

# 多分支结构强化表征能力的 CapsNet 方法<sup>①</sup>



谢海闻, 叶东毅, 陈昭炯

(福州大学 数学与计算机科学学院, 福州 350108)

通讯作者: 叶东毅, E-mail: yiedy@fzu.edu.cn

**摘要:** CapsNet 是一种新的目标识别模型, 通过动态路由和 capsule 识别已知目标的新形态. 针对 CapsNet 的解码器输入层规模随类别数增加而增加, 可延展性较弱的问题, 本文提出多分支自编码器模型. 该模型将各个类别的编码分别传递给解码器, 使解码器规模独立于类别数, 增强了模型的可延展性. 针对单类别图像训练多类别图像识别任务, 本文增加新的优化目标降低非标签类别的编码向量对解码器的激励, 强化了模型的表征能力. MNIST 数据集的实验结果表明, 多分支自编码器具有良好的识别能力且重构能力明显优于 CapsNet, 因而具有更全面的表征能力.

**关键词:** 目标识别; 目标重构; CapsNet; 表征提取; MNIST

引用格式: 谢海闻, 叶东毅, 陈昭炯. 多分支结构强化表征能力的 CapsNet 方法. 计算机系统应用, 2019, 28(3): 111-117. <http://www.c-s-a.org.cn/1003-3254/6815.html>

## Multi-Branches CapsNet Method with Enhanced Representation Capability

XIE Hai-Wen, YE Dong-Yi, CHEN Zhao-Jiong

(College of Mathematics and Computer Science, Fuzhou University, Fuzhou 350108, China)

**Abstract:** A novel neural network for object recognition, CapsNet, uses dynamic routing and capsules to recognize novel state of a known object, while the input layer of CapsNet decoder increases when the number of categories increases, which means a relatively limited scalability. To overcome this weakness, we propose the Multi-branches Auto-Encoder (MAE) which gives coding vectors of every class to the decoder respectively letting the scale of decoder independent from the number of categories enhancing the representation capability of the proposed model. The experiment on MNIST shows that MAE is competitive in recognition and more powerful in reconstruction which means a more complete capability on representation.

**Key words:** object recognition; object reconstruction; CapsNet; representation extracting; MNIST

## 1 引言

卷积神经网络 (Convolutional Neural Network, CNN) 是一类用于处理图像问题的神经网络模型, 最初用于字符识别问题<sup>[1,2]</sup>, 而后扩展到自然图像分类<sup>[3-5]</sup>、目标检测<sup>[6-9]</sup>等更复杂的问题. CNN 的主要结构是卷积层和池化层, 其中卷积层通过卷积得到一定数量的特征图 (feature map), 池化层下采样前一层特征图以获

得具有空间不变性的特征<sup>[10]</sup>. CNN 通过“卷积-池化”获取空间不变特征时很好地捕获了目标实体的共性, 但与此同时目标实体的个性被逐步舍弃.

Sabour、Frosst 和 Hinton<sup>[11]</sup>提出 CapsNet 用于目标识别任务. CapsNet 使用 capsule 表示一个特定的实体 (entity), capsule 包含一个编码向量, 向量的模表示实体的存在概率, 即活跃度, 向量的方向表示实体的存

① 基金项目: 国家自然科学基金 (61672158); 福建省自然科学基金 (2018J1798); 福建省高校产学研合作项目 (2018H6010)

Foundation item: National Natural Science Foundation of China (61672158); Natural Science Foundation of Fujian Province (2018J1798); Industry-University Cooperation Project of Fujian Province (2018H6010)

收稿时间: 2018-09-18; 修改时间: 2018-10-12; 采用时间: 2018-10-18; csa 在线出版时间: 2019-02-22

在形态(个性). 这里所说的实体可以是一个完整的目标, 也可以是目标的一部分. CapsNet 通过动态路由(dynamic routing)的方式实现一致性路由(routing-by-agreement)取代 CNN 使用的最大池化(max-pooling), 使得只有活跃度高的 capsule 受到高层的关注, 避免最大池化引起的位置信息缺失. 具体而言, 在低层级 capsule 中位置信息是不同区域 capsule 的活跃度, 动态路由能够将位置信息转移到高层级 capsule 并以编码向量的码字表示, 因此 CapsNet 在高层级仍然保持了位置信息. 而且动态路由使得 CapsNet 在不增强或变形数据的情况下, 能够识别已知目标的新形态. 在输出层, CapsNet 输出各个类别的编码向量, 级联全部类别的编码向量, 然后屏蔽(mask out)低活跃度类别的编码向量得到输入图像的编码, 即提取了输入图像中目标的表征.

为了促使表征能够更好地表示输入图像中的目标, CapsNet 在训练过程中加入重构损失. 重构损失由重构图像和原始图像的  $L^2$  范数 ( $L^2$  norm) 刻画. 重构图像由解码器根据表得到. 而且文献[11]的实验结果表明加入重构损失提升了 CapsNet 的目标识别效果.

本文从新的角度思考该模型. 将 CapsNet 视作编码器, 与文献[11]重构图像的解码器一同构成一种可以完成目标识别和目标重构的自编码器模型. 我们注意到 CapsNet 解码器输入层的规模随类别数量的增加而增加, 可延展性较差.

针对上述问题, 本文提出多分支自编码器. 结构上, 多分支自编码器的表征是一个二维矩阵, 不屏蔽任何编码向量; 解码器的输入是单一类别的编码向量而非全部类别编码向量的级联, 使解码器规模独立于类别

数量, 增强模型解码器的可延展性. 训练方式上, 加入新的优化目标, 使非标签类别编码向量的重构图像接近全零图像. 这种训练方式减少低活跃度类别的编码向量中所包含的噪声, 提升了正确类别编码向量的表征能力, 使重构目标更加精确. MNIST<sup>[1]</sup>数据集的实验结果表明, 本文模型具有良好的识别能力且重构能力明显优于 CapsNet, 因而具有更全面的表征能力.

## 2 问题描述

设一组单目标图像集合  $\mathbf{X} = \{x_1, x_2, \dots\}$ , 其对应的类别  $\mathbf{C}_X = \{c_{x_1}, c_{x_2}, \dots\}$ , 所有的类别构成集合  $\mathbf{C} = \{c_1, c_2, \dots\}$ . 现有一组复合图像  $\mathbf{Z} = \{z_1, z_2, \dots\}$ , 其对应的类别是  $\mathbf{C}_Z = \{C_{z_1}, C_{z_2}, \dots\}$ , 其中  $C_{z_i}$  是  $\mathbf{C}$  中元素的集合.

本文研究目标是以单目标图像集合  $\mathbf{X}$  和对应的类别  $\mathbf{C}_X$  作为训练数据, 求得多分支自编码器的最优参数, 使得编码器的类别估计  $\hat{\mathbf{C}}_Z$  与真实类别  $\mathbf{C}_Z$  一致并且使得重构的复合图像  $\hat{\mathbf{Z}}$  与原始复合图像  $\mathbf{Z}$  一致.

## 3 CapsNet 回顾

CapsNet 是用于目标识别的神经网络模型, 其结构如图 1 所示. 该模型首先利用一组卷积操作提取原始图像的初级特征, 然后利用编码层将初级特征转化为各个类别的编码向量. 其中, 编码层首先将卷积得到的初级特征整理成低层级 capsule, 然后利用动态路由的方式推断出高层级 capsule. 每个高层级 capsule 对应一个目标类别, 其编码向量是对应类别的编码向量. 类别的存在概率是类别的编码向量的模, 即高级实体的活跃度. 根据类别的存在概率判断原始图像包含的目标类别.

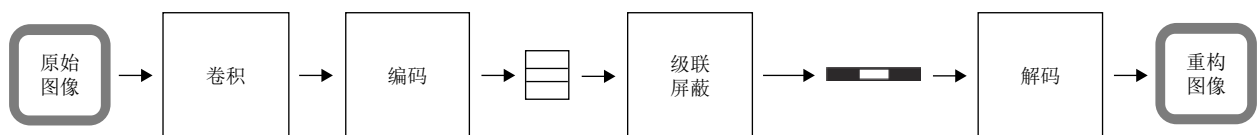


图 1 CapsNet 及其相应的解码器

动态路由是一种确定低层级 capsule 与高层级 capsule 耦合系数(coupling coefficient)的一种迭代算法. 这种算法使得与高层级 capsule 方向接近的低层级 capsule 具有更高的耦合系数. 相比 CNN 使用的最大池化, 动态路由更好地保持了位置信息. 低层级以 capsule 的空间活跃度表示的位置信息, 由动态路由传递到高

层级 capsule 编码向量的码字. 同时, 以 capsule 表征实体相比以量表表征实体而言, 能够更准确地识别出新形态呈现的已知目标.

当原始图像仅有一个类别时, 类别的编码向量经过级联后, 屏蔽活跃度低的编码向量, 仅保留活跃度最高的一个编码向量, 得到原始图像的编码. 编码经过解

码器得到原始图像的重构图像。

当原始图像包含多个类别时, 根据活跃度保留高活跃度的编码向量得到多个编码, 每个编码对应原始图像中的一个类别. 解码器分别解码每个类别的编码, 得到对应类别的重构图像, 即类别重构图像. 叠加这些类别重构图像, 得到原始图像的重构图像.

上述对编码的处理方式使编码变得稀疏, 可以用共同的解码器重构所有类别的编码. 但是这种级联的处理方式, 导致模型的可延展性较差, 解码器输入层的规模随类别数的增加而增加. 当新增类别时, 无法利用已有的解码器, 无法将已有的解码器参数作为初始化信息.

#### 4 多分支自编码器

本文提出的多分支自编码器是用于目标识别和目标重构的神经网络模型, 该模型改进了 CapsNet 的结

构和训练方式, 克服了上节提到的不足, 其结构如图 2 所示. 多分支自编码器在编码部分与 CapsNet 相似, 编码层输出各个类别的编码向量. 与 CapsNet 不同的是, 本文模型分别处理每个类别的编码向量, 而不考虑其活跃度. 每个类别的编码向量分别激励解码器产生各自类别的重构图像. 所有类别重构图像叠加起来获得完整的重构图像. 解码器的输入端神经元数量独立于类别数, 即解码器的输入层规模不随类别的增加而增加, 使得模型具有更好的可延展性. 下面举例说明, 假设  $n$  类编码向量的长度分别为  $a$ , 则 CapsNet 的解码器输入层需要有  $n \times a$  个神经元, 而本文模型的解码器输入层需要  $a$  个神经元. 当类别数增加  $n'$  类时, CapsNet 解码器输入层需要  $(n+n') \times a$  个神经元, 输入层的规模增加; 而本文模型的解码器输入层仍只需要  $a$  个神经元, 输入层的规模不变.

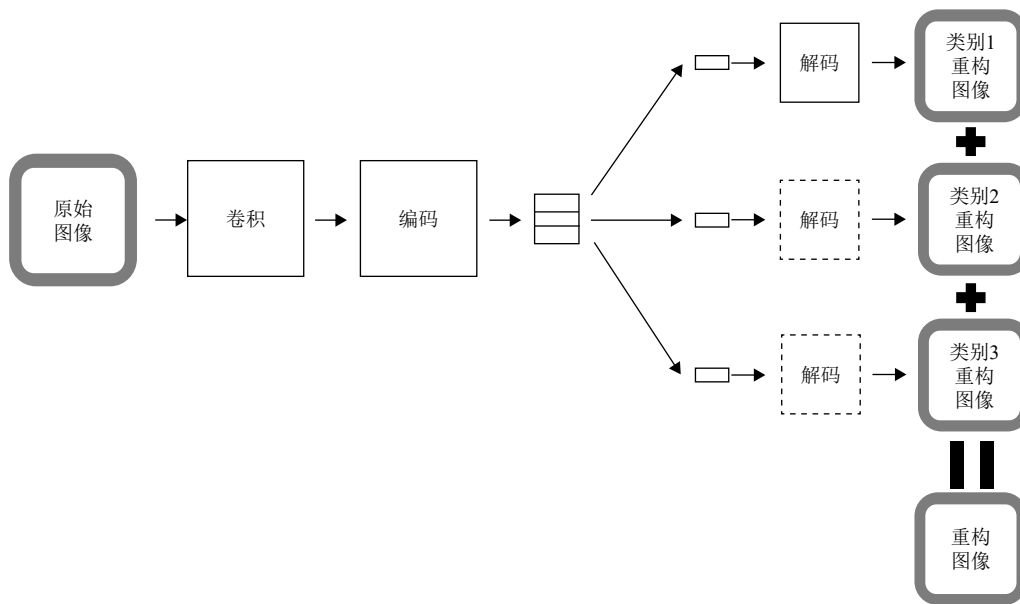


图2 多分支自编码器流程

Capsule 和动态路由机制, 已经能够将原始图像的各类信息蕴含到类别编码向量中, 包括目标位置、形态和目标的类别信息. 即使不采用屏蔽作为编码稀疏化策略, 也能激励解码器产生相应类别的重构图像.

本文模型采用解卷积网络替换文献[11]采用的全连接网络作为解码器. 因为数字图像是二维空间中像素的集合, 像素点之间存在二维空间上的关系, 所以解卷积网络相比全连接网络, 更好地利用空间关系, 能够更精确地重构图像.

我们采用端到端的方式 (end-to-end) 训练网络模

型. 为了更清晰地表述, 我们假设对于  $\forall x \in \mathbf{X}$  多分支自编码器生成一组编码向量  $\{r_1, r_2, \dots\}$ .

为了实现目标识别, 我们引入边际损失 (margin loss), 促使编码层输出的类别编码向量的活跃度 (编码向量的模)  $\{\|r_1\|, \|r_2\|, \dots\}$ , 能够反映相应类别在原始图像中的存在概率:

$$L_m = \sum_k T_k \max(0, m^+ - \|r_k\|)^2 + \lambda(1 - T_k) \max(0, \|r_k\|, -m^-)^2$$

其中, 当第  $k$  个类别存在时,  $T_k = 1$ , 反之  $T_k = 0$ , 概率上

限 $m^+ = 0.9$ , 概率下限 $m^- = 0.1$ .  $\lambda = 0.5$ 防止最初的训练过程抑制所有类别的活跃度<sup>[11]</sup>.

为了实现按类别重构, 标签对应的编码向量(活跃的编码向量)的重构目标是原始图像, 其余编码向量(不活跃的编码向量)的重构目标是全零图像. 我们引入 $L^2$ 范数度量重构误差:

$$L_r = \sum_k T_k \|\hat{x}_k - x\|_2^2 + (1 - T_k) \|\hat{x}_k - \mathbf{0}\|_2^2$$

其中, 当第 $k$ 个类别存在时,  $T_k = 1$ , 反之 $T_k = 0$ .  $\hat{x}_k$ 是第 $k$ 个类别的重构图像,  $x$ 是原始图像. 相比 CapsNet 的训练方式, 本文的训练方式不仅指导解码器将活跃的编码向量映射到原始图像, 而且进一步指导解码器将不活跃的编码向量映射到全零图像. 因此, 从理论上说, 如果能够获得极小化上述损失函数的最优解, 则可以促使活跃的编码向量激励解码器产生与原始图像一致的重构图像, 而促使不活跃的编码向量映射到全零图像, 亦即抑制不活跃的编码向量激励解码器产生任何形式的图形. 所以我们采用的训练方式更加全面, 因而本文模型提取的编码具有更强的表征能力, 能够更精确地重构原始图像. 而且编码向量表征能力的增强, 使得编码向量能够更准确地反映相应类别目标的存在形式, 提升模型的目标识别效果.

为了同时优化两类目标, 取两类误差的加权和作为整体的优化目标

$$\min L = L_m + \alpha L_r \quad (1)$$

其中,  $\alpha$ 是权值系数, 用于调节分类损失和重构损失的重要性.

## 5 基于 Euclidean 距离的不定项分类问题阈值确定方法

因为每个输入图像所包含的目标类别数是不确定的, 所以无法使用排序取最大若干项的方法判断输入图像包含的目标类别. 又因为实验中输入图像类别数小于类别总数, 所以对于多数不存在的类别而言, 存在的类别是相对少数. 因此, 考虑采用离群点检测的方法实现类别估计.

记编码器输入类别存在概率为 $p_i (i = 1 \sim n)$ , 则概率均值为 $m_p = \mathbb{E}[p_i]$ . 记各个类别存在概率与概率均值 $m_p$ 的距离为:

$$d_i = \|p_i - m_p\|_2 \quad (2)$$

则距离均值为:

$$m_d = \mathbb{E}[d_i] \quad (3)$$

判决阈值为:

$$Thread = m_p + m_d \quad (4)$$

当 $p_i > Thread$ 时, 第 $i$ 类是离群点, 认为输入图像中存在第 $i$ 类; 反之, 第 $i$ 类不是离群点, 认为输入图像中不存在第 $i$ 类.

## 6 实验结果与分析

### 6.1 实验设置与评价指标

实验采用主流的深度学习框架 Tensorflow<sup>[12]</sup>构造多分支自编码器. 训练使用的优化器(optimizer)是 Adam 优化器<sup>[13]</sup>, 优化器的参数保持默认设置. 模型参数初始化采用 Xiao 等人<sup>[14]</sup>提出的正交初始化方法. 为了使实验结果客观公平, 实验中 CapsNet 和多分支自编码器的编码器部分采用相同的结构.

为了评价模型的目标识别能力, 实验采用查准率( $P$ )、查全率( $R$ )和  $F1$ -score 作为客观评价指标. 查全率评价模型检出部件的比例; 查准率评价模型检出部件中属于真实部件的比例;  $F1$ -score 是查全率和查准率的调和均值,  $F1$ -score 高意味着模型既不倾向于将负类识别为正类以获得更高的查全率, 也不倾向于将使用过分严格的阈值以获得更高的查准率. 设测试样例数 $N$ ,  $P_i$ 是测试样例的类别估计,  $Y_i$ 是测试样例的真实类别, 则查准率、查全率和  $F1$ -score 的计算方式如下:

$$P = \frac{1}{N} \sum_i \frac{|Y_i \cap P_i|}{|P_i|} \quad (5)$$

$$R = \frac{1}{N} \sum_i \frac{|Y_i \cap P_i|}{|Y_i|} \quad (6)$$

$$F1 = \frac{2PR}{P+R} \quad (7)$$

为直观展现估计和真实标签之间的重合度, 采用交并比 (Intersection-Over-Union, IOU) 展现模型识别的准确率:

$$IOU = \frac{1}{N} \sum_i \frac{|Y_i \cap P_i|}{|Y_i \cup P_i|} \quad (8)$$

模型的目标重构能力可以通过直接观察进行比较, 此外, 实验采用峰值信噪比 (Peak Signal Noise Ratio, PSNR) 作为评价指标, 更加客观地评价模型的重构能力.

$$PSNR = 10 \log \frac{MAX^2}{MSE} \quad (9)$$

其中,  $MAX$  是图像像素的最大值,  $MSE$  是重构图像 $\hat{Z}$ 与

原图像Z的均方误差

### 6.2 MNIST 实验结果分析

训练数据是 MNIST<sup>[11]</sup> 的训练集, 其包含六万幅有数字类别标记的图像, 其每幅图像包含 1 类数字. 原始的 MNIST 图像尺寸是 28×28, 我们将 28×28 的数字随机投入 64×64 的黑色背景中. 测试数据由 MNIST 数据集的测试集制作. 测试数据是多类别的图像, 由 MNIST 的测试集制成. 测试图像是在 64×64 的黑色背景中投入任意多个不同类别的数字, 考虑到画幅的容量, 每张

测试图像至多包含 5 个不同类别的数字, 至少有 1 个数字.

CapsNet 和多分支自编码器的卷积部分设置如表 1 所示, 解码部分设置如表 2 所示.

表 1 MNIST 实验卷积部分结构参数

| 次序 | 卷积核尺寸 | 卷积核数量 | 卷积步长 | 激励函数 |
|----|-------|-------|------|------|
| 1  | 5×5   | 64    | 2    | ReLU |
| 2  | 5×5   | 128   | 2    | ReLU |
| 3  | 5×5   | 256   | 2    | ReLU |

表 2 MNIST 实验解码器结构参数

| 次序 | 卷积核尺寸 | 卷积核数量 | 卷积步长 | 激励函数 | 神经元数量 | 激励函数 |         |
|----|-------|-------|------|------|-------|------|---------|
| 1  | 5     | 256   | 4    | ReLU | 512   | ReLU |         |
| 2  | 本文模型  | 5     | 128  | 4    | ReLU  | 1024 | ReLU    |
| 3  |       | 5     | 1    | 4    | tanh  | 4096 | sigmoid |

图 3 和图 4 展示了模型对多类别图像的重构结果. 图 3 的第 1 行是原始图像, 第 2 行是本文模型的重构结果, 第 3 行是 CapsNet 的重构结果, 每组重构结果的第 1 列是所有类别重构图像的叠加, 第 2 列是仅叠加高活跃度类别的重构图像, 说明本文模型和 CapsNet 的重构能力不随输入图像中包含的类别数量增加而降低, 重构能力较为稳定. 图 4 的“按类别重构”是对每个类别的编码向量分别重构的结果. 当类别不存在时, 本文模型提取的编码向量不会激励解码器产生任何形状, 而 CapsNet 产生了较大的“迷雾状”图形. 出现这种现象的原因是, CapsNet 的训练过程中, 仅有活跃的编码向量到原始图像的训练数据对, 缺少不活跃的编码向量到全零图像的训练数据对, 使得其解码器不知道应该将不活跃的编码向量映射为何种图形, 所以产生了“迷雾

状”图形. 而本文模型在训练过程中增加了不活跃编码向量到全零图像的训练数据对, 使得解码器不仅能够将活跃的编码向量映射到正确的图形, 而且能够将不活跃的编码向量映射到全零图像. 所以如果本文模型提取的编码向量是低活跃度的, 解码器会将其映射到全零图像, 该编码向量不会激励解码器产生任何形状. 由表 3 可得, 在所有类别重构图像和仅活跃类别重构图像上, 本文模型的 PSNR 明显高于 CapsNet, 说明本文模型更好地还原了原始图像. 在所有类别重构图像和仅活跃类别重构图像上, 本文模型的 PSNR 近似相等, 说明本文模型产生的不活跃类别的重构图像是近乎纯净的全零图像, 进一步说明不活跃的编码向量几乎不会激励解码器产生任何图形. 由此可得, 本文模型具有更强、更全面的表征能力, 对目标的重构能力优于 CapsNet.

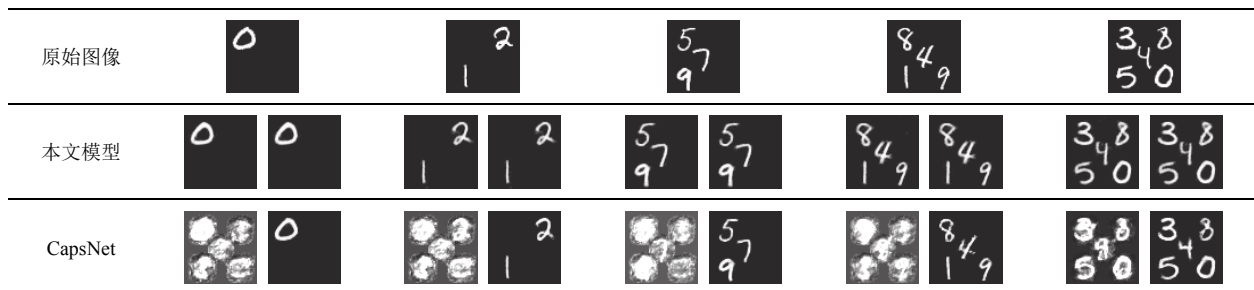


图 3 输入类别数为 1、2、3、4 和 5 时, 本文模型和 CapsNet 的重构结果. 每组的左侧是所有类别重构图像, 右侧是仅活跃类别重构图像

表 3 同时展示了模型对多类别图像的识别结果, 说明本文模型和 CapsNet 在 MNIST 数据集的多目标

识别任务上效果接近,具有良好的识别能力.

图5展示了本文模型输入不同图像时,“数字3”类别编码向量的重构图像.第1行是输入图像,第2行是“数字3”类别的重构图像,呈现的重构图像整体经过归一

化.当输入图像中不含“数字3”时,该类别的编码向量不会激励解码器产生任何图形.一方面说明不论活跃度如何编码向量的表征能力都很强;另一方面说明不采用屏蔽作为编码稀疏化策略,类别重构图像也不会出现混淆.



图4 输入图像包含5类数字时,按类别重构图像

表3 MNIST 实验的结果

| 模型      | PSNR(dB)       |                | 查准率 P(%)       | 查全率 R(%)       | F1 (%)         | IOU (%)        |
|---------|----------------|----------------|----------------|----------------|----------------|----------------|
|         | 所有类别           | 仅活跃类别          |                |                |                |                |
| 本文模型    | <b>20.5006</b> | <b>20.4395</b> | <b>99.4028</b> | 98.4325        | 98.9153        | <b>98.0988</b> |
| CapsNet | 7.2365         | 19.7416        | 99.0307        | <b>98.8447</b> | <b>98.9376</b> | 98.0810        |

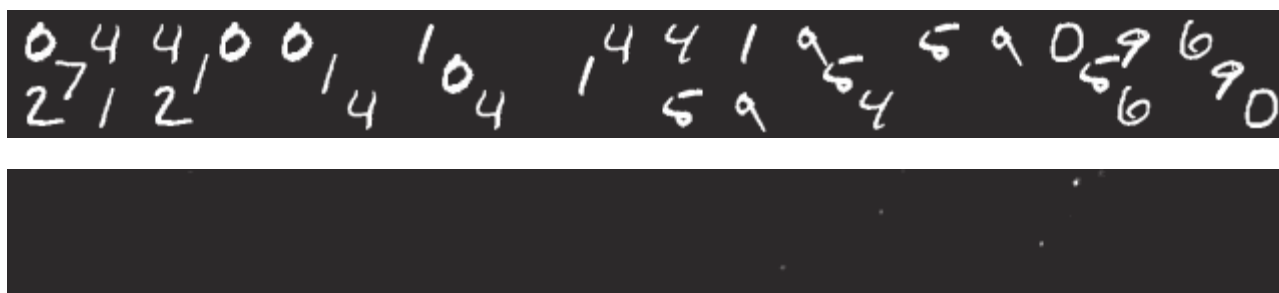


图5 “数字3”对应编码向量在不同输入情况的重构结果

综上所述,经过单类别数字图像训练,在多类别MNIST数据集上,本文模型和CapsNet均展现出良好的识别能力而本文模型的重构能力明显优于CapsNet.

## 7 结论

针对CapsNet解码器的输入层规模随类别数增加而增加,可延展性较弱的问题,本文提出多分支自编码

器.该模型的解码器分别重构每个类别的编码向量,使得解码器的规模独立于类别数量,增强了模型的可延展性.此外在训练方法上,增加全零图像作为非标签类别编码向量的重构目标,提升了编码向量的表征能力,模型的表征能力更全面.实验部分展示了模型经过单一类别图像的训练,识别多类别图像的能力.实验结果表明在多类别MNIST数据集上,多分支自编码器具有

良好的识别能力而重构能力明显优于 CapsNet, 说明本文对 CapsNet 的改进是有效的, 使模型具有更全面的表征能力.

进一步工作, 将研究如何将 CapsNet 应用于汉字字符识别和汉字风格表征.

#### 参考文献

- 1 LeCun Y, Bottou L, Bengio Y, *et al.* Gradient-based learning applied to document recognition. Proceedings of the IEEE, 1998, 86(11): 2278–2324. [doi: [10.1109/5.726791](https://doi.org/10.1109/5.726791)]
- 2 Ranzato M, Huang FJ, Boureau Y L, *et al.* Unsupervised learning of invariant feature hierarchies with applications to object recognition. Proceedings of 2007 IEEE Conference on Computer Vision and Pattern Recognition. Minneapolis, MN, USA. 2007. 1–8.
- 3 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.
- 4 Szegedy C, Liu W, Jia YQ, *et al.* Going deeper with convolutions. Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA, USA. 2015. 1–9.
- 5 Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv: 1409.1556.
- 6 尹宏鹏, 陈波, 柴毅, 等. 基于视觉的目标检测与跟踪综述. 自动化学报, 2016, 42(10): 1466–1489.
- 7 Girshick R. Fast R-CNN. Proceedings of 2015 IEEE International Conference on Computer Vision. Santiago, Chile. 2015. 1440–1448.
- 8 Ren SQ, He KM, Girshick R, *et al.* Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137–1149. [doi: [10.1109/TPAMI.2016.2577031](https://doi.org/10.1109/TPAMI.2016.2577031)]
- 9 He KM, Gkioxari G, Dollár P, *et al.* Mask R-CNN. Proceedings of 2017 IEEE International Conference on Computer Vision. Venice. 2017. 2980–2988.
- 10 周飞燕, 金林鹏, 董军. 卷积神经网络研究综述. 计算机学报, 2017, 40(6): 1229–1251.
- 11 Sabour S, Frosst N, Hinton GE. Dynamic routing between capsules. Proceedings of the 31st Conference on Neural Information Processing Systems. Long Beach, CA, USA. 2017. 3856–3866.
- 12 Abadi M, Barham P, Chen JM, *et al.* Tensorflow: A system for large-scale machine learning. Proceedings of the 12th USENIX Conference on Operating Systems Design and Implementation. Savannah, GA, USA. 2016. 265–283.
- 13 Kingma DP, Ba J. Adam: A method for stochastic optimization. arXiv: 1412.6980.
- 14 Xiao LC, Bahri Y, Sohl-Dickstein J, *et al.* Dynamical isometry and a mean field theory of CNNs: How to train 10, 000-layer vanilla convolutional neural networks. Proceedings of the 35th International Conference on Machine Learning. Stockholm, Sweden. 2018. 5393–5402.