

基于 Siamese 网络的行人重识别方法^①



叶 锋^{1,2}, 刘天璐¹, 李诗颖¹, 华笃伟¹, 陈星宇¹, 林文忠³

¹(福建师范大学 数学与信息学院, 福州 350117)

²(数字福建大数据安全技术研究所, 福州 350117)

³(闽江学院 福建省信息处理与智能控制重点实验室, 福州 350108)

通讯作者: 叶 锋, E-mail: yefeng@fjnu.edu.cn

摘 要: 针对目前行人重识别技术的缺点, 提出一种基于 Siamese 网络的行人重识别方法. 首先使用 Dropout 算法对卷积神经网络进行改良, 降低发生过拟合问题的概率; 而后构造一个 Siamese 网络, 将 CNN (Convolution Neural Network) 中特征提取和检验相融合, 提高图像识别的效率和准确率; 最后利用度量学习算法中的马氏距离作为检索图像匹配相似度的评价指标. 实验结果表明: 针对 Market-1501 数据集, 该方法可以有效提高采用卷积神经网络的行人重识别方法识别效率和准确率.

关键词: 行人重识别; 卷积神经网络; Siamese 网络; Dropout

引用格式: 叶锋, 刘天璐, 李诗颖, 华笃伟, 陈星宇, 林文忠. 基于 Siamese 网络的行人重识别方法. 计算机系统应用, 2020, 29(4): 209-213. <http://www.c-s-a.org.cn/1003-3254/7333.html>

Person Re-Identification Method Based on Siamese Network

YE Feng^{1,2}, LIU Tian-Lu¹, LI Shi-Ying¹, HUA Du-Wei¹, CHEN Xing-Yu¹, LIN Wen-Zhong³

¹(College of Mathematics and Informatics, Fujian Normal University, Fuzhou 350117, China)

²(Digital Fujian Big Data Security Technology Institute, Fuzhou 350117, China)

³(Fujian Provincial Key Laboratory of Information Processing and Intelligent Control, Minjiang University, Fuzhou 350108, China)

Abstract: Aiming at the shortcomings of the current pedestrian re-identification technology, this paper presents a pedestrian re-identification method based on Siamese network. First, Dropout algorithm is used to improve the performance of Convolutional Neural Network (CNN), which can reduce the incidence of the fitting problem. By integration of classification and inspection in the CNN, Siamese network is constructed to improve the efficiency and accuracy of image recognition. Finally, Markov distance for metric learning algorithm is used as the evaluation index of image matching similarity. Experiments are conducted on the Market-1501, and the experimental results show that this method is effective in terms of improving the efficiency and accuracy of pedestrian re-identification algorithm.

Key words: person re-identification; Convolution Neural Network (CNN); Siamese network; Dropout

行人重识别 (Person Re-Identification, Re-Id) 主要是指从视频监控系统中找到特定的行人目标, 根据行人诸如衣着、姿势、发型、外在装饰等特征判定是否

为同一人. 自从 2006 年 CVPR (IEEE Conference on Computer Vision and Pattern Recognition) 会议首次提出行人重识别^[1]的概念后, 越来越多研究者投入到该领

① 基金项目: 福建省自然科学基金面上项目 (2017J01739, 2018J01779); 闽江学院福建省信息处理与智能控制重点实验室开放基金 (MJUKF-IPIC201810); 福州市科技重大项目 (榕科 (2017)325 号)

Foundation item: General Program of Natural Science Foundation of Fujian Province, China (2017J01739, 2018J01779); Open Fund of Fujian Provincial Key Laboratory of Information Processing and Intelligent Control, Minjiang University, (MJUKF-IPIC201810); Science and Technology Major Project of Fuzhou Municipality ((2017)325)

收稿时间: 2019-08-22; 修改时间: 2019-09-09; 采用时间: 2019-09-18; csa 在线出版时间: 2020-04-05

域的研究^[2]. 行人重识别在实际场景中判断时, 存在外在环境、行人姿态、衣着等多方面特征的多变性影响, 从而导致在使用视频图像中提取的特征进行相似度的判定难度很大, 所以行人重识别仍是充满挑战和亟待研究的热点课题.

行人重识别主要分为两大部分: 特征提取和度量学习^[3]. 一是对于行人视频中的帧图像进行行人检测和特征提取; 二是利用提取的特征进行度量学习, 对目标与行人之间的相似性进行度量. 卷积神经网络是深度学习在计算机视觉领域最具代表性的模型之一, 可以直接输入原图像提取特征的特点, 从而可以避免复杂的预处理过程, 进而提高识别效率. 结合深度学习, 诸多学者特征提取和度量学习方面提出了许多新型的行人再辨识方法. 如 Liao 等人^[4]将 HSV 颜色直方图和 SILTP (Scale Invariant Local Ternary Patter) 纹理特征描述子联合起来作为度量学习的特征描述子, 其经过最大池化后更加稳定, 此外他们还提出跨视角二次判别 (Cross-view Quadratic Discriminant Analysis, XQDA) 一种新的度量学习方法; Hinton 等人^[5]提出在 CNN 中应用 Dropout 算法^[6], 可以有效降低过拟合现象发生概率. 同时, Siamese 模型^[7]能有效将特征提取和度量学习融合, 统一进行训练, 以显著提高识别效率. Varior 等^[8,9]在改进型循环神经网络 (Recurrent Neural Networks, RNN) 长短期记忆网络 (Long Short-Term Memory, LSTM)^[10]基础上, 利用其可对视频序列进行处理且具有记忆特征信息的特性, 将其和孪生网络结合进行识别, 可以显著提高重识别效率和检测准确率. 上述方法均在不同程度上对 CNN 作了改进, 比如将特征提取和度量学习相结合, 或者将 Siamese 和 LSTM 相结合, 但是仍然存在一些问题. 同时, 由于各自所采用的数据集样本数量不一, 训练出来的模型也大大不同, 使得识别的准确率有所差异、泛化能力不强.

结合 Siamese 网络, 本文提出了一种改进的行人重识别方法. 首先采用 Dropout 算法改进 CNN 神经网络的全连接层, 以降低过拟合发生概率. 之后利用两个权值共享的 CNN 神经网络的 Siamese 网络, 对输入的两张图片作特征提取. 对提取的图像特征使用马氏距离作为相似度匹配的度量指标, 并输出相应图片对的相似度, 最后在 Market-1501 测试集上进行测试, 实验结果表明, 提出的算法在一定程度上提高了行人重识别的性能.

1 改进的 CNN 方法

1.1 提出的算法总体框架

基于上述分析, 提出一种基于 Siamese 网络的 Re-Id 算法. 其总体框架如下: 首先在 CNN 神经网络的全连接层加入 Dropout 算法, 对神经元的激活值设置停止工作的概率, 改变原来每个神经元顺次连接的结构, 达到连接随机稀疏化的效果, 降低网络的过拟合现象发生率. 再以孪生网络结构为基础, 构建由两个完全相同且参数共享的子网络组成的倒 V 型结构 (包含两个输入和一个输出) 分别用来提取两个行人图片特征并输出相似度度量结果, 此外该并行网络可以同步对输入图片提取特征. 最后, 通过求取马氏距离的算法, 计算输入图像对之间的匹配相似度.

1.2 基于 Dropout 算法改进的 CNN

1.2.1 Dropout 算法

过拟合是在神经网络模型训练学习过程中一个亟待解决的问题. 在 CNN 神经网络中运用 Dropout 算法则可以有效降低模型过拟合发生概率. 具体体现在训练过程中加入 Dropout 平均稀疏模型间的组合^[11], 忽略一般的隐层节点, 从而减少神经网络节点间的相互作用和联结匹配, 减缓过拟合, 可以使行人重识别的准确性得到了较大的提升.

Dropout 就是让一些神经元以一定概率停止工作, 从而增强神经网络模型的鲁棒性, 降低网络过拟合现象发生率. 图 1(a) 为全连接时的神经网络模型, 而图 1(b) 为加入 Dropout 后的网络模型示意图. 可以看出全连接情况下, 神经元间进行无间断的连接, 而加入 Dropout 后神经元间变成间断且稀疏的连接结构.

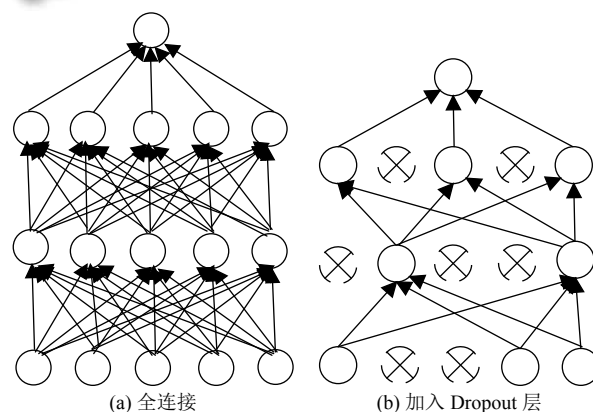


图 1 加入 Dropout 层前后神经网络模型

1.2.2 基于 Dropout 算法的 CNN 模型改进

利用 Dropout 改进 CNN 的流程如图 2 所示. 首先

利用快速排序算法 $Sort(a, b, greater)$ 对各个神经节点输入数组 B 进行降序排序, 排序后记为 a_i ; 然后根据排序由大到小, 利用伯努利概率分布函数, 对节点有条件的分配不同隐藏率(为了简化结构, 加快识别速度, 在图2中, 将所有节点分成了3个部分, 按分配顺序设置不同的隐藏率). 式(1)如下, 其中 n 为神经元数量, $ratio$ 为初始设置的稀疏值.

$$Bernoulli(p) = \begin{cases} P = \min\{1 - ratio + \delta, 1\} & i > n/3 \\ P = 1 - ratio & n/3 \leq i < (2n)/3 \\ P = \max\{1 - ratio - \delta, 0\} & (2n)/3 \leq i < n \end{cases} \quad (1)$$

根据公式(1)生成的结果, 再利用公式 $B_i * Bernoulli(P)_i (B_i = a_i)$ 获得神经节点值. 最后结合输入的稀疏值, 对节点的连接与否进行控制, 从而可以有效控制神经元的发生率, 即随机舍弃某些神经元.

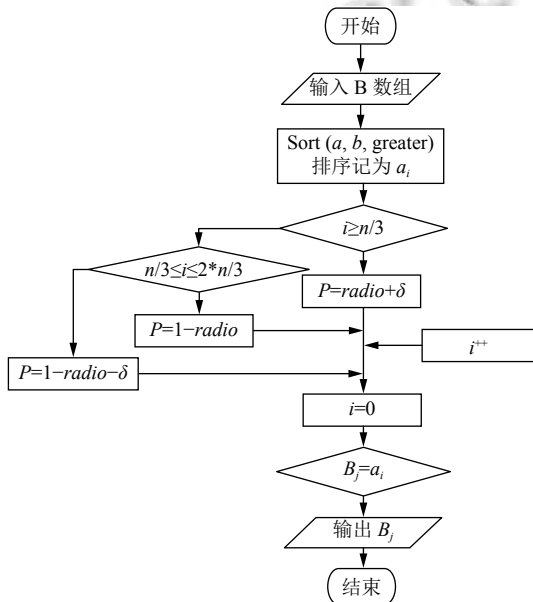


图2 基于 Dropout 算法改进的 CNN 流程

1.3 基于 Siamese 网络的 CNN 结构改进

Siamese Network 也称孪生网络, 是一种相似性度量方法. 它从输入行人图像中提取特征, 再利用特征向量计算相似度.

如图3所示, Siamese 网络的“连体”是通过权值共享来完成的. 孪生神经网络确定两个输入 Input1 和 Input2, 将两个输入分别送入两个神经网络 Network1 和 Network2, 再将对应的输入映射到新的空间, 形成输入行人图像在新的空间中的表示. 最后, 通过在目标空间应用距离度量算法进行相似度损失 (Loss) 的计算, 比较两个输入的相似度.

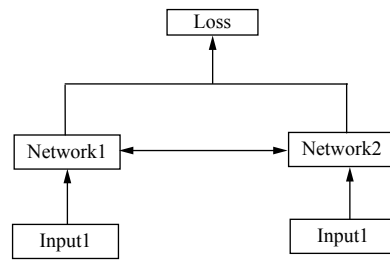


图3 Siamese 网络

本算法具体实现过程中, 两个子网络 Network(1, 2) 都采用 ResNet-50^[6] 预训练网络作特征提取. 首先加载数据集, 用 ImageNet 数据集预训练过的模型 ResNet-50 进行参数微调以适应行人重识别的任务, 训练过程中两个卷积模型权值共享. 利用 Siamese 网络框架将两种卷积网络组合起来并且参数共享, 不仅可以对输入的两个行人图像提取特征进行相似度计算^[11], 判断是否为同一人, 还可以用于分类和特征学习.

1.4 马氏距离度量学习算法

为了提高训练的效率和识别的准确率, 需对全连接层进行替换变更. 由实验设定卷积网络参数, 对输入图像作卷积进行提取特征, 计算已提取的特征向量间的马氏距离进行相似度匹配, 以判断是否为同一行人.

马氏距离度量学习算法^[12] 是一种基于马氏距离的线性学习算法, 通过在样本集中寻找矩阵, 计算两个输入之间的马氏距离, 以作为样本相似度的评价指标. 在学习过程中, 对样本间距离进行线性变换, 用式(2)作为损失函数:

$$W_m^* = \arg \max_{W_m^T W_m = I} \frac{tr(W_m^T S_b W_m)}{tr(W_m^T S_w W_m)} \quad (2)$$

在模型设计过程中, 考虑到姿势、光照、遮挡等多方面原因^[13], 计算马氏距离时利用局部线性嵌入, 对特征相似性度量施加一定程度的限制. 此外, CNN 网络提取的特征还需要经过最大池化实现降维, 一定程度降低整个网络的计算复杂度. 降维后的特征向量用来计算检索图像对的马氏距离, 同时作为在此情况下最小的逼近误差. 利用最大似然估计求最优解, 即:

$$W_m^{(i)} = \frac{\sum_{m=1}^k (Q_{jm}^{(i)})^{-1}}{\sum_{p=1}^k \sum_{q=1}^k (Q_{qp}^{(i)})^{-1}} \quad (3)$$

求解过程中要保持在低位空间重构数据, 满足式

(4) 的限制条件:

$$\frac{1}{N} \sum_{i=1}^N y_i y_i^T = I \quad (4)$$

按升序排序的特征值对应的特征向量, 求得:

$$Y = [y_1 y_2, \dots, y_N] = \sqrt{N} [q_1 q_2, \dots, q_{d_2}]^T \quad (5)$$

计算梯度前, 先按照式 (6) 计算误差:

$$\delta_{ij}^{(M)} = g'(z) l_{ij} (h_i^{(M)} - \delta_i^{(M)}) \odot s'(z_i^{(M)}) \quad (6)$$

之后按照式 (7) 和式 (8) 计算梯度值:

$$\frac{\partial J}{\partial W^{(m)}} = \sum_{i,j} (\delta_{ij}^{(m)} h_i^{(m-1)T} + \delta_{ji}^{(m)} h_i^{(m-1)T}) \quad (7)$$

$$\frac{\partial J}{\partial b^{(m)}} = \sum_{i,j} (\delta_{ij}^{(m)} + \delta_{ji}^{(m)}) \quad (8)$$

在进行行人输入图像对相似度匹配验证时, 先以高维特征为基础计算样本间的马氏距离, 然后计算行人图像对的验证损失. 最后的实验结果表明, 本文所提出的算法确实可以在一定程度上提高行人重识别的精度.

2 实验检测结果分析

本文所提出的算法利用 Pytorch 框架实现, 所采用主要扩展库包括: Numpy, Scipy, Pandas, Matplotlib 等.

2.1 数据集

实验使用清华大学于 2015 年公开的 Person Re-id 数据集: Market-1501^[12]. Market-1501 数据集中的行人图像是在夏季清华大学校园中采集, 总共有 1501 个行人、32 668 个已标注的行人矩形框. 训练集采集了 751 人的数据, 共 12 936 张图像; 测试集采集了 750 人的数据, 共 19 732 张图像, 平均每个人有 26.3 张测试数据.

2.2 实验结果

输入大小为 256×128×3 的样本, 经过神经网络提取特征, 输出大小为 1×1×2048 的特征向量. 最后训练参数设置为: 一次训练的样本数目是 128, 最大迭代次数 70, 学习率 3e-4. 同时在迭代过程中, 利用随机梯度下降算法对模型参数进行更新.

最后对模型进行测试, 输入 Market-1501 数据集的测试数据并验证模型是否有效, Rank-1、mAP 指标如表 1 所示. 其中, Rank-*i* 为搜索结果中最靠前 (置信度最高) 的 *i* 张图有正确结果的概率. AP 为 PR (Precision-

Recall, 查准率和查全率) 曲线下的面积, mAP (如式 (9) 所示) 是对所有类的 AP 值取平均, 其更加全面衡量 ReId 算法效果的指标; mAP 要求将检索图片与被检索图片都匹配正确.

表 1 提出的算法与其他算法的对比

| 方法 | Rank-1 | Rank-5 | mAP | 执行时间 (s) |
|------------------------------|-------------|-------------|-------------|--------------|
| Verif-Identi ^[14] | 79.49 | 93.2 | 58.78 | - |
| GAN ^[15] | 82.96 | - | 66.4 | 38.56 |
| Gate Re-id ^[9] | 64.8 | - | 40.1 | - |
| DNS ^[16] | 55.34 | 80.2 | 31.77 | 41.49 |
| 提出的算法 | 83.2 | 82.2 | 69.2 | 34.21 |

$$mAP = \frac{1}{|Q_R|} \sum_{q \in Q_R} AP(q) \quad (9)$$

本文提出的算法在 Market1501 数据库上的 Rank-1 为 83.2%, Rank-5 为 82.2, mAP 为 69.2%, 与 Verif-Identi^[14]、GAN^[15]、DNS^[16] 等算法相比具有一定的优势.

实际应用时输入检索图像, 模型将输入的行人图像与图库的图像进行相似度匹配, 输出显示可能是同一行人的检索结果, 如图 4 所示, 绿色边框图像代表识别结果为同一个人的行人, 红色边框图像代表识别结果不是同一个人的行人.

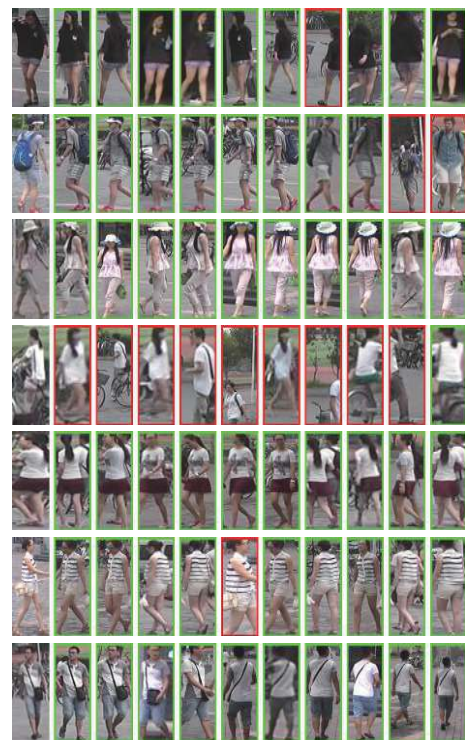


图 4 行人识别结果图

从表 1 可以看出, 算法在本系统的执行运行时间相比于文献[15,16]的算法用时更短. 主要原因在于虽然在训练模型上耗时较长, 但传输用时和提取特征用时更短. 从实验结果图可以看到, 识别正确率较高, 但仍存在错误识别的情况, 本系统仍需要训练更多的数据集以完善模型, 提高正确率.

3 结语

本文提出一种改进 CNN 神经网络进行行人重识别的方法, 先在 CNN 神经网络中加入 Dropout 算法降低过拟合发生概率, 再利用改进后的神经网络与 Siamese 网络框架融合的神经网络对图片进行特征提取, 实验结果表明识别准确率显著提高. 在实验过程中, 基于现有设备的条件, 我们仅在现有的数据集中选择了 Market-1501 数据集进行实验, 因此模型的泛化和结果的准确率有一定的误差. 希望在接下来的学习过程中, 可以改进模型结构增加数据量, 提高模型的泛化性.

参考文献

- Gheissari N, Sebastian TB, Hartley R. Person reidentification using spatiotemporal appearance. Proceedings of 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. New York, NY, USA. 2006. 1528–1535.
- 熊炜, 冯川, 熊子婕, 等. 基于 CNN 的改进行人重识别技术. 计算机工程与科学, 2019, 41(4): 665. [doi: 10.3969/j.issn.1007-130X.2019.04.013]
- Zheng L, Yang Y, Hauptmann AG. Person re-identification: Past, present and future. arXiv: 1610.02984, 2016.
- Liao SC, Hu Y, Zhu XY, *et al.* Person re-identification by local maximal occurrence representation and metric learning. Proceedings of 2005 IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA, USA. 2015. 2197–2206.
- Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Proceedings of the 25th International Conference on Neural Information Processing Systems. Lake Tahoe, NV, USA. 2012. 1097–1105.
- Srivastava N, Hinton G, Krizhevsky A, *et al.* Dropout: A simple way to prevent neural networks from overfitting. The Journal Machine Learning Research, 2014, 15(1): 1929–1958.
- Zagoruyko S, Komodakis N. Learning to compare image patches via convolutional neural networks. Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, MA, USA. 2015. 4353–4361.
- Varior RR, Shuai B, Lu JW, *et al.* A Siamese long short-term memory architecture for human re-identification. Proceedings of the 14th European Conference on Computer Vision (ECCV). Amsterdam, The Netherlands. 2016. 135–153.
- Varior RR, Haloi M, Wang G. Gated Siamese convolutional neural network architecture for human re-identification. Proceedings of the 14th European Conference on Computer Vision (ECCV). Amsterdam, The Netherlands. 2016. 791–808.
- Hochreiter S, Schmidhuber J. Long short-term memory. Neural Computation, 1997, 9(8): 1735–1780. [doi: 10.1162/neco.1997.9.8.1735]
- Zheng L, Shen LY, Tian L, *et al.* Scalable person re-identification: A benchmark. Proceedings of 2015 IEEE International Conference on Computer Vision (ICCV). Santiago, Chile. 2015. 1116–1124.
- 张兴福, 黄少滨. 基于马氏距离度量的局部线性嵌入算法. 模式识别与人工智能, 2012, 25(2): 318–324. [doi: 10.3969/j.issn.1003-6059.2012.02.020]
- 张国云, 向灿群, 罗百通, 等. 一种改进的人脸识别 CNN 结构研究. 计算机工程与应用, 2017, 53(17): 180–185, 191. [doi: 10.3778/j.issn.1002-8331.1612-0136]
- Zheng ZD, Zheng L, Yang Y. A discriminatively learned CNN embedding for person reidentification. ACM Transactions on Multimedia Computing, Communications, and Applications, 2016, 14(1): 13.
- Zheng ZD, Zheng L, Yang Y. Unlabeled samples generated by GAN improve the person re-identification baseline *in vitro*. Proceedings of 2017 IEEE International Conference on Computer Vision. Venice, Italy. 2017. 3774–3782.
- Zhang L, Xiang T, Gong SG. Learning a discriminative null space for person re-identification. Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA. 2016. 1239–1248.