JOURNAL OF IMAGE AND GRAPHICS

主办：中国科学院空天信息创新研究院
中国图象图形学学会
北京应用物理与计算数学研究所

中国图象图形学报

2020
12
VOL.25

ISSN1006-8961
CN11-3758/TB

数字浮雕建模 *P2647*

中国图象图形学报

刊名题字：宋健 | 月刊（1996年创刊）

第25卷第12期（总第296期）

2020年12月16日

中国精品科技期刊
中国国际影响力优秀学术期刊
中国科技核心期刊
中文核心期刊

主管单位 中国科学院
主办单位 中国科学院空天信息创新研究院
中国图象图形学学会
北京应用物理与计算数学研究所

主　　编 吴一戎
编辑出版 《中国图象图形学报》编辑出版委员会
通信地址 北京市海淀区北四环西路19号
邮　　编 100190
电子信箱 jig@radi.ac.cn
电　　话 010-58887035
网　　址 www.cjig.cn

广告发布登记号 京朝工商广登字20170218号
总 发 行 北京报刊发行局
订　　购 全国各地邮局
海外发行 中国国际图书贸易集团有限公司
（邮政信箱：北京399信箱　邮编：100048）
印刷装订 北京科信印刷有限公司

Journal of Image and Graphics

Title inscription: Song Jian | Monthly, Started in 1996

Superintended by Chinese Academy of Sciences
Sponsored by Aerospace Information Research Institute, CAS
China Society of Image and Graphics
Institute of Applied Physics and Computational Mathematics

Editor-in-Chief Wu Yirong
Editor, Publisher Editorial and Publishing Board of Journal of Image and Graphics
Address No. 19, North 4th Ring Road West, Haidian District, Beijing, P. R. China
Zip code 100190
E-mail jig@radi.ac.cn
Telephone 010-58887035
Website www.cjig.cn

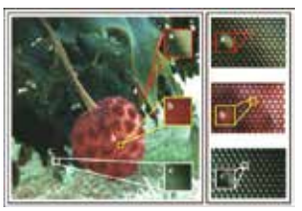
Distributed by Beijing Bureau for Distribution of Newspapers and Journals
Domestic All Local Post Offices in China
Overseas China International Book Trading Corporation
(P.O.Box 399, Beijing 100048,P.R.China))
Printed by Beijing Kexin Printing Co., Ltd.

CN 11-3758/TB
ISSN 1006-8961
CODEN ZTTXFZ

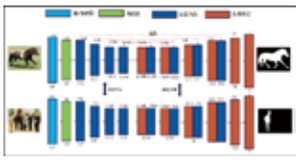
国外发行代号 M1406
国内邮发代号 82-831
国内定价 60.00元

2020年12月 第25卷 第12期（总第296期）
中国图象图形学报
Zhongguo Tuxiang Tuxing Xuebao

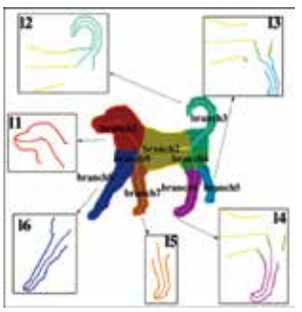
目次



光场显著性检测研究综述(第2465页)



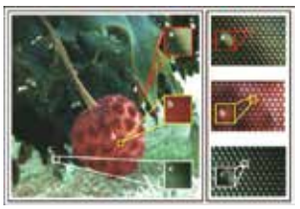
融合注意力机制与知识蒸馏的孪生网络压缩(第2563页)



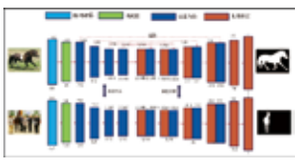
面向可解释性的物体拓扑结构骨架表征方法(第2587页)

Volume 25, Number 12
Published December 16, 2020

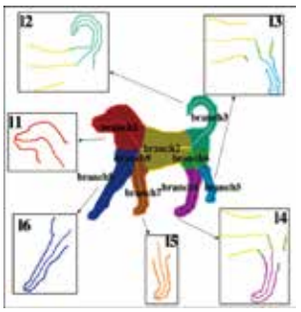
CONTENTS
JOURNAL OF IMAGE AND GRAPHICS



Review of saliency detection on light fields(P2465)



Combining attention mechanism and knowledge distillation for Siamese network compression (P2563)



Skeleton characterization of object topology toward explainability (P2587)

中图法分类号:TP391　　**文献标识码**: A　　**文章编号**: 1006-8961(2020)12-2563-12

论文引用格式: Geng Z M, Yu M Q, Liu X B and Lyu C. 2020. Combining attention mechanism and knowledge distillation for Siamese network compression. Journal of Image and Graphics,25(12):2563-2577(耿增民,余梦巧,刘峡壁,吕超. 2020. 融合注意力机制与知识蒸馏的孪生网络压缩. 中国图象图形学报,25(12):2563-2577)[DOI:10.11834/jig.200051]

融合注意力机制与知识蒸馏的孪生网络压缩

耿增民[1],余梦巧[2],刘峡壁[2],吕超[1]

1. 北京服装学院基础教学部, 北京　100029; 2. 北京理工大学计算机学院, 北京　100081

摘　要: **目的**　使用深度孪生网络解决图像协同分割问题,显著提高了图像分割精度。然而,深度孪生网络需要巨大的计算量,使其应用受到限制。为此,提出一种融合二值化注意力机制与知识蒸馏的孪生网络压缩方法,旨在获取计算量小且分割精度高的孪生网络。**方法**　首先提出一种二值化注意力机制,将其运用到孪生网络中,抽取大网络中的重要知识,再根据重要知识的维度重构原大网络,获取孪生小网络结构。然后基于一种改进的知识蒸馏方法将大网络中的知识迁移到小网络中,迁移过程中先后用大网络的中间层重要知识和真实标签分别指导小网络训练,以获取目标孪生小网络的权值。**结果**　实验结果表明,本文方法可将原孪生网络的规模压缩为原来的1/3.3,显著减小网络计算量,且分割结果接近于现有协同分割方法的最好结果。在MLMR-COS数据集上,压缩后的小网络分割精度略高于大网络,平均Jaccard系数提升了0.07%;在Internet数据集上,小网络分割结果的平均Jaccard系数比传统图像分割方法的最好结果高5%,且达到现有深度协同分割方法的最好效果;对于图像相对复杂的iCoseg数据集,压缩后的小网络分割精度相比于传统图像分割方法和深度协同分割方法的最好效果仅略有下降。**结论**　本文提出的孪生网络压缩方法显著减小了网络计算量和参数量,分割效果接近现有协同分割方法的最好结果。

关键词:孪生网络;网络压缩;知识蒸馏;注意力机制;图像协同分割

Combining attention mechanism and knowledge distillation for Siamese network compression

Geng Zengmin[1], Yu Mengqiao[2], Liu Xiabi[2], Lyu Chao[1]

1. Division of Basic Courses, Beijing Institute of Fashion Technology, Beijing　100029, China;

2. School of Computer, Beijing Institute of Technology, Beijing　100081, China

Abstract: **Objective** Image co-segmentation refers to segmenting common objects from image groups that contain the same or similar objects (foregrounds). Deep neural networks are widely used in this task given their excellent segmentation results. The end-to-end Siamese network is one of the most effective networks for image co-segmentation. However, this network has huge computational costs, which greatly limit its applications. Therefore, network compression is required. Although various network compression methods have been presented in the literature, they are mainly designed for single-branch networks and do not consider the characteristics of a Siamese network. To this end, we propose a novel network compression method specifically for Siamese networks. **Method** The proposed method transfers the important knowledge of a large network to a compressed small network. This method involves three steps. First, we acquire the important knowledge of the large network. To fulfill such task, we develop a binary attention mechanism that is applied to each stage of the

收稿日期:2020-02-18;**修回日期**:2020-03-30;**预印本日期**:2020-04-06

基金项目:北京市教委科技计划一般项目(KM201810012008)

Supported by:Scientific Research Project of Beijing Educational Committee (KM201810012008)

encode module of the Siamese network. This mechanism maintains the features of common objects and eliminates the features of non-common objects in two images. As a result, the response of each stage of the Siamese network is represented as a matrix with sparse channels. We map this sparse response matrix to a dense matrix with smaller channel dimensions through a 1 × 1 kernel size convolution layer. This dense matrix represents the important knowledge of the large network. Second, we build a small network structure. As described in the first step, the number of channels used to represent the knowledge in each stage of a large network can be reduced. Accordingly, the number of channels in each convolution and normalization layers included in each stage can also be reduced. Therefore, we reconstruct each stage of the large network according to the channel dimensions of the dense matrix obtained in the first step to determine the final small network structure. Third, we transfer the knowledge from the large network to the compressed small network. We propose a two-step knowledge distillation method to implement this step. First, the output of each stage/deconvolutional layer of the large network is used as the supervision information. We calculate the Euclidean distance between the middle-layer outputs of the large and small networks as our loss function to guide the training of the small network. This loss function is designed to make sure that the middle-layer outputs of the small and large networks are as similar as possible at the end of the first training stage. Second, we compute the dice loss between the network output and the real label to guide the final refining of the small network and to further improve the segmentation accuracy. **Result** We perform two groups of experiments on three datasets, namely MLMR-COS, Internet, and iCoseg. MLMR-COS has a large scale of images with pixel-wise ground truth. An ablation study is performed on this dataset to verify the rationality of the proposed method. Meanwhile, although Internet and iCoseg are commonly used datasets for co-segmentation, they are too small to be used as training sets for methods based on deep learning. Therefore, we train our network on a training set generated by Pascal VOC 2012 and MSRC before testing it on the Internet and iCoseg to verify its effectiveness. Experimental results show that the proposed method can reduce the size of the original Siamese network by 3.3 times thereby significantly reducing the required amount of computation. Moreover, compared with the existing co-segmentation methods based on deep learning, the proposed method can significantly reduce the amount of computation required in a compressed network. The segmentation accuracy of this compressed network on three datasets is close to the stat of the art. On the MLMR-COS dataset, this compressed small network obtains an average Jaccard index that is 0.07% higher than that of the original large network. Meanwhile, on the Internet and iCoseg datasets, we compare the compressed network with 12 traditional supervised/unsupervised image co-segmentation methods and 3 co-segmentation methods based on deep learning. On the Internet dataset, the compressed network has a Jaccard index that is 5% than the those of traditional image segmentation methods and existing co-segmentation methods based on deep learning. On the iCoseg dataset with relatively complex images, the segmentation accuracy of the compressed small network is slightly lower than those of the other methods. **Conclusion** We propose a network compression method by combining binary attention mechanism and knowledge distillation and apply it to a Siamese network for image co-segmentation. This network significantly reduces the amount of calculation and parameters in Siamese networks and is similar to the state-of-the-art methods in terms of co-segmentation performance.

Key words: Siamese network; network compression; knowledge distillation; attention mechanism; image co-segmentation

0 引 言

图像协同分割是从包含相同或相似对象(前景)的图像组中分割出共同对象。自 Rother 等人(2006)提出基于马尔可夫随机场(Markov random field, MRF)的图像协同分割方法后,协同分割受到越来越多的关注。随着深度学习的迅速发展,深度神经网络在协同分割领域的应用取得巨大成功,其中最具代表性的是孪生网络。用于图像分割的孪生网络通常为双分支结构,包括编码模块、互信息模块和解码模块。编码模块用于抽取图像特征,互信息模块利用两幅图像的相关性定位并强化共同目标的特征,解码模块根据图像特征获取最终分割结果。不同的孪生网络主要区别在于互信息模块的设计。Li 等人(2018)最先提出用端到端的孪生网络解决图像协同分割问题,其互信息模块计算了图像对中两两像素点之间的相关性,用于过滤得到两幅图的公共前景,在效果上取得了较大突破。Chen 等人(2019)在互信息模块中引入注意力机制,用于保留

两幅图像共同的语义信息,压缩非共同语义信息,得到了更好的分割效果。

与其他深度神经网络一样,孪生网络在提高分割精度的同时,也带来巨大的计算量,使其应用受到限制。为了解决这一共性问题,涌现出大量网络压缩方法。纪荣嵘等人(2018)将其划分成参数剪枝、参数共享、低秩分解、紧性卷积核设计和知识蒸馏 5 大类方法。目前,已有的网络压缩方法都是针对单分支网络设计,没有考虑双分支孪生网络的特点。本文提出针对孪生网络的压缩方法,在不损失网络精度的前提下,压缩网络计算量和参数量,使孪生网络在图像协同分割领域有更广泛的应用。

受 Chen 等人(2019)的启发,本文方法将非共同前景的语义信息完全置为 0,而完全保留共同前景的语义信息,得到响应层稀疏(若干通道全为 0)的孪生网络。稀疏的响应矩阵存在明显的信息冗余,本文通过矩阵映射将稀疏响应矩阵映射到通道数更少的低维稠密矩阵,作为原网络的重要知识,并基于该稠密矩阵的维度对原网络重构得到小网络的结构,实现对大网络计算参数量的压缩。最后,将大网络的重要知识迁移到小网络中,迁移过程中,先充分利用大网络的多层中间结果指导小网络训练,使小网络和大网络有相似的中间层响应,再用分割图像的真实标签继续指导小网络训练,进一步提高小网络的分割精度,使小网络具有大网络的分割精度。本文的主要贡献点如下:1)提出了二值化注意力模型,并运用到孪生网络中得到稀疏的响应矩阵,从中提取大网络的重要知识,并由此重构大网络得到小网络的结构;2)提出了两步知识迁移方法,先用多层大网络知识指导小网络训练,再用真实标签调精,显著提高了小网络的训练精度。

1　相关工作

1.1　图像协同分割

Rother 等人(2006)最先提出对图像对进行协同分割,在单幅图像的 MRF 模型能量函数中加入对图像对前景直方图相似性的约束项,再通过最优化能量函数实现对图像对的协同分割。随着深度学习的发展,深度神经网络广泛用于协同分割领域。起初,深度神经网络结合传统图像处理方法的思路居多,主要思想是借助神经网络获取图像间各像素点/超像素的共现图,再结合传统 grab-cut 方法获取最终的分割结果(Wang 等,2017;Yuan 等,2017;Quan 等,2016)。Wang 等人(2017)通过全卷积神经网络获取目标区域候选集,并基于此构建图像间像素点的共现图;Yuan 等人(2017)将稠密条件随机场融入神经网络,获取各候选区域的分割掩码图和属于共同目标区域的概率;Quan 等人(2016)运用卷积神经网络提取超像素的高级语义特征,联合超像素低级视觉特征生成超像素属于前景的概率图,即超像素的共现概率图,再结合 grab-cut 方法获取最终的分割结果。Mukherjee 等人(2018)提出可用于图像分割的 ANNOY(approximate nearest neighbor)库,将候选区域的低级特征联合孪生网络获取的高级语义特征共同输入 ANNOY 库中完成图像协同分割任务。而后,学者们倾向于完全用深度神经网络解决图像分割问题。Li 等人(2018) 提出用端到端的孪生网络解决图像协同分割问题。孪生网络包含编码模块、相关性计算模块和解码模块。输入两幅原图,编码模块通过卷积提取两幅图的特征,相关性计算模块做特征的匹配融合,解码模块通过反卷积获取网络分割结果。将网络分割结果与原图的真实标签图对比,求出损失,用于指导网络训练,是典型的监督学习方法。该方法实现了端到端的图像分割,在效果上取得了较大突破。Chen 等人(2019)提出用注意力学习模块代替 Li 等人(2018)方法中的相关性计算模块,增强两幅图像共同的语义信息,压缩非共同语义信息,以获取更好的分割效果。Hsu 等人(2018)提出非监督的协同分割网络,包含共同注意力图生成模块和语义特征提取模块。对于一组协同分割图像,先通过共同注意力生成模块获取每幅图的注意力图,并由此计算出掩码图、前景图和背景图,再通过语义特征提取模块获取前景图和背景图的语义特征。运用前景图之间的相似性、前背景图之间的差异性以及预选掩码图和当前掩码图之间的相似性构建目标函数,并通过最优化目标函数解决图像分割问题。

1.2　注意力机制

注意力机制常用于图像处理任务中。Zagoruyko 和 Komodakis(2016)提出基于激活值的空间注意力机制。对特征图 $\boldsymbol{F} \in \mathbf{R}^{C\times H\times W}$ (C 、H 、W 分别表示特征图的通道数、高和宽),求解 C 维通道上激活值的 p 范数 ($p \geqslant 1$) ,得到 $H \times W$ 大小的空间注意力

图,通过该空间注意力图将一个网络的知识转移到另一个网络。训练过程中,根据大网络和小网络的特征图计算空间注意力图,然后将空间注意力图之间的欧氏距离加到损失函数中,通过训练使两个网络有相似的空间注意力图和输出结果,达到知识迁移的目的。Zhang 等人(2017)认为上述方法在利用空间内的像素位置的激活值时,没有考虑像素的类别信息,提出类中心的概念描述每个类别的全局性特征,并将类特征融合到特征图中以提高图像处理效果。Chen 等人(2019)将注意力机制用到图像协同分割中,提出一种针对孪生网络的语义信息注意力机制,包括基于通道的注意力机制、基于通道融合的注意力机制和基于空间的注意力机制,充分利用空间和通道信息寻找图像间的共同语义信息,并增强共同语义信息,削弱非共同语义信息,提高了图像分割效果。

1.3 网络压缩

参数剪枝通过剪除网络中冗余的权值、通道和神经元压缩网络的参数量和计算量。参数剪枝最先由 LeCun 等人(1989)提出,通过 Hassien 矩阵计算各个权值的显著性,然后剪除显著性较低的权值。Han 等人(2015)认为权重小的权值对最终的网络决策影响较小,可将权重小于一定阈值的权值剪枝。Li 等人(2016)认为通道内权值绝对值和小的通道重要性较低,可以剪除。而 Luo 和 Wu(2017)认为权重的大小不能完全体现通道的重要性,提出一种基于熵值度量通道重要性的方法,将重要性较低的通道剪枝。Hu 等人(2016)对神经元进行剪枝,认为对于大部分不同输入都产生 0 激活值的神经元重要性较低,可剪枝这些神经元。Srinivas 和 Babu(2015)提出单层网络中权重相同的神经元只需保留其中一个,剪除其他冗余的神经元。

参数共享通过设计一种映射关系使多个参数共享同一个值,权值量化、哈希映射、聚类是最常见的表现方式。Dettmers(2015)将 32 bit 的高精度梯度值和激活值转成 8 bit,在损失微小精度的情况下明显减少了网络计算量,实现网络加速。Courbariaux 等人(2016)进一步将梯度值和权值限制为 -1 和 1,极大减少了网络的存储量和计算量。Rastegari 等人(2016)在 Courbariaux 等人(2016)方法的基础上,将网络输入也转化为二值的,进一步提高了网络计算效率,但也明显地损失了精度。为了解决这一问题,Li 等人(2017)将网络各层输入进行相对高精度的二阶二值量化,提高了网络精度。Chen 等人(2015)通过哈希函数将参数映射到多个哈希桶内,同一个哈希桶内的参数用同一个值表示,减小了网络存储量。Gong 等人(2014)利用 K 均值聚类算法将权值聚类成 K ($K \in \mathbf{N}$) 个簇,同一簇中的权值均用聚类中心值表示,实现了网络压缩。

低秩分解利用矩阵或张量分解技术估计并分解网络中的卷积核,用规模更小的多个卷积核表示原卷积核,有效实现了网络参数量和计算量的压缩。Jaderberg 等人(2014)和 Lebedev 等人(2014)思路相似,将原始卷积核分解成若干个秩为 1 的小卷积核相乘,有效地减少了网络的参数量和计算量。Yu 等人(2017)发现直接对各层卷积核进行低秩分解会明显损失网络精度,提出利用卷积核分解前后的特征图的重构误差辅助估计低秩分解,在参数量和计算量压缩方面取得了显著效果。

紧性卷积核设计方法在构建网络结构时,用低秩的小卷积核代替大卷积核,直接形成了规模较小的网络结构,SqueezeNet、MobileNet、ShuffleNet 等轻量网络是该方法的典型代表。SqueezeNet(Iandola 等,2016)提出用复杂度更低的 fire module 模块代替原始卷积层。fire module 模块由卷积核为 1×1 的压缩层和卷积核为 1×1 、3×3 的扩展层构成,替代后得到的 SqueezeNet 达到了 AlexNet 的分类精度,但参数量只有 AlexNet 的约 1/50 左右。MobileNet(Howard 等,2017)提出深度可分离的卷积模式,将标准卷积分解成深度卷积和 1×1 的逐点卷积,减小了网络参数量和计算量。ShuffleNet(Zhang 等,2018)通过卷积分组和通道打乱来实现网络压缩,其中卷积分组用于减少 1×1 卷积的计算量,通道打乱用于恢复卷积分组带来的精度损失,ShuffleNet 在分类精度和计算效率上均优于 AlexNet 和 MobileNet。

知识蒸馏将大网络的知识迁移到小网络中,使小网络具有与大网络相近的处理能力,是网络压缩的常用方法,由 Ba 和 Caruana(2014)和 Hinton 等人(2015)提出。Ba 和 Caruana(2014)发现相比于真实标签信息,softmax 层的输入含有更丰富的类别信息,于是提出将大小网络 softmax 层的输入特征做一一对比求最小平方误差,以此作为损失函数来指导小网络的训练,使小网络得到与大网络相近的处理

能力。Hinton 等人(2015)对 Ba 和 Caruana(2014)提出的方法做了泛化,提出通过修改 softmax 函数的温度参数 T 来生成类别信息更丰富的软标签,然后用软标签和硬标签共同作为监督信息来指导小网络的训练。Romero 等人(2014)发现当网络模型较深时,小网络很难直接根据 softmax 层的输入模拟大网络的效果,因此提出 Fitnets 模型,先用大网络中间层的信息监督训练小网络前半部分的参数,再用大网络 softmax 层的输入监督整个小网络的训练。该方法对深层大网络的模拟取得显著效果。Zhang 等人(2017)的方法与 Romero 等人(2014)的方法有异曲同工之妙,区别在于 Zhang 等人(2017)的方法是基于损失最小的准则迭代选择中间层作为监督信息,同时中间层的输出不仅用于监督小网络前半段的训练,也用于监督整个小网络的训练。Huang 和 Wang(2017)认为对于图像分割问题,将教师/学生网络的激活层信息进行一一对比求损失的方法丢失了空间和语义信息,因此提出利用大小网络 softmax 层输入分布的一致性来指导小网络的训练,使得大小网络得到的网络响应分布尽可能相似,从而使小网络模拟出大网络的分割效果。Lin 等人(2019)和 Shu 等人(2019)提出用生成对抗的方法进行知识蒸馏,对大小网络的输出求最小平方损失和对抗损失,在最大化对抗损失的同时最小化最小二乘损失,使小网络得到与大网络相似的输出。

2　本文方法

本文方法分为两个阶段。第 1 阶段从大网络中抽取重要知识并重构大网络得到小网络结构,第 2 阶段将大网络的重要知识迁移到小网络中,得到与大网络分割精度相当的目标小网络。

2.1　知识抽取

相较于小网络,大网络能够表达更丰富的特征,但也存在明显的特征冗余,应该从大网络中提取重要的知识,用于指导小网络的训练。本文提出通过二值化注意力机制来利用两幅图的相关性,以获取大网络的重要知识。对于网络某层的特征图 $\boldsymbol{F} \in \mathbf{R}^{C\times H\times W}$,通过二值化注意力机制得到其通道稀疏的特征图,仅 c ($c < C$) 个通道上有激活值,其他均为 0。为 0 的通道值不表示任何语义,是大网络中不重要的信息,可以去除,而仅保留有值的信息作为大网络的重要知识。

Chen 等人(2019)提出在孪生网络中加入注意力机制,将两幅图像共同的语义信息尽可能保留或增强,同时压缩非共同的语义信息。受此启发,本文提出的二值化注意力模型将两幅图像的非共同语义信息完全消除,共同语义信息完全保留。以基于通道的注意力机制模型为例,如图 1 所示,网络当前层响应矩阵为 $\boldsymbol{F}_A$ 和 $\boldsymbol{F}_B$,对 $\boldsymbol{F}_A$ 和 $\boldsymbol{F}_B$ 进行全局池化,得到两个向量 $\boldsymbol{\alpha}_A$ 和 $\boldsymbol{\alpha}_B$,再将 $\boldsymbol{\alpha}_A$ 和 $\boldsymbol{\alpha}_B$ 二值化,得到二值化向量 $\boldsymbol{\alpha}'_A$ 和 $\boldsymbol{\alpha}'_B$,表示两幅输入图像的语义信息。二值化过程中引入阈值 γ,当 $\boldsymbol{\alpha}'_A$ 和 $\boldsymbol{\alpha}'_B$ 第 i 维值 $\geqslant\gamma$,则将其置为 1,否则置为 0。$\boldsymbol{\alpha}'_A$ 的第 i 维值为 1 表示第 i 维通道表示了图像的重要语义信息,$\boldsymbol{\alpha}'_A$ 的第 i 维值为 0 表示第 i 维通道不表示图像的重要语义信息。将 $\boldsymbol{F}_A$ 与 $\boldsymbol{\alpha}'_B$ 相乘、$\boldsymbol{F}_B$ 与 $\boldsymbol{\alpha}'_A$ 相乘得到基于注意力模型的响应矩阵 $\boldsymbol{F}'_A$ 和 $\boldsymbol{F}'_B$,其中 $\boldsymbol{F}_A$ 接收了来自 $\boldsymbol{\alpha}'_B$ 的语义信息,$\boldsymbol{F}_B$ 接收了来自 $\boldsymbol{\alpha}'_A$ 的语义信息。若表示 $\boldsymbol{F}_B$ 语义信息的 $\boldsymbol{\alpha}'_B$ 在第 i 维通道上的激活响应值为 0,则第 i 维通道不表示 $\boldsymbol{B}$ 图像的重要语义信息,那么第 i 维就不表示两幅图的共同语义信息,相应地,$\boldsymbol{F}_A$ 第 i 维通道的语义信息将全部消除,即 $\boldsymbol{F}'_A$ 第 i 维通道的响应值全部为 0。若表示 $\boldsymbol{F}_B$ 语义信息的 $\boldsymbol{\alpha}'_B$ 在第 i 维通道上的激活响应值为 1,则第 i 维通道表示 $\boldsymbol{B}$ 图像的重要语义信息,那么对 $\boldsymbol{A}$ 图像而言,第 i 维就表示两幅图的共同语义信息,相应地,$\boldsymbol{F}_A$ 第 i 维通道的语义信息将完全保留,即 $\boldsymbol{F}'_A$ 第 i 维通道的响应值保持与 $\boldsymbol{F}_A$ 一致。将以上二值化注意力模型用到孪生网络中,会形成响应层稀疏的孪生网络。

本文用于协同分割的 U 型孪生网络结构由编码模块和解码模块构成,如图 2 所示。编码模块为去掉全连接层的 Resnet50 结构,包含一层卷积层和一层池化层,分为 4 段(stage),各段分别包含 3、4、6、3 个瓶颈(bottleneck)结构;解码模块由 6 层反卷积层构成。整个网络所包含的卷积层、反卷积层、归一化层、激活层、池化层共 190 层,其中卷积层和反卷积层共 59 层。若在每一层的卷积层或反卷积层之后加入二值化注意力模型,会大幅增加网络的复杂性,且会造成过度稀疏。因此,本文方法在每段之后加入二值化注意力模型,表示含注意力的段,如图 1 所示,在控制网络复杂性的同时,保持了合适的网络稀疏性和很高的网络精度。

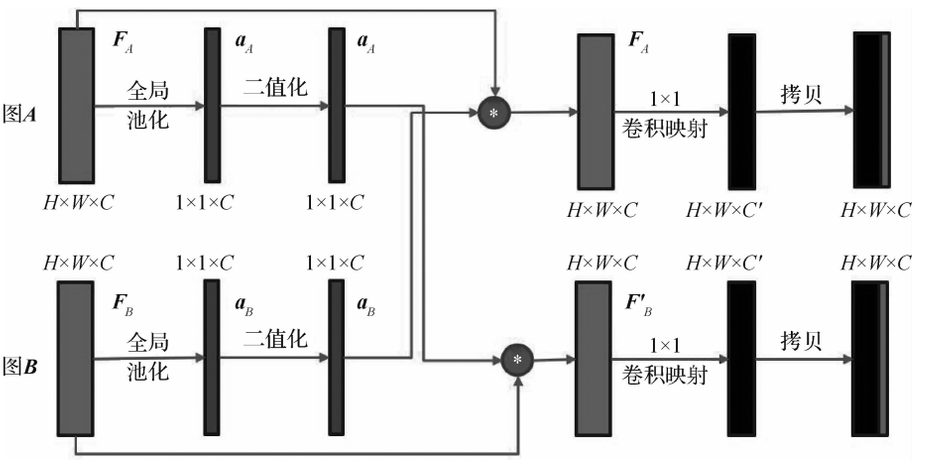


图 1　本文提出的二值化注意力模型以及矩阵映射方法

Fig. 1　The proposed binary attention model and matrix mapping method

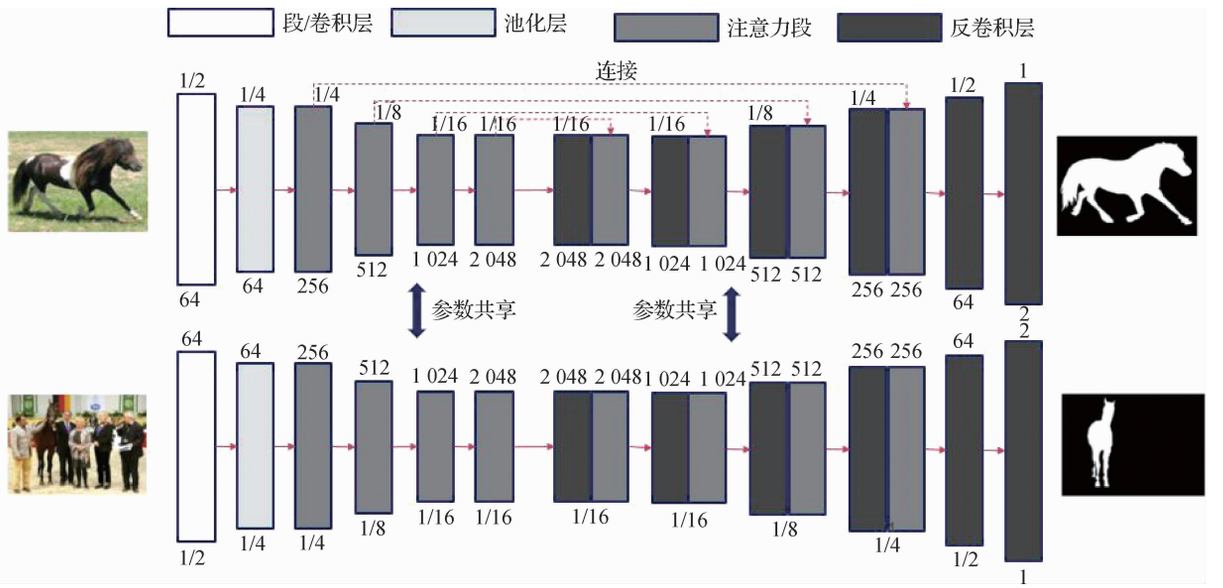


图 2　U 型孪生网络结构

Fig. 2　The architecture of U-shaped Siamese network

稀疏网络的响应矩阵中，激活值全为 0 的通道不表示图像的任何语义信息，为非重要信息，去掉这些通道不会对网络的最终分割结果产生明显影响。也就是说，该层网络可以用更少的通道来表示图像的重要语义信息。本文方法通过矩阵映射，用通道数更少的低维稠密矩阵来表示大网络的信息。常见的稀疏矩阵映射方法有局部线性嵌入法（Roweis 和 Saul，2000）、等度量映射法（Balasubramanian 和 Schwartz，2002）、主成成分分析法（Wold 等，1987）和局部保留投影法（He 和 Niyogi，2003）等。考虑到本问题的特殊性以及执行效率，本文方法采用卷积核为 1 × 1 的卷积运算实现稀疏矩阵的映射。如图 1 所示，原稀疏矩阵的尺寸为 $H \times W \times C$，1 × 1 卷积运算将稀疏矩阵从 $H \times W \times C$ 映射到 $H \times W \times c$，$c < C$。为了保证网络正常训练，需要维持每段原有的维度。于是，在网络中新添一层，将 $H \times W \times c$ 的低维映射矩阵拷贝到 $H \times W \times C$ 的零矩阵中，填充零矩阵的前 c 个通道，保证编码部分正常计算。

2.2　网络重构

知识蒸馏方法中的小网络结构通常由人为选择。小网络的结构对最终的结果有直接影响，本文的小网络结构通过重构大网络得到。由 2.1 节可

知，大网络存在的信息冗余具体表现为通道冗余。本文方法直接重构大网络每段的结构。比如，大网络某段的输出特征为 $\boldsymbol{F} \in \mathbf{R}^{C\times H\times W}$，其中 c（$c < C$）个通道激活值不为 0，其余 $C - c$ 个激活值为 0 的通道均为无效冗余通道，那么就将该段输出通道的维数重构为 C'（$C' < C$，C' 近似 c）维。与之对应，构成该段的若干个瓶颈结构中的卷积层或归一化层也存在通道冗余，对这些卷积层或归一化层一并重构，修改通道数，使之与该段的输出维度相匹配。本文采用的孪生网络编码模块为 ResNet50 结构，结构中 4 个段分别由若干个瓶颈结构堆叠而成，段内各个瓶颈结构相互联系，形成一个整体。4 个段的输入通道分别是 64、128、256 和 512，输出通道分别是 256、512、1 024 和 2 048，即各个段最后的输出通道数是输入通道数的 4 倍。根据这一客观情况，本文方法在重构时直接修改该段输入通道的数目为 $C'/4$（保证 $C'/4$ 为整数），即完成了对整段的重构。结合实验数据，本文方法将小网络编码模块各段的输入通道数从 64、128、256 和 512 重构为 56、112、128 和 256。解码模块的反卷积层与编码模块每段相连接，为了保证维度匹配，将解码模块中反卷积层的输入通道数分别从 2 018、1 024、512、256 重构为 1 024、512、448 和 224。以上重构方法实现了网络结构压缩，得到了小网络结构。

2.3　知识迁移

得到大网络的重要知识后，将大网络中重要的知识迁移到小网络中。Ba 和 Caruana（2014）提出让小网络模拟大网络决策层前一层的输出，使小网络得到与大网络尽可能一致的输出，但效果并不理想。就现实生活中的教师—学生关系而言，学生不仅要学习问题的答案，更重要的是学习老师解决问题的过程。受此启发，本文提出两步学习法，先学习解决问题的过程，再学习问题的答案。就本问题而言，即先学习大网络的各段/层的中间结果，待小网络和大网络的中间结果相似后，再学习真实结果。图 3 为小网络的两步训练过程，图中的大/小网络均为孪生网络，为了方便，表示为单分支网络。具体而言，第 1 阶段（图 3 左侧矩形区域）用大网络的每一段和每一个反卷积层的响应信息作为监督信息指导小网络的学习，将大小网络各段/反卷积层的输出一一对比，求得最小平方误差之和作为损失函数，具体为

$$L_1 = \sum_{i\in K}\frac{1}{2}\left\|\boldsymbol{S}_i(\boldsymbol{x},\boldsymbol{W}) - \boldsymbol{T}_i(\boldsymbol{x},\boldsymbol{W})\right\|^2 \quad (1)$$

式中，$\boldsymbol{K}$ 是大网络中用于监督小网络训练的段和反卷积层构成的集合，$\boldsymbol{x}$ 表示网络输入，$\boldsymbol{W}$ 表示网络权值，$\boldsymbol{S}_i(\boldsymbol{x},\boldsymbol{W})$，$\boldsymbol{T}_i(\boldsymbol{x},\boldsymbol{W})$ 分别表示小网络和大网络在 $\boldsymbol{K}$ 集合中的第 i 段/反卷积层的输出，通过最小化损失函数 L_1 优化小网络的参数。第 2 阶段（图 3 右侧矩形区域）用原图的真实标签信息指导小网络继续训练，其目标函数可以表示为

$$L_2 = \Phi(\boldsymbol{Y}_{\text{true}} - \boldsymbol{S}(\boldsymbol{x},\boldsymbol{W})) \quad (2)$$

式中，Φ 表示 dice 损失函数，$\boldsymbol{Y}_{\text{true}}$ 表示分割图像的真实标签，$\boldsymbol{S}(\boldsymbol{x},\boldsymbol{W})$ 表示小网络的预测结果，通过最小化损失值继续优化小网络参数。

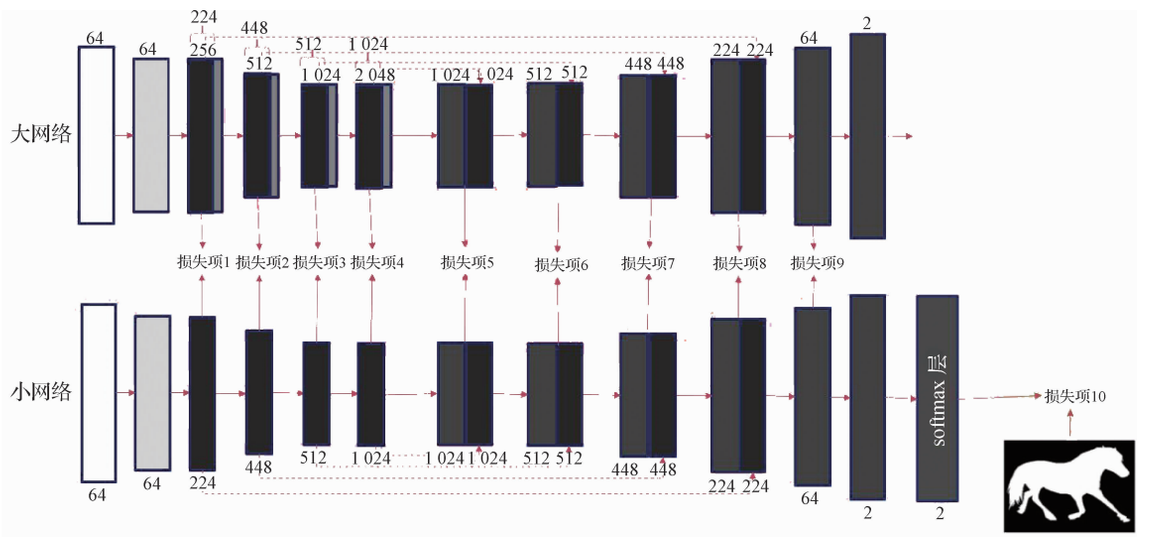


图 3　小网络的两步训练过程

Fig. 3　Two-step training process for small networks

2.4 算法流程

综合以上内容,本文提出孪生网络压缩方法,旨在获取计算量小且精度高的小网络。整体思想是先获取大网络中的重要知识,再根据这些重要知识的维度重构得到小网络结构,最后将大网络中的重要知识迁移到小网络中。具体过程为:1)改造大网络结构。在编码部分的每段之后加入二值化注意力机制,形成稀疏的响应矩阵 $\boldsymbol{F} \in \mathbf{R}^{C\times H\times W}$,并映射到低维 $\boldsymbol{F}' \in \mathbf{R}^{c\times H\times W}$,仅保留大网络的重要知识。为了保证大网络正常训练,在低维矩阵 $\boldsymbol{F}'$ 后添加一个与 $\boldsymbol{F}$ 维度一致的零矩阵,并将 $\boldsymbol{F}'$ 矩阵拷贝到零矩阵的前 c 个通道,然后训练改造后的大网络直到收敛。收敛后的大网络保持参数不变,用于后续指导小网络训练。2)根据大网络映射后低维矩阵的维度重构各段的通道数,并相应地重构解码部分反卷积的通道数,得到小网络的结构。3)分两步将大网络的知识迁移到小网络中。首先,训练时随机初始化小网络,将大网络各段/层的中间输出作为监督信息,并据此求出欧氏距离作为损失指导训练,使小网络尽可能多地获取大网络中的知识。其次,将监督信息换成真实标签,进一步提高小网络分割精度。本文方法的具体步骤如下:

输入:大网络结构 $\boldsymbol{T}$。

输出:训练好的小网络结构 $\boldsymbol{S}$。

1)改造大网络结构 $\boldsymbol{T}$,在各段之后加入二值化注意力机制并映射到低维,得到 $\boldsymbol{T}'$。

2)训练大网络 $\boldsymbol{T}'$ 至收敛。

3)重构 $\boldsymbol{T}'$ 得到小网络 $\boldsymbol{S}$。

4)随机初始化小网络的权值 $\boldsymbol{W}_S$。

5) $\boldsymbol{W}_S = \arg\min_{\boldsymbol{W}_S}(L_1)$

6) $\boldsymbol{W}_S = \arg\min_{\boldsymbol{W}_S}(L_2)$

L_1 和 L_2 为 2.3 节定义的目标函数。

3 实 验

本文共做了 3 组对比实验,第 1 组实验在大规模数据集 MLMR-COS (http://www.iscbit.org/source/MLMR-COS.zip)上进行,旨在验证本文压缩方法设计的合理性。后两组实验分别在常见协同分割数据集 Internet(Rubinstein 等,2013)和 iCoseg(Batra 等,2010)上进行,将通过本文方法压缩得到的小网络与原大网络、传统有监督、无监督和深度协同图像分割方法的分割结果进行对比,以验证本文压缩方法的有效性。

3.1 数据集

MLMR-COS 数据集包含 32 129 幅图像,由灯箱图像、vr_maker40 图像和发光转盘图像混合而成,共同点是背景简单、前景均为常见商品。各自的特点是:灯箱图像是在内壁为纯白色的方型灯箱中拍摄,顶部有白炽灯,物体和接触面易形成阴影。vr_maker40 图像的拍摄环境与灯箱图像相同,但前景多数含有透明区域,因此前背景非常相似。相比于灯箱图像,拍摄发光转盘图像时,白炽灯在底部,因此物体和接触面不易形成阴影,相对简单。MLMR-COS 数据集中每幅图像都有准确的前背景标注信息,本文实验从中随机选择 16 000 幅图像作为训练集训练孪生网络,然后在未见过的 68 幅灯箱图像、108 幅 vr_maker 40 图像和 108 幅发光转盘图像上进行测试。

在公开数据集上的实验采用 MSRC(Microsoft Research Cambridge Object Recognition Image Database)(Shotton 等,2006)数据集和 Pascal VOC (visual object classes)(Everingham 等,2010)数据集作为训练集,然后在 iCoseg 和 Internet 数据集上进行测试。MSRC 数据集包含 23 个类别共 591 幅图像,每幅图像粗略地标记了可用于指导图像分割的标签。VOC 2012 数据集包含 17 125 幅带物体检测框的图像和 2 913 幅含准确前背景标签的图像,仅 2 913 幅含有前背景标签信息的图像可用于图像分割的训练集。考虑到前景的显著性以及图像对之间的相关性,部分特征模糊的图像(例如,MSRC 数据集中部分图像没有明确前景,仅包含草地;VOC 2012 数据集中某两幅图像难以判断是否含有共同前景)被排除在外,最终从 MSRC 和 VOC 2012 数据集分别选取 507 和 1 743 幅图像,并将含有同类前景的图像两两配对生成训练集。训练过程中,随机采样 13 120 对图像对作为训练集训练孪生网络,然后在 Internet 和 iCoseg 数据集上进行测试。Internet 数据集仅包含车、马和飞机 3 个类别,其中 2 746 幅图像含有准确前背景标签,通常从每类中抽取 100 幅图像作为测试集。iCoseg 数据集包含 38 个类别共 634 幅图像,每幅图像都有准确的类别标签,全集和子集常用做测试集。

3.2　评价指标

图像分割领域常用的评价指标是准确率(precision,P)和 Jaccard 系数(J 系数),本文也采用这两种评价指标。准确率是正确分类的像素点占所有像素点的比例,即 $P = N_{\text{correct}}/N_{\text{all}}$,其中, N_{correct} 表示分割结果中正确分类的像素点数目, N_{all} 表示分割图像中像素点的总数。Jacarrd 系数表示分割图和标签图的前景覆盖率,定义为 $J = \frac{\boldsymbol{S} \cap \boldsymbol{G}}{\boldsymbol{S} \cup \boldsymbol{G}}$,其中, $\boldsymbol{S}$ 和 $\boldsymbol{G}$ 分别表示分割结果和真实标签的二值图,其中前景表示为 1,背景表示为 0。

3.3　参数设置

从大网络中抽取重要知识时,引入了二值化注意力机制,二值化阈值 γ 的设置直接影响到大网络的稀疏性和分割精度,确定合适的阈值 γ 至关重要。以在 MLMR-COS 数据集上的 MON(modify original network)实验为例,前 3 段的 γ 值作用在特征抽取过程中, γ 值较大会损失有效的语义信息,影响分割精度,本实验将前 3 段的 γ 均设置为 0.05。第 4 段的阈值是对原图特征的最后一次过滤,调参时将 γ 分别设置为 0.2、0.3、0.4、0.5、0.6、0.7、0.8,统计各自训练结束时通道的稀疏度和网络分割精度。实验发现,当 $\gamma \leqslant 0.6$ 时,网络精度损失微小;当 $\gamma > 0.6$ 后,第 4 段有值的通道非常少,精度损失较明显。对比 $\gamma \leqslant 0.6$ 的 4 组参数,发现当 $\gamma = 0.5$ 时,网络精度与大网络一致,且网络稀疏度很高。综合考虑网络的稀疏性和准确性,将第 4 段的 γ 设置为 0.5。 γ 值的设置与数据集密切相关,对相对复杂的 Pascal VOC 2012 和 MSRC 数据集,需要保留更多的通道来表达更丰富的语义信息,因此将 γ 设置为较小的值,分别为 0.005 和 0.05。其他重要参数设置如下:迭代次数均为 100;批量大小均为 4;训练 MON 和 RCN(retrain compressed network)时,学习率 lr 为 $1e^{-5}$,权重衰减 $decay$ 为 $5e^{-8}$;用本文提出的两步重训练方法训练小网络时,第 1 阶段 lr 为 $1e^{-5}$, $decay$ 为 $5e^{-8}$;第 2 阶段 lr 为 $1e^{-3}$, $decay$ 为 $1e^{-5}$。

3.4　在 MLMR-COS 数据集上的实验

为了验证本文方法的合理性和有效性,对图 2 所示的协同分割网络做了相应改变,形成了 5 组对比实验:1)基础实验(base line, BL)。将 Chen 等人(2019)提出的通道注意力模型加到孪生网络编码模块第 4 段之后计算两幅图像的相关性,所形成的网络为大网络;2)改造大网络(MON)。在编码模块的第 4 段之后分别加入二值化注意力模型,并将每段得到的稀疏响应矩阵映射到低维矩阵,再根据低维矩阵的维度重构解码模块。3)重训练小网络(RCN)。根据 MON 实验中大网络的结构重构,得到小网络结构,并用真实标签作为监督信息训练小网络。4)知识迁移第 1 步(knowledge transfer step1, KT1)。用多层大网络的重要知识指导小网络训练。5)知识迁移第 2 步(knowledge transfer step2,KT2)。在 KT1 的基础上,用真实标签继续指导小网络训练。实验结果如表 1 所示。

表 1　MLMR-COS 数据集上的实验结果

Table 1　Experimental results of each method on MLMR-COS dataset

实验	灯箱图像		vr_maker40 图像		发光转盘图像		平均值	
	J 系数	精确度/%	J 系数	精确度/%	J 系数	精确度/%	J 系数	精确度/%
基础实验(BL)	0.984 7	99.84	0.970 8	99.85	0.990 8	99.85	0.982 1	**99.85**
改造大网络(MON)	0.983 8	99.83	0.971 4	99.85	**0.991 2**	**99.86**	0.982 1	**99.85**
重训练小网络(RCN)	0.972 3	99.70	0.957 1	99.78	0.983 7	99.74	0.971 0	99.74
知识迁移第 1 步(KT1)	0.984 0	99.83	0.967 8	99.83	0.987 7	99.81	0.979 8	99.82
知识迁移第 2 步(KT2)	**0.985 3**	**99.84**	**0.974 6**	**99.87**	0.988 6	99.83	**0.982 8**	**99.85**

注:加粗字体表示各列最优结果。

基础实验在 3 个测试数据集上的平均 Jaccard 系数和准确率分别为 0.982 1 和 99.85% 。相比于 BL,MON 在编码模块的 4 段之后分别加入二值化注意力模型,完全保留图像对的共同语义信息,完全消

除其他语义信息,有助于网络分割出公共前景区域,对于相对简单的发光转盘图像效果尤其明显。最终MON方法在极大地增加网络稀疏性的同时,达到与BL一致的分类效果。

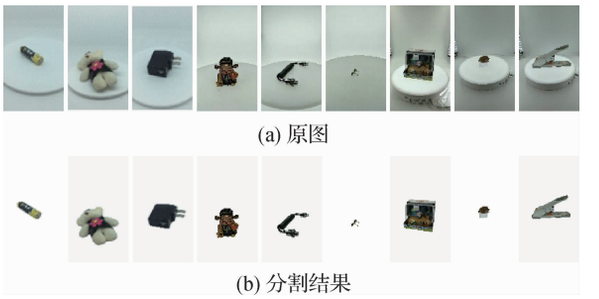
表2是编码模块每段的响应矩阵稀疏性统计结果。可以看出,该实验在每段之后产生了通道稀疏的响应矩阵,其中为0的通道不表示任何语义信息,说明大网络在解决背景简单的MLMR-COS数据集的分割问题时,存在明显的通道冗余,较少的通道即可表示大网络中的重要知识。因此,可以根据表2对大网络进行重构,得到规模更小的小网络结构。RCN得到的小网络平均Jaccard系数和准确率相比大网络分别损失了0.011 1和0.08%。尽管损失很小,但MLMR-COS数据集本身属于简单的分割问题,从精确分割的角度来看,细小的精度损失都会造成分割边缘等细节的分割效果明显下降。KT1用大网络中间层信息指导小网络训练,极大程度上恢复了网络精度,证明大网络的中间层知识对于小网络的精度恢复有重要作用;KT2进一步提高小网络的分割精度,使小网络超越了大网络的分割效果,平均Jaccard系数提升0.000 7,准确率与大网络一致。对比RCN与本文提出的两步知识迁移方法KT1和KT2可以看出,本文提出的知识迁移方法显著提高了小网络的分割精度,得到了更精确的分割结果。整体来看,本文压缩方法将大网络规模从185.26 MB减小为56.03 MB,压缩为原来的1/3.3。处理一对1 920×1 024像素的JPG图像的网络浮点运算量由1.36×10^{12}减小为6.67×10^{11},压缩为原来的1/2.03,且在3个数据集上的平均Jaccard系数提升了0.000 7,证明了本文方法的有效性。

表2　编码模块每段的响应矩阵稀疏性统计结果

Table 2　Sparsity statistics of the response matrix for each stage of the encoding module

段	原响应矩阵非零通道数	MON中响应矩阵非零通道数
1	256	190～210
2	512	250～375（300左右居多）
3	1 024	130～270（200左右居多）
4	2 048	5～30（20左右居多）

图4是小网络在MLMR-COS数据集上的分割结果,可以看出,小网络在MLMR-COS数据集上得到了完整且准确的分割结果。



(a) 原图

(b) 分割结果

图4　小网络在MLMR-COS数据集上的分割结果

Fig. 4　Segmentation results of small network on MLMR-COS dataset ((a) original images; (b) segmentation results)

3.5　在Internet数据集上的实验

为了验证本文压缩得到的小网络在减小计算量的同时保持了很高的分割精度,在Internet数据集的3个类别中各选100幅图像进行实验,对有监督图像分割方法、无监督图像分割方法、深度协同图像分割方法、大网络和小网络的图像分割结果进行对比,结果如表3所示。

从表3可知,大网络在Internet数据集上分割结果的平均Jaccard系数是0.800,显著高于无监督和有监督图像分割方法,略高于深度协同分割方法。相较于大网络,小网络规模被压缩为原来的1/3.3,图像分割的Jaccard系数明显下降,平均Jaccard系数为0.727,但高于有监督方法中Jerripothula等人(2016)方法的0.643和无监督方法中效果最好的Yuan等人(2017)方法的0.677,且与深度协同分割方法的分割效果基本持平。

表3还对比了小网络和深度协同分割方法分割一对512×512像素的JPG图像的浮点运算次数(floating-point operations per second,FLOPs),用于衡量算法效率。对同一GPU机器而言,FLOPs越低,效率越快。传统协同分割方法常用于CPU机器,分割效率明显低于深度协同分割方法,为了公平对比各方法的效率,未将其与深度协同分割方法进行对比。从表3中的FLOPs计算量可以看出,本文方法压缩得到的小网络计算量显著低于深度协同分割方法,分割效率达到了深度协同分割方法的最好效果,充分验证了本文方法的有效性。

小网络在Internet数据集上的分割结果如图5所示。可以看出,小网络在Internet数据集的3个子集上都能得到相对完整的分割前景,验证了小网络

表 3　各方法在 Internet 数据集上的实验结果

Table 3　Experimental results of each method on Internet dataset

	图像分割方法	马		飞机		汽车		平均值		FLOPs
		J 系数	精确度/%	J 系数	精确度/%	J 系数	精确度/%	J 系数	精确度/%	图尺寸:512×512 像素
无监督	Joulin 等人(2012)	0.30	64.20	0.12	47.50	0.35	59.20	0.243	56.97	–
	Rubinstein 等人(2013)	0.52	82.80	0.56	88.00	0.64	85.40	0.427	82.73	–
	Chen 等人(2014)	0.58	89.30	0.40	90.20	0.65	87.60	0.543	89.03	–
	Chang 和 Wang(2015)	0.36	79.70	0.27	72.60	0.36	75.90	0.330	76.07	–
	Lee 等人(2015)	0.39	70.10	0.36	52.80	0.42	64.70	0.392	62.53	–
	Jerripothula 等人(2016)	0.61	88.30	0.61	90.50	0.71	88.00	0.643	88.93	–
	Quan 等人(2016)	0.58	89.30	0.56	91.00	0.67	88.50	0.603	89.60	–
	Hati 等人(2016)	0.20	73.80	0.33	77.70	0.43	62.10	0.320	71.20	–
	Sun 和 Ponce(2016)	0.55	87.60	0.36	88.60	0.73	87.00	0.547	87.73	–
	Tao 等人(2017)	0.55	85.70	0.43	79.60	0.66	84.80	0.547	83.43	–
	Jerripothula 等人(2017)	0.50	81.30	0.48	81.80	0.69	84.70	0.556	82.60	–
有监督	Yuan 等人(2017)	0.65	89.70	0.66	92.60	0.72	90.40	0.677	91.07	–
	Li 等人(2018)	0.69	**92.40**	0.65	94.10	0.83	**93.90**	0.723	**93.50**	5.54×10^{11}
深度协同	Chen 等人(2019)	0.71	–	0.68	–	0.80	–	0.730	–	3.36×10^{11}
	Hsu 等人(2018)	0.61	89.70	0.67	**94.20**	0.82	93.00	0.700	92.30	2.31×10^{11}
大网络		0.78	91.20	**0.77**	93.10	0.84	92.20	**0.800**	92.17	1.81×10^{11}
小网络		0.69	86.50	0.70	90.90	0.79	89.50	0.727	88.97	8.90×10^{10}

注:加粗字体表示各列最优结果,“–”表示原文献没有给出数据。

(a) 原图

(b) 分割效果

图 5　小网络在 Internet 数据集上的分割效果

Fig. 5　Segmentation results of small network on Internet dataset ((a) original images; (b) segmentation results)

分割效果的有效性。主要源于两方面因素:1)相较于传统方法中人工提取的特征,深度神经网络的高维度特征包含的语义信息更丰富,能更好地表达图像信息,得到更好的分割效果。这也是大网络有比小网络更好的分割效果的原因,因为大网络的网络规模更大,每一层表示特征的维度更大,包含的信息比小网络中更丰富。2)实验采用的孪生网络引入了二值化注意力机制,能更有效地识别并保留两幅图像的共同前景信息,去除非共同语义信息,有利于充分利用图像对的协同信息,获取更好的分割效果。

3.6　在 iCoseg 数据集上的实验

为了进一步验证压缩后小网络的分割精度,在包含 38 组数据的完整 iCoseg 数据集以及包含 16 组数据的 iCoseg 子集上做分割实验,对比有监督图像分割方法、无监督图像分割方法、深度协同分割方法、大网络和小网络的分割结果,结果如表 4 所示。

由表 4 可知,大网络在 iCoseg 全集上分割结果的平均 Jaccard 系数是 0.82,与有监督方法中 Yuan 等人(2017)方法的结果一致,比无监督方法的最好

表 4　各方法在 iCoseg 数据集上的实验结果

Table 4　Experimental results of each method on iCoseg dataset

	图像分割方法	iCoseg 子集		iCoseg 全集		FLOPs 图尺寸:512×512 像素
		J 系数	精确度/%	J 系数	精确度/%	
无监督	Jerripothula 等人(2016)	–	–	0.720	91.8	–
	Quan 等人(2016)	0.820	94.8	0.760	93.3	–
	Tao 等人(2017)	0.704	90.8	–	–	–
	Wang 等人(2017)	0.770	93.8	–	–	–
有监督	Yuan 等人(2017)	0.860	**96.0**	0.820	94.4	–
	Li 等人(2018)	0.840	95.1	–	–	5.54×10^{11}
深度协同	Chen 等人(2019)	0.860	–	–	–	23.36×10^{11}
	Hsu 等人(2018)	–	–	**0.840**	**96.5**	2.31×10^{11}
大网络		**0.870**	94.8	0.820	93.4	1.81×10^{11}
小网络		0.780	91.0	0.730	89.5	8.90×10^{10}

注:加粗字体表示各列最优结果。“–”表示原文献没有给出数据。

结果(Quan 等,2016)高 6%,但比深度协同分割算法的最好效果(Hsu 等,2018)低 2%;在 iCoseg 子集上分割结果的平均 Jaccard 系数是 0.87,比有监督方法的最好结果(Yuan 等,2017)高 1%,比无监督方法的最好结果(Quan 等,2016)高 5%,比深度协同分割算法的最好效果(Chen 等,2019)高 1%。小网络规模压缩为原来的 1/3.3 后,处理一对 512×512 像素的 JPG 图像 FLOPs 只有 8.90×10^{10},显著低于深度协同分割方法,但存在一定的精度损失。相较于大网络,在完整 iCoseg 数据集和 iCoseg 子集上,Jaccard 系数都有明显下降,分别下降 9%,低于深度协同分割算法的最好结果,但仍优于无监督方法(Jerripothula 等,2016;Tao 等,2017;Wang 等,2017)的结果。小网络精度明显下降的原因可能是 iCoseg 数据集背景相对复杂,前景多变,包含的信息复杂,压缩后的小网络在通道数上明显减少,不足以充分获取训练集上图像的多样性信息,导致小网络在 iCoseg 数据集上的泛化性能下降。观察表 3 和表 4 还可以发现,在传统图像处理方法中,有监督方法中 Yuan 等人(2017)方法的处理效果最好。严格来说,Yuan 等人(2017)方法采用了传统方法结合深度神经网络的方法,先通过启发式搜索(van de Sande 等,2011)获取每幅图像的候选区域,再用端到端的神经网络获取每个候选区域在图像间的共现图,将协同分割转换成单幅图像的分割问题,再借助传统图像分割方法获取最终的分割结果。该方法受益于深度神经网络,效果普遍优于其他传统图像处理方法,接近于深度方法的效果。

小网络在 iCoseg 数据集上分割结果如图 6 所示。

(a) 原图

(b) 分割效果

图 6　小网络在 iCoseg 数据集上的分割效果

Fig. 6　Segmentation results of small network on iCoseg dataset ((a) original images; (b) segmentation results)

4 结 论

本文提出一种融合注意力机制与知识蒸馏的孪生网络压缩方法，先将二值化注意力机制运用到孪生网络中得到稀疏的响应矩阵，然后映射得到大网络的重要知识，再通过两步知识迁移方法将大网络中的重要知识充分应用到小网络中，得到目标小网络。目标小网络在 MLMR-COS 数据集上的实验结果表明，本文提出的二值化注意力机制准确提取了大网络的重要知识，且两步知识迁移方法有效地将大网络中的重要知识迁移到小网络中。在公开数据集 Internet 和 iCoseg 上的实验结果表明，本文方法压缩得到的小网络在显著减少计算量的同时，基本保持了原有的分割精度。由此可见，本文方法有较好的实验效果和应用价值。未来将对二值化注意力机制中的阈值选择方法做进一步探索，实现在不同数据集上能自动确定阈值，进一步提高小网络的分割结果。

参考文献(References)

Ba L J and Caruana R. 2014. Do deep nets really need to be deep? // Proceedings of the 27th International Conference on Neural Information Processing Systems. Cambridge, USA: MIT Press: 2654-2662

Balasubramanian M and Schwartz E L. 2002. The isomap algorithm and topological stability. Science, 295(5552): #7 [DOI: 10.1126/science.295.5552.7a]

Batra D, Kowdle A, Parikh D, Luo J B and Chen T. 2010. iCoseg: interactive co-segmentation with intelligent scribble guidance//Proceedings of 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Francisco: IEEE: 3169-3176 [DOI: 10.1109/CVPR.2010.5540080]

Chang H S and Wang Y C F. 2015. Optimizing the decomposition for multiple foreground cosegmentation. Computer Vision and Image Understanding, 141: 18-27 [DOI: 10.1016/j.cviu.2015.06.004]

Chen H, Huang Y F and Nakayama H. 2019. Semantic aware attention based deep object co-segmentation//Proceedings of the 14th Asian Conference on Computer Vision. Perth, Australia: Springer: 435-450 [DOI: 10.1007/978-3-030-20870-7_27]

Chen W L, Wilson J T, Tyree S, Weinberger K and Chen Y X. 2015. Compressing neural networks with the hashing trick//Proceedings of the 32nd International Conference on Machine Learning. Lille, France: JMLR: 2285-2294

Chen X L, Shrivastava A and Gupta A. 2014. Enriching visual knowledge bases via object discovery and segmentation//Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, USA: IEEE: 2027-2034 [DOI: 10.1109/CVPR.2014.261]

Courbariaux M, Hubara I, Soudry D, El-Yaniv R and Bengio Y. 2016. Binarized neural networks: training deep neural networks with weights and activations constrained to +1 or -1 [EB/OL]. [2020-01-15]. https://arxiv.org/pdf/1602.02830.pdf

Dettmers T. 2015. 8-bit approximations for parallelism in deep learning [EB/OL]. [2020-01-15]. https://arxiv.org/pdf/1511.04561.pdf

Everingham M, Van Gool L, Williams C K I, Winn J and Zisserman A. 2010. The PASCAL visual object classes (VOC) challenge. International Journal of Computer Vision, 88(2): 303-338 [DOI: 10.1007/s11263-009-0275-4]

Gong Y C, Liu L, Yang M and Bourdev L. 2014. Compressing deep convolutional networks using vector quantization [EB/OL]. [2020-01-15]. https://arxiv.org/pdf/1412.6115.pdf

Han S, Mao H Z and Dally W J. 2015. Deep compression: compressing deep neural networks with pruning, trained quantization and Huffman coding [EB/OL]. [2020-01-15]. https://arxiv.org/pdf/1510.00149.pdf

Hati A, Chaudhuri S and Velmurugan R. 2016. Image co-segmentation using maximum common subgraph matching and region co-growing//Proceedings of the 14th European Conference on Computer Vision. Amsterdam, the Netherlands: Springer: 736-752 [DOI: 10.1007/978-3-319-46466-4_44]

He X F and Niyogi P. 2003. Locality preserving projections//Proceedings of the 16th International Conference on Neural Information Processing Systems. Vancouver, Canada: MIT Press: 153-160

Hinton G, Vinyals O and Dean J. 2015. Distilling the knowledge in a neural network [EB/OL]. [2020-01-15]. https://arxiv.org/pdf/1503.02531.pdf

Howard A G, Zhu M L, Chen B, Kalenichenko D, Wang W J, Weyand T, Andreetto M and Adam H. 2017. MobileNets: efficient convolutional neural networks for mobile vision applications [EB/OL]. [2020-01-15]. https://arxiv.org/pdf/1704.04861.pdf

Hsu K J, Lin Y Y and Chuang Y Y. 2018. Co-attention CNNs for unsupervised object co-segmentation//Proceedings of the 27th International Joint Conference on Artificial Intelligence. Stockholm, Sweden: [s.n.]: 748-756. [DOI: 10.24963/ijcai.2018/104]

Hu H Y, Peng R, Tai Y W and Tang C K. 2016. Network trimming: a data-driven neuron pruning approach towards efficient deep architectures [EB/OL]. [2020-01-15]. https://arxiv.org/pdf/1607.03250.pdf

Huang Z H and Wang N Y. 2017. Like what you like: knowledge distill via neuron selectivity transfer [EB/OL]. [2020-01-15]. https://arxiv.org/pdf/1707.01219.pdf

Iandola F N, Han S, Moskewicz M W, Ashraf K, Dally W J and Keutzer K. 2016. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and < 0.5 MB model size [EB/OL]. [2020-01-15]. https://arxiv.org/pdf/1602.07360.pdf

Jaderberg M, Vedaldi A and Zisserman A. 2014. Speeding up convolutional neural networks with low rank expansions//Proceedings of British Machine Vision Conference. Nottingham, UK: BMVA Press: 1-7 [DOI: 10.5244/C.28.88]

Jerripothula K R, Cai J F, Lu J B and Yuan J S. 2017. Object co-skeletonization with co-segmentation//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 3881-3889 [DOI: 10.1109/CVPR.2017.413]

Jerripothula K R, Cai J F and Yuan J S. 2016. Image co-segmentation via saliency co-fusion. IEEE Transactions on Multimedia, 18(9): 1896-1909 [DOI: 10.1109/tmm.2016.2576283]

Ji R R, Lin S H, Chao F, Wu Y J and Huang F Y. 2018. Deep neural network compression and acceleration: a review. Journal of Computer Research and Development, 55(9): 1871-1888 (纪荣嵘, 林绍辉, 晁飞, 吴永坚, 黄飞跃. 2018. 深度神经网络压缩与加速综述. 计算机研究与发展, 55(9): 1871-1888) [DOI: 10.7544/issn1000-1239.2018.20180129]

Joulin A, Bach F and Ponce J. 2012. Multi-class cosegmentation//Proceedings of 2012 IEEE Conference on Computer Vision and Pattern Recognition. Providence, USA: IEEE: 542-549 [DOI: 10.1109/CVPR.2012.6247719]

LeCun Y, Denker J S and Solla S A. 1989. Optimal brain damage//Proceedings of the 2nd International Conference on Neural Information Processing Systems. Denver, USA: MIT Press: 598-605

Lebedev V, Ganin Y, Rakhuba M, Oseledets I and Lempitsky V. 2014. Speeding-up convolutional neural networks using fine-tuned CP-decomposition [EB/OL]. [2020-01-15]. https://arxiv.org/pdf/1412.6553.pdf

Lee C, Jang W D, Sim J Y and Kim C S. 2015. Multiple random walkers and their application to image cosegmentation//Providence of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA: IEEE: 3837-3845 [DOI: 10.1109/CVPR.2015.7299008]

Li H, Kadav A, Durdanovic I, Samet H and Graf H P. 2016. Pruning filters for efficient convnets [EB/OL]. [2020-01-15]. https://arxiv.org/pdf/1608.08710v3.pdf

Li W H, Jafari O H and Rother C. 2018. Deep object co-segmentation//Proceedings of the 14th Asian Computer Vision - ACCV 2018. Perth, Australia: Springer: 638-653 [DOI: 10.1007/978-3-030-20893-6_40]

Li Z F, Ni B B, Zhang W J, Yang X K and Gao W. 2017. Performance guaranteed network acceleration via high-order residual quantization//Proceedings of 2017 IEEE International Conference on Computer Vision. Venice, Italy: IEEE: 2584-2592 [DOI: 10.1109/ICCV.2017.282]

Lin S H, Ji R R, Yan C Q, Zhang B C, Cao L J, Ye Q X, Huang F Y and Doermann D. 2019. Towards optimal structured CNN pruning via generative adversarial learning//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE: 2790-2799 [DOI: 10.1109/CVPR.2019.00290]

Luo J H and Wu J X. 2017. An entropy-based pruning method for CNN compression [EB/OL]. [2020-01-15]. https://arxiv.org/pdf/1706.05791.pdf

Mukherjee P, Lall B and Lattupally S. 2018. Object cosegmentation using deep Siamese network [EB/OL]. [2020-01-15]. https://arxiv.org/pdf/1803.02555.pdf

Quan R, Han J W, Zhang D W and Nie F P. 2016. Object co-segmentation via graph optimized-flexible manifold ranking//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: 687-695 [DOI: 10.1109/cvpr.2016.81]

Rastegari M, Ordonez V, Redmon J and Farhadi A. 2016. XNOR-net: ImageNet classification using binary convolutional neural networks//Proceedings of the 14th European Conference on Computer Vision. Amsterdam, the Netherlands: Springer: 525-542 [DOI: 10.1007/978-3-319-46493-0_32]

Romero A, Ballas N, Kahou S E, Chassang A, Gatta C and Bengio Y. 2014. FitNets: hints for thin deep nets [EB/OL]. [2020-01-15]. https://arxiv.org/pdf/1412.6550.pdf

Rother C, Minka T, Blake A and Kolmogorov V. 2006. Cosegmentation of image pairs by histogram matching - incorporating a global constraint into MRFs//Proceedings of 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. New York: IEEE: 993-1000 [DOI: 10.1109/CVPR.2006.91]

Roweis S T and Saul L K. 2000. Nonlinear dimensionality reduction by locally linear embedding. Science, 290(5500): 2323-2326 [DOI: 10.1126/science.290.5500.2323]

Rubinstein M, Joulin A, Kopf J and Liu C. 2013. Unsupervised joint object discovery and segmentation in internet images//Proceedings of 2013 IEEE Conference on Computer Vision and Pattern Recognition. Portland: IEEE: 1939-1946 [DOI: 10.1109/CVPR.2013.253]

Shotton J, Winn J, Rother C and Criminisi A. 2006. *TextonBoost*: joint appearance, shape and context modeling for multi-class object recognition and segmentation//Proceedings of the 9th European Conference on Computer Vision. Graz: Springer: 1-15 [DOI: 10.1007/11744023_1]

Shu C Y, Li P, Xie Y, Qu Y Y, Dai L Q and Ma L Z. 2019. Knowledge squeezed adversarial network compression [EB/OL]. [2020-01-15]. https://arxiv.org/pdf/1904.05100.pdf

Srinivas S and Babu R V. 2015. Data-free parameter pruning for deep neural networks//Proceedings of 2015 British Machine Vision Conference. Swansea: BMVA: 31.1-31.12 [DOI: 10.5244/c.29.31]

Sun J and Ponce J. 2016. Learning dictionary of discriminative part detectors for image categorization and cosegmentation. International Journal of Computer Vision, 120(2): 111-133 [DOI: 10.1007/

s11263-016-0899-0]

Tao Z Q, Liu H F, Fu H Z and Fu Y. 2017. Image cosegmentation via saliency-guided constrained clustering with cosine similarity//Proceedings of the 31 st AAAI Conference on Artificial Intelligence. San Francisco, USA: AAAI: 4285-4291

van de Sande K E A, Uijlings J R R, Gevers T and Smeulders A W M. 2011. Segmentation as selective search for object recognition//Proceedings of 2011 International Conference on Computer Vision. Barcelona, Spain: IEEE: 1879-1886 [DOI: 10.1109/ICCV.2011.6126456]

Wang C, Zhang H, Yang L, Cao X C and Xiong H K. 2017. Multiple semantic matching on augmented *N*-partite graph for object co-segmentation. IEEE Transactions on Image Processing, 26(12): 5825-5839 [DOI: 10.1109/TIP.2017.2750410]

Wold S, Esbensen K and Geladi P. 1987. Principal component analysis. Chemometrics and Intelligent Laboratory Systems, 2(1-3): 37-52 [DOI: 10.1016/0169-7439(87)80084-9]

Yu X Y, Liu T L, Wang X C and Tao D C. 2017. On compressing deep models by low rank and sparse decomposition//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE: 7370-7379 [DOI: 10.1109/CVPR.2017.15]

Yuan Z H, Lu T and Wu Y R. 2017. Deep-dense conditional random fields for object co-segmentation//Proceedings of the 26th International Joint Conference on Artificial Intelligence. Macau, China: AAAI Press: 3371-3377 [DOI: 10.24963/ijcai.2017/471]

Zagoruyko S and Komodakis N. 2016. Paying more attention to attention: improving the performance of convolutional neural networks via attention transfer [EB/OL]. [2020-01-15]. https://arxiv.org/pdf/1612.03928.pdf

Zhang X Y, Zhou X Y, Lin M X and Sun J. 2018. ShuffleNet: an extremely efficient convolutional neural network for mobile devices//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE: 6848-6856 [DOI: 10.1109/cvpr.2018.00716]

Zhang Z, Ning G H and He Z H. 2017. Knowledge projection for deep neural networks [EB/OL]. [2020-01-15]. https://arxiv.org/pdf/1710.09505.pdf

作者简介

耿增民，1968 年生，男，教授，主要研究方向为 Web 信息处理、机器学习、服装数字媒体技术与应用。
E-mail: jsjgzm@bift.edu.cn

余梦巧，女，硕士研究生，主要研究方向为机器学习、计算机视觉。E-mail: yumengqiao@bit.edu.cn

刘峡壁，男，副教授，主要研究方向为机器学习、模式识别、计算机视觉、信息检索。E-mail: liuxiabi@bit.edu.cn

吕超，男，讲师，主要研究方向为服装计算机应用、机器学习、模式识别。E-mail: jsjlc@bitf.edu.cn