

基于三元组约束的深度距离度量学习

刘 铭

院 (系): 英才学院

专 业: 计算机科学与技术

学 号: 7130310111

指导教师: 左旺孟教授

2017 年 6 月

哈爾濱工業大學

毕业设计（论文）

题 目 基于三元组约束的

深度距离度量学习

专 业 计算机科学与技术

学 号 7130310111

学 生 刘铭

指 导 教 师 左旺孟教授

答 辩 日 期 2017.06.24

摘要

相似性学习是计算机视觉和模式识别领域中最基本的问题之一。相似性学习的目标是学习到合适的距离（相似性）度量，使得相似的图像间具有较小的距离或较大的相似度，不相似的图像间有较大的距离或较小的相似度。这样，我们就可以通过对相似度排序，进行相似性匹配，在图片集中找到与查询图片最相似的一个或几个图片。

传统相似性匹配研究主要集中于同样模态的数据，为了便于优化，通常采用线性映射与马氏距离或者余弦相似度相结合的方式。对不同领域的数据进行相似性匹配比同模态数据更具有挑战性，在现实中也有很高的应用价值，比如，对人的身份证照片和监控视频进行匹配。跨领域相似性匹配的传统方法一般分为两步：1) 将不同领域的样本映射到同一个公共子空间，2) 在这个公共子空间学习距离（相似性）度量。

我们用广义相似性度量，结合了马氏距离和余弦相似度，并将传统的线性变换拓展为仿射变换，实现了跨领域相似性匹配算法。然后我们在三元组约束下，基于深度学习框架 Caffe 实现了深度特征表达和广义相似性度量的联合端到端学习。我们在 CUHK-03 数据集上进行了多组实验，实验结果表明，我们的模型在该数据集上取得了较高的识别准确率。

关键词：深度学习；度量学习；跨领域匹配；三元组；

Abstract

Similarity learning is one of the most fundamental problems in the field of computer vision and pattern recognition. Similarity learning aims to learn a suitable distance (similarity) metric, on which similar images have smaller distance or bigger similarity, and dissimilar images have bigger distance or smaller similarity. Thus we can do similarity matching by sorting the similarity, then find one or more images from the image set which are most similar to the query image.

The study on traditional similarity matching mainly focus on data of same modality, and for the convenience of optimization, the method of combining linear projection and Mahalanobis distance or Cosine similarity is usually taken. It's more challenging when we do similarity matching on cross-domain visual data compared data of same modality, and it's pretty useful to do cross-domain data matching, e.g., matching the photo from the ID card with the surveillance video. The traditional cross-domain similarity matching methods usually takes two steps: 1) project the data from different domain into a common space, 2) learn distance (similarity) metric on the common space.

We implemented the cross-domain similarity matching algorithm via generalized similarity measure, combining Mahalanobis distance and Cosine similarity, and extend traditional linear projection into affine transformation. Then we implemented the joint end-to-end learning of deep feature representation and generalized similarity measure under the triplet constraint, based on the deep learning framework Caffe. We conducted several groups of experiments on CUHK-3 dataset, experiment results show that our model .get a pretty high matching accuracy on this dataset.

Keywords: deep learning, metric learning, cross-domain matching, triplets

目 录

摘要	I
ABSTRACT	II
第 1 章 绪论	- 1 -
1.1 课题背景及研究的目的和意义.....	- 1 -
1.2 跨领域相似性匹配及其相关理论的发展概况.....	- 2 -
1.2.1 相似性（距离）度量学习	- 2 -
1.2.2 传统跨领域相似性匹配	- 3 -
1.2.3 基于深度特征表达的跨领域相似性匹配	- 3 -
1.2.4 三元组约束	- 4 -
1.3 综述与分析.....	- 5 -
第 2 章 广义相似性度量	- 6 -
2.1 引言	- 6 -
2.2 度量学习	- 6 -
2.2.1 欧氏距离和马氏距离	- 6 -
2.2.2 余弦相似度	- 7 -
2.2.3 马氏距离与余弦相似度的结合	- 8 -
2.3 广义相似性度量	- 8 -
2.3.1 仿射变换	- 9 -
2.3.2 广义相似性度量的推导	- 9 -
2.3.3 广义相似性度量与已有相似性模型的联系	- 11 -
2.4 本章小结	- 12 -
第 3 章 基于三元组约束的跨领域相似性匹配	- 13 -
3.1 引言	- 13 -
3.2 三元组约束	- 13 -
3.2.1 构造三元组	- 13 -

3.2.2 单个三元组的目标函数	- 14 -
3.3 深度神经网络实现.....	- 15 -
3.3.1 损失函数的构造	- 15 -
3.3.2 网络结构	- 16 -
3.3.3 网络的训练	- 17 -
3.4 实验结果.....	- 18 -
3.5 本章小结.....	- 20 -
第 4 章 实验拓展及相关分析	- 21 -
4.1 引言	- 21 -
4.2 对正负样本距离间隔的调整.....	- 21 -
4.3 对网络结构的调整.....	- 22 -
4.4 改变样本选取方式.....	- 22 -
4.4 本章小结.....	- 23 -
结论	- 24 -
参考文献	- 25 -
哈尔滨工业大学本科毕业设计（论文）原创性声明	- 29 -
致谢	- 30 -

第1章 绪论

1.1 课题背景及研究的目的和意义

视觉是人类获取外界信息的最重要来源之一，是人类认识客观世界、了解外界情况的最重要凭借，人类约有 80% 的信息获取自视觉。从信息形式上进行分析，我们能够发现视觉数据中包含的信息往往比其他来源得到的信息更加丰富。同样，计算机视觉也是计算机科学研究的重要组成部分，基于计算机视觉的应用，也具有很高的价值。

计算机视觉领域的最终研究目标是使计算机能像人一样通过视觉观察、感知和理解世界，进而拥有自主适应环境的能力，这是在长期努力、不断发展的基础上才能做到的。就计算机视觉领域目前的发展阶段而言，我们努力的目标是建立各种具有特定应用的视觉系统，使计算机在此基础上，能依靠针对特定视觉任务的具有一定能力的智能完成一定的任务。想要让计算机拥有模仿人脑对视觉信息进行处理和解释的能力，甚至进一步让计算机对视觉信息处理、理解、解释、反应的能力超越人类，首先要做的，就是让计算机对视觉信息具有一定认识、判断的能力，例如相似性学习。相似性学习是计算机视觉和模式识别领域中最基本的问题之一，其要求是，在某种距离（相似性）度量上，使得对于人类来说视觉上相似的图像间具有较小的距离（或较大的相似度），视觉上不相似的图像间有较大的距离（或较小的相似度）。

跨领域视觉数据的相似性匹配则是相似性匹配任务中更加具有挑战性的部分。跨领域视觉相似性匹配有静止图像和视频间人脸检索^[1]、不同年龄人脸的匹配^[2]、人脸照片和素描图像之间的匹配^[3]等多种价值非常高的应用。很多时候我们无法保证获取到的数据是同一模态的，尤其是在数据量较小的应用场合，或者只能获取特定类型数据的情况下，如案件侦破、嫌疑人锁定等，对跨领域数据进行相似性匹配是必不可少的，具有非常高的应用价值。并且，对跨领域数据进行相似性匹配，可以促使我们提取数据更加本质和更加泛化的特征，对计算机视觉和模式识别的研究是有积极作用的。

不仅跨领域相似性匹配本身具有非常高的应用价值，基于此，我们还可以进行更多的研究和应用探索。例如，在跨领域相似性匹配性能足够好的前提下，我们可以将其作为对抗生成网络中的判别器，进而使得对抗生成网络的生成器具有

生成不同模态数据的能力，如根据照片生成动漫头像或者素描像，又如根据素描像生成照片。

卷积神经网络是近年来发展起来的一种高效识别方法，具有局部感知、参数共享、网络复杂度低、易于并行化等特点，而且能够隐式地从训练数据中学习样本特征，避免了对图像复杂的前期预处理，在计算机视觉领域得到了广泛的应用。

我们在卷积神经网络框架下对跨领域数据相似性匹配进行研究，基于三元组约束实现深度特征表达和广义相似性度量的联合端到端学习，并基于深度学习框架 Caffe^[4]实现相似性匹配算法，验证算法的有效性。

1.2 跨领域相似性匹配及其相关理论的发展概况

我们基于三元组约束实现深度特征表达和广义相似性度量的联合端到端学习以完成跨领域数据的相似性匹配，因此在本节中，相关安排如下：1.2.1 简单介绍了相似性度量学习，1.2.2 回顾了几种传统跨领域相似性匹配方法，1.2.3 列举了几个深度特征表达和相似性匹配结合的例子，1.2.4 则对三元组进行了简单讨论。

1.2.1 相似性（距离）度量学习

相似性（距离）度量学习通过学习一个合适的距离度量，使得同类样本之间的距离减小、相似度增大，异类样本之间距离增大、相似度减小，进而提高分类器的判别能力，获得更高的分类性能。度量学习被广泛应用于分类、聚类、检索、身份认证等问题中，并取得了良好的效果，是计算机视觉和机器学习领域重要的研究方向，在模式识别和机器学习中，K 近邻算法、K 均值算法等，都和相似度或者距离有关。这些算法中，相似度（距离）的计算发挥着巨大的作用，直接影响着算法的性能。

传统度量学习主要在特征空间确定（即已经对原始数据完成预处理并提取特征）的情况下，学习马氏距离或者余弦相似性，比如大间隔最近邻（LMNN）^[5]、信息理论度量学习（ITML）^[6]等方法。目前，对于两个样本之间距离的度量形式有很多种，常见的如闵可夫斯基距离（包含曼哈顿距离、欧氏距离、切比雪夫距离等）、马氏距离、海明距离等，其中较为常用的是马氏距离。

我们以马氏距离为例，说明度量学习中距离的概念。在学习到马氏距离度量矩阵 \mathbf{M} 后，两个样本 \mathbf{x} 和 \mathbf{y} 之间的距离可以表示为：

$$d_{\mathbf{M}}^2(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|_{\mathbf{M}}^2 = (\mathbf{x} - \mathbf{y})^T \mathbf{M} (\mathbf{x} - \mathbf{y}) \quad (1-1)$$

其中 \mathbf{M} 是距离度量矩阵。 \mathbf{M} 被限制为半正定矩阵时（此时距离非负），存在一个

分解 $\mathbf{M} = \mathbf{L}^T \mathbf{L}$, 其中 \mathbf{L} 可以看作一个线性相似性映射矩阵, 则公式(1-1)可以写成:

$$d_{\mathbf{M}}^2(\mathbf{x}, \mathbf{y}) = (\mathbf{x} - \mathbf{y})^T \mathbf{L}^T \mathbf{L} (\mathbf{x} - \mathbf{y}) = \|\mathbf{L}(\mathbf{x} - \mathbf{y})\|_2^2 \quad (1-2)$$

特别地, 当 \mathbf{L} 与 \mathbf{M} 是单位阵时, 马氏距离等价于欧氏距离, 也就是说, 欧氏距离是马氏距离的一个特例。在学习马氏距离时, 通常分为学习 \mathbf{M} 和学习 \mathbf{L} 。

在相似性变换方面, 人们常用线性变换以便于优化, 这样在保留目标形状方面具有很好的优势, 但是在复杂形变方面受到极大的限制, 而仿射变换可以表示平移、缩放、翻转、旋转、错切等多种变换, 具有更好表示复杂变换的能力, 因此仿射变换越来越受到欢迎。如公式(1-2)所示, \mathbf{L} 为线性相似性映射, 且对 \mathbf{x} 和 \mathbf{y} 进行的是同样的线性映射 \mathbf{L} 。而对于仿射变换, 我们可以对 \mathbf{x} 和 \mathbf{y} 分别进行 $\mathbf{L}_A \mathbf{x} + \mathbf{a}$ 和 $\mathbf{L}_B \mathbf{y} + \mathbf{b}$ 操作, 这样既对复杂性变有了更好的表示能力, 也拥有了针对不同样本提取不同特征的能力, 更加适合于跨领域(模态)数据。

1.2.2 传统跨领域相似性匹配

对于跨领域视觉数据的相似性匹配, 常用的做法是首先将不同领域的数据映射到同一特征子空间, 然后将一种距离度量方法(通常是欧式距离或者内积)应用到该空间。例如典型相关分析(CCA)通过最大化空间中两个随机变量关于基向量投影的相关性来学习公共子空间^[7,8], 偏最小二乘法(PLS)^[9]通过最小化误差平方和寻找最佳函数匹配, Duan 等人提出的领域适应机(DAM)^[10]可以用于多领域自适应, 其限制在于需要一组预训练的基分类器。

有人提出一些基于类别标签的寻找具有判别性的公共子空间的方法, 如带有组结构的有监督耦合字典学习(SCDL)^[11]被用于多模态检索, 又如多视图判别分析方法(MvDA)^[12]通过同时优化类内和类间瑞利商以寻找公共子空间。有一些与度量学习相结合的方法也被提出, 如 Li 等人在局部自适应函数(LADF)^[13]方法中改造二阶决策函数作为距离度量进行身份验证。Kang 等人^[14]通过加速近端梯度(APG)算法学习一个低秩矩阵, 进行参数化相似性测量。

1.2.3 基于深度特征表达的跨领域相似性匹配

用深度神经网络代替人工进行特征的提取, 可以使提取出的特征更加符合算法需要, 不仅免去了对原始数据进行大量预处理的麻烦, 还能够大幅度提高算法的性能。利用卷积神经网络(CNN)^[15]学习特征表达和上下文关系已经在很多问题中发挥出巨大的作用, 如在目标识别^[16]和语义分割^[17]中有良好的表现, 并且在人脸识别^[18]和行人重识别^[19-21]方面也取得了极好的效果。

Andrew 等人^[22]提出了一个包含多个堆叠非线性变换的多层 CCA 模型；Li 等人^[23]通过深度网络学习一个滤波器，并用学习到的滤波器处理样本中的未对齐情况和光学、几何形变，在行人重识别中取得了较好的效果；Chen 等人^[21]提出了一种深度排序框架，并将其中的相似性模型替换为全连接模型，用于学习图像对的联合表示，然后直接给出相似性评分。

Lin 等人^[24]提出广义相似性度量，对马氏距离和余弦相似度实现了联合表示，将局部自适应函数(LADF)^[13]和联合贝叶斯学习^[25]等已有模型在表现形式上统一，最后实现基于二元组约束的深度特征表达与相似性度量联合学习的深度神经网络，大大提高了识别的性能。

1.2.4 三元组约束

在度量学习中，很多方法基于二元组约束实现，即直接比较两个样本相同或者不同。基于二元组约束的方法，用一个变量（值为-1 或 1）表示样本是否同类，优化的目标为该变量乘以原本的目标函数，这样可以使同类样本相似性为正，异类样本相似性为负。

近期，基于三元组约束的研究也流行起来，并且在相似性匹配中已经取得了喜人的效果。如谷歌的 FaceNet^[26]，就是基于卷积神经网络的三元组约束下完成的工作，刷新了野外带标记人脸（LFW, Labeled Faces in the Wild）数据库^[27]上人脸验证的效果记录，达到 99.63%，超过此前最好结果 30%。

从训练集中随机选取一个样本（如图 1-1 中 Anchor，记作 x_i^a ），然后再随机选取一个与 Anchor 同一类别的样本（如图 1-1 中 Positive，记作 x_i^p ）和一个与 Anchor 不同类别的样本（如图 1-1 中 Negative，记作 x_i^n ），则 (x_i^a, x_i^p, x_i^n) 为一个三元组。其训练的目标是，获得一个相似性度量，使得样本 x_i^a 与 x_i^n 之间的距离足够大， x_i^a 与 x_i^p 之间的距离足够小，两个距离之差最小为 α ，即如公式(1-3)所示：

$$\|f(x_i^a) - f(x_i^p)\|_2^2 + \alpha < \|f(x_i^a) - f(x_i^n)\|_2^2 \quad (1-3)$$

其中， $f(x)$ 表示对样本 x 提取的特征。

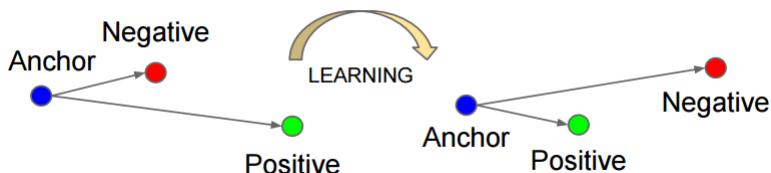


图 1-1 三元组约束示意图^[26]

1.3 综述与分析

本文主要研究基于三元组约束下的广义相似性度量学习方法。我们首先在卷积神经网络下，对不同领域（模态）的样本分别进行深度特征提取，然后将每次训练中的样本的特征组成三元组，基于广义相似性度量学习方法进行学习，以完成对跨领域视觉数据的相似性匹配。各章内容安排如下：

第一章主要介绍了研究的背景和主要意义，总结了国内外在相似性度量学习、跨领域匹配、三元组约束方面的研究现状和进展，对本文研究内容进行了综述。

第二章首先介绍了常见的相似性度量，包括马氏距离、余弦相似度以及它们的结合等，然后简单介绍了仿射变换相比于线性变换的优势及其在相似性度量中的应用，最后基于前两点，给出了广义相似性度量的推导和广义相似性度量与一些其他模型之间的关系。

在第三章中，我们提出了三元组约束下的跨领域相似性匹配学习方法，基于深度神经网络完成深度特征表达和广义相似性度量的联合端到端学习。在本章中，我们首先介绍了三元组的构造以及对于单个三元组的目标函数，然后介绍了深度神经网络的损失函数构造、网络结构以及训练方法，最后给出了我们初步的实验结果。

在第四章中，我们给出了一系列的拓展性实验结果。我们对一些想法和推断进行了验证性实验，对实验结果进行了简单分析。最终给出算法实现的性能和其他相关方法的对比结果。

第2章 广义相似性度量

2.1 引言

相似性度量学习或称距离度量学习，是计算机视觉和机器学习领域一个十分活跃的研究方向，其主要目的是学习一个距离度量，通过该度量，使同类样本之间的间隔更小，异类样本之间的间隔更大，以达到提高识别性能的目的。许多数据挖掘和分析任务中也需要进行样本的比较并给出它们的相似性，如聚类、最近邻搜索、相关性分析等。此外，还有很多实际应用需要计算样本之间的相似性或者距离，如推荐系统、文本分类、信息检索等。

在本章中，我们对度量学习进行简单阐述，并由此引出广义相似性度量。

2.2 度量学习

在本节中，我们介绍了几种常见的距离度量（相似性度量），如欧氏距离、马氏距离、余弦相似度等，以及它们的一些组合和发展。

2.2.1 欧氏距离和马氏距离

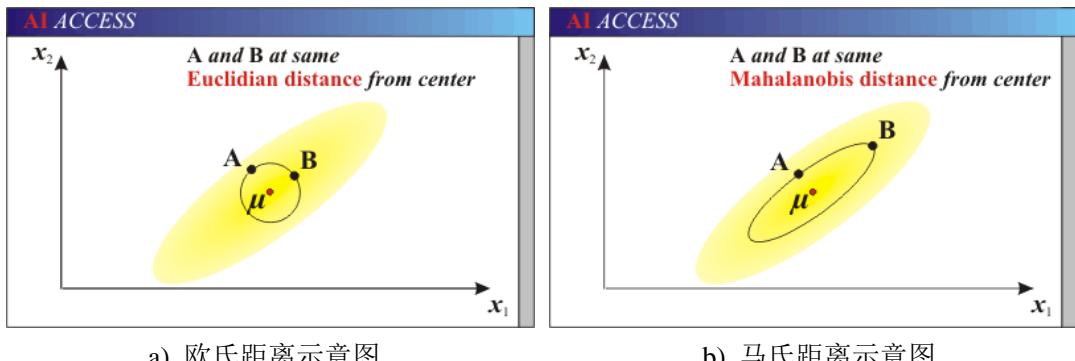
度量学习的核心问题在于学习一个合适的距离度量，如绪论中 1.2.1 条所述，马氏距离是距离度量中常用的一种，具有较好的性质，能取得不错的效果。我们将公式(1-1)所示的马氏距离表达式写成矩阵形式，并用 $\mathbf{f}_1(\mathbf{x})$ 和 $\mathbf{f}_2(\mathbf{y})$ 表示分别对样本 \mathbf{x} 和 \mathbf{y} 用 $\mathbf{f}_1(\cdot)$ 和 $\mathbf{f}_2(\cdot)$ 提取的特征，如公式(2-1)所示：

$$\begin{aligned} d_M^2(\mathbf{f}_1(\mathbf{x}), \mathbf{f}_2(\mathbf{y})) &= \|\mathbf{f}_1(\mathbf{x}) - \mathbf{f}_2(\mathbf{y})\|_M^2 \\ &= (\mathbf{f}_1(\mathbf{x}) - \mathbf{f}_2(\mathbf{y}))^T \mathbf{M} (\mathbf{f}_1(\mathbf{x}) - \mathbf{f}_2(\mathbf{y})) \\ &= \begin{bmatrix} \mathbf{f}_1(\mathbf{x})^T & \mathbf{f}_2(\mathbf{y})^T \end{bmatrix} \begin{bmatrix} \mathbf{M} & -\mathbf{M} \\ -\mathbf{M} & \mathbf{M} \end{bmatrix} \begin{bmatrix} \mathbf{f}_1(\mathbf{x}) \\ \mathbf{f}_2(\mathbf{y}) \end{bmatrix} \end{aligned} \quad (2-1)$$

马氏距离的实质是利用 Cholesky 变换消除样本特征不同维度之间的相关性，并能处理不同维度尺度不同的问题。计算马氏距离的时候，要求不共面的总体样本数大于样本的维度，否则总体样本协方差矩阵的逆不存在，即无法得到度量矩阵 \mathbf{M} 。

相比于马氏距离，我们对欧氏距离更加熟悉，现实生活中我们最常使用的就是欧氏距离。但是欧式距离存在明显的缺点，即欧氏距离对样本不同维度的数据

之间同等看待，在实际应用中，很多时候无法满足我们的需求，样本不同属性对于区分不同个体有着不同的重要性。比如进行人脸识别的时候，表情所起到的作用和鼻子的形状的贡献度并不能直接划等号，我们要突出更具有判别性的属性和维度。马氏距离的优势就在于可以不受量纲的影响，两个样本之间的马氏距离与原始数据的量纲无关，但是马氏距离存在一个明显的缺点，就是可能会过于突出数据变化较小的维度的作用，可能导致在训练集上过拟合。欧氏距离与马氏距离示意如图 2-1 所示：

图 2-1 欧式距离与马氏距离示意^[28]

在如图 2-1 所示的正态分布图中，对于中心点 μ 我们认为 b) 图中 **A** 和 **B** 两点是相同的，但是显然欧式距离无法做到这一点，而马氏距离可以做到。

2.2.2 余弦相似度

余弦相似度用向量空间中两个向量夹角的余弦值度量两个样本之间的差异。不同于距离度量，余弦相似度更加关注两个向量在方向上的差异，而非两者之间的距离。余弦相似度写成矩阵乘积的形式如公式(2-2)所示：

$$\begin{aligned} S(\mathbf{f}_1(\mathbf{x}), \mathbf{f}_2(\mathbf{y})) &= (\mathbf{f}_1(\mathbf{x}))^T \mathbf{H} \mathbf{f}_2(\mathbf{y}) \\ &= \begin{bmatrix} \mathbf{f}_1(\mathbf{x})^T & \mathbf{f}_2(\mathbf{y})^T \end{bmatrix} \begin{bmatrix} \mathbf{0} & \frac{1}{2}\mathbf{H} \\ \frac{1}{2}\mathbf{H} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{f}_1(\mathbf{x}) \\ \mathbf{f}_2(\mathbf{y}) \end{bmatrix} \end{aligned} \quad (2-2)$$

其中，经过学习可以得到一个余弦相似性度量矩阵 \mathbf{H} ，使得相似样本的余弦相似度尽量大，差异较大的样本之间余弦相似度尽量小。特别地，当 \mathbf{H} 为单位阵时，余弦相似度退化为标准向量内积。余弦相似度的最常见应用就是计算文本的相似度，而在相似性匹配领域，余弦相似度也得到了较为广泛的应用，可以较好地计算两个样本之间统计学上的相似度。

2.2.3 马氏距离与余弦相似度的结合

我们反复强调了在度量学习中，学习一个好的距离度量的重要性。相较于提出全新的距离度量，很多人将已有的模型进行了拓展，或者是将几种相似性模型进行了融合，以提高度量的性能和表现。

Chen 等人在 2010 年对马氏距离进行了拓展，将公式(2-1)中相似性矩阵的副对角线元素进行了替换，形式如公式(2-3)所示：

$$\begin{aligned} D(\mathbf{f}_1(\mathbf{x}), \mathbf{f}_2(\mathbf{y})) &= \begin{bmatrix} \mathbf{f}_1(\mathbf{x})^T & \mathbf{f}_2(\mathbf{y})^T \end{bmatrix} \begin{bmatrix} \mathbf{M} & -\mathbf{G} \\ -\mathbf{G} & \mathbf{M} \end{bmatrix} \begin{bmatrix} \mathbf{f}_1(\mathbf{x}) \\ \mathbf{f}_2(\mathbf{y}) \end{bmatrix} \\ &= (\mathbf{f}_1(\mathbf{x}) - \mathbf{f}_2(\mathbf{y}))^T \mathbf{M} (\mathbf{f}_1(\mathbf{x}) - \mathbf{f}_2(\mathbf{y})) + 2\mathbf{f}_1(\mathbf{x})^T (\mathbf{M} - \mathbf{G}) \mathbf{f}_2(\mathbf{y}) \end{aligned} \quad (2-3)$$

将其展开后可以发现，这种形式相当于结合了马氏距离度量和余弦相似性度量。此后，Li 等人^[13]在 2013 年对其进行了进一步修改，将公式(2-3)展开式中类似余弦相似性度量的项 $2\mathbf{f}_1(\mathbf{x})^T (\mathbf{M} - \mathbf{G}) \mathbf{f}_2(\mathbf{y})$ 直接改为余弦相似性度量 $\mathbf{f}_1(\mathbf{x})^T \mathbf{G} \mathbf{f}_2(\mathbf{y})$ ，即如下公式(2-4)所示的形式：

$$\begin{aligned} S(\mathbf{f}_1(\mathbf{x}), \mathbf{f}_2(\mathbf{y})) &= \mathbf{f}_1(\mathbf{x})^T \mathbf{G} \mathbf{f}_2(\mathbf{y}) - (\mathbf{f}_1(\mathbf{x}) - \mathbf{f}_2(\mathbf{y}))^T \mathbf{M} (\mathbf{f}_1(\mathbf{x}) - \mathbf{f}_2(\mathbf{y})) \\ &= \begin{bmatrix} \mathbf{f}_1(\mathbf{x})^T & \mathbf{f}_2(\mathbf{y})^T \end{bmatrix} \begin{bmatrix} -\mathbf{M} & \frac{1}{2}\mathbf{G} + \mathbf{M} \\ \frac{1}{2}\mathbf{G} + \mathbf{M} & -\mathbf{M} \end{bmatrix} \begin{bmatrix} \mathbf{f}_1(\mathbf{x}) \\ \mathbf{f}_2(\mathbf{y}) \end{bmatrix} \end{aligned} \quad (2-4)$$

我们考虑 $\mathbf{f}_1(\mathbf{x})$ 和 $\mathbf{f}_2(\mathbf{y})$ 为来自两个不同领域的数据，则我们可以对公式(2-3)进行改造，对两个领域的数据分别用不同的马氏距离度量矩阵，如公式(2-5)所示：

$$S(\mathbf{f}_1(\mathbf{x}), \mathbf{f}_2(\mathbf{y})) = \begin{bmatrix} \mathbf{f}_1(\mathbf{x})^T & \mathbf{f}_2(\mathbf{y})^T \end{bmatrix} \begin{bmatrix} \mathbf{M}_1 & -\mathbf{G} \\ -\mathbf{G} & \mathbf{M}_2 \end{bmatrix} \begin{bmatrix} \mathbf{f}_1(\mathbf{x}) \\ \mathbf{f}_2(\mathbf{y}) \end{bmatrix} \quad (2-5)$$

我们令 $\mathbf{M}_1 = \mathbf{A}^T \mathbf{A}$, $\mathbf{M}_2 = \mathbf{B}^T \mathbf{B}$, $\mathbf{G} = \mathbf{C}^T \mathbf{D}$ ，可以得到：

$$S(\mathbf{f}_1(\mathbf{x}), \mathbf{f}_2(\mathbf{y})) = \begin{bmatrix} \mathbf{f}_1(\mathbf{x})^T & \mathbf{f}_2(\mathbf{y})^T \end{bmatrix} \begin{bmatrix} \mathbf{A}^T \mathbf{A} & -\mathbf{C}^T \mathbf{D} \\ -\mathbf{D}^T \mathbf{C} & \mathbf{B}^T \mathbf{B} \end{bmatrix} \begin{bmatrix} \mathbf{f}_1(\mathbf{x}) \\ \mathbf{f}_2(\mathbf{y}) \end{bmatrix} \quad (2-6)$$

公式(2-6)为下一步广义相似性度量的推导打下了基础，我们将在 2.3 节中详述广义相似性度量。

2.3 广义相似性度量

在本节中，我们首先在 2.3.1 条引入了仿射变换，并对其进行了简单介绍，之后，我们详细介绍广义相似性度量的推导和性质。

2.3.1 仿射变换

正如我们在 1.2.1 条中所述，在相似性变换方面，人们常使用线性变换以便于优化，而且可以较好的保持样本的形状和空间位置关系，但是在复杂形变方面，线性变换有极大的限制，无法很好的表示复杂的形变，如平移、错切和它们的组合等。

相比于线性变换，仿射变换具有更好的表示复杂变换的能力，可以表示平移、缩放、翻转、旋转、错切等多种变换，因此仿射变换越来越受到欢迎。我们可以通过向线性变换 \mathbf{Ux} 和 \mathbf{Vy} 添加变换向量将其拓展为仿射变换，即用 $\mathbf{L}\mathbf{Ax}+\mathbf{a}$ 和 $\mathbf{L}\mathbf{By}+\mathbf{b}$ 分别代替 \mathbf{Ux} 和 \mathbf{Vy} 来对样本 \mathbf{x} 和 \mathbf{y} 进行变换。仿射变换不要求保留线性空间中的原点，是线性变换的一种推广。

Chang 等人^[29]在欧氏距离度量上对样本应用了仿射变换以提高图片检索性能。他们定义的距离度量如公式(2-7)所示：

$$\begin{aligned} D(\mathbf{f}_1(\mathbf{x}), \mathbf{f}_2(\mathbf{y})) &= \|(\mathbf{Af}_1(\mathbf{x}) + \mathbf{a}) - (\mathbf{Af}_2(\mathbf{y}) + \mathbf{b})\|^2 \\ &= \begin{bmatrix} \mathbf{f}_1(\mathbf{x})^T & \mathbf{f}_2(\mathbf{y})^T & 1 \end{bmatrix} \begin{bmatrix} \mathbf{A}^T \mathbf{A} & -\mathbf{A}^T \mathbf{A} & \mathbf{A}^T(\mathbf{a} - \mathbf{b}) \\ -\mathbf{A}^T \mathbf{A} & \mathbf{A}^T \mathbf{A} & \mathbf{A}^T(\mathbf{b} - \mathbf{a}) \\ (\mathbf{a} - \mathbf{b})^T \mathbf{A} & (\mathbf{b} - \mathbf{a})^T \mathbf{A} & \|\mathbf{a} - \mathbf{b}\|^2 \end{bmatrix} \begin{bmatrix} \mathbf{f}_1(\mathbf{x}) \\ \mathbf{f}_2(\mathbf{y}) \\ 1 \end{bmatrix} \end{aligned} \quad (2-7)$$

同样，公式(2-4)也可以写成仿射变换的形式^[13]，如公式(2-8)所示：

$$\begin{aligned} F(\mathbf{f}_1(\mathbf{x}), \mathbf{f}_2(\mathbf{y})) &= \mathbf{f}_1(\mathbf{x})^T \mathbf{M} \mathbf{f}_1(\mathbf{x}) + \mathbf{f}_2(\mathbf{y})^T \mathbf{M} \mathbf{f}_2(\mathbf{y}) + \mathbf{f}_1(\mathbf{x})^T \mathbf{G} \mathbf{f}_2(\mathbf{y}) \\ &\quad + \mathbf{d}^T (\mathbf{f}_1(\mathbf{x}) + \mathbf{f}_2(\mathbf{y})) + f \\ &= \begin{bmatrix} \mathbf{f}_1(\mathbf{x})^T & \mathbf{f}_2(\mathbf{y})^T & 1 \end{bmatrix} \begin{bmatrix} \mathbf{M} & \frac{1}{2} \mathbf{G} & \mathbf{d} \\ \frac{1}{2} \mathbf{G} & \mathbf{M} & \mathbf{d} \\ \mathbf{d}^T & \mathbf{d}^T & f \end{bmatrix} \begin{bmatrix} \mathbf{f}_1(\mathbf{x}) \\ \mathbf{f}_2(\mathbf{y}) \\ 1 \end{bmatrix} \end{aligned} \quad (2-8)$$

2.3.2 广义相似性度量的推导

我们已经在 2.2 节中介绍了相似性度量，在 2.3.1 条中简单介绍了仿射变换。接下来，我们希望在仿射变换的形式下对马氏距离和余弦相似度进行融合，将其推广为较为一般的形式。首先，我们将仿射变换下的马氏距离和余弦相似度写成同样的形式，然后用一组参数将二者组合在一起。

在前面所述马氏距离、余弦相似性度量和仿射变换的基础上，参考公式(2-1) –

(2-8)的推导和展开，我们可以将仿射变化形式下的马氏距离改写成如下形式，如公式(2-9)所示：

$$\begin{aligned} D_M &= \|(\mathbf{L}_A \mathbf{x} + \mathbf{a}) - (\mathbf{L}_B \mathbf{y} + \mathbf{b})\|_2^2 \\ &= [\mathbf{x}^T \quad \mathbf{y}^T \quad 1] \mathbb{S}_M \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \\ 1 \end{bmatrix} \end{aligned} \quad (2-9)$$

其中矩阵 \mathbb{S}_M 可以被进一步展开为公式(2-10)的形式：

$$\mathbb{S}_M = \begin{bmatrix} \mathbf{L}_A^T \mathbf{L}_A & -\mathbf{L}_A^T \mathbf{L}_B & \mathbf{L}_A^T (\mathbf{a} - \mathbf{b}) \\ -\mathbf{L}_B^T \mathbf{L}_A & \mathbf{L}_B^T \mathbf{L}_B & \mathbf{L}_B^T (\mathbf{b} - \mathbf{a}) \\ (\mathbf{a}^T - \mathbf{b}^T) \mathbf{L}_A & (\mathbf{b}^T - \mathbf{a}^T) \mathbf{L}_B & \|\mathbf{a} - \mathbf{b}\|_2^2 \end{bmatrix} \quad (2-10)$$

我们用仿射空间中的内积表示仿射余弦相似性，并将其展开为矩阵乘积的形式，如公式(2-11)所示：

$$\begin{aligned} S_I &= (\hat{\mathbf{L}}_A \mathbf{x} + \hat{\mathbf{a}})^T (\hat{\mathbf{L}}_B \mathbf{y} + \hat{\mathbf{b}}) \\ &= [\mathbf{x}^T \quad \mathbf{y}^T \quad 1] \mathbb{S}_I \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \\ 1 \end{bmatrix} \end{aligned} \quad (2-11)$$

其中，相关矩阵 \mathbb{S}_I 可以表示为公式(2-12)的形式：

$$\mathbb{S}_I = \begin{bmatrix} \mathbf{0} & \frac{\hat{\mathbf{L}}_A^T \hat{\mathbf{L}}_B}{2} & \frac{\hat{\mathbf{L}}_A^T \hat{\mathbf{b}}}{2} \\ \frac{\hat{\mathbf{L}}_B^T \hat{\mathbf{L}}_A}{2} & \mathbf{0} & \frac{\hat{\mathbf{L}}_B^T \hat{\mathbf{a}}}{2} \\ \frac{\hat{\mathbf{b}}^T \hat{\mathbf{L}}_A}{2} & \frac{\hat{\mathbf{a}}^T \hat{\mathbf{L}}_B}{2} & \hat{\mathbf{a}}^T \hat{\mathbf{b}} \end{bmatrix} \quad (2-12)$$

我们通过一组权重将 D_M 和 S_I 合并起来，可以得到 S 如公式(2-13)所示：

$$\begin{aligned} S &= \mu D_M - \lambda S_I \\ &= [\mathbf{x}^T \quad \mathbf{y}^T \quad 1] \mathbb{S} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \\ 1 \end{bmatrix} \end{aligned} \quad (2-13)$$

其中矩阵 \mathbb{S} 通过合并 \mathbb{S}_M 和 \mathbb{S}_I 得到，如公式(2-14)所示：

$$\mathbb{S} = \begin{bmatrix} \mathbf{A} & \mathbf{C} & \mathbf{d} \\ \mathbf{C}^T & \mathbf{B} & \mathbf{e} \\ \mathbf{d}^T & \mathbf{e}^T & f \end{bmatrix} \quad (2-14)$$

其中，**A**、**B**、**C**、**d**、**e** 和 f 分别为：

$$\begin{aligned}\mathbf{A} &= \mu \mathbf{L}_A^T \mathbf{L}_A \\ \mathbf{B} &= \mu \mathbf{L}_B^T \mathbf{L}_B \\ \mathbf{C} &= -\mu \mathbf{L}_A^T \mathbf{L}_B - \lambda \frac{\hat{\mathbf{L}}_A^T \hat{\mathbf{L}}_B}{2} \\ \mathbf{d} &= \mu \mathbf{L}_A^T (\mathbf{a} - \mathbf{b}) - \lambda \frac{\hat{\mathbf{L}}_A^T \hat{\mathbf{b}}}{2} \\ \mathbf{e} &= \mu \mathbf{L}_B^T (\mathbf{b} - \mathbf{a}) - \lambda \frac{\hat{\mathbf{L}}_B^T \hat{\mathbf{a}}}{2} \\ f &= \mu \|\mathbf{a} - \mathbf{b}\|_2^2 - \lambda \hat{\mathbf{a}}^T \hat{\mathbf{b}}\end{aligned}\tag{2-15}$$

至此，广义相似性度量矩阵 \mathbb{S} 的推导已经完成。对于两个不同领域的样本 \mathbf{x} 和 \mathbf{y} ，考虑 \mathbf{x} 和 \mathbf{y} 的顺序（即 **A** 和 **B**、**d** 和 **e** 分别时候是对应样本 \mathbf{x} 和 \mathbf{y} 的参数），它们的相似性函数 $S(\mathbf{x}, \mathbf{y})$ 如公式(2-16)所示：

$$S(\mathbf{x}, \mathbf{y}) = \begin{bmatrix} \mathbf{x}^T & \mathbf{y}^T & 1 \end{bmatrix} \begin{bmatrix} \mathbf{A} & \mathbf{C} & \mathbf{d} \\ \mathbf{C}^T & \mathbf{B} & \mathbf{e} \\ \mathbf{d}^T & \mathbf{e}^T & f \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \\ 1 \end{bmatrix}\tag{2-16}$$

其中，矩阵 **A** 和 **B** 为半正定矩阵，表示样本在各自领域中的自相关，**C** 为两个领域的相关性矩阵。

2.3.3 广义相似性度量与已有相似性模型的联系

广义相似性度量模型是对马氏距离、余弦相似度的融合和一般化，很多已有的模型可以看作是广义相似性度量模型在对参数(**A**, **B**, **C**, **d**, **e**, f)上添加额外约束的特例。

如传统相似性模型通常定义为 $S_M = \mathbf{x}^T \mathbf{M} \mathbf{y}$ ，当 $\mathbf{A} = \mathbf{B} = \mathbf{0}$, $\mathbf{C} = \mathbf{M}/2$, $\mathbf{d} = \mathbf{e} = \mathbf{0}$ 并且 $f = 0$ 时，这个形式和广义相似性度量模型是等价的。同样，马氏距离的表示 $D_M(\mathbf{x}, \mathbf{y}) = (\mathbf{x} - \mathbf{y})^T \mathbf{M} (\mathbf{x} - \mathbf{y})$ 也被认为是我们的模型在 $\mathbf{A} = \mathbf{B} = \mathbf{M}$, $\mathbf{C} = -\mathbf{M}$, $\mathbf{d} = \mathbf{e} = \mathbf{0}$ 并且 $f = 0$ 时的特例。

在引文[13]中，Li 等人提出一种通过联合建模距离度量和局部自适应阈值规则的判别函数，这个叫做 LADF（即 Locally-Adaptive Decision Function，局部自适应判别函数）的函数被推导为二阶大间隔正则化问题。

LADF 被定义为：

$$F(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T \mathbf{A} \mathbf{x} + \mathbf{y}^T \mathbf{A} \mathbf{y} + 2 \mathbf{x}^T \mathbf{C} \mathbf{y} + \mathbf{d}^T (\mathbf{x} + \mathbf{y}) + f.\tag{2-17}$$

当 $\mathbf{B} = \mathbf{A}$, $\mathbf{e} = \mathbf{d}$ 时, 我们可以看到 $F(\mathbf{x}, \mathbf{y}) = S(\mathbf{x}, \mathbf{y})$ 。

我们应当注意到 LADF 对 \mathbf{x} 和 \mathbf{y} 用同样的矩阵, 即 \mathbf{A} 用于 $\mathbf{x}^T \mathbf{A} \mathbf{x}$ 和 $\mathbf{y}^T \mathbf{A} \mathbf{y}$, \mathbf{d} 用于 $\mathbf{d}^T \mathbf{x}$ 和 $\mathbf{d}^T \mathbf{y}$ 。这个模型对同模态样本的匹配是有效的, 但是可能对 \mathbf{x} 和 \mathbf{y} 来自不同模态的跨领域匹配并不合适。相比于 LADF, 广义相似性模型用 \mathbf{A} 和 \mathbf{d} 计算 $\mathbf{x}^T \mathbf{A} \mathbf{x}$ 和 $\mathbf{d}^T \mathbf{x}$, 用 \mathbf{B} 和 \mathbf{e} 计算 $\mathbf{y}^T \mathbf{B} \mathbf{y}$ 和 $\mathbf{e}^T \mathbf{y}$, 使得我们的模型对跨领域匹配更加有效。

在引文[25]中, Chen 等人通过学习一个样本对的联合分布(即类内变化和类间变化)拓展了经典贝叶斯人脸模型。他们的判别函数形式如下:

$$J(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T \mathbf{A} \mathbf{x} + \mathbf{y}^T \mathbf{A} \mathbf{y} - 2\mathbf{x}^T \mathbf{G} \mathbf{y}. \quad (2-18)$$

这个判别函数也是广义相似性模型的特殊形式, 此时 $\mathbf{B} = \mathbf{A}$, $\mathbf{C} = -\mathbf{G}$, $\mathbf{d} = \mathbf{0}$, $\mathbf{e} = \mathbf{0}$ 且 $f = 0$ 。

2.4 本章小结

在本章中, 我们主要介绍了广义相似性度量, 包括广义相似性度量的产生背景、推导过程和一些相关工作。首先是对常见的相似性度量, 包括欧氏距离、马氏距离、余弦相似度等进行了简单介绍, 然后给出了仿射变换相比子线性变换的优势及其在相似性度量中的应用, 这些都是广义相似性度量模型的产生基础。在此基础上, 我们给出了广义相似性度量的推导过程, 同时介绍了广义相似性度量与一些其他模型之间的关系。

在本章的基础上, 我们完成了基于三元组约束的深度特征表达和广义相似性度量的联合端到端学习。

第3章 基于三元组约束的跨领域相似性匹配

3.1 引言

相似性匹配是计算机视觉和模式识别领域的基本问题，目前相似性匹配研究主要是基于同领域（模态）数据。如本文1.1节所述，跨领域相似性匹配在特定场景下具有很高的应用价值，我们的研究就是对跨领域视觉数据进行相似性匹配。

在本章中，我们提出了基于三元组约束的跨领域相似性匹配学习方法，基于深度神经网络完成深度特征表达和广义相似性度量的联合端到端学习。我们首先介绍了三元组的构造以及对于单个三元组的目标函数，然后介绍了深度神经网络的损失函数构造、网络结构以及训练方法，最后给出了我们初步的实验结果。

3.2 三元组约束

在本节中，我们介绍了针对跨领域数据的三元组构造方法、构造情况分析以及针对单个三元组设计的目标函数。

3.2.1 构造三元组

对训练集构造三元组($\mathbf{x}, \mathbf{y}, \mathbf{z}$)，其中 \mathbf{x}, \mathbf{y} 为同类样本， \mathbf{x}, \mathbf{z} 为异类样本， \mathbf{y}, \mathbf{z} 属于同一领域， \mathbf{x} 属于另一领域。我们希望通过两个不同的神经网络学习它们的特征（其中一个网络学习 \mathbf{x} 的特征，另一个网络学习 \mathbf{y}, \mathbf{z} 的特征，这两个网络有一些底层共享参数）。

在同一个 mini-batch 中，假定第一个领域中有 N 个样本，第二个领域中有 M 个样本。对于第一个领域中的样本 δ_h ($\delta_h \in Domain1, h = 1, \dots, N$)，假定其同领域异类样本有 $\{\mathbf{x}_1^\delta, \dots, \mathbf{x}_k^\delta, \dots, \mathbf{x}_n^\delta\}$ ，不同领域同类样本有 $\{\mathbf{y}_1^\delta, \dots, \mathbf{y}_i^\delta, \dots, \mathbf{y}_p^\delta\}$ ，不同领域异类样本有 $\{\mathbf{z}_1^\delta, \dots, \mathbf{z}_j^\delta, \dots, \mathbf{z}_q^\delta\}$ ；对于第二个领域中的样本 σ_g ($\sigma_g \in Domain2, g = 1, \dots, M$)，假定其同领域异类样本有 $\{\mathbf{x}_1^\sigma, \dots, \mathbf{x}_k^\sigma, \dots, \mathbf{x}_n^\sigma\}$ ，不同领域同类样本有 $\{\mathbf{y}_1^\sigma, \dots, \mathbf{y}_i^\sigma, \dots, \mathbf{y}_p^\sigma\}$ ，不同领域异类样本有 $\{\mathbf{z}_1^\sigma, \dots, \mathbf{z}_j^\sigma, \dots, \mathbf{z}_q^\sigma\}$ 。

对样本 δ_h 可构造的三元组有三种，分别是 $(\delta_h, \mathbf{y}_i^\delta, \mathbf{z}_j^\delta)$ ($i = 1, \dots, p; j = 1, \dots, q$)， $(\mathbf{y}_i^\delta, \delta_h, \mathbf{x}_k^\delta)$ ($i = 1, \dots, p; k = 1, \dots, n$) 和 $(\mathbf{z}_j^\delta, \mathbf{x}_k^\delta, \delta_h)$ ($j = 1, \dots, q; k = 1, \dots, n$)，其中第三种要

求 $[type(\mathbf{z}_j^\delta) = type(\mathbf{x}_k^\delta)]$ ，对第二个领域中的样本 σ_g 可以构造的三元组同样有三种情况，分别是 $(\sigma_g, \mathbf{y}_i^\sigma, \mathbf{z}_j^\sigma)(i=1, \dots, p; j=1, \dots, q)$ ， $(\mathbf{y}_i^\sigma, \sigma_g, \mathbf{x}_k^\sigma)(i=1, \dots, p; k=1, \dots, n)$ 和 $(\mathbf{z}_j^\sigma, \mathbf{x}_k^\sigma, \sigma_g)(j=1, \dots, q; k=1, \dots, n)$ ，同样第三种要求 $[type(\mathbf{z}_j^\sigma) = type(\mathbf{x}_k^\sigma)]$ 。对样本 δ_h 构成的三个三元组进行分析，发现 $(\delta_h, \mathbf{y}_i^\delta, \mathbf{z}_j^\delta)$ 所能表示的三元组等价于 $(\mathbf{y}_i^\sigma, \sigma_g, \mathbf{x}_k^\sigma)$ ， $(\mathbf{y}_i^\delta, \delta_h, \mathbf{x}_k^\delta)$ 所能表示的三元组等价于 $(\sigma_g, \mathbf{y}_i^\sigma, \mathbf{z}_j^\sigma)$ ；而 $\sum_{h,j,k} (\mathbf{z}_j^\delta, \mathbf{x}_k^\delta, \delta_h)$ 与 $\sum_{h,i,j} (\mathbf{y}_i^\delta, \delta_h, \mathbf{x}_k^\delta)$ 一致， $\sum_{g,j,k} (\mathbf{z}_j^\sigma, \mathbf{x}_k^\sigma, \sigma_g)$ 则与 $\sum_{g,i,j} (\mathbf{y}_i^\sigma, \sigma_g, \mathbf{x}_k^\sigma)$ 一致。

因此，我们可以对第一个领域中的样本 δ_h 进行遍历，并且构造其中两个三元组 $(\delta_h, \mathbf{y}_i^\delta, \mathbf{z}_j^\delta)(i=1, \dots, p; j=1, \dots, q)$ 和 $(\mathbf{y}_i^\delta, \delta_h, \mathbf{x}_k^\delta)(i=1, \dots, p; k=1, \dots, n)$ ，这样就可以保证覆盖了所有符合条件的三元组。

总结上述的分析和推理，我们可以对跨领域数据构造三元组可以遵循以下步骤：

- 1) 遍历第一个领域，将遍历到的样本记作 \mathbf{x}_1 ，对第二个领域中与 \mathbf{x}_1 同类别的样本进行遍历，将遍历到的样本记作 \mathbf{y}_1 ；
- 2) 对两个领域与 \mathbf{x}_1 和 \mathbf{y}_1 不同类别的样本分别进行遍历，对于遍历到的样本，分别记作 \mathbf{x}_2 和 \mathbf{y}_2 ；
- 3) $(\mathbf{x}_1, \mathbf{y}_1, \mathbf{y}_2)$ 和 $(\mathbf{y}_1, \mathbf{x}_1, \mathbf{x}_2)$ 就是我们构造的样本。

3.2.2 单个三元组的目标函数

对训练集构造的三元组 $(\mathbf{x}, \mathbf{y}, \mathbf{z})$ ，我们希望它们的相似性函数 $S(\mathbf{x}, \mathbf{y})$ 和 $S(\mathbf{x}, \mathbf{z})$ 之间满足样本 \mathbf{x} 到异领域异类样本 \mathbf{z} 的距离和到异领域同类样本 \mathbf{y} 的距离之差足够大，将这个距离最小值设置为1，我们构造三元组 $(\mathbf{x}, \mathbf{y}, \mathbf{z})$ 的目标函数如下：

$$S(\mathbf{x}, \mathbf{z}) - S(\mathbf{x}, \mathbf{y}) > 1 \quad (3-1)$$

因此，对每个三元组，损失函数如下：

$$G(\mathbf{x}, \mathbf{y}, \mathbf{z}) = (1 + S(\mathbf{x}, \mathbf{y}) - S(\mathbf{x}, \mathbf{z}))_+ \quad (3-2)$$

其中， $S(\mathbf{x}, \mathbf{y})$ 参考公式(2-16)，考虑 \mathbf{x} 、 \mathbf{y} 和 \mathbf{z} 可能来自不同的领域，形式如公式(3-3)：

$$S(\mathbf{x}, \mathbf{y}) = \begin{cases} \tilde{S}(\mathbf{x}, \mathbf{y}), & \mathbf{x} \text{属于Domain1, 且} \mathbf{y} \text{属于Domain2} \\ \tilde{S}(\mathbf{y}, \mathbf{x}), & \mathbf{y} \text{属于Domain1, 且} \mathbf{x} \text{属于Domain2} \end{cases} \quad (3-3)$$

$$\tilde{S}(\mathbf{x}, \mathbf{y}) = \begin{bmatrix} \mathbf{x}^T & \mathbf{y}^T & 1 \end{bmatrix} \begin{bmatrix} \mathbf{A} & \mathbf{C} & \mathbf{d} \\ \mathbf{C}^T & \mathbf{B} & \mathbf{e} \\ \mathbf{d}^T & \mathbf{e}^T & f \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \\ 1 \end{bmatrix}$$

3.3 深度神经网络实现

本节主要介绍了本文所实现深度神经网络的损失函数的构造、网络结构的设计以及网络参数的训练方法。该网络进行深度特征表达和广义相似性度量的联合端到端学习，最终实现跨领域数据的相似性匹配。

3.3.1 损失函数的构造

我们在网络的训练中对样本采取分批处理的训练策略，即每轮迭代随机选取训练样本中的一小部分进行网络的训练。由于每轮训练迭代中有效三元组的数目不一样，因此，我们将有效三元组的损失函数求和，然后除以有效三元组的数目作为整个网络的损失函数。如 3.1.1 条对三元组构造情况的分析，我们将有效三元组 $(\delta_h, \mathbf{y}_i^\delta, \mathbf{z}_j^\delta) (i=1, \dots, p; j=1, \dots, q)$ 和 $(\mathbf{y}_i^\delta, \delta_h, \mathbf{x}_k^\delta) (i=1, \dots, p; k=1, \dots, n)$ 的损失函数求均值，得到网络的损失函数如下：

$$\begin{aligned} L &= \sum_{h=1}^N \sum_{i=1}^p \left(\frac{\sum_{j=1}^q (1 + \tilde{S}(\delta_h, \mathbf{y}_i^\delta) - \tilde{S}(\delta_h, \mathbf{z}_j^\delta))_+}{v} + \frac{\sum_{k=1}^n (1 + \tilde{S}(\delta_h, \mathbf{y}_i^\delta) - \tilde{S}(\mathbf{x}_k^\delta, \mathbf{y}_i^\delta))_+}{v} \right) \\ &= \sum_{h=1}^N \sum_{i=1}^p \left(\sum_{j=1}^q (\phi_{hij})_+ + \sum_{k=1}^n (\psi_{hik})_+ \right) / v \end{aligned} \quad (3-4)$$

其中， v 表示有效三元组的个数， $\tilde{S}(\delta_h, \mathbf{y}_i^\delta)$ 、 $\tilde{S}(\delta_h, \mathbf{z}_j^\delta)$ 和 $\tilde{S}(\mathbf{x}_k^\delta, \mathbf{y}_i^\delta)$ 为公式(3-3)中的相似性函数， ϕ_{hij} 和 ψ_{hik} 分别表示 $1 + \tilde{S}(\delta_h, \mathbf{y}_i^\delta) - \tilde{S}(\delta_h, \mathbf{z}_j^\delta)$ 和 $1 + \tilde{S}(\delta_h, \mathbf{y}_i^\delta) - \tilde{S}(\mathbf{x}_k^\delta, \mathbf{y}_i^\delta)$ ，

另外，我们将 $\sum_{j=1}^q (\phi_{hij})_+$ 记为 Φ ，将 $\sum_{k=1}^n (\psi_{hik})_+$ 记为 Ψ ，以便后续表述。对于相似性矩阵的参数 $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{d}, \mathbf{e}, f)$ ，我们在 1.2.1 条中介绍马氏距离时已经介绍了马氏距离度量矩阵 \mathbf{M} 为半正定矩阵的情况，类似地，我们要求 \mathbf{A} 和 \mathbf{B} 为半正定矩阵。

为了便于表示和将相似性矩阵的参数与深度神经网络输出对应，我们令 $\mathbf{A} = \mathbf{L}_A^T \mathbf{L}_A$, $\mathbf{B} = \mathbf{L}_B^T \mathbf{L}_B$, $\mathbf{C} = -\mathbf{L}_C^T \mathbf{L}_C$ ，那么公式(3-3)中的 $\tilde{S}(\mathbf{x}, \mathbf{y})$ 可以改写为：

$$\begin{aligned} \tilde{S}(\mathbf{x}, \mathbf{y}) &= \|\mathbf{L}_A \mathbf{f}_1(\mathbf{x})\|^2 + \|\mathbf{L}_B \mathbf{f}_2(\mathbf{y})\|^2 + 2\mathbf{d}^T \mathbf{f}_1(\mathbf{x}) \\ &\quad - 2(\mathbf{L}_C^T \mathbf{f}_1(\mathbf{x}))^T (\mathbf{L}_C^T \mathbf{f}_2(\mathbf{y})) + 2\mathbf{e}^T \mathbf{f}_2(\mathbf{y}) + f \end{aligned} \quad (3-5)$$

我们用两个全连接层分别表示两个领域的参数，因此两个全连接层参数分别是：

$$W^x = \begin{bmatrix} \mathbf{L}_A \\ \mathbf{d}^T \\ \mathbf{L}_C^x \end{bmatrix}, W^y = \begin{bmatrix} \mathbf{L}_B \\ \mathbf{e}^T \\ \mathbf{L}_C^y \end{bmatrix} \quad (3-6)$$

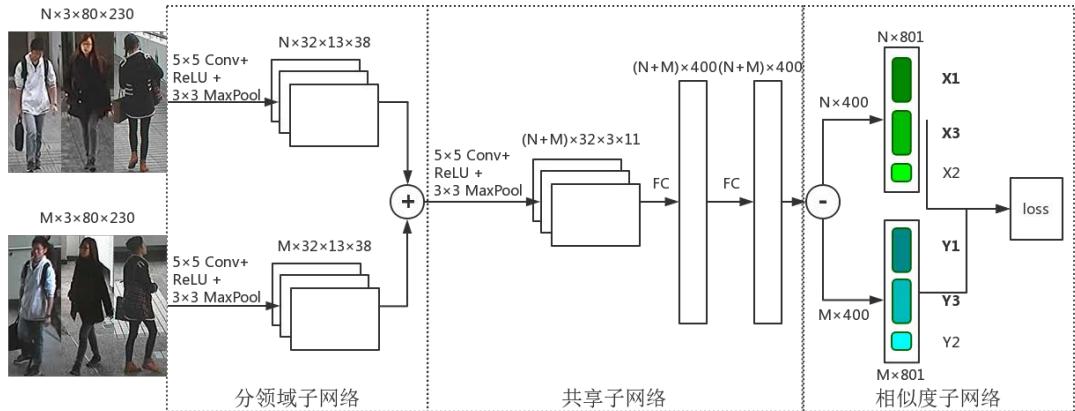
我们令 $\mathbf{x}_1 = \mathbf{L}_A \mathbf{f}_1(\mathbf{x})$, $x_2 = \mathbf{d}^T \mathbf{f}_1(\mathbf{x})$, $\mathbf{x}_3 = \mathbf{L}_C^x \mathbf{f}_1(\mathbf{x})$, $\mathbf{y}_1 = \mathbf{L}_B \mathbf{f}_2(\mathbf{y})$, $y_2 = \mathbf{e}^T \mathbf{f}_2(\mathbf{y})$, $\mathbf{y}_3 = \mathbf{L}_C^y \mathbf{f}_2(\mathbf{y})$ 以方便描述, 因此, 公式(3-5)又可以写成:

$$\tilde{S}(\mathbf{x}, \mathbf{y}) = \|\mathbf{x}_1\|^2 + \|\mathbf{y}_1\|^2 + 2x_2 - 2\mathbf{x}_3^T \mathbf{y}_3 + 2y_2 + f \quad (3-7)$$

那么在公式(3-4)中, 相似性函数 $\tilde{S}(\delta_h, \mathbf{y}_i^\delta)$ 、 $\tilde{S}(\delta_h, \mathbf{z}_j^\delta)$ 和 $\tilde{S}(\mathbf{x}_k^\delta, \mathbf{y}_i^\delta)$ 分别可以写成 $\|\delta_{h,1}\|^2 + \|\mathbf{y}_{i,1}^\delta\|^2 + 2\delta_{h,2} - 2\delta_{h,3}^T \mathbf{y}_{i,3}^\delta + 2y_{i,2}^\delta + f$ 、 $\|\delta_{h,1}\|^2 + \|\mathbf{z}_{j,1}^\delta\|^2 + 2\delta_{h,2} - 2\delta_{h,3}^T \mathbf{z}_{j,3}^\delta + 2z_{j,2}^\delta + f$ 和 $\|\mathbf{x}_{k,1}^\delta\|^2 + \|\mathbf{y}_{i,1}^\delta\|^2 + 2x_{k,2}^\delta - 2\mathbf{x}_{k,3}^T \mathbf{y}_{i,3}^\delta + 2y_{i,2}^\delta + f$ 。经过化简和计算, 我们知道, ϕ_{hij} 可以展开为 $1 + \|\mathbf{y}_{i,1}^\delta\|^2 - \|\mathbf{z}_{j,1}^\delta\|^2 - 2\delta_{h,3}^T \mathbf{y}_{i,3}^\delta + 2\delta_{h,3}^T \mathbf{z}_{j,3}^\delta + 2y_{i,2}^\delta - 2z_{j,2}^\delta$, 而 ψ_{hik} 则可以展开为 $1 + \|\delta_{h,1}\|^2 - \|\mathbf{x}_{k,1}^\delta\|^2 - 2\delta_{h,3}^T \mathbf{y}_{i,3}^\delta + 2\mathbf{x}_{k,3}^T \mathbf{y}_{i,3}^\delta + 2\delta_{h,2} - 2x_{k,2}^\delta$ 。

3.3.2 网络结构

根据上述的推导, 我们确定了深度特征提取部分的网络输出 (即提取的样本特征) 应为 \mathbf{x}_1 、 x_2 、 \mathbf{x}_3 和 \mathbf{y}_1 、 y_2 、 \mathbf{y}_3 。我们采用的网络结构如图 3-1 所示:



注: 左侧图片来自 CUHK-01^[30]行人重识别数据库。该网络结构包含三个部分: 分领域子网络、共享子网络和相似度子网络。前两部分对不同领域的样本进行特征提取, 相似度子网络首先分别输出两个领域样本的样本特征, 然后在 loss 层中进行基于三元组的训练。

图 3-1 深度神经网络结构

如图 3-1, 从左到右, 首先是两个用于不同领域样本的分领域子网络, 用两个分支处理不同领域的样本。接下来, 将两部分合并之后进入共享子网络, 之后再分别送入相似度子网络中, 得到最终的特征向量, 并构造三元组, 计算损失函数。

3.3.3 网络的训练

我们采用 BP 算法对网络进行训练。损失函数如公式(3-4)所示，由 Loss 层的 forward 函数实现。梯度的反向传播由 Loss 层的 backward 函数和其他层的 backward 共同实现，其中 Loss 层实现损失函数对某个参数的梯度计算，其他层共同实现某个参数对整体参数的梯度。

根据链式求导法则，损失函数 L 对网络参数 Ω 的梯度可以写成如下形式：

$$\begin{aligned} \frac{\partial L}{\partial \Omega} &= \sum_{h=1}^N \sum_{i=1}^p \left(\frac{\partial \Phi}{\partial \Omega} + \frac{\partial \Psi}{\partial \Omega} \right) / v \\ &= \sum_{h=1}^N \sum_{i=1}^p \left(\begin{array}{l} \sum_{j=1}^q \hat{\phi}_{hij} \cdot \kappa_\phi \left(\{\delta_{h,1}, \delta_{h,2}, \delta_{h,3}, \mathbf{y}_{i,1}^\delta, y_{i,2}^\delta, \mathbf{y}_{i,3}^\delta, \mathbf{z}_{j,1}^\delta, z_{j,2}^\delta, \mathbf{z}_{j,3}^\delta\} \right) \\ + \sum_{k=1}^n \hat{\psi}_{hik} \cdot \kappa_\psi \left(\{\delta_{h,1}, \delta_{h,2}, \delta_{h,3}, \mathbf{y}_{i,1}^\delta, y_{i,2}^\delta, \mathbf{y}_{i,3}^\delta, \mathbf{x}_{k,1}^\delta, x_{k,2}^\delta, \mathbf{x}_{k,3}^\delta\} \right) \end{array} \right) \quad (3-8) \\ \hat{\phi}_{hij} &= \begin{cases} 1, \phi_{hij} > 0 \\ 0, \text{else} \end{cases}, \hat{\psi}_{hik} = \begin{cases} 1, \psi_{hik} > 0 \\ 0, \text{else} \end{cases}, \kappa_\phi(D) = \sum_{x \in D} \frac{\partial \phi_{hij}}{\partial x} \frac{\partial x}{\partial \Omega}, \kappa_\psi(D) = \sum_{x \in D} \frac{\partial \psi_{hik}}{\partial x} \frac{\partial x}{\partial \Omega} \end{aligned}$$

其中， v 在求导时并入到 $\frac{\partial \phi_{hij}}{\partial \delta_{h,1}}, \dots, \frac{\partial \psi_{hik}}{\partial \mathbf{x}_{k,3}^\delta}$ 中， L 为损失函数， $\frac{\partial \delta_{h,1}}{\partial \Omega}, \dots, \frac{\partial \mathbf{z}_{j,3}^\delta}{\partial \Omega}$ 可以通过全连接层、卷积层的 backward 函数求出，其余需要通过相似性子网络 Loss 层的 backward 函数实现。计算 loss 的时候，直接根据上述公式，对样本构造的三元组，根据 ϕ_{hij} 和 ψ_{hik} 的值确定是否计算偏导和更新误差，它们的表达式分别为：

$$\begin{aligned} \frac{\partial \phi_{hij}}{\partial \delta_{h,1}} &= 0, \quad \frac{\partial \phi_{hij}}{\partial \delta_{h,2}} = 0, \quad \frac{\partial \phi_{hij}}{\partial \delta_{h,3}} = \frac{2(\mathbf{z}_{j,3}^\delta - \mathbf{y}_{i,3}^\delta)}{v} \\ \frac{\partial \phi_{hij}}{\partial \mathbf{y}_{i,1}^\delta} &= \frac{2\mathbf{y}_{i,1}^\delta}{v}, \quad \frac{\partial \phi_{hij}}{\partial y_{i,2}^\delta} = \frac{2}{v}, \quad \frac{\partial \phi_{hij}}{\partial \mathbf{y}_{i,3}^\delta} = -\frac{2\delta_{h,3}}{v} \\ \frac{\partial \phi_{hij}}{\partial \mathbf{z}_{j,1}^\delta} &= -\frac{2\mathbf{z}_{j,1}^\delta}{v}, \quad \frac{\partial \phi_{hij}}{\partial z_{j,2}^\delta} = -\frac{2}{v}, \quad \frac{\partial \phi_{hij}}{\partial \mathbf{z}_{j,3}^\delta} = \frac{2\delta_{h,3}}{v} \\ \frac{\partial \psi_{hik}}{\partial \delta_{h,1}} &= \frac{2\delta_{h,1}}{v}, \quad \frac{\partial \psi_{hik}}{\partial \delta_{h,2}} = \frac{2}{v}, \quad \frac{\partial \psi_{hik}}{\partial \delta_{h,3}} = -\frac{2\mathbf{y}_{i,3}^\delta}{v} \\ \frac{\partial \psi_{hik}}{\partial \mathbf{y}_{i,1}^\delta} &= 0, \quad \frac{\partial \psi_{hik}}{\partial y_{i,2}^\delta} = 0, \quad \frac{\partial \psi_{hik}}{\partial \mathbf{y}_{i,3}^\delta} = \frac{2(\mathbf{x}_{k,3}^\delta - \delta_{h,3})}{v} \\ \frac{\partial \psi_{hik}}{\partial \mathbf{x}_{k,1}^\delta} &= -\frac{2\mathbf{x}_{k,1}^\delta}{v}, \quad \frac{\partial \psi_{hik}}{\partial x_{k,2}^\delta} = -\frac{2}{v}, \quad \frac{\partial \psi_{hik}}{\partial \mathbf{x}_{k,3}^\delta} = \frac{2\mathbf{y}_{i,3}^\delta}{v} \quad (3-9) \end{aligned}$$

3.4 实验结果

我们在 CUHK-03^[23]行人重识别数据库上进行跨领域相似性匹配实验。首先，我们将批大小设置为 128，即每次迭代在每个领域读入 128 张图，进行 15 万次迭代的训练。每次迭代的训练中，首先在前向过程中实现深度特征提取和三元组构建、损失函数计算等任务，并在后向过程中完成梯度反向传播。我们进行了多组实验，并进行了多种初始化方法和训练参数，最终我们选取了三组具有代表性的实验，分别在初始学习率为 0.0001、0.0005 和 0.001 时进行训练，损失函数图像和准确率的累计匹配特性（CMC，Cumulative Matching Characteristic）曲线如下：

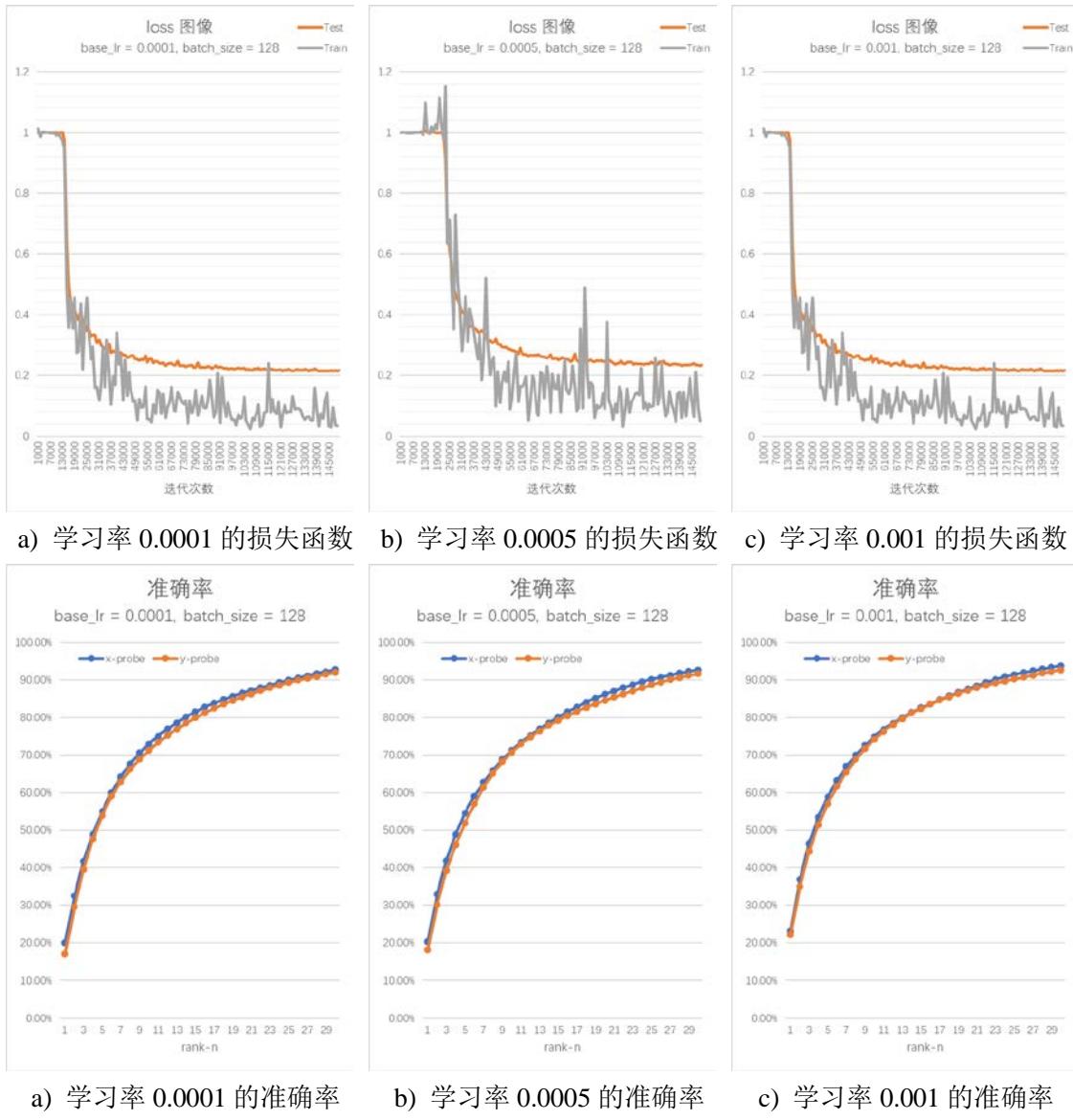


图 3-2 每次迭代批大小为 128 时的典型实验结果

从图 3-2 中可以看出，学习率设置为 0.001 时，训练集上损失函数的下降不如另外两组，但是在精确率上相较于另外两组有一些提升，无论是哪一个领域作为查询集，哪一个领域作为图片集，rank1 准确率都达到了 22% 左右，原因可能是学习率较高时，更容易突破极小值点，从而更好达到全局最优。我们分析认为，在批大小设置为 128 时，实验结果已经达到一个当前限制下较高的水平，因此，我们紧接着进行了每次实验每个领域读取 256 张图的实验，典型的实验结果如下图所示：

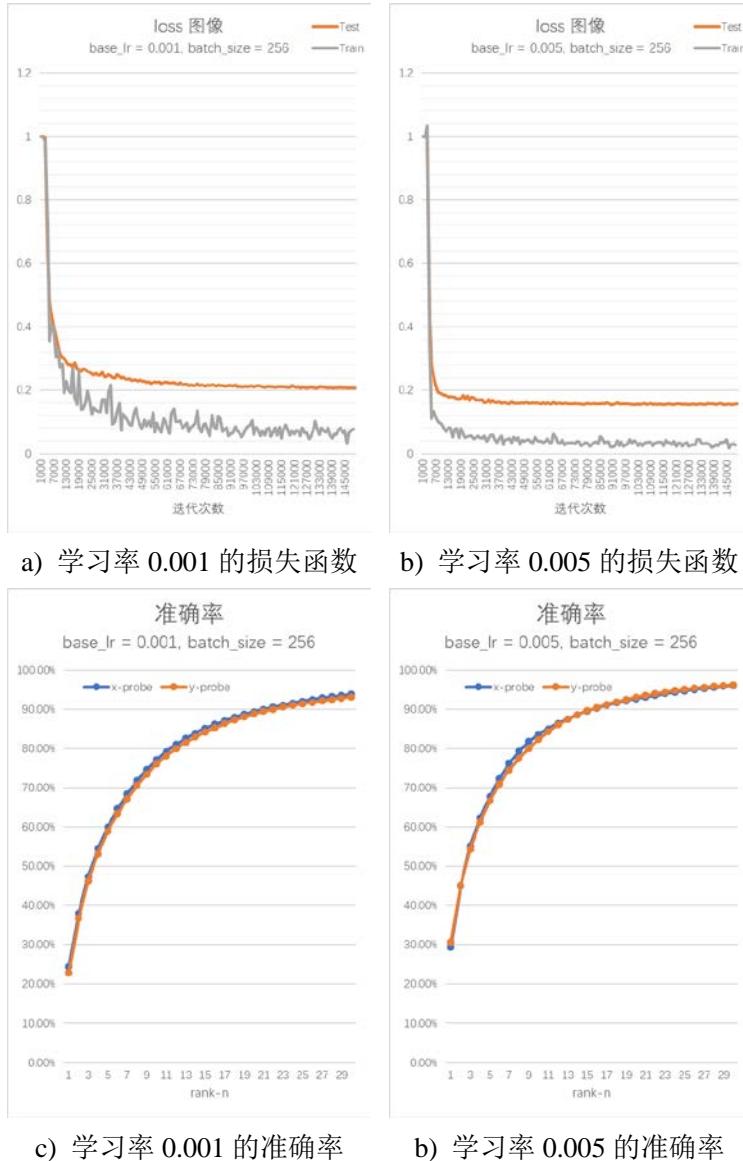


图 3-3 每次迭代批大小为 256 时的典型实验结果

从图 3-3 中可以看出，学习率设置为 0.005 时，rank1 准确率已经达到了 30% 左右，损失函数的稳定性也有了较大提升，整体波动更小。经过分析，我们认为

当每次迭代样本数提升时，该批数据与样本整体的分布更加接近，并且在抵消掉一些偏差较大的变化后，损失函数在训练过程中的稳定性有了较大的提升，并且匹配的准确率有了较大的提高。

3.5 本章小结

本章对三元组的构造情况进行了详细论述，在此基础上，提出了针对单个三元组的目标函数和损失函数。之后，我们将改写后的相似性函数与深度神经网络结构相对应，并完成了网络结构的设计和训练算法的推导。

在理论推导的基础上，我们基于深度学习框架 Caffe 实现了设计的网络结构，在实现 CPU 运行代码的基础上，学习了 CUDA 编程知识，完成了 GPU 版本代码的实现，并最终给出了实验结果。我们对本实验的拓展和进一步探究、以及与其他方法的结果对比在第四章中详细介绍。

第4章 实验拓展及相关分析

4.1 引言

在进行项目研究和实验的过程中，对实验过程进行反思、分析和改进是非常重要的。因此，我们对实验进行了一系列调整和改进，并在本章中介绍。

在本章中，我们将介绍在第三章的基础性实验之外，对跨领域相似性匹配的一些思考、基于这些思考对实验做出的调整以及得到的实验结果，并讨论这些调整后得到的结果的原因以及给我们的启发。

4.2 对正负样本距离间隔的调整

如公式(3-1)所述，我们将三元组($\mathbf{x}, \mathbf{y}, \mathbf{z}$)的相似性函数 $S(\mathbf{x}, \mathbf{y})$ 和 $S(\mathbf{x}, \mathbf{z})$ 之间满足的样本 \mathbf{x} 到异领域异类样本 \mathbf{z} 的距离和到异领域同类样本 \mathbf{y} 的距离之差设置为 1，即我们将正负样本距离间隔设置为 1。我们希望将这个间隔扩大，以观察实验结果。经过实验，我们选出其中两组实验结果，即正负样本距离间隔设置为 1 和设置为 10 的时候前 10 个候选的准确率的对比，如表 4-1 所示：

表 4-1 实验结果对比（正负样本距离间隔设置为 1 和 10）

Rank	1	2	3	4	5	6	7	8	9	10
间隔 1 (%)	30.72	45.03	54.69	61.33	66.54	70.83	74.29	77.33	79.90	82.19
间隔 10 (%)	30.90	46.04	55.30	62.08	67.25	71.45	74.69	77.26	79.29	81.05

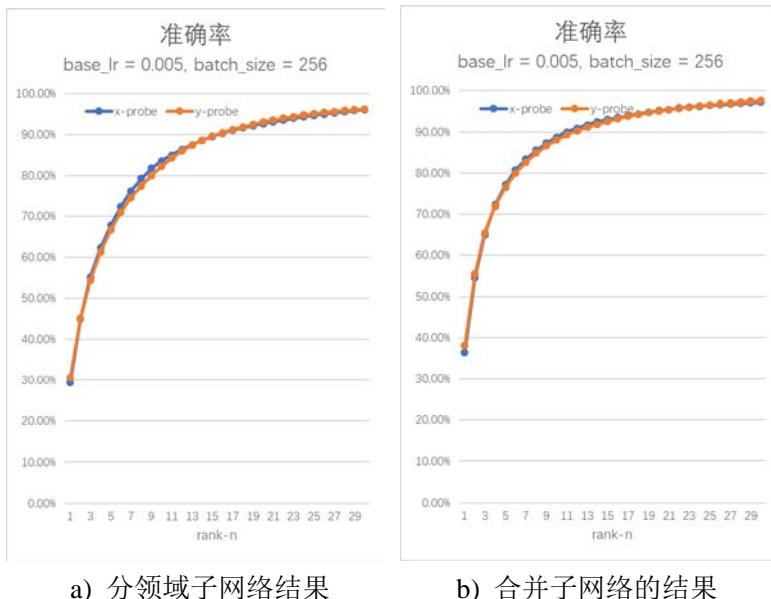
我们可以发现，提升正负样本距离间隔，对提高整体的匹配性能帮助不大，对比在 rank30 的准确率，间隔为 1 达到了 96.01%，而间隔为 10 达到了 96.17%。对结果进行分析，我们发现，广义相似性度量矩阵有马氏距离的特性，在数量级方面做出的改变，并不能对匹配的性能有很好的效果。类似于量纲的变化，正负样本距离间隔的变化被马氏距离所消除。

这对我们的启发是，在对实验方法进行改进和调整时，一定要充分考虑到所使用方法的特性，在合理推断、正确推导的基础上进行改进和进一步实验，不能没有根据地盲目尝试。对样本间隔进行调整虽然有一定道理，但是在考虑到马氏距离和广义相似性度量的特性时，这种做法就失去了意义。

4.3 对网络结构的调整

如图 3-1 所示，我们首先对不同领域的样本在分领域子网络中进行初步特征提取，即针对不同的领域，提取针对性的特征。我们在对实验数据进行分析之后发现，同一个类别不同领域的样本在提取特征方面往往有着同样的要求。比如不同领域的行人，衣服的位置大致相同，跨领域数据的颜色和形状等信息可以用同样的参数提取。

基于以上分析，我们将分领域子网络部分不同领域的子网络合并，然后进行实验。我们将网络修改后，进行了多次实验，在初始学习率为 0.005、批大小为 256 时得到了如下结果：



a) 分领域子网络结果 b) 合并子网络的结果

图 4-1 合并分领域子网络的实验结果对比

可以看到，我们的猜想是合理的，并且在实验结果上得到了体现，rank1 的准确率在第一个领域和第二个领域作为查询集时分别达到了 36.46% 和 38.11%。超出之前的结果 8 个点。

4.4 改变样本选取方式

如 3.3.1 条所述，每轮迭代随机选取训练样本中的一小部分进行网络的训练。我们之前是在样本集中随机抽一定量的样本，经过分析，这样会造成样本的分布比较散乱，有效三元组的个数会特别少。因此我们将样本的选取方式更改为先选择类别，再将被选择的类别中一定量（由于样本数的问题，小于或等于设定的值）的图片送入神经网络，进行三元组的构造。经过这样的调整，我们在 rank1 时的准

确率分别达到 40.51% 和 43.99%。此外，之前的结果是在训练 15 万次迭代后达到的，调整之后，2.5 万次左右，即可达到目前的结果。究其原因，是这样的样本选择方式保证了样本分布较为集中，并且有效三元组的数量得以保证，因此训练效率大大提高，性能也有较大的提升。

我们利用 CUHK-03 数据集官方发布的绘图代码绘制 CMC 曲线。我们与其他几种跨领域相似性匹配方法进行了对比，可以看出，我们的方法在跨领域相似性匹配方面具有较大的优势。其中，LMNN^[5]、ITML^[6]、LDM^[31]、RANK^[32]方法均为基于固定特征表示的距离度量学习方法，SDALF^[33]、KISSME^[34]和 eSDC^[35]均为专为行人重识别问题设计的算法，DFPNN^[23]通过深度网络学习滤波器处理未对齐和几何、光学形变。

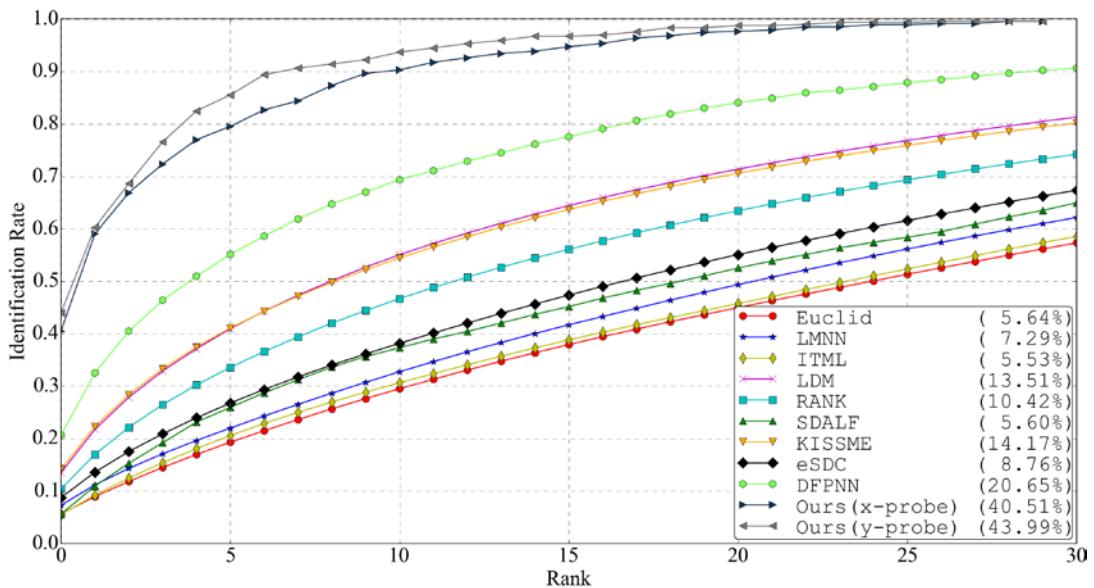


图 4-2 CUHK-03 数据集上的实验结果对比

4.4 本章小结

在本章中，我们进行了一些拓展性试验，对网络的性质进行了探究和分析，在实验过程中学到了很多知识，受到了很多启发，并且在实验结果上有了不错的提升，与图 4-2 中所示其他方法相比，我们的方法在性能上有很大的优势。

结 论

本文在卷积神经网络框架下对跨领域数据相似性匹配进行研究，基于三元组约束推导深度特征表达和广义相似性度量的联合端到端学习，并在深度学习框架 Caffe 的基础上完成验证性实验。

实验结果表明，广义相似性度量对马氏距离、余弦相似度和仿射变换的整合，加上针对跨领域数据进行的调整（如针对不同领域的样本，设置不同的匹配参数），使得匹配性能大大提高。并且三元组的应用，打破了二元组同类样本相似性为正、异类样本相似性为负的限制，只要求参考样本到同类样本的距离与到异类样本的距离之差大于一个特定值（即 4.2 节所指的正负样本距离间隔），放宽了限制，提高了训练的效率和算法的性能。

本课题后续的工作主要有：

(1) 针对实验中遇到的问题，进行进一步的学习和探究。在此基础上，对更多跨领域数据形式应用本文工作，并针对性提高算法性能和执行效率，以期让算法具有更广泛的适用性；

(2) 将本课题取得的成果应用到生成对抗网络（GAN）等的研究中，尝试用 GAN 网络生成不同模态的数据，如根据照片生成素描像或者动漫头像，或者根据侧身照片生成正面照片等。

参考文献

- [1] Dewan M A A, Granger E, Marcialis G L, et al. Adaptive appearance model tracking for still-to-video face recognition[J]. Pattern Recognition, 2016, 49(C):129-151.
- [2] Amal Seralkhatem Osman Ali, Vijanth Sagayan Asirvadam, Aamir Saeed Malik, et al. Age-Invariant Face Recognition Using Triangle Geometric Features[J]. International Journal of Pattern Recognition & Artificial Intelligence, 2015, 29(05):150407184814000.
- [3] Roy H, Bhattacharjee D. Face Sketch-Photo Matching Using the Local Gradient Fuzzy Pattern[J]. Intelligent Systems IEEE, 2016, 31(3):30-39.
- [4] Jia Y, Shelhamer E, Donahue J, et al. Caffe: Convolutional architecture for fast feature embedding[C]//Proceedings of the 22nd ACM international conference on Multimedia. ACM, 2014: 675-678.
- [5] Weinberger K Q, Blitzer J, Saul L K. Distance Metric Learning for Large Margin Nearest Neighbor Classification[C]//NIPS. 2005:207-244.
- [6] Davis J V, Kulis B, Jain P, et al. Information-theoretic metric learning[C]//International Conference on Machine Learning. ACM, 2007:209-216.
- [7] Hardoon D, Szedmak S, Shawe-Taylor J. Canonical Correlation Analysis: An Overview with Application to Learning Methods[J]. Neural Computation, 2004, 16(12):2639.
- [8] Costa P J, Coviello E, Doyle G, et al. On the Role of Correlation and Abstraction in Cross-Modal Multimedia Retrieval[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2014, 36(3):521-535.
- [9] Sharma A, Jacobs D W. Bypassing synthesis: PLS for face recognition with pose, low-resolution and sketch[C]//IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2011:593-600.
- [10] Duan L, Xu D, Tsang I W. Domain adaptation from multiple sources: a

- domain-dependent regularization approach[J]. IEEE Transactions on Neural Networks & Learning Systems, 2012, 23(3):504.
- [11] Zhuang Y, Wang Y, Wu F, et al. Supervised coupled dictionary learning with group structures for multi-modal retrieval[C]//AAAI Conference on Artificial Intelligence. 2013.
- [12] Kan M, Shan S, Zhang H, et al. Multi-view Discriminant Analysis[M]. Computer Vision – ECCV 2012. Springer International Publishing, 2012:808-821.
- [13] Li Z, Chang S, Liang F, et al. Learning Locally-Adaptive Decision Functions for Person Verification[C]//IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2013:3610-3617.
- [14] Kang C, Liao S, He Y, et al. Cross-Modal Similarity Learning: A Low Rank Bilinear Formulation[C]//ACM International Conference on Information and Knowledge Management. ACM, 2015:1251-1260.
- [15] Lecun Y, Boser B, Denker J S, et al. Backpropagation applied to handwritten zip code recognition[J]. Neural Computation, 1989, 1(4):541-551.
- [16] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[C]//International Conference on Neural Information Processing Systems, 2012:1097-1105.
- [17] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2016, 79(10):1337-1342.
- [18] Sun Y, Chen Y, Wang X, et al. Deep learning face representation by joint identification-verification[C]//International Conference on Neural Information Processing Systems, 2014:1988-1996.
- [19] Xiong F, Gou M, Camps O, et al. Person Re-Identification Using Kernel-Based Metric Learning Methods[M]//Computer Vision – ECCV 2014. Springer International Publishing, 2014:1-16.
- [20] Xiao T, Li H, Ouyang W, et al. Learning Deep Feature Representations with

- Domain Guided Dropout for Person Re-identification[C]//IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2016:1249-1258
- [21] Chen S, Guo C, Lai J. Deep Ranking for Person Re-Identification via Joint Representation Learning[J]. IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society, 2016, 25(5):2353.
- [22] Andrew G, Arora R, Bilmes J, et al. Deep canonical correlation analysis[C]// International Conference on International Conference on Machine Learning. JMLR.org, 2013:III-1247.
- [23] Li W, Zhao R, Xiao T, et al. DeepReID: Deep Filter Pairing Neural Network for Person Re-identification[C]//IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2014:152-159.
- [24] Lin L, Wang G, Zuo W, et al. Cross-Domain Visual Matching via Generalized Similarity Measure and Feature Learning[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2016(Accepted).
- [25] Chen D, Cao X, Wang L, et al. Bayesian Face Revisited: A Joint Formulation[C]//European Conference on Computer Vision. 2012:566-579.
- [26] Schroff F, Kalenichenko D, Philbin J. FaceNet: A unified embedding for face recognition and clustering[C]//IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2015:815-823.
- [27] Huang G, Mattar M, Berg T, et al. Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments[J]. Month, 2008.
- [28] Loveloulou, Mahalanobis Distance (马氏距离)[EB/OL], (2011-04-21) [2017-06-24]. <http://blog.csdn.net/loveloulou/article/details/6339976>
- [29] Yeung D Y, Chang H. Locally Smooth Metric Learning with Application to Image Retrieval[C]//IEEE, International Conference on Computer Vision. IEEE, 2008:1-7.
- [30] Li W, Zhao R, Wang X. Human reidentification with transferred metric learning[C]//Asian Conference on Computer Vision. Springer-Verlag,

2012:31-44.

- [31] Guillaumin M, Verbeek J, Schmid C. Is that you? Metric learning approaches for face identification[C]//Computer Vision, 2009 IEEE 12th international conference on. IEEE, 2009: 498-505.
- [32] McFee B, Lanckriet G R. Metric learning to rank[C]//Proceedings of the 27th International Conference on Machine Learning (ICML-10). 2010: 775-782.
- [33] Farenzena M, Bazzani L, Perina A, et al. Person re-identification by symmetry-driven accumulation of local features[C]//IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2010: 2360-2367.
- [34] Koestinger M, Hirzer M, Wohlhart P, et al. Large scale metric learning from equivalence constraints[C]//IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2012: 2288-2295.
- [35] Zhao R, Ouyang W, Wang X. Unsupervised salience learning for person re-identification[C]//IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2013: 3586-3593.

哈尔滨工业大学本科毕业设计（论文）原创性声明

本人郑重声明：在哈尔滨工业大学攻读学士学位期间，所提交的毕业设计（论文）《基于三元组约束的深度距离度量学习》，是本人在导师指导下独立进行研究工作所取得的成果。对本文的研究工作做出重要贡献的个人和集体，均已在文中以明确方式注明，其它未注明部分不包含他人已发表或撰写过的研究成果，不存在购买、由他人代写、剽窃和伪造数据等作假行为。

本人愿为此声明承担法律责任。

作者签名：

日期： 年 月 日

致 谢

时光飞逝如白驹过隙，转眼间本科四年的时光已经接近尾声，站在了一个新的起跑线上。感谢哈工大的培养，感谢“规格严格，功夫到家”校训的指引，让我找到自己前进的方向。

衷心感谢导师左旺孟教授对本人的精心指导。左老师在计算机视觉和机器学习领域有着独到的见解和深度的研究，对我的课题给予了无微不至的关怀。感谢左老师的 support 与理解，让我得以协调好兼职辅导员的工作与本课题的研究和实验。每一封邮件、每一次谈话，都让我感受到左老师的可敬、可亲、可爱，让我感受到左老师对学术前沿的把握和对学生的殷殷期望，他的言传身教将使我终生受益。

感谢实验室的张宏志老师。张老师在学术和工程上都有深入的见解，刚开始本课题研究时张老师对我的帮助和关心，帮我渡过了一开始最困难的那段时光。张老师在实验室组会上对学术的严格要求，告诉我怎样才是一名合格的科研工作者，让我终身难忘，受益无穷。

感谢王法强师兄，在繁忙的工作中抽出时间为我答疑解惑，帮助我快速入门。法强师兄在我倦怠时对我的鼓励，在我困惑时给我的帮助，在科研上树立的榜样，将是我最宝贵的记忆。

同时，还要感谢实验室的所有老师和师兄师姐，他们让我觉得身边有更多可以仰望和学习的榜样，在实验室的日子，温暖而美好。感谢一起做毕业设计的同学，一起奋斗的是我们永恒的青春。

还要感谢我的父母，在远方求学，他们的支持和理解，是我永远最坚实的后盾、最温暖的港湾。

感谢所有给予我帮助和批评的人，因为你们，生命充满奋斗的力量！