

密级：_____

南昌大学

NANCHANG UNIVERSITY

学士学位论文

THESIS OF BACHELOR

(2013—2017 年)



题 目 深度度量学习的理论研究与实际运用

学 院： 信息工程学院 系 计算机科学与技术系

专 业： 网络工程

班 级： 网络工程 131 班

学 号： 6103213008

学生姓名： 骆明远

指导教师： 黄伟

起讫日期： 2017.02.13-2017.05.19

南昌大学

学士学位论文原创性申明

本人郑重声明：所呈交的论文是本人在导师的指导下独立进行研究所取得的研究成果。除了文中特别加以标注引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写的成果作品。对本文的研究作出重要贡献的个人和集体，均已在文中以明确方式表明。本人完全意识到本声明的法律后果由本人承担。

作者签名：

日期：

学位论文版权使用授权书

本学位论文作者完全了解学校有关保留、使用学位论文的规定，同意学校保留并向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅和借阅。本人授权南昌大学可以将本论文的全部内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本学位论文。

保密，在 年解密后适用本授权书。

本学位论文属于

不保密。

（请在以上相应方框内打“√”）

作者签名：

日期：

导师签名：

日期：

深度度量学习的理论研究与实际运用

专 业：网络工程

学 号：6103213008

学生姓名：骆明远

指导教师：黄伟

摘要

近几年来,深度学习技术在计算机视觉和机器学习领域受到的瞩目日益增长,大量深度学习模型和方法不断涌现。一般来说,深度学习技术能够从原始数据中提取特定的潜在特征表示,并将这些特征运用到诸如识别、分类等具体的计算机视觉应用问题中来。然而,在这些问题中,需要对不同数据的特征进行定量比较以确定数据之间的相似程度,这就需要度量学习技术来实现。本论文创新地将深度学习领域最新技术之一的深度残差网络与度量学习相结合,提出了深度残差度量学习这一新方法,并将它运用在行人再识别这一计算机视觉研究领域的热门应用问题中来,以实现不重叠多角度摄像头拍摄图像中的行人识别匹配任务。本论文通过大量的统计实验和对比分析,表明本论文新提出的深度残差度量学习方法与本论文实验部分复现的传统浅层度量学习方法和其他深度度量学习方法相比,都具有更好的实验效果。

关键词：深度度量学习；深度残差网络；行人再识别

Theoretical Investigation and Practical Application of Deep Metric Learning – A Novel Method

Abstract

In recent years, deep learning receives more and more popularity in computer vision and machine learning fields. It is also widely acknowledged that, a variety of deep learning models and methods have been proposed. Deep learning techniques are capable to extract latent features from raw data and incorporate these learned features into various computer vision tasks, including detection, classification, etc. However, the similarity between different features need to be determined, and metric learning techniques are adopted to realize it in these tasks. In this thesis, a novel deep residual network-based metric learning method is proposed. This new method incorporates deep residual network, which is known as one of the latest deep learning techniques, together with conventional metric learning techniques. Also, this new method is utilized to realize the pedestrian re-identification task in this thesis. In the experimental section of this thesis, the newly proposed method has been compared with dozens of conventional shallow learning-based metric learning approaches as well as some recently introduced deep learning-based metric learning methods, statistical analysis turns out that this new method is superior to the compared methods in this pedestrian re-identification task.

Keyword: deep metric learning; deep residual network; pedestrian re-identification

目录

摘要	I
Abstract	II
第一章 引言	1
1.1 机器学习	1
1.2 深度学习	1
1.3 本论文主要研究内容	2
第二章 度量学习	4
2.1 度量	4
2.2 浅层度量学习	4
2.3 深度度量学习	5
第三章 基于深度残差网络的深度度量学习方法	7
3.1 深度残差网络	7
3.2 方法的模型结构与理论推导	8
3.2.1 Mahalanobis 度量层	8
3.2.2 包含三支的深度残差网络部分	9
3.3 方法的实现	11
第四章 实验结果与分析	13
4.1 数据库来源	13
4.2 比较方法及实现手段	13
4.3 实验结果	15
4.4 实验分析	17
第五章 结论与展望	21
5.1 结论	21
5.2 未来研究展望	21
参考文献	22
致谢	26

第一章 引言

1.1 机器学习

机器学习 (Machine Learning, ML) 在近 20 年快速兴起, 被运用到数据搜索、语音识别、金融分析等多个不同领域, 并取得了较好的表现。机器学习涉及到计算机科学、统计学、概率论、神经科学等多个不同学科, 主要内容包括有监督学习、无监督学习、半监督学习、强化学习等, 其目的在于让机器 (系统) 在一定的经验 (数据) 条件下通过学习 (训练) 获得一定的模式 (模型), 用于提高完成特定任务的性能 (又名: 狭义人工智能)。

“机器学习”这个概念由 *A. L. Samuel* 在二十世纪五十年代提出, 当时 *Samuel* 将其运用在一个跳棋程序^[1]上, 使其具有自我学习能力, 能够通过分析大量不同棋局的形势来提高程序自身的弈棋水平。五十年代中后期, 基于神经网络的机器学习理论开始出现, 主要包括 *F. Rosenblatt* 于 1957 年提出的感知机 (Perceptron)^[2]和 *B. Widrow* 于 1960 年提出的自适应线性元件 (Adaline)^[3]。六七十年代, 基于逻辑表示的机器学习理论快速发展起来, 主要有 *E. B. Hunt* 等人于 1961 年提出的概念学习系统^[4]、*P. H. Winston* 于 1970 年提出的结构学习系统^[5]和 *R. S. Michalski* 等人于 1983 年提出的归纳学习系统^[6]等。八十年代之后, *D. E. Rumelhart* 于 1986 年重新发明的误差反向传播算法 (Back Propagation, BP)^[7]使得人工神经网络重新受到关注。与此同时, 基于决策理论的机器学习技术、强化学习技术和统计学习理论也得到了相应的发展, 主要包括 *J. R. Quinlan* 于 1986 年提出的 ID3 决策树算法^[8]和由 *C. Cortes* 和 *V. Vapnik* 于 1995 年提出的支持向量机 (Support Vector Machine, SVM)^[9]等。

近些年, 机器学习技术蓬勃发展, 涌现出许多新的技术, 例如深度信念网络 (Deep Belief Network, DBN)^[10]、限制波尔兹曼机 (Restricted Boltzmann Machine, RBM)^[11]、卷积神经网络 (Convolutional Neural Network, CNN)^[12]等。随着这些新技术的快速发展, 并且伴随着计算机硬件的存储、传输和计算能力不断提高, 机器学习在多个领域内取得了可观的应用效果, 比较著名的有科大讯飞的自动语音识别技术^[13], Tesla 的自动驾驶技术和 DeepMind 的围棋程序 AlphaGo^[14]等。这其中许多应用给我们的生活带来了极大的便利。

1.2 深度学习

深度学习 (Deep Learning, DL) 由机器学习中的神经网络发展而来, 包

含多隐层的神经网络即深度学习结构,能够从数据中学习潜在的抽象高层次特征。1957年提出的感知机^[2]是最早的单层神经网络,可用于线性分类,但是解决不了经典异或(XOR)问题^[15]。1986年提出的误差反向传播算法^[7]使得训练多层感知机成为可能,采用 Sigmoid 激励函数进行非线性映射解决了非线性分类问题。1989年, *G. Cybenko* 证明了一个包含有限个数简单神经元单隐层的前馈神经网络能够任意逼近任何闭区间内的连续函数^[16]。1991年, *S. Hochreiter* 指出误差反向传播算法存在着梯度消失问题^[17],使得前层的网络难以训练。1997年, *S. Hochreiter* 提出长短时记忆网络(Long Short Term Memory Network, LSTM)^[18]用于解决梯度消失问题。1998年, *Y. LeCun* 设计了 LeNet 模型^[12],这是一个卷积神经网络,在 MNIST 数字数据库^[12]上有很好的表现。

深度学习起始于 2006 年。*G. E. Hinton* 提出了使用单层波尔兹曼机自编码预训练深层神经网络来解决其中的梯度消失问题^[19],这使得深度学习成为可能。2012 年, *Hinton* 课题组使用设计的卷积神经网络模型 AlexNet^[20]在当年的 ImageNet 大规模视觉识别挑战大赛(ILSVRC-2010)^[21]上夺得冠军,使得卷积神经网络受到关注。之后深度学习进入爆发期,各种深度学习模型(VGG^[22]、GoogleNet^[23]和 ResNet^[24]等)在 ImageNet 各个比赛上取得越来越好的成绩,使得深度学习受到广泛关注。近几年,深度学习出现了许多新的方法和模型,都能够一定领域内取得较好的效果。例如,2014年, *N. Srivastava* 提出一种防止神经网络模型过拟合的简单方法(Dropout)^[25], *I. Goodfellow* 提出了一个通过对抗过程评估生成模型的新框架——生成对抗网络(Generative Adversarial Network, GAN)^[26];2015年, *J. Long* 提出了一种新的卷积结构——全卷积网络(Full Convolution Network, FCN)^[27], *S. Ioffe* 提出一种通过减少内部协变量转变加速深度网络训练的批标准化方法(Batch Normalization, BN)^[28], *K. He* 提出深度残差网络(Deep Residual Network, DRN)^[24]来解决深度网络训练难的问题。

1.3 本论文主要研究内容

通常来说,深度学习的目的在于通过分层的低层次特征代表来构建高层次的潜在特征信息,并通过不同对象的高层次特征来计算它们之间的相似度(距离)。然而,如何度量相似度(距离)往往是不易确定的,而度量学习(Metric Learning, ML)能够很好地解决这个问题。度量学习能够根据不同的任务来自学习出针对某个特定任务的度量距离函数。

结合机器学习时下最流行的深度学习技术与传统浅层学习研究中的经典度量学习问题,即可以产生深度度量学习(Deep Metric Learning, DML)这一研究

方向。同时，深度度量学习可以被应用在行人识别、图像分类等多个计算机视觉领域的研究问题中，其目的可以概括为：通过学习到合适的度量函数，使得同类别对象之间的相似程度变大（即同类别对象之间的距离变小），而不同类别对象之间的相似程度变小（即不同类别对象之间的距离变大），从而获得更好的识别分类效果。具体来说，本论文将基于深度残差网络这个近年来在深度学习领域获得广泛瞩目的模型，将度量学习运用在基于视觉信息的行人再识别（**Pedestrian Re-identification**）研究问题中，在实现行人识别的同时完成本论文拟提出的新颖的深度残差度量学习方法（**Deep Residual Metric Learning, DRML**）。该方法在国际、国内深度学习和度量学习领域尚未正式提出，本论文的工作属于开创性研究。

第二章 度量学习

2.1 度量

在统计科学及相关领域里,度量(Metric)用于确定两个对象之间的相似度,一般为实值函数。通常没有统一的标准来确定对象之间的相似度,但是根据经验可知:相似度与距离负相关,即相似的对象之间距离小,不同的对象之间距离大。通常在具体的实际运用中,挑选一个适合该运用问题的合适的度量函数是一个具有挑战性的研究问题,度量函数选取的是否合适将直接决定该实际运用问题最终效果的好坏。通过度量学习,可以根据对象实例的先验知识学习出具体的度量函数,以获得较好的实际运用效果。因此,度量学习从理论和实际两方面都具备重要的研究意义。

从另一方面来看,度量学习也可以看作是有监督学习(Supervised Learning)的一个分支,与回归和分类有紧密联系,可以被广泛使用于排序推荐、视觉跟踪、人脸识别和声音验证等领域。度量学习获得的度量一般需满足非负性、对称性和可加性等三个限制。具体来说,当对象 x_i 是空间 R^d 中的向量,对于对称半正定锥 S_+^d 中的任一个矩阵 W , x 所在空间内的距离伪度量由下式定义

$$D_w(x_1 - x_2)^2 = (x_1 - x_2)^T W (x_1 - x_2) \quad (1)$$

当 W 是对称正定矩阵时, D_w 是度量。此外,由于任意一个对称半正定矩阵 $W \in S_+^d$ 都可以分解为 $W = L^T L$,其中 $L \in R^{e \times d}$ 且 $e \geq \text{rank}(W)$,那么度量 D_w 可写为

$$D_w(x_1 - x_2)^2 = (x_1 - x_2)^T L^T L (x_1 - x_2) = \|L(x_1 - x_2)\|_2^2 \quad (2)$$

这与映射变换之后的特征向量 $x'_1 = Lx_1$ 和 $x'_2 = Lx_2$ 之间的 *Euclidean* 距离的平方 $\|x'_1 - x'_2\|_2^2$ 相一致。因此,公式(1)中的距离可以看作是公式(2)中加权 *Euclidean* 距离的推广。

2.2 浅层度量学习

浅层学习(Shallow Learning)是相对于深度学习而言,指的是深度学习研究时期之前大量提出的传统机器学习方法。浅层度量学习(Shallow Metric Learning)基于的是传统度量学习方法,按照训练数据有无标签可分为有监督度量学习(Supervised Metric Learning)和无监督度量学习(Unsupervised Metric Learning)。

有监督度量学习可分为全局度量学习(Global Metric Learning)和局部度量学习(Local Metric Learning)。全局度量学习是在全局约束条件下进行度量学习,

例如满足同一类中所有对象之间的约束关系较紧密,不同类之间对象的约束关系较松散。比较著名的全局度量学习方法包括 *E. P. Xing* 于 2002 年提出的概率全局距离度量学习 (Probabilistic Global Distance Metric Learning, PGDM)^[29]、*A. Bar-Hillel* 于 2002 年提出的相关成分分析 (Relevant Components Analysis, RCA)^[30]、*T. Hertz* 于 2004 年提出的 DistBoost^[31]、*S. C. H. Hoi* 于 2006 年提出的判别成分分析 (Discriminative Component Analysis, DCA)^[32]、*L. Yang* 于 2012 年提出的有效距离度量学习 (Active Distance Metric Learning, BAYES+VAR)^[33]等。而局部度量学习是在局部约束条件下进行度量学习。与全局约束不同的是,它只需要满足局部的约束条件。比较著名的局部度量学习方法包括 *J. Goldberger* 于 2004 年提出的邻近成分分析 (Neighborhood Component Analysis, NCA)^[34]、*L. Yang* 于 2006 年提出的局部距离度量学习 (Localized Distance Metric Learning, LDM)^[35]、*M. Sugiyama* 于 2007 年提出的局部 Fisher 判别分析 (Local Fisher Discriminant Analysis, LFDA)^[36]、*K. Q. Weinberger* 于 2009 年提出的大范围最近邻分类 (Large Margin NN Classifier, LMNN)^[37]等。

无监督度量学习一般又被称为流形学习 (Manifold Learning),它可以在两个方面进行分类:一个是基于学习的映射是否线性,另一个是基于它的约束是否是全局的。比较著名的无监督度量学习方法包括 *J. B. Kruskal* 于 1964 年提出的多维尺度变换 (Multidimensional Scaling, MDS)^[38]、*S. Wold* 于 1987 年提出的主成分分析 (Principal Component Analysis, PCA)^[39]、*J. B. Tenenbaum* 于 2000 年提出的等距映射算法 (ISOMAP)^[40]、*S. T. Roweis* 于 2000 年提出的局部线性嵌入算法 (Locally Linear Embedding, LLE)^[41]、*M. Belkin* 于 2001 年提出的 Laplacian 特征映射算法 (Laplacian Eigenmap, LE)^[42]、*X. He* 于 2003 年提出的局部保持映射算法 (Locality Preserving Projections, LPP)^[43]和 *X. He* 于 2005 年提出的邻近保持嵌入算法 (Neighborhood Preserving Embedding, NPE)^[44]等。

2.3 深度度量学习

深度度量学习 (Deep Metric Learning) 是指将深度学习技术与度量学习问题结合,利用深度学习的技术特点学习到数据的潜在高层次特征,并通过度量学习获得高层次特征之间的距离度量,从而实现一种端对端的学习方式 (End-to-End Learning)。深度度量学习通常被用于识别或分类问题,近期研究结果表明:深度度量学习在这些问题中的求解效果较好。

近几年来,深度度量学习被应用在计算机视觉研究中的诸多方面特定问题的求解中。例如:2014 年,*S. Z. Li* 构建 DML 模型^[45]用于行人再识别,该工作将

卷积网络与 *Mahalanobis* 距离度量^[46]相融合; *J. Lu* 通过构建 DDML 模型^[47]用于原始人脸验证。2015 年, *S. Z. Li* 构建 CDML 模型^[48]并在深度度量模型中加入权重限制, 从而实现行人再识别; *J. Lu* 构建 MMDML 模型^[49]用于图像集分类, 将流形学习引入深度度量学习, 构建 DTML 模型^[50]用于迁移学习, 结合粒子滤波构建 DML Tracker 模型^[51]用于视觉跟踪; *E. Hoffer* 构建 Triplet Network 模型^[52]用于图像分类, *H. Liu* 构建 SLM+DeepML 模型^[53]用于行人再识别。2016 年, *J. H. Lai* 构建 Deep Ranking Framework 模型^[54]用于行人再识别。

就目前的发展趋势来看, 度量学习对于不同特征之间相关性的学习所具有的优势能够高效地融入深度学习, 使得深度学习技术能够在人脸识别、行人识别、物体识别、图像集分类、目标跟踪和视觉搜索等多个领域的实际问题运用中获得效果的再提高。

第三章 基于深度残差网络的深度度量学习方法

3.1 深度残差网络

随着卷积神经网络的不断发展,人们期望更深层的卷积网络能够具备更好的泛化性能,从而达到更好的实验效果。然而,更深层次的深度卷积网络的训练过程出现了许多问题。一个是梯度消失/发散问题^[55,56],这导致了深层卷积网络训练难以收敛;另一个是退化问题^[24],即随着网络层数增加,训练误差增加,网络正确率达到饱和并且快速下降,并且这种现象并不是由过拟合产生。从目前的最新研究成果来看,梯度消失/发散问题能够由标准初始化方法(Normalized Initialization)^[57]和批标准化方法(BN)^[28]很好地解决。

另一方面,深度残差网络(Deep Residual Network, DRN)^[24]是 *K. He* 于 2015-2016 年提出的深度学习领域最新技术概念,可以用于退化问题求解并加快深层网络的收敛速度。该方法的核心思想是:在原始网络层的基础上增加并行的恒等映射层,形成一个残差学习结构,如图 1 所示。具体来说:假设需要学习的潜在映射为 $H(x)$,那么构造非线性堆叠网络层来对应另一个残差映射 $F(x) := H(x) - x$,这样的话潜在映射就可写为 $F(x) + x$,而原始网络层优化残差映射比优化潜在映射更容易,映射 $F(x) + x$ 可由加入“直连连接”(Shortcut Connection)的前馈神经网络实现。

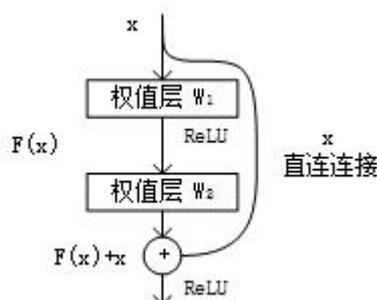


图 1 残差学习结构

残差学习结构的数学形式由下式给出:

$$y = F(x, \{W_i\}) + x \quad (3)$$

其中 x 和 y 分别为残差学习结构的输入和输出向量,函数 $F(x, \{W_i\})$ 代表需要学习的残差映射, W_i 代表第 i 层网络的权值(偏置省略的写法)。如图 1 给出的例子,其中 $F = W_2 \sigma(W_1 x)$, σ 代表 ReLU 激励函数。运算 $F(x) + x$ 表示“直连连接”和按元素相加。在公式(3)里, x 与 F 的输出的维度必须相同,但许多情况下是不同的,这里可以将一个线性映射 W_s 代替直连连接来保证维度同一。因此,公式(3)

的数学闭式可以进一步变化为： $y = F(x, \{W_i\}) + W_s x$ 。

3.2 方法的模型结构与理论推导

本论文针对行人再识别问题 (Pedestrian Re-identification) 构造基于深度残差网络的深度度量学习方法 (Deep Residual Metric Learning, DRML)。行人再识别问题的目的是：从两张行人图像中提取特征 (可做验证性评估, 即预测身份未知的行人图像身份并统计正确率) 并用一定的度量计算它们之间的相似度 (可做识别性评估, 即计算某张身份未知的行人图像与所有身份已知的行人图像之间的相似度并将所有身份已知的行人图像按照相似度递减排序, 最后计算 CMC 曲线或其他曲线)。为了获得更好的效果, 来自同一个行人的图像应该有较小的距离, 来自不同行人的图像应该有较大的距离。因此, 本论文使用能够从大量不同的图像里提取合适潜在特征的深度卷积神经网络来构造模型。深度残差度量学习 (DRML) 的网络结构如图 2 所示。网络可划分为两个部分, 深度残差网络部分 (DRN part) 和 Mahalanobis 度量层 (Mahalanobis metric layer)。第一部分由两个共享权值的孪生深度残差网络构成, 用于提取行人图像的特征。第二部分用于计算特征之间的相似度。将两个部分合并, 使得特征提取部分 (深度残差网络部分) 和度量学习部分 (Mahalanobis 度量层) 能够使用梯度下降方法 (Gradient Descent, GD) 共同训练, 从而实现端对端学习 (End-to-End Learning)。

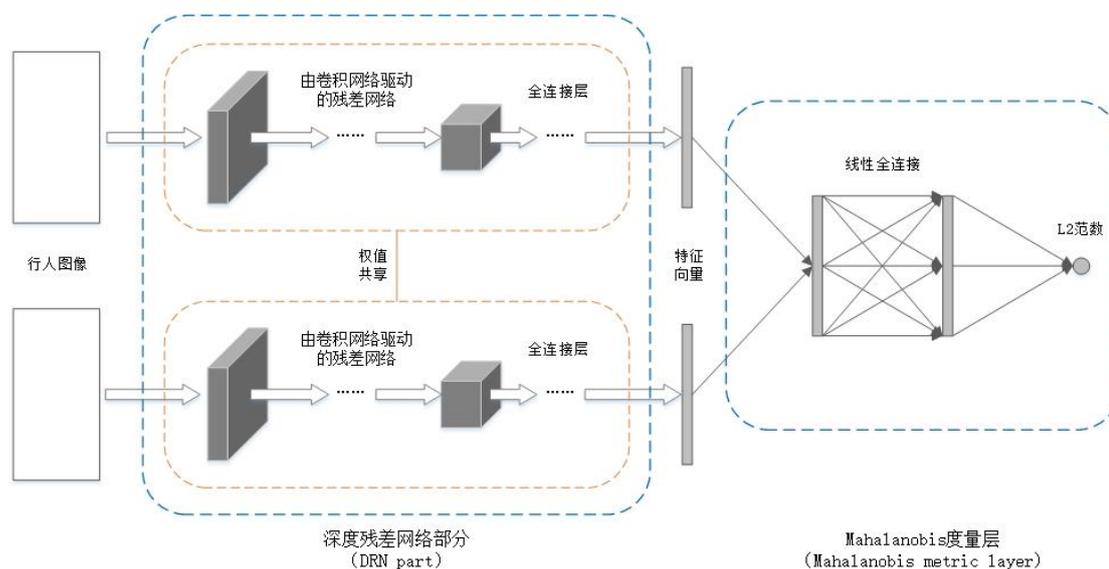


图 2 深度残差度量学习模型结构

3.2.1 Mahalanobis 度量层

给出两个来自不同角度的行人图像集 Γ_1 和 Γ_2 , X_1 和 X_2 分别是 Γ_1 和 Γ_2 通过深度残差网络部分 (DRN part) 提取的特征集, 定义 $x_1 \in X_1$ 和 $x_2^p \in X_2$ 作为正例 (来自

同一个行人), $x_1 \in X_1$ 和 $x_2^n \in X_2$ 作为负例 (来自不同行人)。使用 *Mahalanobis* 度量^[46]计算它们的距离 (相似度), 目标是学习到一个 *Mahalanobis* 度量能够最小化正例内的距离并且最大化负例内的距离, *Mahalanobis* 距离公式由下式给出

$$d(x_1, x_2) = \sqrt{(x_1 - x_2)^T M (x_1 - x_2)} \quad (4)$$

其中 $x_2 \in \{x_2^p, x_2^n\}$, M 是协方差矩阵。在传统度量分析问题里, 使用何种特征已知, 矩阵 M 可以根据数据分布假设确定。在本论文模型里, 特征 x_1 和 x_2 由深度残差网络部分学习得到且分布未知。因此, 矩阵 M 和深度残差网络部分使用误差反向传播算法共同训练。

定义 ψ 为深度残差网络部分的前向传播函数, $I_1 \in I_1$ 和 $I_2 \in I_2$ 分别作为特征 x_1 和 x_2 的原始行人图像, 那么 $x_1 = \psi(I_1)$ 和 $x_2 = \psi(I_2)$ 。由于矩阵 M 对称且半正定, 所以可做分解 $M = WW^T$ 。因为直接学习矩阵 M 比学习矩阵 W 难, 并且 WW^T 总是半正定的, 所以距离公式 (4) 可写为

$$\begin{aligned} d(x_1, x_2) &= \sqrt{(\psi(I_1) - \psi(I_2))^T WW^T (\psi(I_1) - \psi(I_2))} \\ &= \sqrt{\left(W^T (\psi(I_1) - \psi(I_2))\right)^T \left(W^T (\psi(I_1) - \psi(I_2))\right)} \\ &= \|W^T (\psi(I_1) - \psi(I_2))\|_2 \end{aligned} \quad (5)$$

其中内积 $W^T (\psi(I_1) - \psi(I_2))$ 可以在一个权重矩阵为 W^T 的全连接层实现, 全连接层的输出可由下式计算

$$y = f(W^T x + b) \quad (6)$$

其中 b 是偏置项, 这里可以置为 0 向量, f 是全连接层的激励函数。在深度残差网络部分之后使用 *Mahalanobis* 度量层来定量计算特征向量的距离 (相似度)。该网络的训练误差可以由下式定义

$$L = \|d(\psi(I_1) - \psi(I_2^p)) - d(\psi(I_1) - \psi(I_2^n))\|_2 \quad (7)$$

其中 I_2^p 和 I_2^n 作为特征 x_2^p 和 x_2^n 的原始行人图像。在每一次前向传播后, 训练误差的第一项和第二项都会被计算, 然后计算梯度并且反向传播, 这个误差公式能够最小化正例距离并且最大化负例距离。

3.2.2 包含三支的深度残差网络部分

在这节的开始, 展示了共享权值的孪生深度残差网络 (DRN part)。实际上, 每个深度残差网络包含如图 3 所示的 3 个分支 (branch)。输入的图像首先标准化为 128×64 的 RGB 图像, 然后划分为三个重叠的 64×64 的 RGB 部分并被分别输入到每个分支里。每个分支由一个包含卷积网络的残差网络构成, 每个分支之

间不共享权值。之后，3个分支连接到一个使用 ReLU 激励函数的全连接层，最后连接到一个使用线性激励函数的全连接层。深度残差网络提取输入图像的特征向量，随后特征向量被输入度量学习部分。

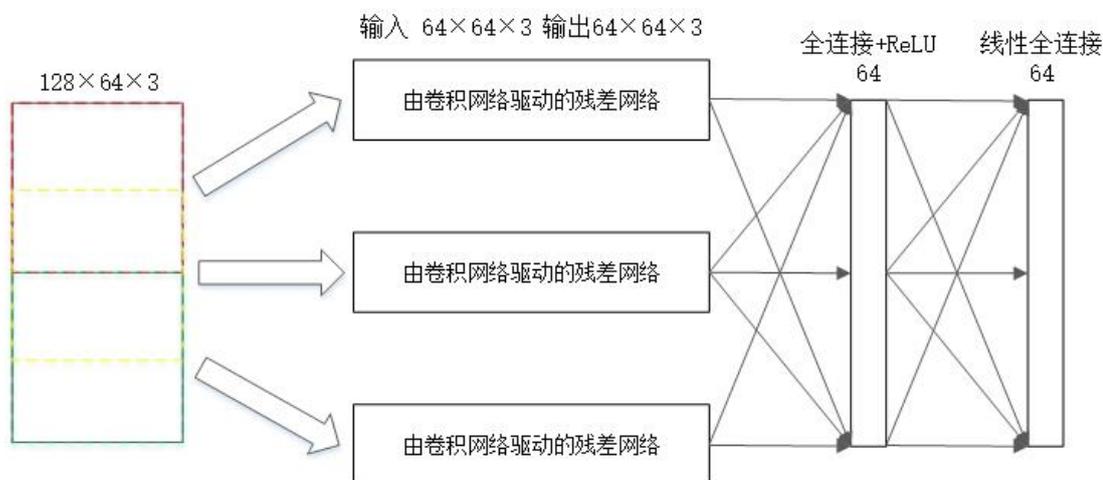


图3 深度残差网络 (DRN part) 部分

每个分支都是一个由卷积网络驱动的残差网络 (ResNet with CNN)，如图4所示。分支由多个相同的卷积网络结构和“直连连接” (1×1 卷积、BN 和 ReLU 激励函数) 组成，每个卷积网络结构和一个“直连连接”构成一个并行的残差网络结构，多个相同的残差网络结构依次相连构成深层的网络。卷积网络结构的输入与输出具有相同的大小，在其输入与输出之间连接“直连连接”使得整个网络能够更好地进行优化。

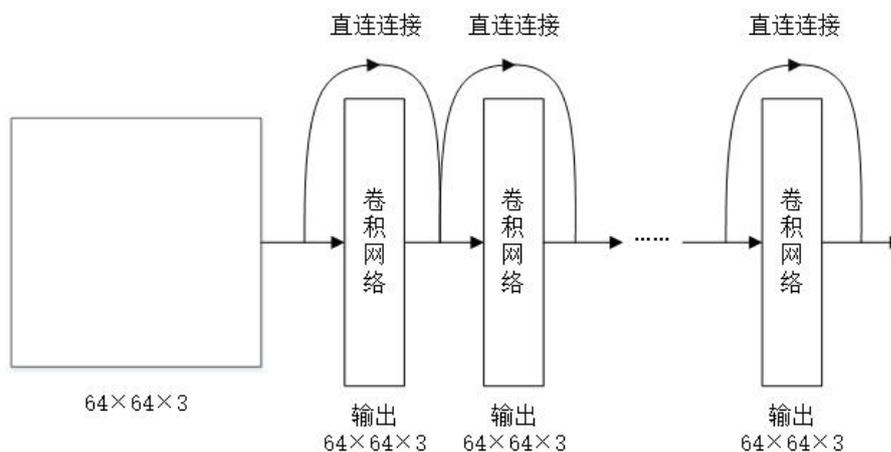


图4 由卷积网络驱动的残差网络 (ResNet with CNN) 结构

分支中的用于提取部分特征的卷积网络 (CNN) 结构如图5所示，这个结构是在 Hinton 提出的卷积网络模型 AlexNet 的基础上增加一层全卷积层构成。多层卷积层和池化层能够提取图像的高层次特征，这些特征图的大小比输入图像小，为了保证卷积网络结构能够结合“直连连接”，那么必须保证输出特征图的大小与输入图像相同，所以在卷积网络的最后加入全卷积层来增大特征图的大小。

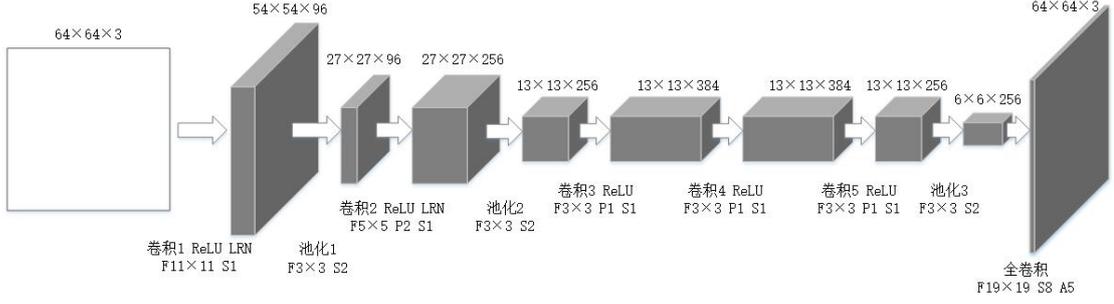


图 5 卷积网络 (CNN) 结构, 其中 F 代表卷积核大小、P(A)代表填补、S 代表步长

3.3 方法的实现

本论文的实验环境为 NVIDIA 提供的 CUDA 并行计算架构, cuDNN 深度学习加速库和 Facebook 提供的 Torch 深度学习框架, 使用 GT 630 GPU 进行运算。深度残差度量学习方法 (DRML) 使用在 ImageNet 数据集^[58]上预训练好的 AlexNet 模型构造 (parameters transferring), 之后将深度残差网络部分 (DRN part) 中的卷积网络在 CUHK03 数据集^[59]上微调 (pre-training), 最后将整个模型在 CUHK01 数据集^[60]上进行微调 (fine-tuning) (识别性评估)。在 CUHK03 数据集上进行微调的中间过程可以进行验证性评估。第四章的实验将表明, 本论文构造的新方法能够达到突出的验证和识别效果。

将深度残差网络部分在 CUHK03 数据集上进行微调的网络结构由深度残差网络部分中的卷积网络、带有 ReLU 激励函数的全连接层、普通全连接层等部分构成, 该网络结构在 CUHK03 数据集上使用 Softmax 分类器进行分类学习。在 CUHK01 数据集上微调的网络结构即深度残差度量学习模型, 其中深度残差网络部分的卷积网络由在 CUHK03 数据集上微调之后的网络结构中的卷积网络构成, 每个由卷积网络驱动的残差网络包含 3 个卷积网络 (内存限制)。此外, 由 3.2.1 节的叙述可知: Mahalanobis 度量层的目标是最小化同类之间的距离并且最大化异类之间的距离。相比于 Mahalanobis 度量, Euclidean 度量区分性较差但普遍性较高, 因为 Euclidean 度量并不需要考虑异类之间的关系。因此, 在 CUHK03 数据集上微调时, 将 Mahalanobis 度量层中的矩阵 M 用单位矩阵 I 进行初始化。

将深度残差度量学习方法的训练误差 (7) 式应用到训练集中的每对正例图像和每对负例图像, 可得到方法的优化问题如下式所示

$$\min_{\psi, d} \sum_{\substack{I_1 \in I_1 \\ I_2^p, I_2^n \in I_2}} \left\| d(\psi(I_1) - \psi(I_2^p)) - d(\psi(I_1) - \psi(I_2^n)) \right\|_2 \quad (8)$$

具体的微调实现中, 通常将正例和负例的误差 d 通过一个函数 g 区分, 当输入距离 d 由正例获得时 $g(d) = d$, 当输入距离 d 由负例获得时 $g(d) = \max(0, margin -$

d), 其中 $margin$ 为限定的边界, 那么优化问题可改为下式

$$\min_{\psi, d} \sum_{\substack{I_1 \in \Gamma_1 \\ I_2 \in \Gamma_2}} g(d(\psi(I_1) - \psi(I_2))) \quad (9)$$

上式可采用误差反向传播算法进行优化, 具体实现时采用批量梯度下降算法 (Batch Gradient Descent, BGD) 进行优化学习, 学习率取 0.001。

第四章 实验结果与分析

4.1 数据库来源

本论文在两个行人数据集上进行实验评估，CUHK03 数据集^[59]上进行验证性评估，CUHK01 数据集^[60]上进行识别性评估。具体来说，CUHK03 数据集分为 detected 集和 labeled 集，detected 集是原始行人图像集，包括 1467 个行人对象共计 14096 张行人图像，labeled 集的每张图像是在 detected 集对应的图像上人工选择行人区域部分构成的图像，包括 1467 个行人对象共计 14097 张行人图像。本论文使用 CUHK03 数据集的所有图像来微调深度残差网络部分，将 labeled 集作为训练集，detected 集作为测试集，进行验证性评估。CUHK01 数据集包含 971 个行人对象，每个行人 4 张图像共计 3884 张行人图像。深度残差度量学习（DRML）模型在 CUHK01 数据集上微调时，随机选择 485 个行人对象的所有图像作为训练集，其余 486 个行人对象的所有图像作为测试集，进行识别性评估。

4.2 比较方法及实现手段

在 CUHK03 数据集上做微调（验证性评估）时，将微调的网络结构中的卷积网络替换为其他卷积网络进行对比实验，其中的卷积网络包括：（1）AlexNet+全卷积（本论文构造的卷积网络）、（2）AlexNet（本论文构造的卷积网络除去全卷积层）、（3）3 层卷积层+2 层池化层（浅层卷积网络）。这些不同的卷积网络构造的微调网络结构使用相同的实验环境进行训练与测试，它们在 CUHK03 上不同的表现能够表明浅层与深层卷积网络在特征提取方面性能的不同。

在 CUHK01 数据集上做微调（识别性评估）时，除了在深度残差度量学习方法（DRML）上进行实验外，本文方法还与当前最先进的深度学习方法和传统的浅层学习方法进行对比实验。所有方法的技术细节如表 1 所示（其中的灰度共生矩阵（Gray-Level Co-occurrence Matrix, GLCM）^[61]是一种图像特征，能反映图像灰度的纹理方向、亮暗变化等信息，在对比方法中作为浅层学习的输入特征），其中基于深度学习的对比方法使用相同的实验环境进行训练与测试，传统的浅层学习对比方法使用 MATLAB 进行训练与测试。

表 1 CUHK01 微调对比方法

方法	来源	介绍	深度学习/浅层学习
DRML	本论文	DRML 方法是一个用于行人再识别的方法，将残差网络和度量学习与深度卷积网络相结合。	深度学习
DML	本论文	DML 方法与 DRML 方法类似，除了没有残差网络。	深度学习
Norm+X-Corr ^[62]	NIPS2016	Normalized X-Corr 方法是一个用于行人再识别的方法。它包括共享权值的卷积层（用于提取图像特征），标准化相关层（计算两个特征之间的相似度），交叉特征聚合层（包含像素的局部领域信息）和全连接层（综合较远的像素信息）。	深度学习
Fused ^[62]	NIPS2016	Fused 方法与 Normalized X-Corr 方法相似。共享权值的卷积层与 Normalized X-Corr 方法中的相同，然后将输出分别输入到两个并行的通道：一个是标准化相关层，另一个是交叉输入领域层，然后将输出分别输入到交叉特征聚合层和汇总层，最后是全连接层。	深度学习
GLCM+SVM ^[9]	传统	SVM 是一个有监督学习方法，包括相关的数据分析学习算法，用于分类或回归分析。对于给定的训练集，其中每个数据属于两个类别中的一个。SVM 算法创建的模型是非概率线性二分类模型，可以区分新数据属于两个类别中的哪一个。SVM 将原始特征空间中的点映射到高维空间以便能够区分不同类别的数据并尽可能地让间距更大。新的数据映射到相同的高维空间，根据数据在间距的两侧来预测该数据所属类别。	浅层学习
BPNN ^[7]	传统	BP 是一个训练人工神经网络的方法，通常与类似梯度下降的算法相关联。该算法不断重复两个过程，传播和权值更新。当一个向量被输入到网络中时，先一层一层地进行前向传播，直到传播到最后的输出层。然	浅层学习

		后使用误差函数将网络的输出与目标输出相比较，误差值会从输出层开始反向传播，与每一个神经元相关的误差值代表其对输出的贡献。	
GLCM+DCA [32]	传统	DCA 的基本思想是学习到一个最佳的数据映射，能够最大化不同类别数据之间的距离度量并同时最小化同一类别数据之间的距离度量。	浅层学习
GLCM+RCA [30]	传统	RCA 旨在识别数据并小规模化数据中的无关变量。该方法用线性映射来进行特征空间转换。该线性映射在“相关维度”上具有较大的值，在“不相关维度”上具有较小的值。	浅层学习
GLCM+NCA [34]	传统	NCA 是一种有监督学习方法，根据给定的距离度量进行多元数据分类。该方法提供了一个与 K-近邻算法相同的功能，并直接使用了一个称为随机近邻的概念。	浅层学习
GLCM+LM NN ^[37]	传统	LMNN 从 K 近邻分类学习一个伪度量。该算法基于半定规划（凸优化的一个子集）。	浅层学习

4.3 实验结果

在 CUHK03 数据集上进行微调实验（验证性评估）时，训练集为 CUHK03 数据集中的 labeled 集，测试集为 CUHK03 数据集中的 detected 集，实验结果如表 2 所示。

表 2 CUHK03 实验结果

卷积网络	最低误差	训练集正确率	测试集正确率
3 层卷积层+2 层池化层	2.048e-05	100%	61.12%
AlexNet ^[20]	4.225e-05	100%	76.61%
AlexNet+全卷积	0.011	99.88%	77.02%

在 CUHK01 数据集上进行微调实验（识别性评估）时，训练集为随机选取的 485 个行人对象的所有图像，测试集为其余 486 个行人对象的所有图像，结果采用查准率-查全率曲线（Precision-Recall curve, P-R curve）进行对比，如图 6 所示。

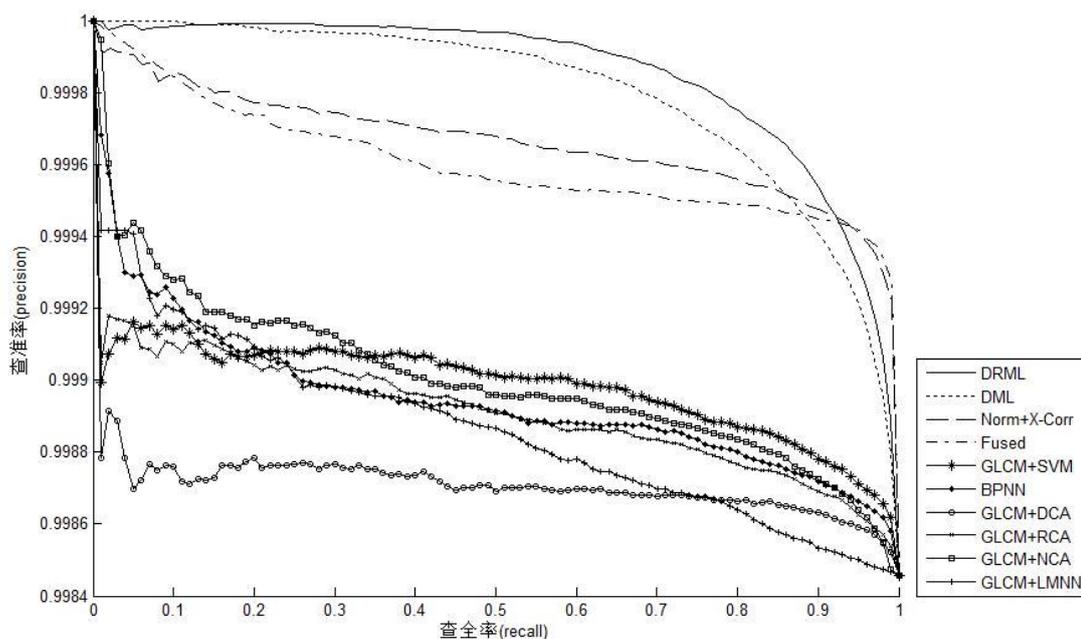


图 6 CUHK01 实验结果的查准率-查全率曲线 (P-R curve)

图 6 是通过 P-R 曲线的形式对参与对比的 10 种方法进行了定性描述，每种方法 P-R 曲线下的实际面积反应了该种方法的平均准确性。接下来，一种具备统计学意义的定量比较分析方法将在这 10 种方法所获取的所有查准率 (precision) 和查全率 (recall) 结果上进行。

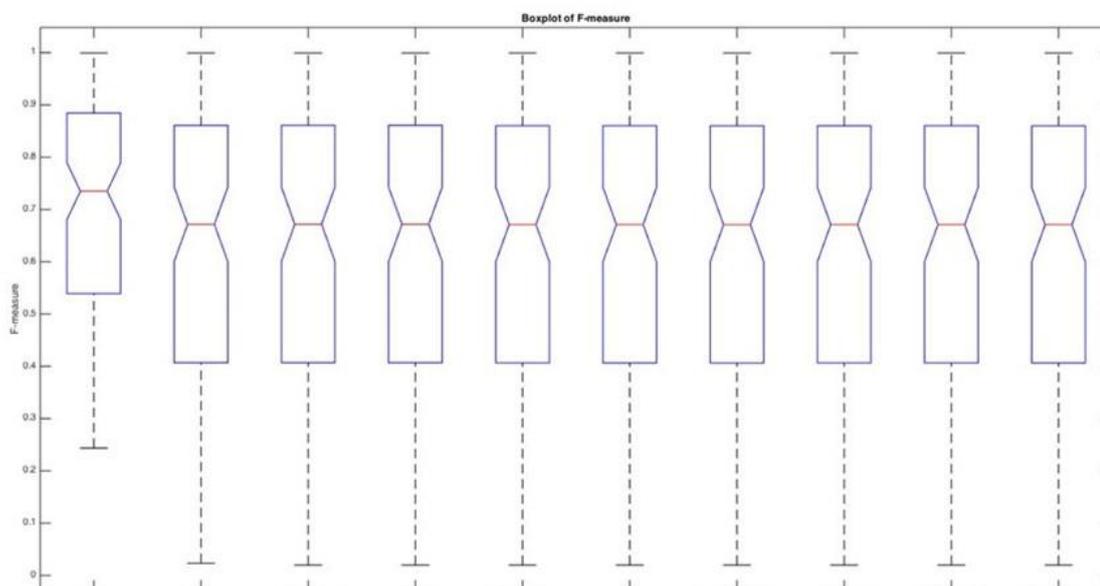


图 7 CUHK01 实验结果的 F-measure 箱状图统计结果，从左到右依次为 DRML、DML、Norm+X-Corr、Fused、GLCM+SVM、BPNN、GLCM+DCA、GLCM+RCA、GLCM+NCA、GLCM+LMNN

图 7 描绘了 CUHK01 实验结果的 F-measure 箱状图统计结果。其中，F-measure 按照 $F\text{-measure} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$ 的无偏定义进行计算。可以看出，本

论文新提出的 DRML 方法的统计箱比其他比较方法的统计箱更高（在每个统计箱中，中间的横线表示该方法的 F-measure 中位数，统计箱的上、下边界表示该方法 F-measure 的上四分位数和下四分位数），这说明 DRML 方法 F-measure 的结果要优于其他比较方法。此外，表 3 定量描述了不同方法之间两两比较的细节。在比较不同方法时，表 3 采用了点估计和区间估计两种方式。其中，点估计描绘的是两种方法 F-measure 的直接差异（该直接差异=方法 1-方法 2），区间估计描绘的是两种方法 F-measure 直接差异有可能落入的区间。可以看出 DRML 方法和其他 9 种方法比较时，点估计都是正数，这说明从点估计的角度看，DRML 方法的 F-measure 要高于其他方法。从区间估计来看，这些区间的上界恒为正且幅值均大于对应的下界幅值，这说明从区间估计来看，两种方法 F-measure 的直接差异更有可能为正数（即 DRML 方法的 F-measure 更高）。此外，另外一个有趣的现象是：DML、Norm+X-Corr、Fused 等深度学习的方法效果接近，这说明不带有深度残差网络思想的度量学习方法可以获得和现有 Norm+X-Corr、Fused 等深度学习方法类似的实验效果，而深度学习方法的实验效果从总体上优于浅层学习方法的实验效果。

表 3 CUHK01 实验的 F-measure 点估计与区间估计定量结果

方法 1	方法 2	点估计	区间估计
DRML	DML	0.0839	[-0.0381, 0.2058]
DRML	Norm+X-Corr	0.0839	[-0.0380, 0.2059]
DRML	Fused	0.0839	[-0.0381, 0.2059]
DRML	GLCM+SVM	0.0844	[-0.0375, 0.2064]
DRML	BPNN	0.0845	[-0.0375, 0.2064]
DRML	GLCM+DCA	0.0846	[-0.0374, 0.2065]
DRML	GLCM+RCA	0.0845	[-0.0375, 0.2064]
DRML	GLCM+NCA	0.0844	[-0.0375, 0.2064]
DRML	GLCM+LMNN	0.0845	[-0.0374, 0.2065]

4.4 实验分析

从在 CUHK03 数据集上进行微调（验证性评估）的实验结果（在实验测试集上的正确率）可以看到，本论文构造的卷积网络结构（AlexNet+全卷积）的表现优于 AlexNet 卷积网络结构，而 AlexNet 卷积网络结构的表现优于浅层卷积网络（3 层卷积层+2 层池化层）。本论文认为这个实验结果可以显著地说明深层的

卷积网络比浅层的卷积网络具有更好的实验效果,同时在卷积网络没有足够深的情况下,更深的卷积网络可以具有更好的实验效果,这也与卷积网络的特征相符。同时表明本论文构造的卷积网络结构能够在行人再识别问题上达到显著的验证效果。

从在 CUHK01 数据集上进行微调(识别性评估)的实验结果(查准率-查全率曲线)可以看到,本论文构造的深度残差度量学习方法(DRML)的表现显著优于其他对比方法。深度残差度量学习方法优于不含残差的深度度量学习方法(DML)表明残差网络与深度卷积度量学习相结合的表现优于深度卷积度量学习;优于其他深度学习方法或者浅层学习方法表明,深度残差度量学习方法在行人再识别问题上可以达到令人满意的识别效果。由于硬件条件(GPU 并行计算单元核数量)的限制,若进一步增加实验时间,预期可以达到更好的实验效果。

就深度残差度量学习方法(DRML)的训练过程来看,随着训练轮次的增加,测试集中正例图像距离与负例图像距离的分布差异逐渐增大,即在测试集上的表现越来越好,如图 8 所示。

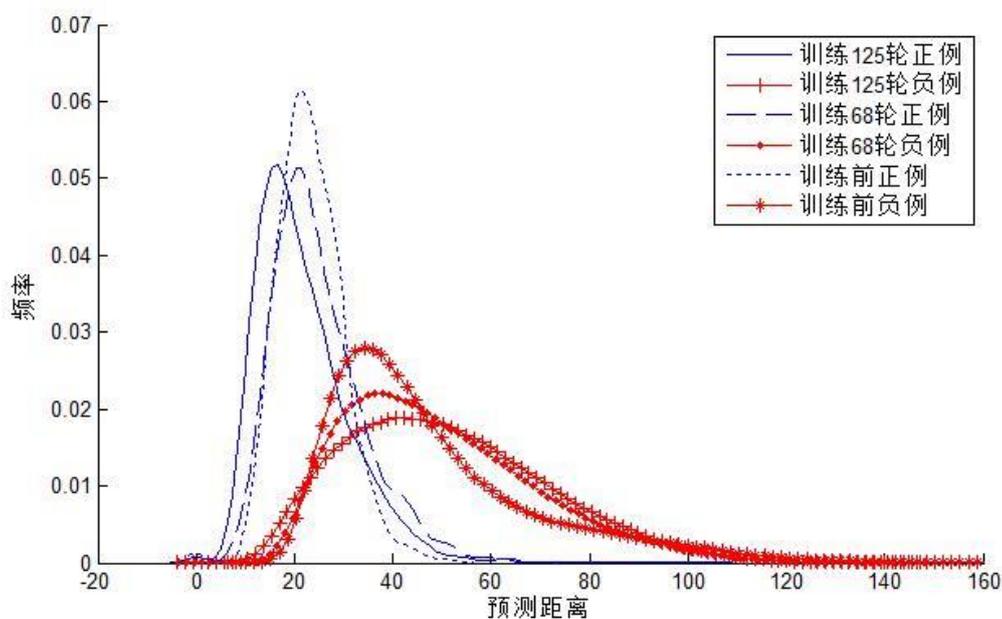


图 8 DRML 测试集正负例距离分布曲线

训练之后深度残差度量学习方法(DRML)中 *Mahalanobis* 度量层的参数 $W(64 \times 64)$ 的数值主要分布在对角线上,近似对角矩阵,即主对角线上的值都在 1 附近,其他的值都在 0 附近。主对角线取值如图 9 所示,在 $W_{3,3}$ 取得最大值 1.004,在 $W_{31,31}$ 取得最小值 0.9956,说明两张图像的特征向量在相同维上差异的增加对图像相似度的降低在第 3 维最大,第 31 维最小。

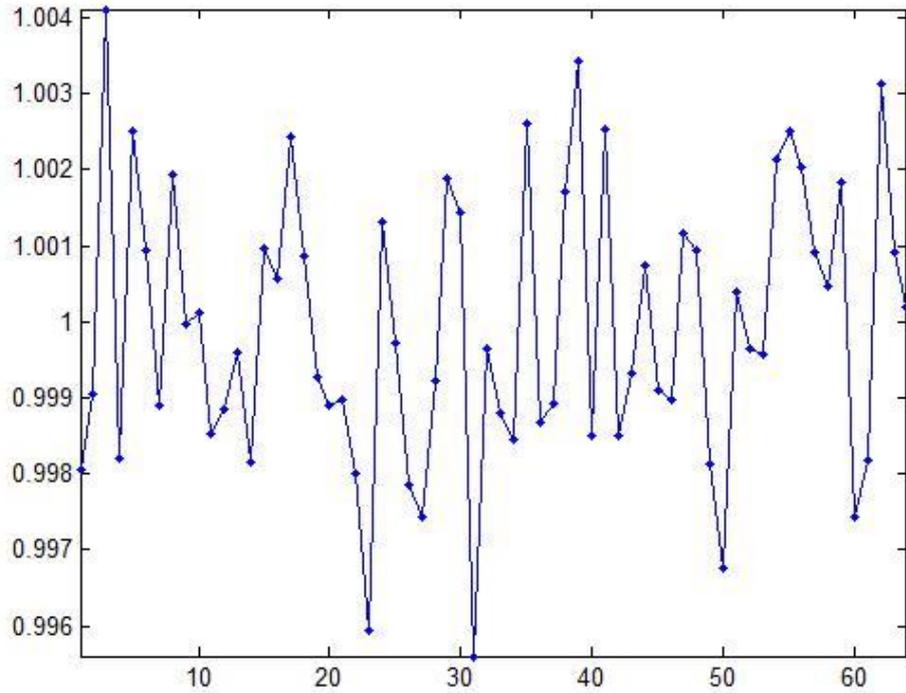


图9 $W(64 \times 64)$ 的主对角线数值分布曲线

若将主对角线上的值置为 0，则可以观察到其他值的分布近似对称矩阵，如图 10 所示，在 $W_{35,62}$ 和 $W_{62,35}$ 上取得最大值 0.00609，在 $W_{56,3}$ 和 $W_{3,56}$ 上取得最小值 -0.00528，说明两张图像的特征向量在不同维上差异的增加对图像相似度的降低在第 35 维与第 62 维之间最大，第 56 维与第 3 维之间最小。

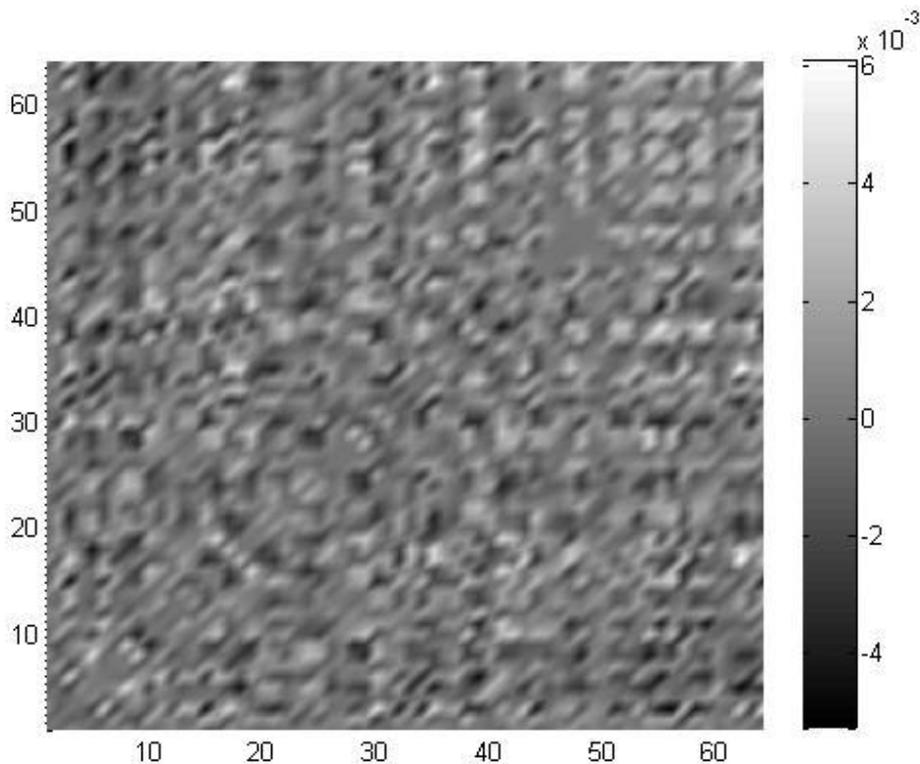


图10 $W(64 \times 64)$ 的非主对角线数值分布热点图，主对角线数值置为 0

图 11 展示了深度残差度量学习方法 (DRML) 在 CUHK01 数据集上预测错误的一些例子。从这些例子可以看出, 预测错误的正例中两张图像的颜色明亮程度不同且背景差异大, 预测错误的负例中两张图像的行人形态相似。这个现象表明网络结构中学习到的特征主要集中在形态上, 这使得给具有相似形态的两张负例图像预测变得困难。



图 11 DRML 在 CUHK01 数据集上预测失败的一些例子, 第一行: 预测错误的正例图像, 第二行: 预测错误的负例图像

第五章 结论与展望

5.1 结论

本论文构造了深度残差度量学习方法（DRML）来进行行人再识别，并获得了显著的实验效果。该新方法是将深度残差网络与度量学习相结合，从而同时具有残差网络与度量学习的优势，可以实现使用较深层次的卷积网络自动提取潜在的抽象特征信息，并且通过度量学习找出适宜本论文解决问题的度量函数。此外，本论文提出的新方法能实现特征提取部分和度量学习部分的共同优化学习，从而实现深度学习下的端对端自动学习。在实验部分，本论文提出的深度残差度量学习新方法在 CUHK03 数据集和 CUHK01 数据集上都能获得令人满意的实验效果。

5.2 未来研究展望

通过本论文深度残差度量学习方法（DRML）的优秀表现可以看到，将残差网络与度量学习相结合蕴含着巨大的优势，这些优势体现在特征提取部分运用残差网络后可以包含更深层次的网络，度量学习部分能够根据不同的数据学习到一个适应数据的度量以便更好地比较数据相似度。随着深度学习和度量学习不断发展，不难预见，深度残差度量学习方法能够运用到其他许多领域上并获得较好的效果。

参考文献

- [1] Samuel A L. Some studies in machine learning using the game of checkers[J]. IBM Journal of research and development, 1959, 3(3): 210-229.
- [2] Rosenblatt F. The perceptron: A probabilistic model for information storage and organization in the brain[J]. Psychological review, 1958, 65(6): 386.
- [3] Widrow B, Hoff M E. Adaptive switching circuits[C]//IRE WESCON convention record. 1960, 4(1): 96-104.
- [4] Hunt E B, Hovland C I. Programming a model of human concept formulation[C]//Papers presented at the May 9-11, 1961, western joint IRE-AIEE-ACM computer conference. ACM, 1961: 145-155.
- [5] Winston P H. Learning structural descriptions from examples[J]. 1970.
- [6] Mitchell, Tom M. Machine learning : an artificial intelligence approach.[M]. Springer, 1984.
- [7] Rumelhart D E, Hinton G E, Williams R J. Learning internal representations by error propagation[R]. California Univ San Diego La Jolla Inst for Cognitive Science, 1986.
- [8] Quinlan J R. Induction of decision trees[J]. Machine learning, 1986, 1(1): 81-106.
- [9] Cortes C, Vapnik V. Support-vector networks[J]. Machine learning, 1995, 20(3): 273-297.
- [10] Hinton G E, Osindero S, Teh Y W. A fast learning algorithm for deep belief nets[J]. Neural computation, 2006, 18(7): 1527-1554.
- [11] Smolensky P. Information processing in dynamical systems: Foundations of harmony theory[R]. COLORADO UNIV AT BOULDER DEPT OF COMPUTER SCIENCE, 1986.
- [12] LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.
- [13] Zhang S, Liu C, Jiang H, et al. Feedforward sequential memory networks: A new structure to learn long-term dependency[J]. arXiv preprint arXiv:1512.08301, 2015.
- [14] Silver D, Huang A, Maddison C J, et al. Mastering the game of Go with deep neural networks and tree search[J]. Nature, 2016, 529(7587): 484-489.
- [15] Minsky M, Papert S. Perceptrons[J]. 1969.
- [16] Cybenko G. Approximation by superposition of sigmoidal functions[J]. Mathematics of Control, Signals and Systems, 1989, 2(4): 303-314.
- [17] Hochreiter S. Untersuchungen zu dynamischen neuronalen Netzen[D]. diploma thesis, institut für informatik, lehrstuhl prof. brauer, technische universität münchen, 1991.
- [18] Hochreiter S, Schmidhuber J. Long short-term memory[J]. Neural computation,

1997, 9(8): 1735-1780.

[19] Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks[J]. science, 2006, 313(5786): 504-507.

[20] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[C]//Advances in neural information processing systems. 2012: 1097-1105.

[21] Olga Russakovsky*, Jia Deng*, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg and Li Fei-Fei. (* = equal contribution) ImageNet Large Scale Visual Recognition Challenge. IJCV, 2015.

[22] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.

[23] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 1-9.

[24] 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.

[25] Srivastava N, Hinton G E, Krizhevsky A, et al. Dropout: a simple way to prevent neural networks from overfitting[J]. Journal of Machine Learning Research, 2014, 15(1): 1929-1958.

[26] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[C]//Advances in neural information processing systems. 2014: 2672-2680.

[27] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 3431-3440.

[28] Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[J]. arXiv preprint arXiv:1502.03167, 2015.

[29] Xing E P, Ng A Y, Jordan M I, et al. Distance metric learning with application to clustering with side-information[C]//NIPS. 2002, 15(505-512): 12.

[30] Bar-Hillel A, Hertz T, Shental N, et al. Learning via equivalence constraints, with applications to the enhancement of image and video retrieval[C]//In: Proc. IEEE Conference on Computer Vision and Pattern Recognition. 2002.

[31] Hertz T, Bar-Hillel A, Weinshall D. Learning distance functions for image retrieval[C]//Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on. IEEE, 2004, 2: II-II.

[32] Hoi S C H, Liu W, Lyu M R, et al. Learning distance metrics with contextual constraints for image retrieval[C]//Computer vision and pattern recognition, 2006 IEEE

- computer society conference on. IEEE, 2006, 2: 2072-2078.
- [33] Yang L, Jin R, Sukthankar R. Bayesian active distance metric learning[J]. arXiv preprint arXiv:1206.5283, 2012.
- [34] JacobGoldberger S R, GeoffHinton R S. Neighbourhood components analysis[J]. NIPS'04, 2004.
- [35] Yang L, Jin R, Sukthankar R, et al. An efficient algorithm for local distance metric learning[C]//AAAI. 2006, 2: 543-548.
- [36] Sugiyama M. Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis[J]. Journal of machine learning research, 2007, 8(May): 1027-1061.
- [37] Weinberger K Q, Saul L K. Distance metric learning for large margin nearest neighbor classification[J]. Journal of Machine Learning Research, 2009, 10(Feb): 207-244.
- [38] Kruskal J B. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis[J]. Psychometrika, 1964, 29(1): 1-27.
- [39] Wold S, Esbensen K, Geladi P. Principal component analysis[J]. Chemometrics and intelligent laboratory systems, 1987, 2(1-3): 37-52.
- [40] Tenenbaum J B, De Silva V, Langford J C. A global geometric framework for nonlinear dimensionality reduction[J]. science, 2000, 290(5500): 2319-2323.
- [41] Roweis S T, Saul L K. Nonlinear dimensionality reduction by locally linear embedding[J]. science, 2000, 290(5500): 2323-2326.
- [42] Belkin M, Niyogi P. Laplacian eigenmaps and spectral techniques for embedding and clustering[C]//NIPS. 2001, 14(14): 585-591.
- [43] He X, Niyogi P. Locality preserving projections[C]//NIPS. 2003, 16(2003).
- [44] He X, Cai D, Yan S, et al. Neighborhood preserving embedding[C]//Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on. IEEE, 2005, 2: 1208-1213.
- [45] Yi D, Lei Z, Liao S, et al. Deep metric learning for person re-identification[C]//Pattern Recognition (ICPR), 2014 22nd International Conference on. IEEE, 2014: 34-39.
- [46] Mahalanobis P C. On the generalized distance in statistics[J]. Proceedings of the National Institute of Sciences (Calcutta), 1936, 2: 49-55.
- [47] Hu J, Lu J, Tan Y P. Discriminative deep metric learning for face verification in the wild[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014: 1875-1882.
- [48] Shi H, Zhu X, Liao S, et al. Constrained deep metric learning for person re-identification[J]. arXiv preprint arXiv:1511.07545, 2015.

- [49] Lu J, Wang G, Deng W, et al. Multi-manifold deep metric learning for image set classification[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 1137-1145.
- [50] Hu J, Lu J, Tan Y P. Deep transfer metric learning[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 325-333.
- [51] Hu J, Lu J, Tan Y P. Deep metric learning for visual tracking[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2016, 26(11): 2056-2068.
- [52] Hoffer E, Ailon N. Deep metric learning using triplet network[C]//International Workshop on Similarity-Based Pattern Recognition. Springer International Publishing, 2015: 84-92.
- [53] Liu H, Ma B, Qin L, et al. Set-label modeling and deep metric learning on person re-identification[J]. Neurocomputing, 2015, 151: 1283-1292.
- [54] Chen S Z, Guo C C, Lai J H. Deep ranking for person re-identification via joint representation learning[J]. IEEE Transactions on Image Processing, 2016, 25(5): 2353-2367.
- [55] Bengio Y, Simard P, Frasconi P. Learning long-term dependencies with gradient descent is difficult[J]. IEEE transactions on neural networks, 1994, 5(2): 157-166.
- [56] Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks[C]//Aistats. 2010, 9: 249-256.
- [57] Y. LeCun, L. Bottou, G. B. Orr, and K.-R. Muller. Efficient backprop."In Neural Networks: Tricks of the Trade, pages 9–50. Springer, 1998.
- [58] Deng J, Dong W, Socher R, et al. Imagenet: A large-scale hierarchical image database[C]//Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. IEEE, 2009: 248-255.
- [59] Li W, Zhao R, Xiao T, et al. Deepreid: Deep filter pairing neural network for person re-identification[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014: 152-159.
- [60] Li W, Zhao R, Wang X. Human reidentification with transferred metric learning[C]//Asian Conference on Computer Vision. Springer Berlin Heidelberg, 2012: 31-44.
- [61] Haralick R M, Shanmugam K. Textural features for image classification[J]. IEEE Transactions on systems, man, and cybernetics, 1973, 3(6): 610-621.
- [62] Subramaniam A, Chatterjee M, Mittal A. Deep Neural Networks with Inexact Matching for Person Re-Identification[C]//Advances in Neural Information Processing Systems. 2016: 2667-2675.

致谢

通过这一阶段的努力，我的毕业设计论文《深度度量学习的理论研究与实际运用》终于完成了，这意味着我的本科生活即将结束。在我学习深度度量学习的阶段，我不仅在学习上获得了很多知识，在思想上也得到了很多启发，这不仅与我自身的努力有关，还与各位老师、同学的鼓励、支持和帮助分不开。在此也十分感谢指导我完成这篇论文的黄伟老师，再次感谢！