基于胶囊网络的小量垃圾图像识别及分类

缪金志 李丹
四川大学锦城学院 四川 成都 611730

【摘 要】虽然现主流的深度学习网络架构学习能力强且具有完美的拟合能力，但其大多需要大量数据的支持且网络较为庞大，而由于某些训练数据样本的标签大量缺失，许多小数量数据集很难从这些优秀的深度学习网络架构中得到有效的训练。

【关键词】网络数据；图像

我们希望通过从现有样本中产生新的训练样本以解决这些问题。根据 [1] 中的理论，我们可以通过对样本的相关实例化参数添加随机可控的噪声，生成较为真实的扩充数据，这些扩充的样本可反映在人类世界中该物体的真实变化。

我们的训练结果仅仅使用每一个类别其中的 190 条训练样本，就基本达到了全数据集每个类别 900 条训练样本的结果。而 [2] 中的实验结果显示，该数据生成技术的改良应用于 MNIST 等数据集上也能够达到优秀的结果。在图像生成部分，我们选择了一种可高效利用损失函数组合的策略去提升重构精度。这种组合策略使得那些缺少大量标签的数据在训练中十分有效。

1 介绍

社会的发展不可避免地带来了“垃圾围城”，因此近年来国家开始下达强制垃圾分类任务，智慧垃圾分类技术备受追捧。虽然针对主流垃圾的模型已经搭建起来，然而对于更多生活中不常见的垃圾，由于缺乏充足的带有标签的数据，垃圾识别仍遗留着一些难题 [4]。

我们在训练中应用一些传统模型时，例如线性回归，K 近邻，非线性分类器和 SVM 等，它们无法达到像深度学习一样近似于人类表现的水平。虽然得益于卷积神经网络对于深度特征和空间解释的编码能力，神经网络能够很好地理解图像中低层次和高层次的特征，但大量的池化操作却使得许多有价值的信息丢失，对于一些数据是巨大的损失。Hinton 曾公开表示池化的使用就是一个大灾难。他认为生物对图像的处理，应该具备比较完备的旋转、平移、放缩不变性，卷积神经网络显然不具备这点功能。

Hinton 提出胶囊的初衷，便是为了解决神经网络中池化的缺陷。神经网络模型强调的是特征提取，而胶囊网络模型则强调网络中构建矢量的方向大小属性及其内部联系。当前的一些卷积神经网络，极有可能将鼻子与嘴巴位置互换的人脸也识别为人脸，这显然是错误的。以人类的视觉识别来说，人脸的各个部位间是有相对空间位置关系的，因此论文使用了前人提出的胶囊网络概念。

在这篇论文中，我们的数据集包含六个大类分别为玻璃、金属、纸板、纸张、塑料和其他垃圾，每个类别平均含有 900 张图片，共 5427 张彩色图像。我们通过添加胶囊网络 [3]，丢弃池化操作从而保留相对位置信息，基于 [2] 等人的想法使用转置卷积代替原本的解码器网络，使用微调实例化参数的方法添加一些数量可控的、可代表数据本质的噪声进行新数据集的生成，构建了一种可处理只有少量带标签样本的数据集的技术，达到了在小规模垃圾数据下的精准识别及分类。

图 1 部分数据集（从左至右为玻璃、金属、纸张、塑料）

2 研究现状

现有已发表的文献中提出了很多图像生成技术。普通的数据扩充技术例如抖动和翻转仅能提供少量的样本扩充。最基础的生成式对抗网络不能产生带标签的新数据。而另一个较好的图像生成技术是变分自动编码器，变分自动编码器描绘所有的图像为一维的向量，然而胶囊网络对于每一个类别都有专用的维度。

因此当变分自动编码器的一维向量发生混乱时，很有可能影响到多个类别。

3 整体架构

这个模块介绍了整体的架构。在模块 3.1 部分，在缩减数据集的实验之前，我们尝试使用手写字符集进行简单测试。随后，我们打算使用缩减数据集（如每个类别仅 190 个样本）的垃圾图像数据进行训练并对其精确调参，将其与手写字符集的结果进行对比。

为了解决以小数量样本训练分类器时面对的不利因素，在 3.2 部分引入了一种增加数据样本的技术。

3.1 基于胶囊网络的垃圾图像识别

对于垃圾分类识别任务，选择使用胶囊网络层和解码器网络层共同完成，如图 2 所示。在对垃圾图像进行分类前，首先尝试在 EMNIST 手写数据集上简单测试，随便再将其迁移至垃圾分类数据集上训练。

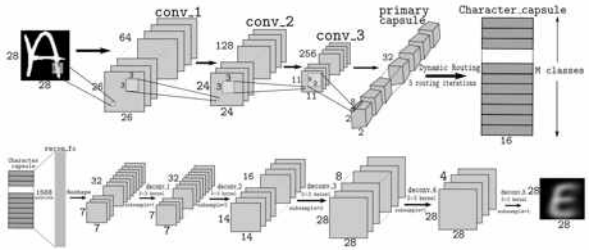

图 2　网络结构

“胶囊”这个思想最早在 2011 年作为一个转译自动编码器在 [1] 中提出。在胶囊网络中，首先堆积了 3 层卷积层，其中分别包含了步长为 1 的 64 个 3*3 卷积核，步长为 1 的 128 个 3*3 卷积核和步长为 2 的 256 个 3*3 卷积核。第四层是一层基本胶囊网络，由卷积而得到通道数为 32 的 8 维胶囊，每一个基本胶囊网络包含 8 个步长为 2 的 9*9 内核的卷积单元。第五层为字符胶囊层，对于每一个类别设置 16 维的全连接胶囊层。卷积层间的权重由反向传播调整，基本胶囊网络和字符胶囊网络间的权重则使用动态路由调整。

解码器网络由一个全连接层组成，随后紧接着 5 个转置卷积网络层，转置卷积层可将由胶囊网络得到的 16 维胶囊还原为原始数据等同大小的图像。所有的网络层均使用 sigmoid 非线性映射。

我们可以将从胶囊网络得到的 16 维胶囊放入解码器网络，便可以还原它的图像。对结果分析，虽然胶囊网络在图像分类识别上已经达到了很好的性能，但解码器网络的重构图像较为模糊。因此我们选择将还原的图像锐化，将其和原始数据集结合，对解码器网络再训练，使其拥有锐化图像的能力。

3.2 基于对实例化参数微调的图像数据生成技术

Hinton[1] 提出，“在胶囊网络中可以使用 16 维的向量表示任一字符”。因此只要我们对某一实例化参数进行较为真实的微调，就可以生成有效的新数据。我们可以从图 3 中发现，解码器重构的图像虽然有些许变化，但十分地模糊，特别是垃圾分类样本。导致其主要原因很可能是训练集数量过小或图像信息丢失过多。因此我们尝试运用 Keras 的锐化算法对每一个重构后的图像进行锐化操作，然后将锐化后的图像和原始图像结合起来，对解码器网络进行多轮重训练，从而得到一个拥有锐化功能的解码器网络。

图 3 未经锐化训练的解码器与原始图像比较

解码器网络训练完成以后，我们便可以通过微调实例化参数来生成新的数据集了。通过实验发现，方差越高，我们观察到的变化则越大。在没有任何限制条件下对实例化参数增加噪声点会导致重建图像产生不同的变形，所以我们设计了一种控制机制以避免图像失真 -- 记录每个增加到实例化参数的噪声点，使实例化参数调整的值不能超过方差最大值。使形成的新数据和原始数据相结合，它们可有效地增加样本数量，以此解决我们小样本的问题。随后，我们使用结合的数据集再训练就得到了最终的分类模型。

3.3 损失函数

本文的损失函数拥有两个部分。对于胶囊输出部分的损失函数选择 Margin loss。而对于解码器重构部分我们选择了一种组合损失函数。

“组合预测关键即是获得各单项预测的权系数值”[5]。如果只是线性地将两个损失函数组合设计为一个新的损失函数，那么获得的重建图像相对较弱。因而我们选择使用两个解码器，每个解码器的损失函数会生成两个单独的重建输出。然后我们比较了两个重建的输出和测试图像之间各像素差的绝对值，并分配最接近测试图像的像素值到最终的重建输出。我们测试损失函数组合 MSE & DSSIM， MSE & BCE 和 BCE & DSSIM。最终得到 BCE & DSSIM 组合重建精度较高。

4 实验结果及分析

对于垃圾分类数据集，我们从训练集（数据集的85%）中，用每类 190 个训练样本（训练集的 25%）训练网络，以及测试集（数据集的 15%-- 每类 135 个样本）进行测试。为了测试网络结构的表现，我们依然会使用全样本集去训练测试以评估它。得到仅在 190 个样本下就可达到全数据每类 900 条数据下的不产生新数据的测试准确率—92.8%，且在全数据 + 产生新数据的情况下可达到 94.2% 的准确率。

数据集	类别	训练集大小（每类）	测试准确率
手写字符集	47	300	0.871
垃圾图像集	6	190	0.928
手写字符集	47	112800（全集）	0.893
垃圾图像集	6	5427（全集）	0.942

表 1　手写字符与垃圾图像结果比较

4.1 垃圾分类识别结果

少量数据情况下，在手写字符上可达到 87% 以上的准确率，而在垃圾图像分类数据集上最终可达到 92.8% 的准确率，与手写字符数据集相比略高一些。对数据本身进行分析，认为手写字符集类别多数据量大，因此训练难度也更大。

此胶囊架构相较于其他网络更加轻量级，虽然能够较好地提取空间特征，但在 RGB 彩色图处理为灰度图时会丢失颜色信息，使得对垃圾图像数据训练时很容易混淆玻璃和塑料两个类别（这两个类别多为瓶罐），这两类对于特征的提取十分相似，在一定程度上降低准确率，后续考虑加深网络架构，消耗更多的计算机资源处理 RGB 彩色图像以保证信息完整性。

4.2 图像数据生成的结果

在这一部分，我们讨论解码器重建技术和微调技术的结果。可以看到，图像经过锐化重建后要比原始重建的图片更加清晰形象。因此解码器重新训练技术在锐化重建图像中获得了巨大成功。

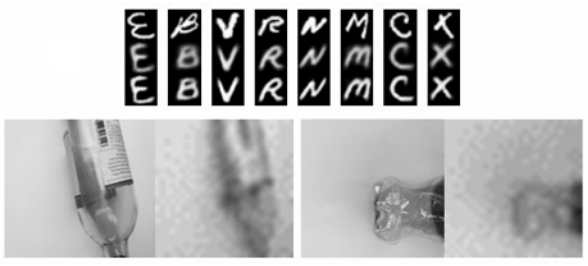

图 4 原始图片与最终实例化参数调整重构图片比较

即使对解码器重训练在锐化重建图像上获得巨大成功，但依然没有得到我们所期待的变化。因此，我们通过微调技术在重建的图片上进行新的图像数据生成。从图 4 观察到，对解码器网络的重训练后的重建图片现在捕获了我们需要的细微变化，比如手写字符中 R 的倾斜，V 的不对称，垃圾图像中瓶身的曲线，瓶底的凹凸性，并且相比未经过锐化图像训练的重建图像更接近测试集的图像，在解码器性能的提升上获得了明显的成功。

5 结束语

在这篇文章，我们通过胶囊网络介绍了一种在小数据集上进行识别分类的技术。我们演示了在垃圾集和知名手写字符集上的表现。算法通过微调图像对应的实例化参数去创建新的训练样本，与常规数据增强技术相比，我们技术所生成的图像具有更多真实的细节。为了更好地提升重建图像，我们对不同的损失函数组合进行了分析比较。

我们的方法在垃圾图像分类识别上表现得很好，在字符识别上也达到了不错的效果，该网络响应速度快，准确率高，在小数量集数据上即可拥有较高准确率，可达到市场上智慧垃圾分类标准。此外，我们打算更改网络架构使其能够应用在 RGB 彩色图像上。

【参考文献】

[1]Hinton G E, Krizhevsky A, Wang S D. Transforming auto-encoders[C]//International conference on artificial neural networks. Springer, Berlin, Heidelberg, 2011: 44-51.

[2]Jayasundara V, Jayasekara S, Jayasekara H, et al. Textcaps: Handwritten character recognition with very small datasets[C]//2019 IEEE winter conference on applications of computer vision (WACV). IEEE, 2019: 254-262.

[3]Sabour S, Frosst N, Hinton G E. Dynamic routing between capsules[C]//Advances in neural information processing systems. 2017: 3856-3866.

[4] 黄国维 . 基于深度学习的城市垃圾桶智能分类研究 [D]. 淮南 : 安徽理工大学 ,2019.

[5] 殷和俊 , 杨桂元 . 基于一类损失函数的组合预测模型构建 [J]. 哈尔滨商业大学学报 (自然科学版),2018,34(01):107-112.