

引用格式:沈贵华,张炼,姜博.基于语义一致性约束的中草药数据关联标注算法[J].中国传媒大学学报(自然科学版),2022,29(04):33-40.

文章编号:1673-4793(2022)04-0033-08

# 基于语义一致性约束的中草药数据关联标注算法

沈贵华<sup>1\*</sup>,张炼<sup>2</sup>,姜博<sup>2</sup>

(1. 中国文物信息咨询中心,北京 100029; 2. 北京邮电大学人工智能学院,北京 100876)

**摘要:**中草药数据具有丰富的数字化资源,利用标注的方法对中草药的类别和功能进行分析是一种能够实现文化识别、文化解读、文化传承的科学技术方法。本文聚焦于中草药领域的书籍图像数据,实现了对图文混排的中草药书籍数据的数字图文处理与关联标注。本文提出了基于语义一致性约束的中草药关联标注算法,针对中草药图片存在的类间差距较小,图像多形态等问题综合使用了特征提取模型,注意力机制等结构,并针对中草药数据的特点对经典标注模型结构进行优化,实现中草药图像的标注任务。

**关键词:**语义一致性约束;注意力机制;图像标注

**中图分类号:**TP391 **文献标识码:**A

## Chinese herbal medicine data association annotation algorithm based on semantic consistency constraint

SHEN Guihua<sup>1\*</sup>, ZHANG Lian<sup>2</sup>, JIANG Bo<sup>2</sup>

(1. China Cultural Heritage Information And Consulting Center, Beijing 100029, China; 2. College of Artificial Intelligence, Beijing University of Posts and Telecommunications, Beijing 100876, China)

**Abstract:** Chinese herbal medicine data is rich in digital resources. Using annotation to analyze the categories and functions of Chinese herbal medicine is a scientific and technological method that can realize cultural identification, cultural interpretation and cultural inheritance. This paper focuses on the book image data in the field of Chinese herbal medicine, and realizes the digital image and text processing and associated annotation of the Chinese herbal medicine book data with mixed text and text. In this paper, a Chinese Herbal Medicine Association annotation algorithm based on semantic consistency constraint is proposed. Aiming at the problems of small gap between classes and multi morphology of images in Chinese herbal medicine images, the structure of feature extraction model and attention mechanism are comprehensively used. According to the characteristics of Chinese herbal medicine data, the structure of classical annotation model is optimized to realize the annotation task of Chinese herbal medicine images.

**Keywords:** Semantic consistency constraint; attention mechanism; image annotation

### 1 引言

随着计算机的快速发展,图像标注技术得到了较大的进步,除了传统的标注方法外,以人工智能为核心

心的算法模型被广泛应用到了标注任务中,常用的有SVM模型<sup>[1]</sup>,逻辑回归模型<sup>[2]</sup>,KNN<sup>[3]</sup>算法等,这些算法在标注领域都取得了很好的效果,但是在一些细粒

基金项目:揭榜挂帅重点研发课题(课题编号:2021YFF0901702)

作者简介(\*为通讯作者):沈贵华(1980-),男,硕士,高级工程师,主要从事文物数据资源汇聚与智能处理研究。Email: 306865283@qq.com

度的标注任务中这些基于统计的浅层模型无法很好地区分相似度较高的类间图像,因此研究人员又把目光聚焦于能够更好地提取图像特征的神经网络模型。2015年,ResNet<sup>[4]</sup>网络模型通过引入残差模块解决了网络过深导致的网络退化问题,使得神经网络能够通过更深层的网络提取图像中的高层语义特征,并在实验中取得了极佳的效果。2017年Google团队在此基础上提出了Inception-V3<sup>[5]</sup>网络模型,使得网络可以并行执行多个卷积计算和池化计算,并且进一步加深了网络的深度,使得识别效果进一步提升。

这些经典的深度学习网络模型已能够很好地完成一般的图像标注任务,但对相似度较高的细粒度图像标注任务其识别效果并不理想,原因是在提取特征时并没有对图像区域进行优先度的区分,因此无法对只有细微区别的细粒度图像进行识别,也就是说这些经典算法模型往往忽略了图像中信息的关键程度。在人类视觉系统中人们观察事物的方式往往不是全局的,而是将注意力聚焦于某几个区域而忽略一些无效信息,通过这种方式可以大大加快信息获取速率且可以提高信息的密度。因此研究人员受其启发便提出了注意力机制模块,通过引入感兴趣区域的方式使模型更好地将关注点转移到关键区域上,从而大大提高模型的识别效果。

中草药一般是野外的动植物经过药化之后的形态,其中以植物居多,由于中草药的标注更加类似于植物的细粒度标注任务,其存在类间区分度低的特点,所以本文参考使用细粒度标注模型中相关模型结构对算法模型进行建立。由于数据集中中草药在同一类别下存在着两种形态的图像,分别是中草药的原始植物形态和药用形态,其在外在表现形式上具有很大的区分度,但其应被标识为同一种药材。因此如何在同一类别存在多种形态的情况下,依旧能够以较高的准确率对其进行标注,这也是本文需要解决的核心问题。

## 2 引入注意力机制的关联标注模型

### 2.1 Inception 结构

Google公司在2014年的ILSVRC比赛中使用GoogLeNet<sup>[6]</sup>模型取得了冠军,其论文中指出,对于神经网络模型来说为了获得更好的模型效果,最简单也是最有效的方式就是通过增加网络模型的宽度和深度去提高模型的特征提取能力,但对神经网络来说过

深或者过宽的网络会导致以下几个问题:

(1)网络模型过大导致参数过多,可能导致过拟合的问题,尤其在数据集较小时更容易出现此类问题。

(2)网络复杂性的增大会影响模型的使用,以及网络模型过深容易导致梯度弥散问题。

GoogLeNet为解决上述问题提出了Inception V1模块结构,如图1所示,目标是将卷积层和全连接层的密集结构进行化简,转化为稀疏连接。通过并行的多组卷积以及池化层对输入进行处理,经过拼接之后再输出,这样原有的卷积层就会被分解为多组卷积核较小( $3 \times 3$ 和 $5 \times 5$ )的卷积层,同时由于卷积的计算量过大,在卷积之前还可以通过一组 $1 \times 1$ 的卷积核对输入进行降维减少计算量,通过这种方式网络的宽度和深度都会对应程度的增大,而参数规模反而会降低,从而提升网络的性能。此处Inception使用多组卷积操作获取不同大小卷积核的卷积结果,得到不同大小的感受野并将其进行拼接,从而获取到图像中不同尺度大小特征图的整合结果。

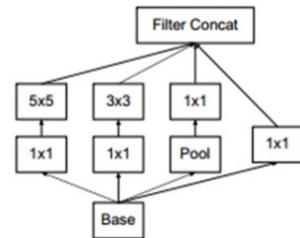


图1 Inception V1模块结构

在此基础上,研究人员进一步认为,对于卷积核大小为 $5 \times 5$ 的卷积核,可以通过两个连续的卷积核大小为 $3 \times 3$ 的卷积层代替,通过这种方式在不影响卷积层的感受野的前提下,可以进一步减少参数数量,所以最终的Inception V3模块结构如图2所示。

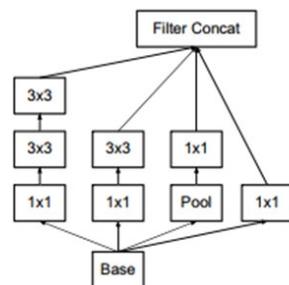


图2 Inception V3结构

在应用了Inception模块的GoogLeNet中全部使用了Inception模块代替卷积模块,并且在此基础上做

了以下几点优化,首先是使用平均池化层代替全连接层,第二,为了避免出现过拟合现象,在网络中使用了 Dropout 层,第三,在网络中加入了两个 Softmax 分支作为辅助分类器,并且分别给定一个权重增加到最终结果上,其作用是使模型在前向传导过程中防止梯度消失,产生梯度消失的原因是模型深度过深导致在反向传播过程中任一求导结果为 0,那么最终结果一定为 0,通过引入两个辅助分类器可以引入新的梯度,从而避免这种情况的发生。还有一个原因就是相关人员认为在深层网络中,不仅最后一层的特征可以作为输出结果,其中间某些层的输出也可以用作分类器的输入影响最终结果,也就是起到了多层模型融合的作用,但需要注意的是这个结构仅用于在训练阶段提升模型效果,测试时这个结构会被去掉。GoogLeNet 结构如图 3 所示。

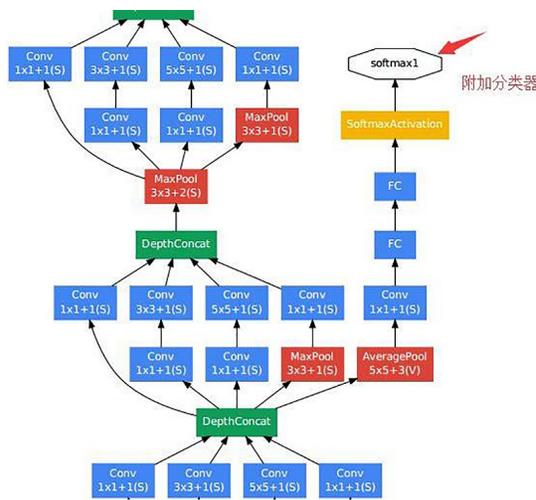


图 3 GoogLeNet 局部网络结构

## 2.2 Xception 结构

### 2.2.1 Depthwise Separable Convolution:

Depthwise Separable Convolution 中文叫做深度可分离卷积,是一种特征提取模型,其结构为一个深度卷积 (Depthwise Convolution) 后连接一个逐点卷积 (Pointwise Convolution)。深度卷积是一个具有 Channel-wise 的  $n \times n$  的空间卷积模型,比如有一个 5 通道的输入,那么深度卷积就应该具有 5 个  $n \times n$  的空间卷积结构,这种结构在每个通道上提取输入的特征图,但由于对每个通道独立计算,导致其无法提取出跨通道的特征,所以在这个结构后面应连接一个逐点卷积层引入通道间特征。逐点卷积就是一个用于维度调整的简单的  $1 \times 1$  卷积结构。这种结构与传统的卷积结构相比,不

需要在所有的通道下执行卷积操作,同时通过逐点卷积也保留了通道之间的相关特征,其结构如图 4 所示。一般情况下可以用深度可分离卷积去代替传统的卷积层,同时这个结构最大优点就是能够极大程度降低模型的参数,使得模型更轻量(将参数个数由  $n \times n \times m \times k$  减少为  $n \times n \times m + n \times n \times k$ ,其中  $m$  和  $k$  分别为输入维度和输出维度)。

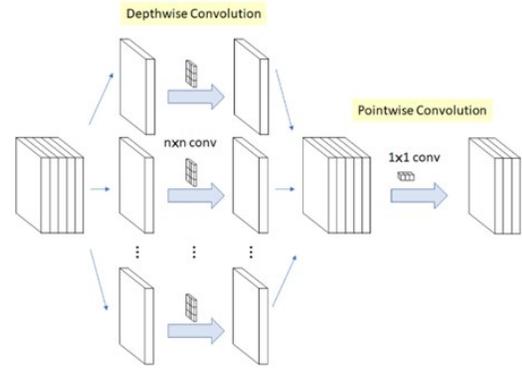


图 4 深度可分离卷积结构

### 2.2.2 Xception 网络结构

Xception 网络结构是在 Inception 网络结构的基础上引入了深度可分离卷积的思想,进一步简化了模型参数,提高计算速度,同时保证了模型的效果。为了方便介绍,此处将 Inception v3 结构进行化简,考虑模型只通过一组  $3 \times 3$  的卷积结构组成,并且不包含池化层,结构如图 5 所示:

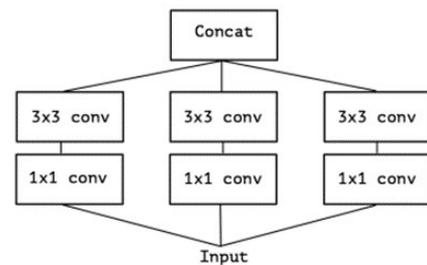


图 5 简化后的 Inception 结构

对于简化后的模型结构可以发现并行的每组卷积计算首先都是通过一个  $1 \times 1$  的卷积层之后连接一个  $3 \times 3$  大小的卷积层,最后通过拼接的方式对特征进行合并,其中  $1 \times 1$  卷积对应的是深度可分离卷积中的逐点卷积,目的是提取跨通道的相关性,而  $3 \times 3$  卷积对应的是深度卷积,目的是提取通道特征。在这种结构的基础上可以对结构进一步调整为图 6 所示:

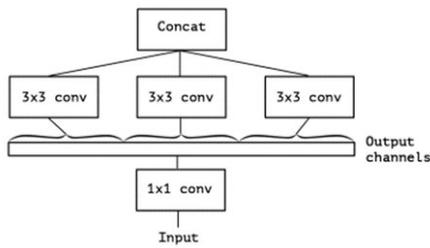


图6 调整后的模型结构

因为每组卷积操作首先都要通过1个1×1的逐点卷积,可以将第一步的卷积操作进行简化,将多次逐点卷积运算合并为一次,所有3×3的深度卷积运算共用这一次的结果,并且每个深度卷积结构只关注于某几个通道,通过这种方式模型的参数进一步缩小,并且在极端情况下,每一个深度卷积只关心一个通道,那么这种极端模型将其称作 Extreme 版本的 Inception 结构,或者叫做 Xception 结构,如图7所示。

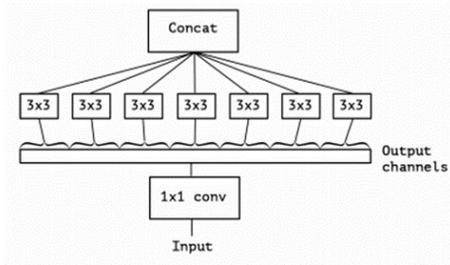


图7 Xception 结构单元

Xception 结构与 Inception 结构主要还有两点不同:

(1)计算顺序不同:这里指的是深度卷积操作和逐点卷积操作顺序不同,在 Xception 结构中调换了这两者的顺序,先通过逐点卷积提取跨通道特征,再经过深度卷积提取单通道特征。这里可以这么做的原因是在网络结构中,Xception 结构一般是处于堆叠情况的,也就是说在全局视野中,每个深度卷积结构之后一般都会有一组逐点卷积操作,所以这么处理模型并不会影响网络的结果。

(2)是否加入非线性结构:在使用 Inception 结构时,需要在第一步处理和第二步处理之间加入非线性函数,一般选择 ReLU 激活函数,但是在 Xception 结构中并不需要在两步计算之间加入非线性函数。通过实验验证,去除两步操作之间的非线性处理不会影响最终的准确率,甚至在一定程度上还会提高网络的性能。

Xception 网络结构中全部使用了深度可分离卷积代替普通卷积结构作为网络的特征提取模块提取图像特征,整体网络被分为3个 flow,包括 Entry flow, Middle flow, Exit flow,其中 Entry flow 包含了4个可分离卷积模块,Middle flow 包括了8个模块并且是由连续的8个相同的模块拼接组成,Exit flow 包括2个模块并将提取好的特征作为结果输出,整个网络共由36个卷积结构和14个模块组成,每个模块都通过 ReLU 进行激活,并且在每个模块结构之间都引入了一个线性残差结构防止训练过程中的梯度弥散问题。最终的网路结构如图8所示。

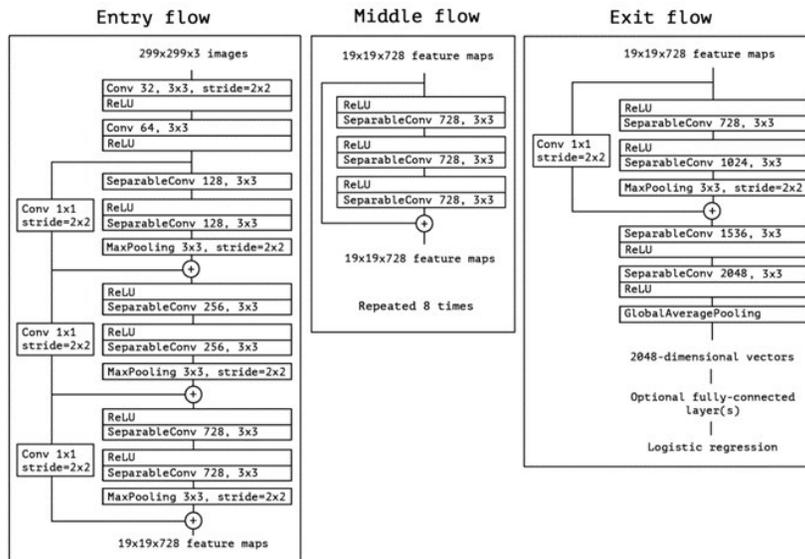


图8 Xception 网络结构

### 2.3 注意力机制模块

受人类视觉系统的启发,研究人员认为在图像处理时并不需要传统卷积运算提取出全局特征,因为一般情况下关注的是图像中某个部分,或者说叫做前景部分,而应当忽略背景信息。由此提出了注意力机制,其目的就是用于区分图像中的关注区域和非关注区域。注意力机制被分为硬注意力机制和软注意力机制,硬注意力机制指的是对关注区域进行区分时使用的是0-1编码,某个区域要么是关注区要么是非关注区,模型更加关注每个点的情况而忽略区域的划分,这种方法存在的主要问题是其网络模型的参数不可微分,需引入分数评估器进行模型参数的学习。而软注意力机制是一个连续分布的问题,其阈值分布在0到1之间,通过不同的分数判断某个区域的被关注程度,可以看出这是一个天然可微的模型,可以通过反向传播来学习模型的参数,所以通常使用软注意力机制处理图像任务。而从注意力机制关注的维度来区分,可以将其分为空间域注意力,通道域注意力,层域注意力,混合域注意力,时间域注意力几个类别。本文注意力机制采用的是基于混合域的注意力结构,此处简单介绍常用的空间域,通道域和混合域结构。

#### 2.3.1 空间域

在基于空间域的注意力模型中,并不关注图像中的所有区域,关注的是图像中与任务有关的感兴趣区,并且通过特殊的模型结构去提取特征网络中的关键区域。比较经典的是Google提出的STN<sup>[7]</sup>网络结构,其通过Localisation Net结构去学习网络的仿射变换系数,也就是学习形变系数,并通过系数对目标进行旋转等操作,使得输入的样本被处理为规则图像之后通过网络定位关键区域。结构如图9所示。

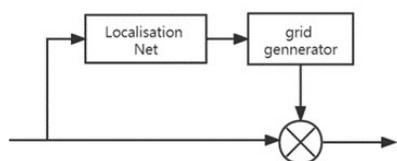


图9 空间域注意力模型

#### 2.3.2 通道域

通道域与空间域相对,空间域关注图像的尺度空间,而通道域关注图像的通道特征。经典的通道域注意力模型是SENet<sup>[8]</sup>网络结构,其通过Squeeze过程将每个通道上的空间特征进行压缩,之后通过Excitation过程将

每个通道的特征赋予一个权重,由此来学习通道之间的特征,最后通过scale过程将特征图中不重要的通道特征减弱,重要的通道特征进行加强。结构如图10所示。

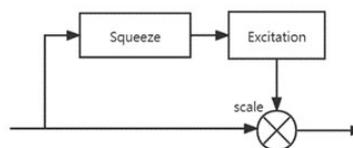


图10 通道域注意力模型

#### 2.3.3 混合域

空间域提取图像的区域特征,但是其提取的特征局限在单通道上,全局的信息提取能力较弱,而通道域提取图像的通道间特征,忽略了单通道内的信息。由此可以想到是否存在一种综合了两种结构的模型,也就是混合注意力结构。经典的混合域注意力模型为Residual Attention Network<sup>[9]</sup>网络,其通过提出的残差注意力学习结构,完成了空间域和通道域特征的整合输出,其主要思想参考了残差网络,通过堆叠多层注意力模块的方式提取不同维度的注意力特征,并且通过将Mask之后的特征张量和Mask之前的特征张量进行拼接,同时作为下一层的输出(残差结构),解决了同时提取空间注意力和通道注意力时可能导致的网络特征信息损失问题。

### 2.4 基于语义一致性约束的关联标注算法结构

在处理中草药图片数据的过程中发现中草药数据存在着双形态的特点,如图11所示,由于两个形态之间图像差别很大,无法归纳为同一标注结果下的数据,同时两者还具有较强的语义一致性,若是完全分离识别会导致精度下降。为了解决这个问题本文提出了基于语义一致性约束的中草药关联标注算法,将两形态的中草药数据分别处理提取其特征,并最后联合到同一个特征矩阵上,通过这种方式不仅可以分别提取不同形态图片中的特征,同时也满足了语义一致性约束条件,提高了识别标注的准确性。



图11 人参的植物形态(左),人参的药材形态(右)

2.4.1 主网络模型

中草药数据集标注的主要问题是解决中草药数据存在的多形态问题。在经典的标注算法中由于这种问题很少出现,所以在这种情况下会被标注算法区分为完全不同的两种类别,由此导致标注结果产生误差,本文提出一种基于语义一致性约束的算法模型去解决这个问题。本文在经典网络模型的基础上引入语义一致性提高模型的识别效果,语义一致性约束可以描述为:

(1)在单一形态图像中,在对不同语义类别的类进行识别时,需要最小化其他具有相同语义的类的类中心与本类的类中心之间的距离。

(2)在具有相同语义但形态不同的图像上,两者的类中心应该小于其他具有不同语义的不同形态的图像之间的类中心距离。

以上两点可以通过公式进行数学表示,分别为:

$$d_1 = \frac{1}{k} \sum_{j=1}^k [ \frac{1}{n_j} \sum_{i=1}^{n_j} ( \|c_j^v - v_i^j\|_2 + \|c_j^t - t_j\|_2 ) ] \quad (1)$$

$$d_2 = \frac{1}{k} \sum_{j=1}^k \|c_j^t - c_j^v\|_2 \quad (2)$$

式中, $k$ 为语义标签的总数, $v$ 表示算法输入的图片, $t$ 表示算法的语义标签, $n$ 是图像-标签集的总数。由此我们可以求出语义一致性约束的损失函数表达:

$$\text{Loss}_{\text{constraint}} = d_1 + d_2 \quad (3)$$

在此基础上构建模型的主网络结构为双层特征提取网络,并且在特征提取之后对其进行向量拼接,再将结果通过解码器进行解码,最终对结果进行标注。通过这种方式可以分别提取同一类别下不同形态图像的特征,并经过神经网络的反向传播特性对模型参数进行训练,这里经过对比分析,特征提取结构选择 Xception 网络模型。同时为了获得更好的特征表达本文加入双线性模块对特征图进行处理,具体结构选择使用一个卷积核大小为  $1 \times 1 \times m$  的卷积层和一个卷积核大小为  $1 \times 3 \times n$  和  $3 \times 1 \times n$  的串连卷积层,如图 12 所示。此处为了使得到的特征矩阵具有更好的分布效果,可以对特征进行归一化,对得到的聚合特征矩阵进行展开得到向量( $g$ ),之后再带符号的平方根正则化( $s$ )和归一化( $c$ )操作,最后将结果送至解码器中进行解码。

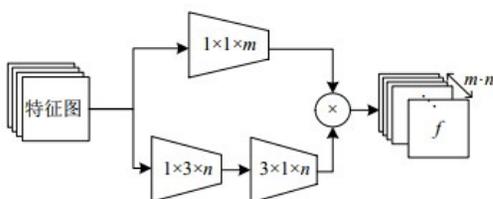


图 12 用于特征融合的双线性模块

通过这种结构将不同形态的图像放在同一特征空间中进行训练,并在网络训练过程中联合使用了具有语义一致性的图像特征,由此得到了联合了多形态特征的提取结果,最终的网络模型如图 13 所示:

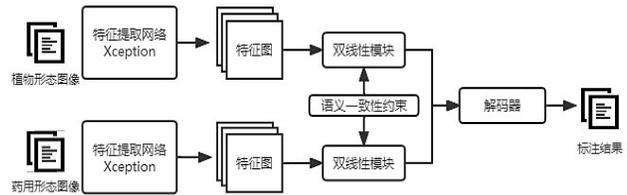


图 13 主网络结构图

2.4.2 引入注意力机制的特征提取模块

主网络结构已经解决了标注中草药数据的第一个问题——中草药多形态标注问题。在此基础上进一步改进解决中草药图像标注的第二个问题,即中草药的细粒度标注问题,对于中草药图像数据来说其类间图像相似度较高,经典的标注模型无法很好的对其进行区分,由此本文在上文提出的主网络结构基础上引入注意力机制,通过注意力机制获取图像上的关键区域特征,从而解决针对中草药图像的细粒度标注问题。

由于主网络结构中已经通过卷积网络提取到了图像的特征图,在此基础上将特征图引入注意力结构中,获取图像的注意点,并进一步通过注意力机制提取图像中关键区域的细节特征,并将得到的局部特征图与全局特征图进行合并,最后共同通过解码器进行处理,引入注意力机制的特征提取结构如图 14 所示,并用如图 15 所示的热力图对关注区域进行可视化。



图 14 引入注意力模块的特征提取结构

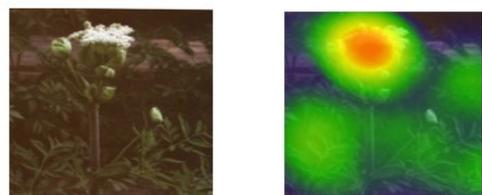


图 15 原图(左)和注意力网络特征图的热力图(右)

将注意力机制引入主网络结构时,需要将引入注

注意力机制的特征提取结构代替主网络结构中的特征提取结构,并且在训练时主网络结构的损失函数和注意力网络的损失函数可以共同训练优化:

$$\text{Loss} = \alpha \text{Loss}_{\text{main}} + \beta \text{Loss}_{\text{attention}} + \gamma \text{Loss}_{\text{constraint}} \quad (4)$$

其中 $\alpha, \beta, \gamma$ 为调节模型效果的超参数。

最后Loss的计算公式为:

$$\text{Loss}(p, q) = -\sum p(x_i) \log_2(q(x_i)) \quad (5)$$

$$q(x_i) = \frac{\exp(x_i)}{\sum_{i=1}^n \exp(x_i)} \quad (6)$$

公式中 $x$ 代表图像类别信息, $p(x)$ 代表 $x$ 的目标值。

模型在梯度下降方法上采用Adam和SGD联合使用进行训练,因为Adam收敛较快但准确度低,而SGD收敛速度较慢,但是最终模型的准确率较高,所以先通过Adam提高模型的收敛速度,直到模型的准确率不再继续提升时,通过SGD优化模型的收敛效果。

### 3 实验结果分析

本节通过对比实验来验证本文提出的模型在中草药图文数据集标注任务中的有效性。首先对实验的参数设置进行介绍,并叙述了实验流程,之后对实验结果进行详细分析。

#### 3.1 实验参数设置

本文使用了PyTorch深度学习框架对模型进行训练和测试,选择优化后的Xception网络结构作为特征提取的骨干网络,并在此基础上引入注意力机制构建本文提出的基于语义一致性约束的关联标注模型。本文使用的主要测试数据为中草药图文数据集,并且在制作训练集之前,将图像统一裁剪为 $512 \times 512$ 尺寸大小的图像,之后对其进行随机的水平翻转和归一化处理。模型的初始学习率设置为0.001,动量设置为0.9,权重衰减率设置为0.0004, batch\_size设置为32。

实验是在Linux下的Python 3.6, Tensorflow 1.12.0和一块16g大小的NVIDIA Tesla GPU下进行。

#### 3.2 实验流程

因为提出的算法模型较大,如果通过随机参数的方式从头训练模型会花费较长的时间,同时为了提高模型的效果,本文采用了迁移学习的方式去训练模型,将预先在ImageNet数据集上训练好的参数迁移到本文的模型中继续训练。具体训练过程如下:

(1)首先分别在单一形态的算法模型上加载在

ImageNet上训练好的Xception模型参数,并加入全连接层,并首先锁定除全连接层外的所有层进行训练(训练数据预先进行随机的旋转和仿射变换等扩充数据集的预处理),early-stop的monitor设置为val-loss,最终结果在val-accuracy上进行收敛,第一次收敛之后再解锁模型的后八层(实验得出效果最好)重新进行训练。

(2)搭建本文提出的基于语义一致性约束的标注算法模型并分别加载除全连接层外的上一步中训练好的参数,将参数导入后锁定除最后一层全连接层外的层级进行训练,训练数据采取随机抽样的方式抽取两种不同模态的数据图片(也要进行随机的旋转和仿射变换等),并且为了使得模型能够适应单形态数据集,在数据抓取的过程中还会随机加入空白图片,概率取0.2,之后解锁模型的后6层重新对模型进行训练,收敛后得到最终的结果,收敛过程如图16所示。

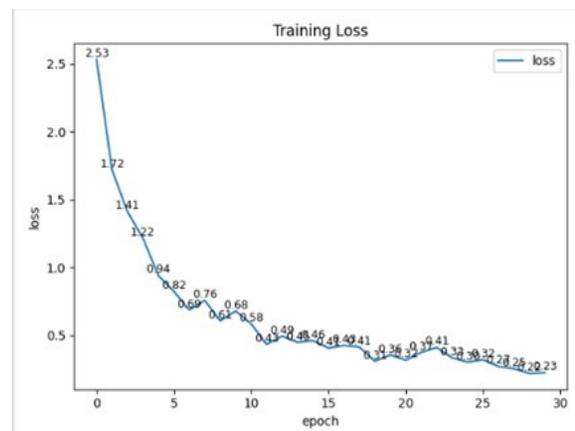


图16 语义一致性模型收敛过程

#### 3.3 实验结果与分析

本文共做了两组对比实验,分别为针对本文提出的语义一致性模型的消融实验和针对其他算法的对比实验。

实验1:首先在本文数据集下,根据本文提出的语义一致性模型中具有的双层模型,用于特征提取的Xception网络结构和注意力网络三个组件进行实验,分别探索每个组件对结果的影响,结果见表1。

表1 不同组件对算法效果的影响

结构组件	Top-1 准确率(%)
Xception	62.1
Xception+双层模型	74.2
Xception+双层模型+注意力网络	83.6
Inception-V3	60.2
Inception-V3+双层模型	69.1
Inception-V3+双层模型+注意力网络	79.7

可以从表1中看出使用Xception提取特征的骨干网能够更好地提取图像特征,并且在引入注意力机制和双层模型结构之后,会使模型的效果进一步提高,最优情况下达到83.6%的准确率。

实验2:在第二个对比实验中,将本文模型和一些常用的标注模型进行对比分析,并且在公共数据集CUB-200-2011和本文数据下分别进行对比实验,结果如表2所示。

表2 不同模型在CUB 200-2011下的准确度

模型	Top-1 准确率(%)	Top-5 准确率(%)
VGG-16	77.8	83.1
ResNet-101	83.5	89.7
Inception-V3	83.7	93.2
BCNN	84.1	95.4
MA_CNN	86.5	93.1
本文模型	85.1	92.0

可以从表2看出本文算法在CUB 200-2011数据集下效果并没有明显优势,但是仍旧可以产生较好的效果,因为本文算法主要是应用于本文提出的针对中草药数据的标注任务,并且根据数据集的特点进行了模型调整,所以会产生这种情况。接下来使用本文制作的中草药数据集对模型进行分析,结果见表3。

表3 不同模型在本文数据集下的准确度

模型	Top-1 准确率(%)	Top-5 准确率(%)
VGG-16	60.1	76.5
Resnet-101	62.4	71.1
Inception-V3	60.2	78.6
BCNN	68.5	80.1
MA_CNN	79.6	89.3
本文模型	83.6	92.6

可以从表3看出本文根据中草药数据的特点进行调整后的模型与其他模型相比不论是top1准确率还是top5准确率都有明显优势,产生这种情况的原因是经典的标注算法并不完全适用于本文的中草药数据,而本文根据中草药数据特点并且参考了细粒度标注算法模型结构,对经典标注算法进行改进,引入了注意力机制并通过双层模型解决中草药图像中存在的多形态细粒度标注问题。最终实验结果也证明了本文提出的语义一致性约束模型的效果。

#### 4 总结与展望

本文提出了中草药图文数据关联标注算法,解决了

中草药图像数据存在的类间差距较小,图像多形态等问题,为针对中草药图像数据的标注方法提供了一个新思路,具有一定的实用价值,算法以图像数据为基础通过深度学习技术对图像内容进行识别和分析,通过这种方式我们可以方便准确地了解到某一中草药图像中所蕴含的文化内容,同时因为算法针对中草药数据做了特殊调整,其识别效果和准确率也进一步提高。

未来的主要工作有:基于语义一致性约束的中草药数据关联标注算法模型虽然可以有效地解决中草药标注中的问题,但在一些公共数据集上的尝试并没有取得与本文数据相似的效果,因此针对这点我们可以尝试一些新的模型结构去使模型的泛化性得到进一步的提高。

#### 参考文献(References):

- [1] Ciocca G, Cusano C, Santini S, et al. Halfway through the semantic gap: prosemantic features for image retrieval[J]. Information Sciences, 2011, 181(22): 4943-4958.
- [2] Redmon J, Farhadi A. Yolov3: An incremental improvement [DB/OL]. arXiv:1804.02767.
- [3] Zhang M L, Zhou Z H. ML-KNN: A lazy learning approach to multi-label learning [J]. Pattern Recognition, 2007, 40(7): 2038-2048.
- [4] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.
- [5] Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the inception architecture for computer vision [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 2818-2826.
- [6] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015: 1-9.
- [7] Jaderberg M, Simonyan K, Zisserman A, et al. Spatial transformer networks [C]//NIPS'15: Proceedings of the 28th International Conference on Neural Information Processing Systems, 2015: 2017-2025.
- [8] Hu J, Shen L, Sun G. Squeeze-and-excitation networks [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7132-7141.
- [9] Wang F, Jiang M, Qian C, et al. Residual attention network for image classification [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 3156-3164.