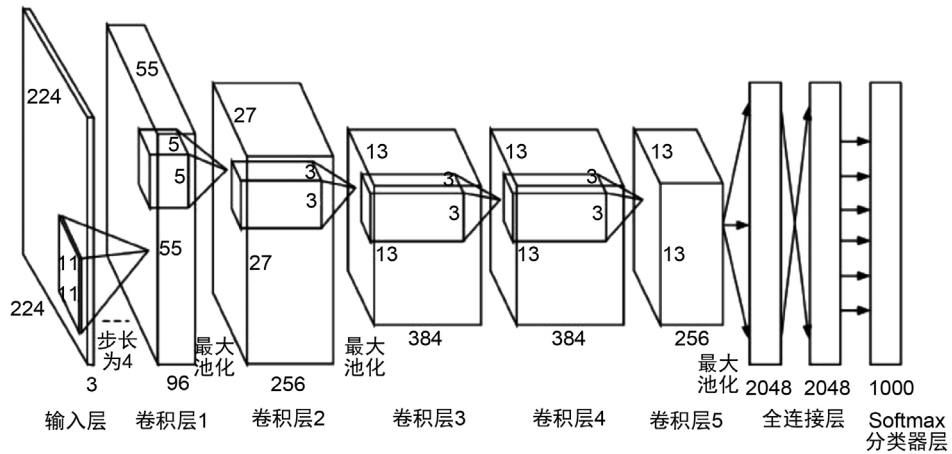


Figure 2. Simplified AlexNet model [21]
图 2. 简化的 AlexNet 模型[21]

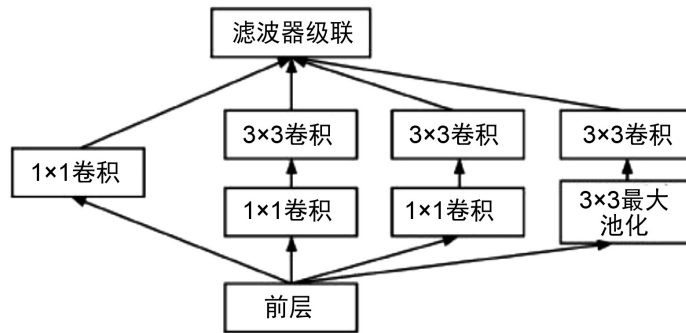


Figure 3. Simplified inception module structure [24]
图 3. 简化的 Inception 模块结构[24]

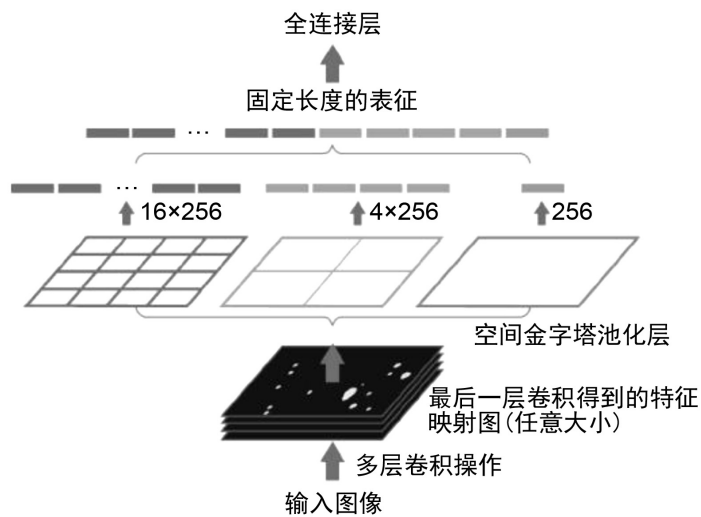


Figure 4. Space pyramid pool model structure [25]
图 4. 空间金字塔池化模型结构[25]

数决定的，而卷积层的输出维数由卷积层的输入维数决定，所以需要固定输入图像的大小。从而需要对图像进行剪裁或者长宽调整，这样会导致原始图像信息丢失或者图像扭曲变形。空间金字塔池化方法将

将输入图像划分为个数固定的局部空间块并进行最大池化,使得输出维数固定。同时,采用多层次空间块划分,保障提取到不同尺度的特征。这种方法使得该网络模型能够使用任意大小的图像作为输入。

微软亚洲研究院在 2015 年年初提出了 PreLU-Nets [26]模型,该模型在 ILSVRC 图像分类数据集上以 4.9%的错误率成为首次超越人眼识别(错误率 5.1% [23])效果的模型。模型以参数化修正线性单元(PReLU)作为激活函数,以微小的积算成本法大幅提高识别准确率。此外,在修正线性单元(ReLU/PReLU)的建模过程中,推导出了一种可以使层数较多的模型收敛的初始化方法。

不久之后,Google 提出的新一代 GoogleNet 模型,该模型在 ILSVRC2012 数据集上取得了 4.82%错误率的成绩[27]。该模型在将归一化方法用于网络内部的激活函数中,对层与层之间传输的数据进行归一化,由于采用随机梯度下降法训练,归一化只能在每个 mini-batch 内进行。此方法取得了很好的效果,例如高学习率、准确率、减少过拟合等。

卷积神经网络对于图像学习的能力日益增强,然而其对于图像空间不变性尤其是旋转不变性的学习能力还是不尽人意。为了解决这个问题 Google DeepMind 提出了 Spatial transformer [28]模块,该模块可以在任意位置加入,对输入数据进行空间变换,使得提取到的特征易于学习与分类。此外,在训练过程中对于所需的变换参数,该模块有着自主学习的能力,无需额外的监督训练。

ILSVRC2015,微软亚洲研究院团队提出的深达 152 层的深层残差网络以 3.57% [29]错误率的绝对优势取得图像分类冠军。随着网络深度的增加,训练难度不断提高,对于图像识别分类的准确率达到饱和甚至开始下降。该团队提出了残差学习思想,即由网络训练取得的效果无法在进一步的时候,让网络层学习值为 0 的残差函数相较于恒等函数更容易取得好的效果。如图 5 所示,将 shortcut connection 方法用于网络中部分层间的链接,以实现残差学习,从而适当准确率不会因网络层数的增加而下降。再此之后的 ILSVRC2016 与 ILSVRC2017 的比赛中 ensemble models 成为主流,在模型创新方面并没有较大的突破。

自从 2012 年卷积神经网络首次应用在 ILSVRC 图像分类比赛并取得令人瞩目的成绩以来,卷积神经网络被广泛应用于图像识别与分类领域。人们孜孜不倦的研究使得不断改进的网络模型一一涌现,刷新着 ILSVRC 比赛记录,也使得卷积神经网络对于图像特征的学习提取能力日新月异的发展。同时,借由 ImageNet, MSCOCO 等大规模数据集的出现,卷积神经网络的训练强度不断提升,使得模型有着更强的泛化能力,提升在实际图像分类问题中的应用效果。

5. 总结与展望

相较于传统的图像分类方法,卷积神经网络拥有特征自主提取、自主学习的能力,并通过权值共享

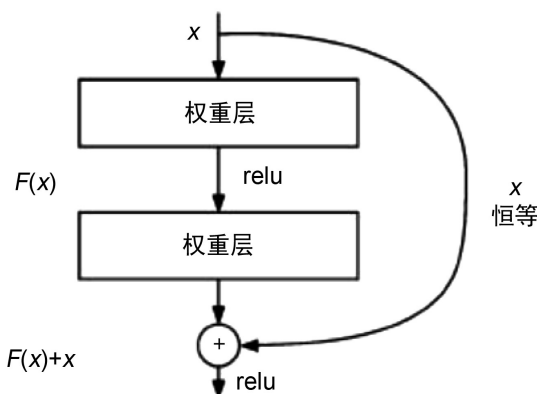


Figure 5. Residual learning module [29]

图 5. 残差学习模块[29]

的方式大大减少了全连接层所需神经元的数量, 简化了网络结构使其所需的计算量明显下降。此外, 卷积神经网络有着学习迁移的能力, 经过训练的网络可以将之前所学到的特征应用于一项新的图像分类任务中, 从而有效改善传统图像分类方法通用性差的问题, 并且能大大提高图像分类的准确率及效率。

随着基于深度卷积神经网络在各类图像分类系统中的应用越来越广泛, 识别效果越来越好, 其研究工作一直深受研究者的重视。但是, 仍有一些问题还没有较好的解决方案, 主要表现在以下几方面:

卷积神经网络的理论研究相对落后, 对于图像特征提取、分类的具体机理的理解尚不透彻, 导致了网络结构与网络参数的设置需要一定的经验, 且随着网络层次的不断加深容易出现网络退化、过拟合等问题。

对于图像分类问题来说, 网络的训练需要大量的已标注的数据集来提高其泛化能力, 而现有的数据集已经不能满足其发展需求。这是目前制约卷积神经网络发展推广的主要因素。

第三, 卷积神经网络尚存在一些缺陷。其完备性的相关理论与证明研究相对匮乏, 在图像分类上存在着“欺骗” [30]等现象。

卷积神经网络在图像分类领域中取得了巨大的成功, 其研究仍有广阔的发展前景。目前, 进一步理解其工作原理、优化网络结构、发展无监督式学习方法以及借鉴生物视觉系统的机理是其未来发展的主要方向。

基金项目

北京高校物流技术工程研究中心项目(BJLE2010)。

参考文献 (References)

- [1] 许可. 卷积神经网络在图像识别上的应用的研究[D]: [硕士学位论文]. 杭州: 浙江大学, 2012.
- [2] 席晓聪. 图像分类方法研究[D]: [硕士学位论文]. 济南: 山东大学, 2013.
- [3] Swain, M.J. and Ballard, D.H. (1991) Color Indexing. *International Journal of Computer Vision*, **7**, 11-32. <https://doi.org/10.1007/BF00130487>
- [4] Haralick, R.M. (1973) Texture Features for Image Classification. *IEEE Transactions on Systems Man & Cybernetics*, **3**, 610-621.
- [5] 章志勇, 潘志庚, 张明敏, 等. 基于多尺度通用傅里叶描述子的灰度图像检索[J]. 中国图象图形学报, 2005, 10(5): 611-615.
- [6] 刘三军. 基于图像特征空间学习的图像分类方法研究[D]: [硕士学位论文]. 西安: 西安电子科技大学, 2014.
- [7] Kirby, M. and Sirovich, L. (1990) Application of the Karhunen-Loeve Procedure for the Characterization of Human Faces. *IEEE Computer Society*.
- [8] Swets, D.L. and Weng, J. (1996) Using Discriminant Eigenfeatures for Image Retrieval. *IEEE Computer Society*.
- [9] Friedman, J.H. and Tukey, J.W. (2006) A Projection Pursuit Algorithm for Exploratory Data Analysis. *IEEE Transactions on Computers*, **C-23**, 881-890. <https://doi.org/10.1109/T-C.1974.224051>
- [10] Chiang, S.S., Chang, C.I. and Ginsberg, I.W. (2001) Unsupervised Target Detection in Hyperspectral Images Using Projection Pursuit. *IEEE Transactions on Geoscience & Remote Sensing*, **39**, 1380-1391. <https://doi.org/10.1109/36.934071>
- [11] Burges, C.J.C. (1998) A Tutorial on Support Vector Machines for Pattern Recognition. Kluwer Academic Publishers, Dordrecht.
- [12] Zhang, D., Chen, S. and Zhou, Z.H. (2006) Recognizing Face or Object from a Single Image: Linear vs. Kernel Methods on 2D Patterns. *Structural, Syntactic, and Statistical Pattern Recognition, Joint IAPR International Workshops, SSPR 2006 and SPR 2006*, Hong Kong, 17-19 August 2006, 889-897.
- [13] Mika, S., Smola, A. and Scholz, M. (1999) Kernel PCA and De-Noising in Feature Spaces. In: *Conference on Advances in Neural Information Processing Systems II*, MIT Press, Cambridge, 536-542.
- [14] LeCun, Y.L., Bottou, L., Bengio, Y., et al. (1998) Gradient-Based Learning Applied to Document Recognition. *Proceedings of the IEEE*, **86**, 2278-2324. <https://doi.org/10.1109/5.726791>

- [15] Hubel, D.H. and Wiesel, T.N. (1962) Receptive Fields, Binocular Interaction and Functional Architecture in the Cat's Visual Cortex. *Journal of Physiology*, **160**, 106-154. <https://doi.org/10.1113/jphysiol.1962.sp006837>
- [16] Fukushima, K. (1980) Neocognitron: A Self-Organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position. *Biological Cybernetics*, **36**, 193-202. <https://doi.org/10.1007/BF00344251>
- [17] Hinton, G.E. and Salakhutdinov, R.R. (2006) Reducing the Dimensionality of Data with Neural Networks. *Science*, **313**, 504-507. <https://doi.org/10.1126/science.1127647>
- [18] Deng, J., Dong, W., Socher, R., et al. (2009) ImageNet: A Large-Scale Hierarchical Image Database. *Computer Vision and Pattern Recognition*, Miami, 20-25 June 2009, 248-255.
- [19] Ahonen, T., Hadid, A. and Pietikainen, M. (2006) Face Description with Local Binary Patterns: Application to Face Recognition. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, **28**, 2037-2041. <https://doi.org/10.1109/TPAMI.2006.244>
- [20] Russakovsky, O., Deng, J., Su, H., et al. (2015) ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision*, **115**, 211-252. <https://doi.org/10.1007/s11263-015-0816-y>
- [21] Krizhevsky, A., Sutskever, I. and Hinton, G.E. (2017) ImageNet Classification with Deep Convolutional Neural Networks. *Communications of the ACM*, **60**, 84-90.
- [22] Zeiler, M.D. and Fergus, R. (2013) Visualizing and Understanding Convolutional Networks. *Computer Vision*, Zurich, 6-12 September 2014, 818-833.
- [23] Szegedy, C., Liu, W., Jia, Y., et al. (2015) Going Deeper with Convolutions. *Computer Vision and Pattern Recognition*, Boston, 7-12 June 2015, 1-9.
- [24] Lin, M., Chen, Q. and Yan, S. (2013) Network in Network. *International Conference on Learning Representations*, Scottsdale, 2-4 May 2013, 10.
- [25] He, K., Zhang, X., Ren, S., et al. (2015) Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, **37**, 1904-1916. <https://doi.org/10.1109/TPAMI.2015.2389824>
- [26] He, K., Zhang, X., Ren, S., et al. (2015) Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. *IEEE International Conference on Computer Vision*, Santiago, 7-13 December 2015, 1026-1034.
- [27] Ioffe, S. and Szegedy, C. (2015) Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *International Conference on Machine Learning*, Lille, 6-11 July 2015, 448-456.
- [28] Jaderberg, M., Simonyan, K., Zisserman, A., et al. (2015) Spatial Transformer Networks. *Neural Information Processing Systems*, Montreal, 7-12 December 2015, 2017-2025.
- [29] He, K., Zhang, X., Ren, S., et al. (2015) Deep Residual Learning for Image Recognition. *Computer Vision and Pattern Recognition*, Las Vegas, 26 June-1 July 2016, 770-778.
- [30] Nguyen, A., Yosinski, J. and Clune, J. (2014) Deep Neural Networks Are Easily Fooled: High Confidence Predictions for Unrecognizable Images. *Computer Vision and Pattern Recognition*, Boston, 7-12 June 2015, 427-436.

知网检索的两种方式:

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>
下拉列表框选择: [ISSN], 输入期刊 ISSN: 2326-3415, 即可查询
2. 打开知网首页 <http://cnki.net/>
左侧“国际文献总库”进入, 输入文章标题, 即可查询

投稿请点击: <http://www.hanspub.org/Submission.aspx>

期刊邮箱: airr@hanspub.org