深度 EM 胶囊网络全重叠手写数字识别与分离

姚红革 [1] 董泽浩 [1] 喻 钧 [1] 白小军 [1, 2]

摘 要 基于胶囊网络的向量神经元思想和期望最大算法 (Expectation-maximization, EM), 设计了一种以 EM 为向量聚类算法的深度胶囊网络 (Deep capsule network, DCN), 实现了重叠手写数字的识别与分离. 该网络由两部分组成, 第 1 部分是 "识别网络", 将 EM 算法改为 EM 向量聚类算法, 以替换原胶囊网络 CapsNet 中的迭代路由部分, 这一改动优化了网络的运算过程, 实现了重叠数字识别. 第 2 部分是 "重构网络", 由结构完全相同的两个并行网络组成, 对双向量进行并行重构, 实现了重叠数字的分离. 实验结果显示, 对于 100% 全重叠手写数字图片本网络识别率达到了 96%, 对比 CapsNet 在 80% 的重叠率下 95% 的识别率, 本文网络在难度提升的情况下, 识别率有明显提高, 能够将完全重叠的两张手写数字进行图片进行准确地分离.

关键词 深度胶囊网络, 重叠数字识别, 重叠数字分离, EM 向量聚类

引用格式 姚红革, 董泽浩, 喻钧, 白小军. 深度 EM 胶囊网络全重叠手写数字识别与分离. 自动化学报, 2022, **48**(12): 2996−3005

DOI 10.16383/j.aas.c190849

Fully Overlapped Handwritten Number Recognition and Separation Based on Deep EM Capsule Network

YAO Hong-Ge[1] DONG Ze-Hao[1] YU Jun[1] BAI Xiao-Jun[1, 2]

Abstract Based on the idea of vector neuron of capsule network and expectation maximization algorithm (EM), a deep capsule network (DCN) with EM as the vector clustering algorithm is designed to recognize and separate overlapping handwritten digits The network consists of two parts. The first part is "identification network". The EM algorithm is changed to EM vector clustering algorithm to replace the iterative routing part of the original capsule network CapsNet. This change optimizes the network operation process and realizes overlapping number recognition. The second part is the "reconstruction network", which is composed of two parallel networks with identical structure. The bi-vector are reconstructed in parallel to realize the separation of overlapping digits. The experimental results show that for 100% full overlap handwritten digit, the recognition rate of the network reaches 96%. Compared with the 95% recognition rate of CapsNet at 80% overlap rate, the recognition rate of the network in this paper is significantly improved in the case of increased difficulty, and can accurately separate two completely overlapping handwritten digits.

Key words Deep capsule network (DCN), overlapping number recognition, overlapping number separation, EM vector clustering

Citation Yao Hong-Ge, Dong Ze-Hao, Yu Jun, Bai Xiao-Jun. Fully overlapped handwritten number recognition and separation based on deep EM capsule network. *Acta Automatica Sinica*, 2022, **48**(12): 2996−3005

识别并分离高度重合数字对象的问题由 Hinton 等[1] 于 2002 年提出, 多年来也有其他研究者在该领域进行了研究, 如 Goodfellow 等[2] 使用深度卷积网络, Ba 等[3] 使用视觉注意力机制和 Greff 等[4] 使用深度无监督分组进行尝试. 他们均是利用对象形状的先验知识进行分离. 在性能最好的 Ba 等[3] 的研究中虽然实现了 95% 的识别率, 但图片也只是 4% 的重叠率.

直到 Sabour 等[5] 所研究的胶囊网络 CapsNet 面世, 重叠手写体识别成功率才有了大幅提高, 当重叠率 80% 时识别率可达 95%. 胶囊网络的主要特征是, 使用胶囊神经元代替了普通神经元, 使用向量代替了在网络中流通的标量. 胶囊神经元除了承载着网络权值的联系之外, 其向量内部也存在着维度上的联系, 丰富了图像特征的表达与提取能力. 在 CapsNet 中使用了迭代路由算法, 该算法用向量内积来表示向量方向的同向程度, 动态路由通过迭

收稿日期 2019-12-18 录用日期 2020-04-16
Manuscript received December 18, 2019; accepted April 16, 2020
本文责任编委 金连文
Recommended by Associate Editor JIN Lian-Wen
1. 西安工业大学计算机科学与工程学院 西安 710021 2. 电子信息现场勘验应用技术公安部重点实验室 西安 710121
1. College of Computer Science and Engineering, Xi′an Technological University, Xi′an 710021 2. Key Laboratory of Electronic Information Processing with Applications in Crime Scene Investigation, Ministry of Public Security, Xi′an 710121

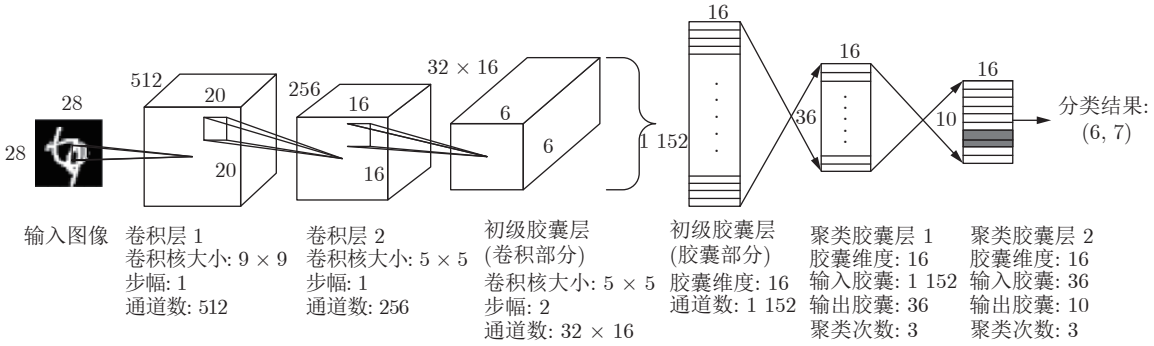
代来实现. CapsNet 将最突出的向量作为分类结果输出, 向量的突出程度跟胶囊内与输出向量方向相近的向量数目和模长正相关. 为避免在使用内积作为衡量手段出现无上界的情况, 对向量进行了输出前的压缩.

CapsNet 的优势是简单易实现, 但在使用它进行全重叠数字图片识别时发现, 由于网络深度宽度不足, 中间向量的规模太小, 同时内积路由算法效率低. 这些因素降低了网络的速度, 影响了网络的聚类效果, 从而使网络对图像特征提取不够充分, 在分类时表现不佳, 导致重构出来的分离图片不够准确和清晰. 为了提高对全重叠手写数字的识别精度, 基于 CapsNet, 本文提出以下改进方法:

1) 首先对胶囊网络 CapNet 进行加深. 在它的 Conv1 层之后加入一层卷积层 “卷积层 2”, 提高目标特征提取能力; 另外在 CapNet 的 DigitCap 之后, 对应本文 “初级胶囊层 (胶囊部分)”之后加入一层全连接胶囊层 “聚类胶囊层 1”, 增加聚类能力以增强网络识别能力, 参见图 1(a).

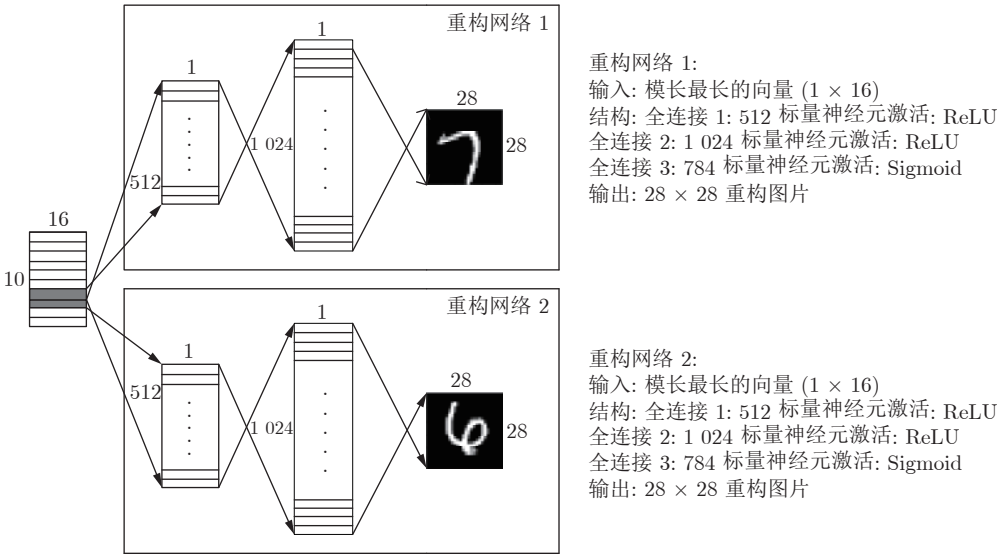
2) 提高胶囊维度为 16 维. 这样使各个胶囊层胶囊统一维度为 16 维, 既提高了胶囊对图片特征表达能力, 减少维度转换时系统消耗和信息的丢失和变异, 也便于各层间信息的传输.

3) 用 EM (Expectation-maximization) 向量聚类取代原路由聚类, 提高聚类效果. 胶囊网络中向量神经元将低级特征预测为高级特征, 输出向量的分布符合以不同高级特征为期望的混合高斯模型[5]. 混合高斯模型是有限混合概率分布模型, 其可用 EM 算法找到最大似然估计[6−7]. 通过假设隐变量的存在, 简化似然函数方程的求解[6−8]. 基于此特点, 本文将 EM 聚类改为 EM 向量聚类, 并用它取代胶囊网络中的迭代路由, 提高了聚类效果. 也减少中间变量的产生, 降低显存以及空间消耗, 总体提高系统的运行效率.



(a) 分类网络结构图

(a) Classification network structure chart



(b) 双并行重构网络结构图

(b) Double-parallel reconstruction network structure diagram

图 1　深度胶囊网络结构图

Fig. 1　Deep capsule network structure diagram

4) 设计了一个并行重构网络. 因为要分离两个重叠的数字, 需要取两个模数最长的向量来进行重构, 因此数字重构网络必须要设计成并行的两个网络结构, 才能对模数最长的两个向量分别并行重构. 依据这一想法, 本文设计了一个双并行重构网络结构, 实现了对两个全重叠手写数字的分离重构, 参见图 1(b).

1 相关工作

胶囊网络的思想最早出现于 Hinton 等[7] 提出的分组神经元. 基于此, Sabour 等[5] 进一步提出胶囊间的动态路由算法, 该算法使胶囊进入了初级应用阶段. 尽管是初级应用, 但它实现了目标属性间的 "等变性" (Equivariance), "等变性"保留有图像各部分信息间的关联. 而在此之前的神经网络只是实现空间不变性, 空间不变性实现的一般方法是卷积神经网络 (Convolutional neural network, CNN) 的池操作, 空间不变性与 "等变性"比较丢失了图像各部分间的关联信息. 它的实现要归因于网络内的动态路由算法, 但动态路由算法优化能力较弱, 于是 Wang 等[9] 通过引入耦合分布 KL (Kullback-Leibler) 散度来优化动态路由, 使胶囊网络性能获得一定的提升. 胶囊网络的又一应用是 CapsGan[10] 网络, 它使用胶囊网络作为生成式对抗网络 (Generative adversarial network, GAN) 中的甄别器, 比 CNN 的 GAN 获得更好的生成效果. 以上方法均是对胶囊网络优化和新领域的应用, 在网络构造上基本没有改变.

LaLonde 等[11] 和 Rajasegaran 等[12] 对胶囊网络结构进行了加深. LaLonde 等通过卷积, 让所有胶囊沿深度方向作为输入进行转换, 包含在较高层的胶囊里. 加深胶囊层必定增加动态路由量, 引起计算复杂度的增加. 为了降低计算复杂度, Rajasegaran 等[12] 采用了如下措施: 在初始阶段减少路由迭代次数; 在路由中间层使用三维卷积, 采用参数共享而减少参数的数目; 同时提出本地化路由代替完全连接的路由. 新加深的胶囊网络具有捕获更细致信息的能力, 增强了它的实际应用能力, 可以处理比 MNIST 数据集更复杂的数据集. 本文对胶囊网络的加深主要体现在前端的特征提取和后端的分离方面, 目的是增强对重叠手写体的识别能力和分离能力.

将 EM 算法应用于胶囊路由也起源于 Hinton 等[13] 的研究之作, 底层胶囊的姿态矩阵通过与转换矩阵相乘而得到高层胶囊的姿态矩阵, 这个过程可以看作是每一个底层胶囊对高层胶囊所表达图像特征的投票. 投票通过分配一个权重系数来实现. 这个系数是由 EM 算法进行循环更新的, 通过 EM 算法系统将底层胶囊的输出路由给高层胶囊. 底层胶囊与高层胶囊的这种联系反应的是图像中实体的整体与部分间关系, 它使胶囊网络具有了对所关注实体的视角不变性. EM 算法在文献 [13] 中是直接应用, 并未改动. 本文依据输入向量的独立性对 EM 算法的 E (Expectation) 步进行了改进, 并依信息熵重新定义了混合度, 优化了胶囊间的迭代, 加速其收敛, 并将其用于手写体数字分离中, 相较于文献 [13] 分离效果有了明显提高.

Mixup[14] 和 Between-Class learning[15] 是两个对类别不同的样本进行重叠的算法, 可以是两个图片按不同混合比的重叠. 其目的是通过丰富训练样本的状态来提高所训练模型的泛化能力. Mixup 和 Between-Class learning 算法说明将不同类别的图片重叠来训练模型能提高模型的分类能力. 这一点与本文方法相同. 但这两种方法目的是分类, 不能将混合的像素按图像本来分离. 本文方法是基于细致识别下的重叠图片的重构分离.

2 DCN 网络

基于胶囊网络 CapsNet, 本文构建了一个以 EM 为向量聚类的深度胶囊网络(Deep capsule network, DCN), 其网络结构如图 1 所示, 由分类网络 (参见图 1(a)) 和重构网络 (参见图 1(b)) 组成.

因为卷积层在神经网络中具有提取多级特征的能力, 而且可以通过卷积核的共享降低运算量. 因此在 DCN 中, 使用了两个卷积层对输入图像的特征进行提取, 其中卷积层 1 使用 512 个 9×9 的卷积核对图像进行卷积在卷积层 2 中使用 256 个 5×5 卷积核进行卷积, 最终得到 $256 \times 16 \times 16$ 的特征图.

然后构建一个初级胶囊层, 其前半部分通过多重卷积获得一组 $32 \times 16 \times 6 \times 6$ 的标量, 由其后半部分的胶囊生成一组由 16 维向量组成的 1 152 个向量神经元, 每个神经元输出一个 16 维的向量. 在每个 6×6 的网格中, 设定权重共享给每一个胶囊, 然后对每个输出向量进行输出.

接下来使用两个聚类胶囊层进行最终的分类, 增加的聚类胶囊层 1 是对初级胶囊中的向量, 通过 EM 向量聚类进行初步筛选, 形成较为高级的有明显倾向性的高级向量给聚类胶囊层 2, 然后再由聚类胶囊层 2 进行第 2 次 EM 向量聚类, 细选出可用于表示不同类别信息的向量. 在每次聚类之后是压缩. 聚类的过程使得高级特征更集中, 压缩的目的是为了限制向量的模长. 模长被限制于 0 ~ 1 之间, 用以表达其所属类别的概率. 再由最后一层产生 10 个 16 维的向量代表 0 ~ 9 的 10 分类结果, 作为输出.

在检测重叠手写数字时, 选取输出模最长的前两个向量作为最可能重叠的结果进行输出. 如果模长第二的向量模长不足 0.1, 就认为是由两个分类相同的数字叠加而成.

重构由重构网络完成, 本文重构网络是由两个结构相同的 3 层全连接网络构成, 详见图 1(b). 重构时选取 “分类网络”输出的模长最长的两个向量, 为避免其余 8 个向量的干扰, 将其全部值置为 “零”. 然后将这 10 个 16 维向量, 首尾接续分别传入两个并行重构网络进行重构.

2.1　姿态变换矩阵

底层胶囊所生成的向量可以认为其代表了某种低级特征, 该低级特征通过姿态变换矩阵可对高级特征进行预测, 这种预测是对向量的方向以及维度的变换, 其表达式为

$$\boldsymbol{U}_{(l+1,\,j)} = \boldsymbol{W}_{(i,\,j)}\boldsymbol{V}_{(l,\,i)} \tag{1}$$

其中, $\boldsymbol{U}_{(l+1,\,j)}$ 表示在第 i 层中第 j 个胶囊的预测向量, 即预测结果. $\boldsymbol{W}_{(i,\,j)}$ 表示由 l 层的 i 胶囊输出到第 $l+1$ 层中第 j 个胶囊特征的姿态变换矩阵. $\boldsymbol{V}_{(l,\,i)}$ 表示第 l 层中第 i 个胶囊的输出向量.

2.2　EM 向量聚类算法

胶囊网络中向量神经元将低级特征预测为高级特征, 输出向量的分布符合以不同高级特征为期望的混合高斯模型[5]. 基于此, 将 EM 聚类改造成为 EM 向量聚类, 用它取代胶囊网络中的迭代路由, 以优化系统, 提高其运行效率.

2.2.1　EM 向量聚类

经过姿态变换方程产生的一组预测向量是符合混合高斯分布的[5], 如式 (2) 所示, 经过多轮迭代获得概率最大的分布函数[6−7], 作为胶囊的输出.

$$p(\boldsymbol{X}) = \sum_j \frac{\alpha_j}{(2\pi)^{\frac{d}{2}}(\det \boldsymbol{\Sigma}_j)^{\frac{1}{2}}} \times \exp\left(-\frac{1}{2}(\boldsymbol{X}-\mu_j)^{\mathrm{T}}\boldsymbol{\Sigma}_j^{-1}(\boldsymbol{X}-\mu_j)\right) \tag{2}$$

其中, j 代表类别, $\boldsymbol{X}$ 为输入向量, α_j 为第 j 类的概率且 $\boldsymbol{\Sigma}_j\alpha_j = 1$, μ_j 为第 j 类的向量期望, $\boldsymbol{\Sigma}_j$ 为协方差矩阵.

因为低级特征来自于输入图像的变换结果, 所产生的向量之间可以认为是相互独立的, 因此协方差矩阵是一个对角阵, 这样就相当于输入 $\boldsymbol{X}$ 在各分量解耦. 所以本文相较于标准 EM 迭代算法进行了改动, 即

$$p(\boldsymbol{X}) = \sum_j \frac{\alpha_j}{\sqrt{2\pi\sigma^d}} \exp\left(-\frac{1}{2\sigma^2}||\boldsymbol{X}-\mu_j||^2\right) \tag{3}$$

因为将输入分布视为混合高斯分布进行聚类, 聚类中心向量是类内向量的加权平均, 无法通过模长来衡量显著性. 所以引入一个标量 a_j 作为缩放尺度来衡量显著性, 并在输出之前代入 asquashing 函数来控制输出向量的模长.

用 EM 聚类结果得到输出高斯分布的方差, 方差越大意味着预测向量分布越接近均匀分布, 说明这个输出胶囊输入的预测结果并不明显接近同一种特征, 此时 a_j 应该小; 方差越小意味着分布越集中, 说明这个输出的输入的预测结果大致相近, 此时 a_j 应该大. 基于这种思想选择使用信息熵 C_j 来辅助 a_j 衡量特征的显著程度[6−7, 16], C_j 表达式可定为

$$C_j = \left(1 + \sum_{l=1}^{d} \ln\sigma_j^l\right) \sum_i r_{ij} \tag{4}$$

其中, σ_j^l 为第 l 层 j 胶囊均方差, r_{ij} 为底层 i 胶囊转变为高层 j 胶囊的模长占比, C_j 表达第 j 个胶囊所表达的特征的显著性. 基于 C_j 得到缩放尺度 a_j 可定为

$$a_j = \mathrm{sigmoid}\,(1 - C_j) \tag{5}$$

当分布的方差越小时 C_j 的值越小, 因此通过最大化 C_j 的方式实现迭代优化. 为防止无上限的情况, 在此采用 sigmoid 激活函数.

2.2.2　算法流程

EM 向量聚类算法的流程如图 2 所示.

在已知 $\boldsymbol{U}_{ij}$, $\boldsymbol{S}_j$, a_j 的情况下, 其中 $\boldsymbol{U}_{ij}$ 表示 l 层的第 i 个胶囊输出经过姿态转换矩阵处理后向 $l+1$ 层的第 j 个高层胶囊输出的向量预测; $\boldsymbol{S}_j$ 表示 $l+1$ 层胶囊的输出方向; σ_j^2 为 $l+1$ 层胶囊的输出方向的方差; a_j 表示其特征的显著程度. EM 向量聚类的具体算法流程如下.

算法 1. EM 向量聚类算法

初始化变量 $\boldsymbol{U}_{ij}$, a_j, $\boldsymbol{S}_j$, σ_j^2

For 低级胶囊 i **to** 高级胶囊 j

E-Step.

$$p_{ij} \leftarrow N\left(\boldsymbol{U}_{ij}; \boldsymbol{S}_j, \sigma_j^2\right);$$

$$R_{ij} \leftarrow \frac{a_j p_{ij}}{\sum_{j=1}^{k} a_j p_{ij}};$$

$$r_{ij} \leftarrow \frac{||\,\boldsymbol{U}_{ij}\,||\,R_{ij}}{\sum_{i=1}^{n} ||\,\boldsymbol{U}_{ij}\,||\,R_{ij}}$$

M-Step-1.

$$S_j \leftarrow \sum_{i=1}^{n} r_{ij}\boldsymbol{U}_{ij}; \sigma_j^2 \leftarrow \sum_{i=1}^{n} r_{ij}(\boldsymbol{U}_{ij} - \boldsymbol{S}_j)^2;$$

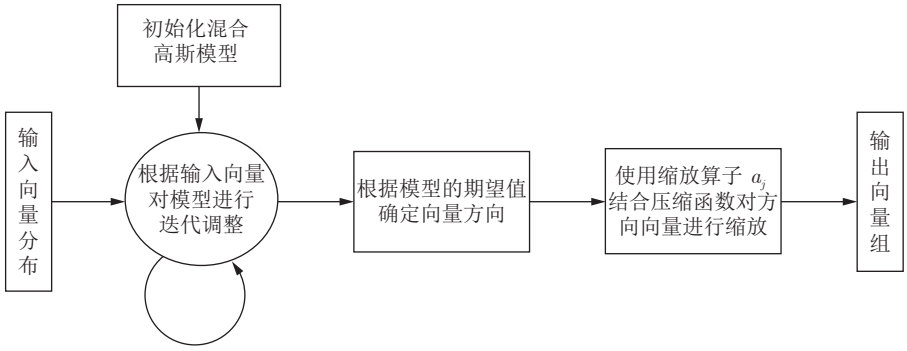


图 2 EM 向量聚类算法流程图

Fig. 2 Flow chart of EM vector clustering algorithm

M-Step-2.

$$C_j \leftarrow \sum_i (1 + \sum_{l=1}^{d} \ln \sigma_j^l) r_{ij};$$

$$a_j \leftarrow \text{sigmoid}(1 - C_j)$$

重复 E-Step, M-Step-1, M-Step-2 共 z 次, z 即聚类次数.

Endfor.

EM 向量聚类算法流程说明如下.

1) 初始化变量 $\boldsymbol{U}_{ij}$, a_j, $\boldsymbol{S}_j$, σ_j^2.

2) 步骤 E-Step. 根据参数 $\boldsymbol{S}_j$, σ_j^2 计算每个输入向量对输出胶囊的先验概率 p_{ij}, 再使用贝叶斯公式与 a_j 得到输出胶囊的后验概率 R_{ij}, 然后依据输入向量的长度对每一个预测向量进行加权平均, 得到一个高级胶囊的加权概率 r_{ij}.

3) 步骤 M-Step-1. 根据计算得到的 r_{ij} 以及预测向量分布 $\boldsymbol{U}_{ij}$, 求出迭代优化后的高斯混合模型参数 $\boldsymbol{S}_j$, σ_j^2.

4) 步骤 M-Step-2. 因为要进行重叠数字检测, 不希望网络发生类间竞争, 所以不用原归一化的先验概率, 而采用基于信息熵的缩放尺度 a_j. 跟据式 (4) 与式 (5), 由 $\boldsymbol{S}_j$, σ_j^2, r_{ij} 得 a_j, 并且可知, a_j 与原 EM 算法中的类别概率 α_j 同收敛, 为了简化运算, 将 a_j 与之替换.

在进行多次迭代之后, 以概率 a_j 作为 j 胶囊的输出尺度, 对输出的方向向量进行缩放, 得到最终的输出向量, 即

$$\boldsymbol{V}_j = \text{asquashing}\,(a_j, \boldsymbol{S}_j) \tag{6}$$

式中, asquashing 为压缩函数, 参见第 2.3 节.

2.3 压缩函数

为了防止胶囊向量在后续运算中无限增长导致网络 "爆炸", 同时又能用其模长表示分类概率, 使用一个非线性函数对这些向量进行压缩, 并使模长维持在 0 ~ 1 之间. 这也在一定程度上抑制了与当前高级特征相关性小的向量. 压缩函数 asquashing 为

$$\boldsymbol{V}_j = \frac{a_j \,||\, \boldsymbol{S}_j||^2}{1 + ||\, \boldsymbol{S}_j||^2} \frac{\boldsymbol{S}_j}{||\, \boldsymbol{S}_j\,||} \tag{7}$$

其中, $\boldsymbol{V}_j$ 表示最终输出向量, $\boldsymbol{S}_j$ 表示在进行压缩之前的原始输出, a_j 是缩放尺度.

2.4 并行重构损失

为了实现对并行重构网络 (见图 1(b)) 的训练, 构建了一个并行重构损失函数 L_{recon}, 通过使用均方误差计算输入图片与输出图片的差来实现, 即

$$L_{\text{recon}} = \delta \sum MSELoss(Image_{\text{GT}}, Image_{\text{recon}}) \tag{8}$$

式中, $Image_{\text{GT}}$ 为叠加前的真值图像, $Image_{\text{recon}}$ 为重构后的结果, δ 为重构损失的缩放倍数. 当重构损失在总损失中占比过大时会导致网络的过拟合, 本文使用 δ 取值为 0.0005 对重构损失进行缩放.

训练时选取模长最大的两个向量, 同时放入两个重构网络进行训练. 将上式的重构误差加入总损失函数中, 参见第 2.5 节, 可以使重构网络与分类网络一起进行共同训练.

2.5 代价函数

因为重叠手写数字识别, 需要进行两分类, 也即需要最后输出的向量中有两个模长较长的向量. 由于是双向量结果, 所以要避免这两个向量间的竞争. 在此选择使用 Margin Loss 作为代价函数, 它适用于双分类, 在不同类识别结果之间不进行竞争, 其具体形式为

$$\begin{aligned} L_{\text{cls}} = &\sum T_k \max(0,\, m^+ - \|v_k\|)^2 + \\ &\lambda\,(1 - T_k)\max(0,\, \|v_k\| - m^-)^2 \end{aligned} \tag{9}$$

式中, L_{cls} 表示 k 个分类的胶囊的分类误差, T_k 表示第 k 分类的标签值.

为防止过优化, 将式 (9) 中的 m^- 和 m^+ 分别设

定为 0.1 与 0.9. 若为正标签, 则式 (9) 的前半部分有效, 希望正标签的胶囊输出的向量的模长 v_k 保持在 0.9 以上; 若为负标签, 则式 (9) 的后半部分有效, 希望负标签的胶囊输出的向量的模长 v_k 保持在 0.1 以下. 最后, 将每个分类的损失函数值进行相加, 再与重构损失 L_{recon} 联合起来形成最终的损失函数值 L_{total}, 即

$$L_{\text{total}} = L_{\text{cls}} + L_{\text{recon}} \tag{10}$$

通过式 (10) 进行分类网络与重构网络的联合训练.

3　实验与数据分析

3.1　数据集

本实验采用的数据集为 3 种: 1）MNIST 原数据集; 2）全重叠数据集; 3）前两种数据集的混合集. 其中第 2 种是由 MNIST 原数据集生成, 生成方式是将 MNIST 数据集的一半 (30 000 幅图像), 与另一半 (30 000 幅图像) 进行叠加生成, 重叠率为 100%, 也即全重叠生成, 叠加后效果如图 3 所示.

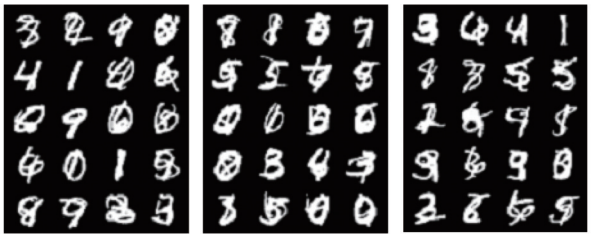


图 3　全重叠数据集

Fig. 3　Full-overlapping dataset

标签是对原 one-hot 标签进行处理后得到的, 如表 1 所示. 若由两个不同数字叠加, 将这两个数字的位置置为 1, 其他位置置为 0; 如果是由相同数字叠加, 将其位置置为 2, 其他位置置为 0.

表 1　数据集标签

Table 1　Dataset label

输入图像	标签	说明
	(0, 0, 0, 0, 0, 0, 0, 1, 0, 0)	无叠加
	(0, 0, 0, 0, 0, 0, 0, 0, 0, 2)	两个相同数字叠加
	(0, 0, 0, 1, 0, 0, 0, 1, 0, 0)	两个不同数字叠加

3.2　EM 向量聚类效果实验

3.2.1　EM 向量聚类模长

输出向量的模长是对分类概率的度量, 模长越长属于该类的概率越高. 它也是聚类效率和效果的反映, 因为聚类是将正确的类别向量进行放大, 提示降低不正确类别向量模长, 所以越快达到高模长, 说明所用聚类形式的效率越高, 效果越好.

在 DCN 结构上分别用 MNIST 数据集、全重叠手写数字数据集以及混合数据集进行训练. 对全重叠图片进行测试, 以测试不同聚类迭代次数 R 下 EM 向量聚类的模长, 见图 4 所示. 以下本文实验不做特别说明时其值均为 3, 并与 CapsNet 路由模长进行对比, 如表 2 所示, 其为分别进行 10 次测量的均值.

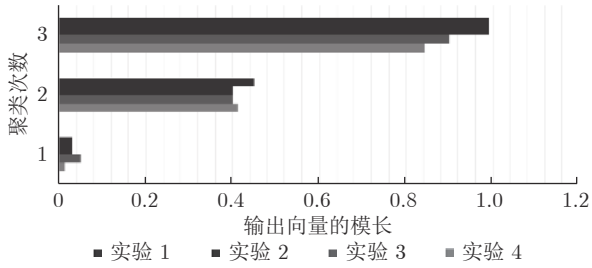


图 4　不同聚类次数下输出向量的模长

Fig. 4　Module length of output vector under different clustering times

从表 2 和图 4 可发现, 对重叠数字识别时, 不同的聚类次数对输出向量的模长有着相当大的影响. 在进行聚类之前, 最长的输出向量只有不到 0.1 的长度, 而在进行了 3 次聚类之后, 正确的向量的长度已经达到了 0.85 以上 (由于压缩函数的存在, 向量的长度不能超过 1). 从表 2 可见, DCN 所用 EM 向量聚类效果, 在第 3 轮聚类时 ($R = 3$), 在 3 个数据集下模长都明显高于 CapsNet 的路由聚类, 说明 EM 向量聚类效果较路由算法更好.

3.2.2　EM 向量聚类速度

在 DCN 中一共进行两次 EM 聚类, 分别在初级胶囊层与聚类胶囊层 1 之间和聚类胶囊层 1 与聚类胶囊层 2 之间, 见图 1(a). 因为聚类是一个无监督过程, 该过程并不对学习参数进行保存, 所以在每一次网络进行聚类时, 都先初始化参数然后多次

表 2　在不同聚类次数下的激活向量模长

Table 2　Active vector module length under different clustering times

网络结构及聚类形式	所用训练集	$R = 1$	$R = 2$	$R = 3$
DCN EM 聚类/CapsNet 路由聚类	MNIST数据集	0.0413/0.0536	0.5241/0.4122	0.9800/0.8792
	全重叠数据集	0.0332/0.0423	0.4342/0.5865	0.9943/0.8653
	混合数据集	0.0323/0.0354	0.4543/0.3252	0.9923/0.9173

迭代. 迭代过程无论在训练还是测试中都会进行, 是整个网络中最耗时的部分. 表 3 是在不同的聚类次数之下网络进行一个 Epoch 所花费的时间 (实验平台是单张 titan XP). 从表 3 可知, 每次聚类中每增加一次迭代, 训练时间都会增加近三分之一 (对比进行一次聚类的网络).

表 3 参数量与不同聚类次数下的单 Epoch 消耗时间 (s)
Table 3 Parameter quantity and single epoch consumption time under different clustering times (s)

网络结构	参数量	聚类算法	$R=1$	$R=2$	$R=3$
CapsNet	8 215 568	迭代路由	150±2	210±2	240±2
DCN	20 128 032	EM	240±2	300±2	340±2

因为 DCN 是原 CapsNet 网络的加深与扩宽, DCN 的参数量达到了原 CapsNet 网络的 2.45 倍 (增加 140%), 所以 DCN 网络相较于 CapsNet 网络能够提取更多细粒度特征, 识别能力更强. 但 DCN 较 CapsNet 的运行时间也增加了 40%, 如表 3 所示.

DCN 较 CapsNet 在增加网络深度与宽度, 从而导致训练参数量增加 140% 的情况下, 对相同训练数据量的训练时间仅增加 40%, 缩短的运行时间可以认为是 EM 向量聚类算法较 CapsNet 迭代路由算法快的时间. 这说明单纯就 DCN 的 EM 向量聚类算法, 与 CapsNet 的向量内积迭代路由算法比较, 在速度上前者有明显优势.

在 DCN 中, 分别用迭代路由和 EM 算法对单 Epoch 消耗时间进行了实验, 结果如表 4 所示. 在相同条件下, 对于不同的迭代次数 R, EM 算法较迭代路由算法消耗时间减少约 30% ~ 40%.

表 4 DCN 不同聚类算法单 Epoch 消耗时间 (s)
Table 4 Single epoch consumption time of different DCN clustering algorithms (s)

聚类算法	$R=1$	$R=2$	$R=3$
迭代路由	350±2	410±2	440±2
EM	240±2	300±2	340±2

3.3 DCN 识别与分离

3.3.1 不同数据集上的识别率及对比

为了检测 DCN 对全重叠手写数字数据集的识别率, 用 MNIST 数据集、全重叠手写数字数据集和这两种混合数据集训练, 对得到的网络模型进行对比实验. 设定了两组实验, 分别对无重叠的字体识别以及对全重叠字体进行识别.

由表 5 可知, DCN 使用 MNIST 与全重叠数据集混合训练得到的网络不仅在重叠目标识别任务上取得了 96.55% 的正确率, 在无重叠的识别上的正确率也提高到了 95.7%.

表 5 DCN 识别手写数字效果对比 (%)
Table 5 Effect comparison of handwritten digits recognized by DCN (%)

所用训练集	无重叠手写数字识别率	全重叠手写数字识别率
MNIST 数据集	99.6	55.2
全重叠手写数字数据集	80.7	96.75
混合数据集	95.7	96.55

值得注意的是, 使用 MNIST 数据集训练的 DCN 模型在全重叠的识别任务上得到了 55.2% 的正确率. 尽管识别率不高, 但这是在简单的数据集上进行训练而对复杂数据集的识别结果. 一定程度反映了 DCN 网络的特征提取, 以及运用所提取低级特征对高级特征进行有效预测的能力.

同时, 使用重叠手写数字数据集进行训练的 DCN 模型, 在进行无重叠识别时, 取得了 80% 的识别率. 这表明在不进行特别的训练集设计时, DCN 网络可以在使用重叠图片进行训练后, 对不重叠的图片进行识别, 即在特征有区别的情况下, 也能保证一定准确度的识别率.

DCN 模型对于全重叠手写数字测试集 5 000 个测试样本的总体识别率达到了 96.75%, 其识别准确率与 loss 值的变化曲线如图 5 所示.

由图 5 可知, 在不到 20 个 Epoch 下测试准确率达到 96% 以上, 损失由 1 开始缩小至低于 0.02,

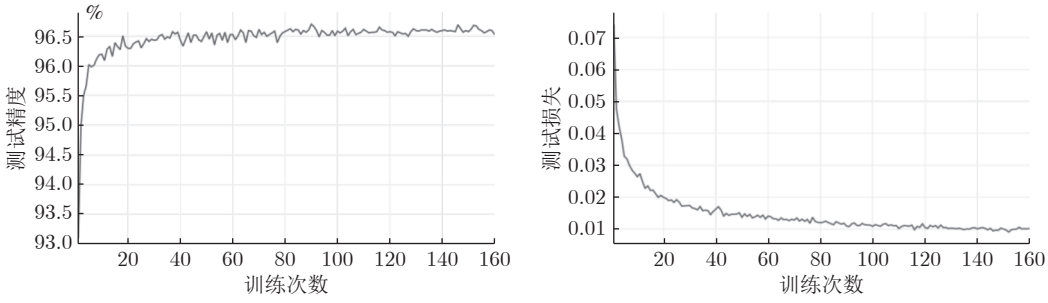


图 5 DCN 对全重叠手写数字的识别率与损失函数值曲线
Fig. 5 Recognition rate and loss value curve of DCN for fully overlapped handwritten digits

且其后没有反复, 说明 DCN “识别网络”运行收敛快且平稳, 能够较好地将重叠的数字进行分类识别.

与 CapsNet 进行对比, CapsNet 在 80% 重叠率的 MutiMNIST 数据集上取得了 95% 的正确率[5], 在全重叠数据集中正确率只有 88%. 本文 DCN 网络结果在全重叠识别正确率达到了 96.75%, 见表 6 所示. 与 CapsNet 对比在全重叠的情况下, DCN 识别准确率高于 CapsNet.

表 6　重叠手写数字识别率对比 ($R = 3$) (%)

Table 6　Comparison of recognition rate of overlapping handwritten digits ($R = 3$) (%)

网络模型	训练集	重叠率	正确率
CapsNet	MutiMNIST	80	95
	全重叠数据集	100	88
DCN	全重叠数据集	100	96.75

3.3.2　分离效果

本文构建的 DCN 重构网络见图 1(b). 其为对分类网络 (图 1(a)) 的输出向量进行重构, 得到分离图片. 在分离训练过程中, 使用掩蔽的方法只把正确的数字胶囊的激活向量保留下来. 然后用两个激活向量通过两个并行的重构网络进行重构, 最终输出两幅 28×28 像素的灰度图片, 显示已经分离的手写体数字的分离效果.

重构时的重构误差是通过计算重构图片的像素亮度以及与叠加之前的图片的像素亮度进行对比, 然后加和得到, 参见式 (9). 把得到的此重构误差按一定的占比放入到总误差中, 参见式 (11), 然后对全网络进行统一训练, 进而得到重构图片.

图 6 显示了在不同缩放数量级的情况下, 总损失函数值 L_{total} 的变化情况. 由图 6(a) 中总损失函数值 L_{total} 升高的情况可以得知, 在重构误差占比大于 0.005 时网络出现了过拟合的情况. 重构损失占比过大抑制了分类的损失 L_{cls}, 导致分类效果的下降. 通过反复试调, 将重构损失 L_{recon} 占比降低至 0.0005 时, 重构损失才不会在训练过程中抑制 L_{cls} 的作用, 得到的总损失 L_{total} 曲线收敛迅速, 在 20 Epoch 时 L_{total} 值下降到了 0.02, 而且下降平稳, 没有反复, 见图 6(b)所示.

图 7 为分离结果, 图 7(a) 为 100 个待分离重叠数字图片, 图 7(b) 和图 7(c) 为分类网络识别后由重构网络所重构的分离结果.

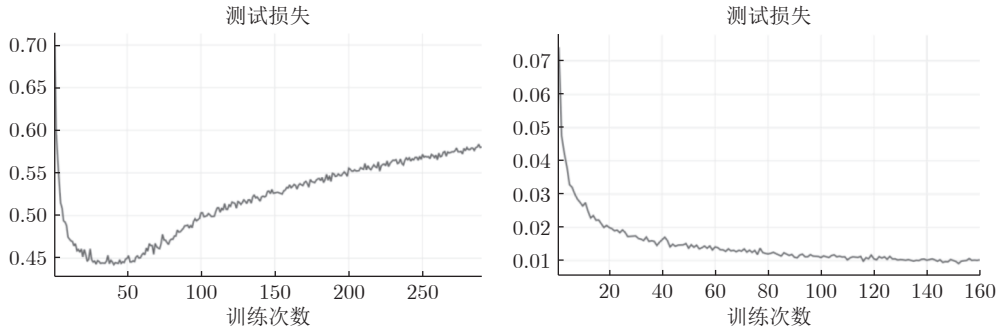
表 7 显示的是 8 个重叠图片的分离情况, 其中标注 “$*$”的 3 组数字 “7”与 “9”的组合中, 3 幅重叠图片均由相同数字不同写法的图片叠加而成, 在进行准确分类之后, 得到的重构结果与原本的数字一样, 这说明整个网络对重叠数字分离准确, 尽管这 3 组数字笔画有些许区别, 但网络进行了准确的识别与重构. (3, 7), (9, 1), (0, 8), (0, 4) 这 4 幅图片, 重叠后图形复杂, “识别网络”识别准确, “分离网络”分离后字体笔划基本清晰.

卷积在标注 “$\bullet$”的数字 “5”与 “9”的组合中, 原图叠加后的特征复杂, 网络分类出现错误. 由重构结果可以得知, DCN 网络依旧将数字 “9”完整地区分出来, 但是将另一个数字 “5”识别成了数字 “8”. 说明网络对于极复杂的图片的识别不够理想, 需要进一步提高.

3.4　对全重叠手写汉字的测试

用 DCN 对 CASIA 汉字手写图片集中的 “不”、“下”、“丑”、“世”、“专”、“王”、“也”、“卫”、“大”、“人”10 个汉字进行全重叠测试. 共进行了 150 个 Epoch 训练, 训练的平均识别率为 92.7%, 如见图 8 所示.

表 8 是部分识别和分离结果. 从中可以看出, 对图片清晰、字体简单的汉字, 识别结果准确, 分离基本清晰. 但对于字体复杂不规整的汉字, 重叠图片识别率低, 如最后两个标签 (王, 丑)、(也, 卫) 识别错误, 分离结果模糊.



(a) 重构误差占比 0.005
(a) Reconstruction error ratio 0.005

(b) 重构误差占比 0.0005
(b) Reconstruction error ratio 0.0005

图 6　重构 loss 函数占比收敛对比

Fig. 6　Comparison of proportion convergence of reconstructed loss function

(a) 输入图片
(a) Original picture

(b) 重构的分离结果 1
(b) Separation result of reconstructing 1

(c) 重构的分离结果 2
(c) Separation result of reconstructing 2

图 7 重构结果

Fig. 7 Reconstructing results

表 7 全重叠手写数字分类与重构的部分结果

Table 7 Partial results of classification and reconstruction of fully overlapped handwritten digits

分类标签	(3, 7)	(9, 1)	(0, 8)	(0, 4)	(9, 7)*	(7, 9)*	(7, 9)*	(5, 9)•
分类结果	(3, 7)	(9, 1)	(8, 0)	(0, 4)	(7, 9)*	(7, 9)*	(7, 9)*	(8, 9)•
输入图片								
重构图片 1								
重构图片 2								

表 8 部分识别和分离结果

Table 8 Partial identification and separation results

分类标签	(不, 专)	(下, 不)	(丑, 下)	(不, 丑)	(下, 世)	(下, 专)	(王, 丑)	(也, 卫)
分类结果	(不, 专)	(下, 不)	(丑, 下)	(不, 丑)	(下, 世)	(下, 专)	(丑, 不能确定)	(不能确定, 不能确定)
输入图片								
重构图片 1								
重构图片 2								

图 8 训练识别率

Fig. 8 Training recognition rate

对重叠汉字测试, DCN 在所选汉字较简单情况下测试识别误差为 15.2%, 相较 MNIST 手写数字重叠识别误差较高. 对于简单图片分离基本清晰, 复杂图片识别与分离误差较高.

4 结束语

本文设计了一种深度胶囊网络模型 DCN, 它具有 6 层网络结构, 使用向量维数为 16 维, 用 EM 的向量聚类算法代替了原路由算法. 同时构建了一个并行重构网络, 以实现对重叠目标的分离重构. 最后用不同的聚类次数与训练集对重叠手写体数字进行了识别实验, 结果显示 DCN 网络对全重叠手写数字识别率达到 96%, 超过了胶囊网络 CapsNet 在 80% 重叠率下识别率 95%, 分离重构图片的效果较好. 但是 DCN 对重叠数字的重构效果还未达到

理想效果, 重构目标还是有一定比例的模糊和近 4% 的识别错误问题, 这将在后期工作中进行完善. 后期工作也将进一步提高该方法应用于重叠手写汉字的识别.

References

1 Hinton G E, Ghahramani Z, Teh Y W. Learning to parse images. In: Proceedings of the 12th International Conference on Neural Information Processing Systems. Denver, USA: MIT Press, 1999. 463−469

2 Goodfellow I J, Bulatov Y, Ibarz J, Arnoud S, Shet V B. Multi-digit number recognition from street view imagery using deep convolutional neural networks. In: Proceedings of the 2nd International Conference on Learning Representations. Banff, Canada: ICLR, 2014.

3 Ba J, Mnih V, Kavukcuoglu K. Multiple object recognition with visual attention. In: Proceedings of the 3rd International Conference on Learning Representations. San Diego, USA: ICLR, 2015.

4 Greff K, Rasmus A, Berglund M, Hao T H, Schmidhuber J, Valpola H. Tagger: Deep unsupervised perceptual grouping. In: Proceedings of the 30th International Conference on Neural Information Processing Systems. Barcelona, Spain: Curran Associates Inc., 2016. 4491−4499

5 Sabour S, Frosst N, Hinton G E. Dynamic routing between capsules. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, USA: Curran Associates Inc., 2017. 3859−3869

6 Gupta M R, Chen Y H. Theory and use of the EM algorithm. *Foundations and Trends in Signal Processing*, 2011, **4**(3): 223−296

7 Xuan G R, Zhang W, Chai P Q. EM algorithms of Gaussian mixture model and hidden Markov model. In: Proceedings of the 2001 International Conference on Image Processing (Cat. No.01CH37205). Thessaloniki, Greece: IEEE, 2001. 145−148

8 Jain A K, Dubes R C. *Algorithms for Clustering Data*. Prentice-Hall Inc., 1988.

9 Wang D L, Liu Q. An optimization view on dynamic routing between capsules. In: Proceedings of the 6th International Conference on Learning Representations. Vancouver, Canada: ICLR, 2018.

10 Jaiswal A, AbdAlmageed W, Wu Y, Natarajan P. CapsuleGAN: Generative adversarial capsule network. In: Proceedings of the 2018 European Conference on Computer Vision. Munich, Germany: Springer, 2018. 526−535

11 LaLonde R, Bagci U. Capsules for object segmentation. arXiv Preprint arXiv:1804.04241, 2018.

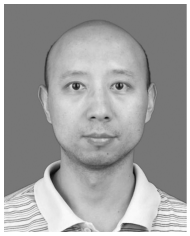
12 Rajasegaran J, Jayasundara V, Jayasekara S, Jayasekara H, Seneviratne S, Rodrigo R. DeepCaps: Going deeper with capsule networks. In: Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE, 2019. 10717−10725

13 Hinton G E, Sabour S, Frosst N. Matrix capsules with EM routing. In: Proceedings of the 6th International Conference on Learning Representations. Vancouver, Canada: ICLR, 2018.

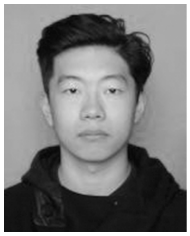
14 Zhang H Y, Cissé M, Dauphin Y N, Lopez-Paz D. mixup: Beyond empirical risk minimization. In: Proceedings of the 6th International Conference on Learning Representations. Vancouver, Canada: ICLR, 2018.

15 Tokozume Y, Ushiku Y, Harada T. Between-class learning for image classification. In: Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE, 2018. 5486−5494

16 Zhu Zhou-Hua. EM algorithm and its application in mixture of Gaussian. *Modern Electronics Technique*, 2003, **26**(24): 88−90 (朱周华. 期望最大(EM)算法及其在混合高斯模型中的应用. 现代电子技术, 2003, **26**(24): 88−90)



姚红革 博士, 西安工业大学计算机科学与工程学院副教授. 主要研究方向为机器学习, 计算机视觉.
E-mail: yaohongge@xatu.edu.cn
(**YAO Hong-Ge** Ph.D., associate professor at the School of Computer Science and Engineering, Xi′an Technological University. His research interest covers machine learning and computer vision.)



董泽浩 西安工业大学计算机科学与工程学院硕士研究生.主要研究方向为深度学习, 胶囊网络.
E-mail: axxddzh@gmail.com
(**DONG Ze-Hao** Master student at the School of Computer Science and Engineering, Xi′an Technological University. His research interest covers deep learning and capsule network.)

喻 钧 西安工业大学计算机科学与工程学院教授. 主要研究方向为图像处理, 模式识别.
E-mail: yujun@xatu.edu.cn
(**YU Jun** Professor at the School of Computer Science and Engineering, Xi′an Technology University. Her research interest covers image processing and pattern recognition.)

白小军 西安工业大学计算机科学与工程学院副教授, 电子信息现场勘验应用技术公安部重点实验室研究员. 主要研究方向为数字图像处理, 人工智能与机器学习. 本文通信作者.
E-mail: baixiaojun@xatu.edu.cn
(**BAI Xiao-Jun** Associate professor at the School of Computer Science and Engineering, Xi′an Technological University, and also a researcher of the Key Laboratory of Electronic Information Processing with Applications in Crime Scene Investigation, Ministry of Public Security. His research interest covers digital image processing, artificial intelligence, and machine learning. Corresponding author of this paper.)