

# 基于卷积神经网络和光谱特征的 孵前种鸭蛋受精信息无损检测

李庆旭<sup>1</sup>, 王巧华<sup>1,2\*</sup>, 顾伟<sup>1</sup>, 高升<sup>1</sup>, 马美湖<sup>3</sup>

1. 华中农业大学工学院, 湖北 武汉 430070
2. 农业部长江中下游农业装备重点实验室, 湖北 武汉 430070
3. 国家蛋品加工技术研发分中心, 湖北 武汉 430070

**摘要** 我国是鸭蛋与鸭肉生产和消费大国, 为此每年需要孵化大量雏鸭才能满足生产需要。由于无精蛋在孵化过程中无法孵化出雏鸭, 所以尽早地将其剔除可避免资源的浪费。国内鸭蛋孵化行业需在种鸭蛋入孵7天左右使用人工照蛋方式才能将无精蛋剔除。针对人工照蛋效率低且剔除的无精蛋已无食用价值等问题, 以入孵前种鸭蛋为研究对象, 将可见-近红外透射光谱技术与卷积神经网络相结合, 用于入孵前种鸭蛋受精信息的无损鉴别。实验中对采集得到的400~1000 nm原始光谱信息进行预处理(去除噪声波段和Savitzky-Golay卷积平滑处理)去除噪声与无关信息, 使用竞争性自适应重加权算法(CARS)与连续投影算法(SPA)选取特征波长, 并将选择的特征波长转换成二维光谱矩阵。二维光谱矩阵既可以表征特征光谱的有效信息, 又可以将光谱信息传入到神经网络进行训练。针对光谱数据的特点, 网络过深容易造成模型的过拟合, 网络过浅则会造成模型的欠拟合。为此构建了一个4层的卷积神经网络(CNN)用于对光谱信息矩阵进行训练, 该网络包括3个卷积层和1个全连接层, 卷积层用于自动提取光谱二维信息矩阵的有效信息, 全连接层通过对卷积层提取的局部特征进行整合进而供输出层决策, 此外在卷积神经网络中引入了局部相应归一化、池化和dropout操作可以加速网络的收敛速度并防止模型过拟合。运用SPA提取的特征波长建立的卷积神经网络模型训练集准确率为97.71%, 测试集准确率为97.41%, 验证集准确率为98.29%; 运用CARS提取的特征波长建立的卷积神经网络模型训练集准确率为97.42%, 测试集准确率为97.41%, 验证集准确率为97.44%, 而使用SPA和CARS提取的特征波长建立的传统机器学习模型测试集精度最高仅为87.39%。研究表明, 利用卷积神经网络与光谱技术相结合可以实现入孵前种鸭蛋受精信息无损鉴别, 可为后续开发动态在线检测设备提供高效、无损、快速的技术支持。

**关键词** 入孵前种鸭蛋; 受精; 卷积神经网络; 无损检测; 可见-近红外光谱

**中图分类号:** TP391.4 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2020)12-3847-07

## 引言

我国是鸭蛋与鸭肉生产和消费大国, 为此每年需要孵化大量雏鸭才能满足鸭肉和鸭蛋生产需要。由于无精蛋无法孵化出雏鸭, 随着孵化时间推移会变质且对受精蛋会造成污染, 因此尽可能早地剔除出无精蛋, 可以有效地减少资源的浪费, 提高鸭蛋孵化业和种鸭养殖业的经济效益。目前国内鸭蛋孵化行业去除无精蛋的方法依然是在种鸭蛋孵化7天以

后进行人工照蛋, 效率低下且挑选出的无精蛋已失去食用价值, 造成资源的巨大浪费。在孵化之前鉴别出来无精蛋是目前禽蛋孵化产业需要解决的老大难问题, 因此研究入孵前种鸭蛋的受精信息无损检测技术, 将会拥有广阔的应用空间。

针对禽蛋孵化, 国内外学者做了相关的研究, 研究人员尝试使用了生物电<sup>[1]</sup>、机器视觉<sup>[2]</sup>、光谱、高光谱成像<sup>[3]</sup>等技术手段实现了禽蛋入孵前及孵化早期受精信息的无损判别。但大都针对的是鸡蛋孵化, 针对鸭蛋孵化的研究很少, 由于种鸭蛋品种差异大、表面脏污多、蛋壳厚度差异大, 故

收稿日期: 2019-10-29, 修订日期: 2020-03-11

基金项目: 国家自然科学基金项目(31871863), 国家科技支撑计划项目(2015BAD19B05), 公益性行业(农业)科研专项(201303084)资助

作者简介: 李庆旭, 1994年生, 华中农业大学工学院博士研究生 e-mail: 1948260008@qq.com

\* 通讯联系人 e-mail: wqh@mail.hzau.edu.cn

与种鸡蛋相比种鸭蛋的检测难度更大。张伟等<sup>[4]</sup>提出利用机器视觉与敲击振动相结合的方法实现了对孵化 5 d 的种鸭蛋受精信息检测, 这种使用敲击振动装置部署到实际生产难度较大。Dong 等<sup>[5]</sup>提出了利用可见-近红外光谱技术初步实现了对入孵前种鸭蛋受精信息的无损判别, 但是要求种鸭蛋颜色和尺寸一致, 此研究证明了运用可见-近红外光谱技术对入孵前种鸭蛋受精信息无损检测的可行性, 但还不适于实际应用。

可见-近红外光谱由有机物的含氢基团振动的合频与倍频组成, 谱图中含有大量能够反映物质差异的信息, 具有快速、无损、简单的特点, 目前被广泛应用于农产品无损检测。文献<sup>[6]</sup>研究结果表明可见-近红外光谱可以穿透禽蛋的表面, 并将禽蛋的内部物质信息在光谱图中反映出来。深度学习中的卷积神经网络(convolutional neural networks, CNN)目前被广泛应用于农产品分类检测中, 卷积神经网络含有更多的隐含层, 通过卷积和非线性变换操作自动提取数据本身的大量特征可用于拟合复杂模型<sup>[7]</sup>。近年有学者利用光谱信息与卷积神经网络相结合检测夏威夷果、烟叶、花椒等, 但是利用两者结合起来检测禽蛋尚未有文献记载。本文利用可见-近红外透射光谱技术与深度学习相结合对入孵前种鸭蛋受精信息进行无损检测, 探索一种更符合实际生产的区分入孵前无精鸭蛋和受精鸭蛋的鉴别技术。

## 1 实验部分

### 1.1 仪器与材料

实验样品为 499 枚缙云麻鸭种蛋, 均为新鲜生产的种蛋。采购于湖北神丹种鸭场。对种鸭蛋表面使用酒精进行清理消毒并编号。

种鸭蛋光谱采集系统如图 1, 由 Maya2000Pro 光纤光谱仪、暗箱、光源、光纤探头等组成。光谱仪能够采集的波长范围在 200~1 100 nm 之间, 光谱仪的最小采样间隔为 0.5 nm, 光源为 100 W 卤素灯。另外还有智能孵化器 1 台。

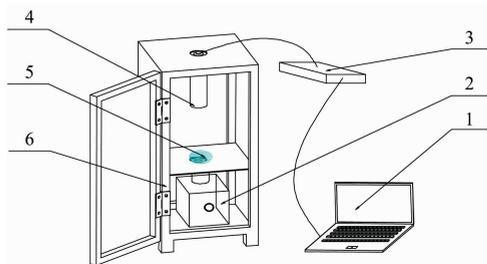


图 1 光谱采集系统

- 1: 计算机; 2: 光源; 3: Maya2000Pro 光纤光谱仪;  
4: 光纤探头; 5: 鸭蛋; 6: 暗箱
- Fig. 1 Spectrum acquisition system
- 1: Computer; 2: Light source;  
3: Maya2000Pro fiber optic spectrometer;  
4: Fiber optic probe; 5: Duck eggs; 6: Dark box

### 1.2 方法

#### 1.2.1 光谱采集

使用可见-近红外透射光谱采集系统对清洗消毒后的入孵前种鸭蛋进行透射光谱采集, 采集光谱时将种蛋竖直放置, 大头向上。考虑到模型后续将部署到生产中, 将光谱仪的采集积分时间设置为 100 ms, 扫描次数设置为 3。

#### 1.2.2 受精结果的人工判别

将智能孵化箱预热 1 h, 把采集数据后的 499 枚种蛋同时放入智能孵化箱中。设置智能孵化箱模式为“鸭”, 智能孵化箱会根据模式自动调节适用于鸭蛋孵化的最佳温湿度。鸭蛋入孵后 7 d 对其进行破壳处理, 蛋壳含有血丝为受精蛋, 得到准确的受精结果。

### 1.3 数据处理与建模

#### 1.3.1 样本集划分

为了避免样本分布不均匀, 利用了 Kennard-Stone 法对样本集进行划分, 将变量空间中的相对欧式距离差异较大的样本划入训练集, 其余样本划入测试集<sup>[8]</sup>。为了更好地验证模型的泛化能力, 从测试集样本中再划出一个验证集。本实验共采集 499 个入孵前种鸭蛋样本, 按照 7:2:1 的比例对应划分出训练集、测试集和验证集样本。其中训练集样本数量为 350 个(无精 197 个, 受精 153 个), 测试集样本数量为 100 个(受精 54 个, 无精 46 个), 验证集样本数量为 49 个(受精 27 个, 无精 22 个)。

#### 1.3.2 谱区选择与光谱预处理

谱区选择对光谱信息分析十分重要, 若选取范围较大则会带入大量噪声和冗余信息导致后续建模出现过拟合现象, 反之选取范围较小会丢失大量信息导致后续建模出现欠拟合现象。采集得到入孵前种鸭蛋原始谱图如图 2 所示, 由于谱图两端的噪音较大无法有效地反映出种鸭蛋内部真实的物质信息, 故截取 400~1 000 nm 范围内的透射光谱信息用于后续的分析与处理, 选择的谱区范围如图 3 所示。

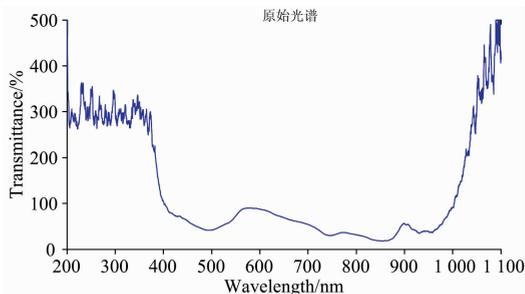


图 2 入孵前种鸭蛋原始光谱图

Fig. 2 Original spectrum of duck eggs before hatched

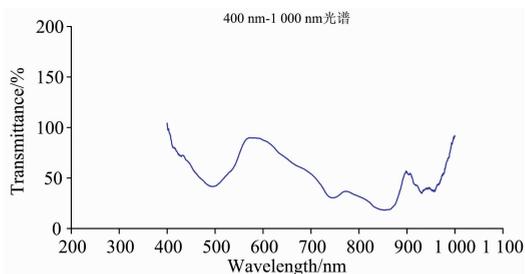


图 3 入孵前种鸭蛋 400~1 000 nm 光谱图

Fig. 3 400~1 000 nm spectrum of duck eggs before hatched

由于仪器操作、温度变化和杂散光的原因会使采集得到的原始光谱数据引入噪声与无关信息,选取的 400~1 000 nm 范围内的透射原始光谱依然含有较为明显的噪声(光谱信号中含有杂波)。为此,采用 SG 卷积平滑处理对截取后的种鸭蛋光谱进行预处理。经反复尝试,本实验使用窗口宽度为 9 的 2 阶微分进行平滑处理。平滑后的曲线如图 4 所示,可以看出 SG 平滑处理后消除了杂波干扰,光谱曲线变得更加平滑<sup>[12]</sup>。

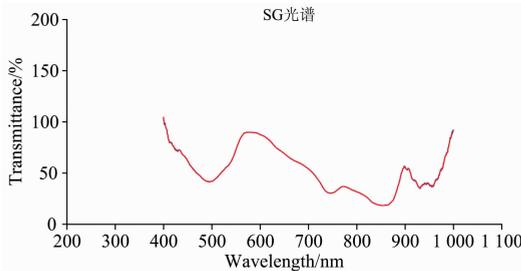


图 4 入孵前种鸭蛋 SG 平滑后光谱图

Fig. 4 SG smoothed spectrum of duck eggs before hatched

### 1.3.3 光谱特征波段选择

预处理后的光谱信息依然高达 1 361 维,部分波段依然存在很强的相关性,高维光谱信息中不仅含有反映内部物质差异的信息还包含大量冗余信息,若直接使用高维光谱信息建立种鸭蛋受精信息判别模型会导致模型出现过拟合现象,进而导致模型的泛化能力变差。同时考虑将光谱信息输入给卷积神经网络,若光谱波段存在很强的相关性则不利于卷积神经网络自动学习光谱的有效特征,从而导致网络过拟合现象的出现,且数据维数太高会大大降低网络的训练速度,不利于后期模型的部署。因此对预处理后的光谱数据进行特征波长选取,降低光谱数据的维数显得十分重要。本文使用了 CARS 与 SPA 这两种较为常用的特征波长筛选算法,筛选出能够反映受精蛋与无精蛋差异的关键波长点。

(1)竞争性自适应重加权(CARS)算法利用自适应重加权采样手段选出 PLS 模型中回归系数绝对值相对较大的波长点,去除权重相对较小的波长点,并使用交叉验证来选择 RMSECV 值最低的子集,可以有效地寻找到变量的最优组合<sup>[9]</sup>。对预处理后的训练集光谱数据使用 CARS 进行特征波长选择,选取入孵前种鸭蛋受精信息预测最优波长组合过程如下:经反复尝试,本实验将蒙特卡罗采样次数设定为 50,采用 10 折交叉验证。由图 5(a)可知,随着取样运行次数的增加,选取变量的数量逐步递减。由图 5(b)可知,RMSECV 值先缓慢递减后递增,RMSECV 值递减,说明种鸭蛋光谱数据中部分无用的信息被剔除,RMSECV 值递增,说明种鸭蛋光谱数据中有部分重要信息被剔除。当 RMSECV 值达到最小时,各变量的回归系数如图 5(c)中竖线处,此时的采样运行次数是 35,CARS 提取的最优波长点数量为 15。

(2)连续投影算法(SPA)是使向量空间的共线性最小化的前向变量选择算法,能够从原始光谱信息中将冗余信息去除,从而解决共线性问题。对于处理后的训练集光谱数据使用 SPA 选择特征波长,由 SPA 的原理可知,利用均方根误

差(RMSE)最小化的原则,选出均方根误差的导数变小的过渡点,在过渡点之前冗余信息被剔除,过渡值之后冗余信息已基本被剔除<sup>[10]</sup>。SPA 选取入孵前种鸭蛋受精信息预测最优波长组合如图 6(a,b)所示,SPA 选取的特征波长数为 11。

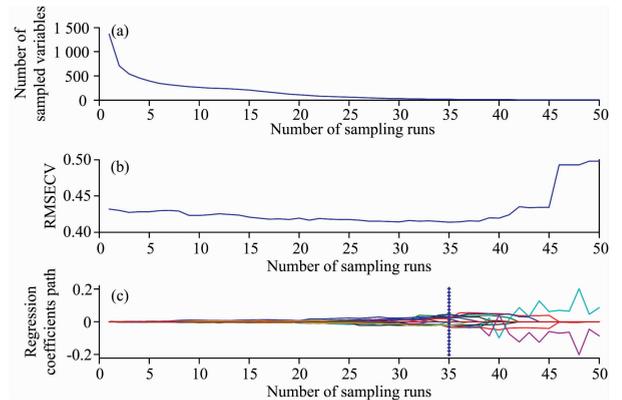


图 5 (a)取样变量数;(b)RMSECV;(c)回归系数路径

Fig. 5 (a) Number of sampling variables; (b) RMSECV; (c) Regression coefficient path

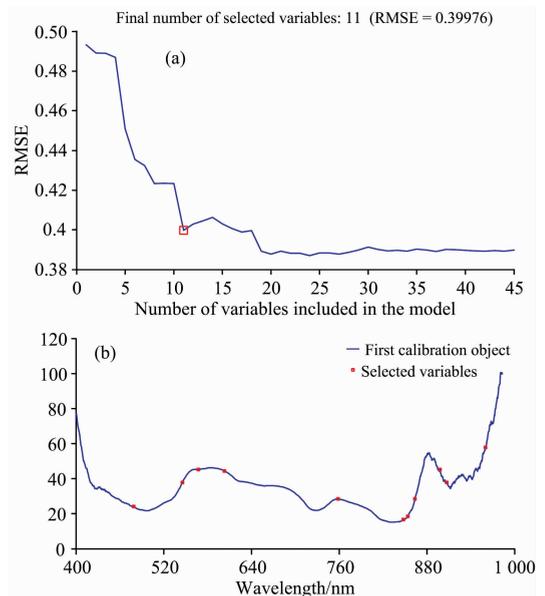


图 6 (a)RMSE;(b)选取的最优波长编号索引

Fig. 6 (a) RMSE; (b) Selected optimal wavelength number index

## 2 结果与讨论

### 2.1 卷积神经网络搭建

经过 CARS 或 SPA 提取特征波长后,数据维数已经降到较低维度,若继续对其进行降维处理可能会导致光谱有用信息的丢失。针对分类问题,一般使用机器学习或者深度学习进行建模。本研究先使用逻辑回归(logistic regression, LR)和支持向量机(support vector machines, SVM)两种常用的机器学习算法并使用十折交叉验证对 CARS 和 SPA 提取的特征波长建立入孵前种鸭蛋受精信息判别模型。为方便起

见, 此处 KS 划分的测试集和验证集合并为测试集, 判别结果如表 1, 不难发现 LR 和 SVM 模型不能完全拟合光谱数据导致模型效果不佳, 但证明了使用 CARS 或 SPA 提取的特征波长建立区分无精蛋和受精蛋的可行性。由于机器学习模型对较为复杂的模型拟合效果不佳, 往往会产生过拟合或欠拟合的现象, 本研究两种机器学习模型则是出现了轻微欠拟合现象。卷积神经网络引入了局部连接和权值共享的机制, 使其能含有更多的隐层, 这种机制使得卷积神经网络在拟合复杂模型时具有明显的优势, 故考虑使用卷积神经网络对入孵前种鸭蛋是否受精进行判别。

表 1 CARS 和 SPA 提取的特征波长建模结果

Table 1 Characteristic wavelength modeling results of CARS and SPA extraction

	SPA		CARS	
	训练集 准确率/%	测试集 准确率/%	训练集 准确率/%	测试集 准确率/%
SVM	87.39	87.39	79.94	79.94
LR	73.35	72.48	79.08	78.52

卷积神经网络主要用于图像的分类与识别, 图像数据一般为 2 维或 3 维矩阵, 而本研究使用 CARS 或 SPA 提取的特征波长数据为 1 维数据。因此需要将一维光谱数据转换为二维光谱信息矩阵, 转换公式如式(1)

$$\mathbf{S} = \mathbf{X}^T \mathbf{X} \quad (1)$$

式(1)中,  $\mathbf{X}$  为一维光谱数据,  $\mathbf{X}^T$  为一维光谱数据的转置。二维光谱信息矩阵包含了一维光谱数据的原有信息, 体现了样本的方差和协方差, 同时适用卷积神经网络的输入结构。以 SPA 提取的种鸭蛋光谱特征矩阵为例, 单个种鸭蛋样本的一维光谱数据为  $\mathbf{x} = [x_1, x_2, \dots, x_{11}]$  二维光谱信息矩阵如式(2)

$$\begin{bmatrix} x_1^2 & x_1 x_2 & x_1 x_3 & \dots & x_1 x_{11} \\ x_2 x_1 & x_2^2 & x_2 x_3 & \dots & x_2 x_{11} \\ \dots & \dots & \dots & \dots & \dots \\ x_{11} x_1 & x_{11} x_2 & x_{11} x_3 & \dots & x_{11}^2 \end{bmatrix} \quad (2)$$

卷积神经网络通常由卷积、池化与全连接等层构成。卷积(conv2d)层用于提取大量特征; 池化层可以减小卷积层提取的特征的维数, 从而能够加速神经网络收敛; 全连接层可将网络前端输出的特征还原给输出层, 最后由输出层输出分类结果。考虑到运用 CARS 和 SPA 提取的二维光谱信息矩阵尺寸较小, SPA 提取的二维矩阵大小为  $11 \times 11$ 、CARS 为  $15 \times 15$ , 若搭建的网络层数过深, 容易出现过拟合现象。结合入孵前种鸭蛋二维光谱信息矩阵的特点, 搭建了 4 层的卷积神经网络, 包括了 3 个卷积层和 1 个全连接层, 网络的结构如图 7 所示, 具体实现过程如下, 其网络参数如表 2 所示(以 CARS 提取的入孵前种鸭蛋特征波长为例, SPA 与之结构相同):

(1) 输入层: 将 CARS 提取的入孵前种鸭蛋特征波长使用式(1)转换为二维光谱信息矩阵作为输入层, 输入层尺寸为  $15 \times 15$ ;

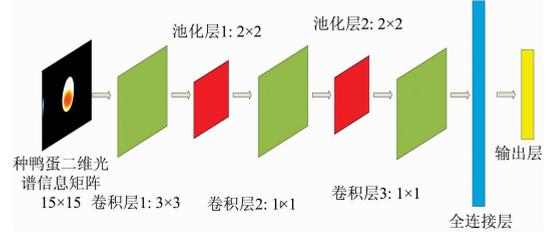


图 7 入孵前种鸭蛋网络结构

Fig. 7 Pre-incubation duck egg network structure

表 2 入孵前种鸭蛋网络参数

Table 2 Network parameters of duck eggs before hatching

输入	操作	卷积核个数	步长
$15 \times 15 \times 1$	conv2d, $3 \times 3$	64	1
$15 \times 15 \times 64$	pool, $2 \times 2$	—	2
$8 \times 8 \times 64$	conv2d, $1 \times 1$	192	1
$8 \times 8 \times 192$	pool, $2 \times 2$	—	2
$4 \times 4 \times 192$	conv2d, $1 \times 1$	384	1
$4 \times 4 \times 384$	FC	—	—
2	—	—	—

(2) 卷积层 1: 考虑到输入的种鸭蛋二维光谱信息矩阵尺寸较小, conv 操作的核尺寸与步长不宜设置过大, 将卷积核尺寸设定为  $3 \times 3$ , 卷积核数量设定为 64, 步长设定为 1。输入层经过卷积操作后使用 ReLU 函数进行激活(可使种鸭蛋光谱矩阵变得更加稀疏), ReLU 函数的实现公式如式(3)

$$y = \begin{cases} 0 & \text{others} \\ x & \text{if } x > 0 \end{cases} \quad (3)$$

ReLU 函数具有加速模型收敛的作用。为了提高模型的性能, 在 ReLU 激活后的光谱特征矩阵执行局部响应归一化(LRN)操作, 卷积操作可以提取光谱矩阵中的大量信息, 种鸭蛋二维光谱信息矩阵经过卷积操作之后输出特征矩阵尺寸为  $15 \times 15 \times 64$ ;

(3) 池化层 1: 将卷积层 1 的输出特征矩阵进行降维操作, 能够提高模型的收敛速度。池化核的尺寸和步长均设置为 2, 池化(pool)处理后的输出特征矩阵尺寸为  $8 \times 8 \times 64$ ;

(4) 卷积层 2: 设置卷积核的尺寸为  $1 \times 1$ , 卷积核的数量设置为 192, 并将步长设置为 1。池化层 1 的输出经过卷积后, 加入 ReLU 和 LRN 操作, 输出尺寸为  $8 \times 8 \times 192$ ;

(5) 池化层 2: 池化核的尺寸和步长均设置为 2, 卷积层 2 的输出特征矩阵经过池化操作后, 输出尺寸为  $4 \times 4 \times 192$ ;

(6) 卷积层 3: 卷积内核的尺寸设置为  $1 \times 1$ , 卷积内核的数量设置为 384, 步长设定为 1。池化层 2 的输出经过卷积后, 加入 ReLU 和 LRN 操作后输出尺寸为  $4 \times 4 \times 384$ ;

(7) 全连接层: 全连接层(FC)的神经元个数设置 512, 通过将卷积层 3 输出的光谱特征矩阵转换为  $1 \times 6144$  的数据, 输入给 512 个全连接的神经元后输出 512 个权值, 为了防止模型出现过拟合现象, 在输出之前使用 dropout 层随机失活部分神经元。

(8) 输出层: 将全连接层的 512 个权值经过 softmax 函数得到一个二维得分矩阵, 分别是受精蛋和无精蛋的得分系数。

## 2.2 卷积神经网络训练与测试

针对入孵前种鸭蛋光谱信息的卷积神经网络搭建完成后开始对其进行训练,虽然卷积神经网络要求有大量数据进行训练,本实验共采集 499 个人孵前种鸭蛋光谱信息样本做二分类,针对光谱信息分类问题样本数量已经足够。卷积神经网络训练的核心思想是通过反向传播连续的迭代,使得模型的预测值不断地与实际值接近。以 CARS 选取的特征波长进行训练为例,使用 Adam 优化器选择训练过程中最佳梯度下降方向,可加速模型的收敛。使用均值平方差作为损失(loss)函数用来计算预测值与实际值的差值。学习率(learning rate)初始化为 0.00001,每次抽选 8 个样本进行训练(BatchSize=8),dropout 设置为 0.5,迭代次数设置为 20 000 后开始训练。训练过程中的损失函数变化趋势如图 8,可以发现分类损失 loss 在训练的前 6000 次一直处于快速下降状态,迭代到 10 000 次之后 loss 一直维持在较低水平,说明模型达到收敛。

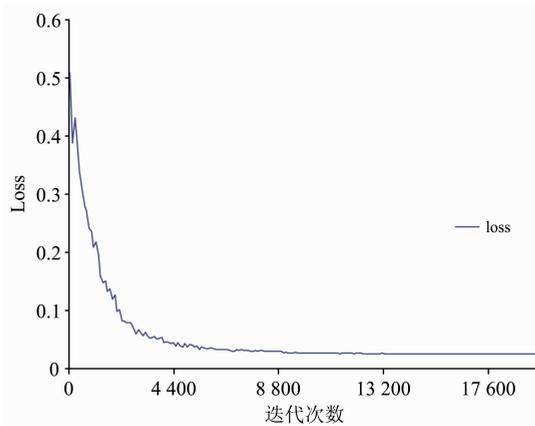


图 8 损失函数变化趋势

Fig. 8 Trend of loss function

模型训练 20 000 次后对其进行保存,然后将训练集、测试集、验证集共 499 个种鸭蛋二维光谱信息矩阵输入到训练好的模型进行测试,验证模型的性能。SPA 提取的种鸭蛋光谱特征矩阵进行卷积操作与 CARS 类似,网络的结构完全相同。模型测试结果如表 3,从测试结果可以看出,SPA 和 CARS 提取的特征波长转换为二维信息矩阵后,使用搭建的 4 层卷积神经网络建立入孵前种鸭蛋受精信息判别模型,从验证集和测试集的效果来看,模型未出现过拟合或者欠拟合现象,模型性能取得理想效果。

表 3 模型测试结果

Table 3 Model test results

模型	SPA			CARS		
	训练集 准确率 /%	测试集 准确率 /%	验证集 准确率 /%	训练集 准确率 /%	测试集 准确率 /%	验证集 准确率 /%
CNN	97.71	97.41	98.29	97.42	97.41	97.44

## 3 结 论

利用可见-近红外透射光谱技术及深度学习研究了种鸭蛋孵化前受精信息的无损检测技术,通过研究分析获得如下结论:

(1)通过对种鸭蛋孵化前光谱信息的分析,发现其特征波长点主要分布在可见-近红外波段(400~1 000 nm)范围内;经过 SG 预处理后,使用 CARS 选取的孵前种鸭蛋特征波长分别建立了逻辑回归、SVM、CNN 判别模型,三者的测试集准确率分别为 78.52%, 79.94%, 97.41%;使用 SPA 选取的孵前种鸭蛋特征波长分别建立了逻辑回归、SVM、CNN 判别模型,三者的测试集准确率分别为 72.48%, 87.39% 和 97.41%。无论使用 CARS 还是 SPA 选取特征波长后,使用卷积神经网络建立判别模型都取得了最佳的效果,与传统的机器学习方法相比性能提升明显;

(2)光谱数据经过预处理和特征波长选取后,将一维光谱数据转换为二维光谱信息矩阵,实现了将卷积神经网络与光谱数据相结合,建立了种鸭蛋受精信息判别模型。可以避免将 1 361 维数据直接用卷积神经网络训练带来的维数爆炸问题,大大加快了模型训练与测试的速度;

(3)采用 3 个卷积层和 1 个全连接层组成的神经网络结构,对入孵前种鸭蛋光谱数据的分析取得了较好的效果。为使用卷积神经网络方法对禽蛋光谱数据进行分析提供了参考;

研究结果利用可见-近红外透射光谱结合卷积神经网络对种鸭蛋入孵前受精信息无损检测具有可行性,使用 CARS 和 SPA 选取的特征波长建立的卷积神经网络模型验证集准确率均为 97.44% 和 98.29%,可以满足部署到实际生产的精度要求,为后续相应装置的研发提供了模型支持。但模型仍有再提高之处,使用的实验样本仅为缙云麻鸭种蛋,后续需要采集其他地区鸭品种的种蛋光谱数据来提高模型的通用性。

## References

- [ 1 ] Ghaderi M, Banakar A, Masoudia A. *Measurement*, 2018, (114): 191.
- [ 2 ] Hashemzadeh M, Farajzadeh N. *International Journal of Computational Intelligence Systems*, 2016, 9(5): 850.
- [ 3 ] ZHU Zhi-hui, LIU Ting, MA Mei-hu(祝志慧, 刘 婷, 马美湖). *Transactions of the Chinese Society of Agricultural Engineering(农业工程学报)*, 2015, 31(15): 285.
- [ 4 ] ZHANG Wei, TU Kang, LIU Peng, et al(张 伟, 屠 康, 刘 鹏, 等). *Transactions of the Chinese Society of Agricultural Machinery(农业机械学报)*, 2012, 43(2): 140.
- [ 5 ] Dong Jun, Dong Xiaoguang, Li Yanlei, et al. *Computers and Electronics in Agriculture*, 2019, 157: 471.
- [ 6 ] ZHU Zhi-hui, XIE De-jun, LI Wan-qing, et al(祝志慧, 谢德君, 李婉清, 等). *Transactions of the Chinese Society of Agricultural Engineering(农业工程学报)*, 2015, 31(2): 312.
- [ 7 ] YUAN Pei-sen, LI Wei, REN Shou-gang, et al(袁培森, 黎 薇, 任守纲, 等). *Transactions of the Chinese Society of Agricultural Engineering(农业工程学报)*, 2018, 34(5): 152.
- [ 8 ] FU Dan-dan, WANG Qiao-hua(付丹丹, 王巧华). *Food Science(食品科学)*, 2016, 37(22): 173.
- [ 9 ] GAO Sheng, WANG Qiao-hua, FU Dan-dan, et al(高 升, 王巧华, 付丹丹, 等). *Acta Optica Sinica(光学学报)*, 2019, 10(39): 1.
- [ 10 ] ZHU Zhi-hui, HONG Qi, WU Lin-feng, et al(祝志慧, 洪 琪, 吴林峰, 等). *Spectroscopy and Spectral Analysis(光谱学与光谱分析)*, 2019, 39(9): 2780.

# Non-Destructive Testing of Fertilization Information of Pre-Incubation Duck Eggs Based on Convolutional Neural Network and Spectral Features

LI Qing-xu<sup>1</sup>, WANG Qiao-hua<sup>1, 2\*</sup>, GU Wei<sup>1</sup>, GAO Sheng<sup>1</sup>, MA Mei-hu<sup>3</sup>

1. College of Engineering, Huazhong Agricultural University, Wuhan 430070, China

2. Key Laboratory of Agricultural Equipment in Mid-Lower Yangtze River, Ministry of Agriculture and Rural Agriculture, Wuhan 430070, China

3. National Egg Research and Development Center, Wuhan 430070, China

**Abstract** China is a big country for the production and consumption of duck eggs and duck meat. For this reason, a large number of ducklings need to be hatched each year to meet production needs. Since the Infertile egg cannot hatch the duckling during the hatching process, it can be avoided by eliminating it as early as possible. At present, it is necessary to carry out artificial photographing eggs in the country for about 7 days after hatching, so that the infertile eggs can be removed. For the low-efficiency and elimination of artificial eggs, there is no edible value problem. In this paper, the duck eggs before hatching were taken as research objects, and the visible/near-infrared transmission spectroscopy combined with the convolutional neural network was used to realize the duck eggs before hatching Non-destructive identification of fertilization information. The method preprocesses the acquired 400~1 000 nm raw spectral information (removing the noise band and Savitzky-Golay convolution smoothing) to eliminate noise and extraneous information, and applies the competitive adaptive re-weighting algorithm (CARS) and continuous projection algorithm (SPA) to select the characteristic wavelengths, and converts the selected characteristic wavelengths into spectral two-dimensional information matrix. The spectral two-dimensional information matrix can not only represent the effective information of the characteristic spectrum, but also can transmit the spectral information to the neural network for training. For the characteristics of spectral data, the network is too deep to cause over-fitting of the model, and the shallow network will cause under-fitting of the model. A four-layer convolutional neural network (CNN) is built to train the spectral information, including three convolutional layers and one fully-connected layer. The convolutional layer is used to extract the spectral two-dimensional information matrix automatically. Effective information, the fully connected layer is integrated for the output layer by integrating the local features extracted by the convolutional layer. In addition, the introduction of local corresponding normalization layer, pooling layer and dropout in the convolutional neural network can accelerate the convergence speed of the network and prevent model overfitting. The accuracy of the model training set established by the SPA extracted feature wavelength is 97.71%, the test set accuracy rate is 97.41%, and the verification set accuracy rate is 98.29%. The model training set accuracy rate established by CARS extraction is 97.42%, the test set accuracy rate 97.41%, the verification set accuracy rate is 97.44%, the traditional machine learning model test set established using the characteristic

wavelengths extracted by SPA and CARS has a precision of only 87.39%. The results show that the combination of deep learning and spectroscopy can realize the non-destructive identification of the fertilization information of duck eggs before hatching, which can provide efficient, non-destructive and rapid model support for the subsequent development of dynamic online detection equipment.

**Keywords** Pre-incubation duck eggs; Fertilization; Convolutional neural network; Nondestructive testing; Visible/near infrared spectroscopy

(Received Oct. 29, 2019; accepted Mar. 11, 2020)

\* Corresponding author

欢迎投稿

欢迎订阅

欢迎刊登广告

## 《光谱学与光谱分析》2021年征订启事

国内邮发代码：82-68

国外发行代码：M905

《光谱学与光谱分析》1981年创刊，国内统一刊号：CN 11-2200/O4，国际标准刊号：ISSN 1000-0593，CODEN 码：GYGFED，国内外公开发行人，大 16 开本，332 页，月刊；是中国科协主管，中国光学学会主办，钢铁研究总院、中国科学院物理研究所、北京大学、清华大学共同承办的学术性刊物。北京大学出版社出版，每期售价 95 元，全年 1140 元。刊登主要内容：激光光谱测量、红外、拉曼、紫外、可见光谱、发射光谱、吸收光谱、X 射线荧光光谱、激光显微光谱、光谱化学分析、国内外光谱化学分析领域内的最新研究成果、开创性研究论文、学科发展前沿和最新进展、综合评述、研究简报、问题讨论、书刊评述。

《光谱学与光谱分析》适用于冶金、地质、机械、环境保护、国防、天文、医药、农林、化学化工、商检等各领域的科学研究单位、高等院校、制造厂家、从事光谱学与光谱分析的研究人员、高校有关专业的师生、管理干部。

《光谱学与光谱分析》为我国首批自然科学核心期刊，中国科协优秀科技期刊，中国科协择优支持基础性、高科技学术期刊，中国科技论文统计源刊，“中国科学引文数据库”，“中国物理文摘”，“中国学术期刊文摘”，同时被国内外的 CJCR, CNKI, CSCD, SCI, AA, CA, Ei, AJ, PJK, MEDLINE, Scopus 等文献机构收录。根据中国科学技术信息研究所发布信息，中国科技期刊物理类影响因子、引文量及综合评价总分《光谱学与光谱分析》都居前几位。欢迎国内外厂商在《光谱学与光谱分析》发布广告（广告经营许可证：京海市监广登字 20170260 号）。

《光谱学与光谱分析》的主编为高松院士。

欢迎新老客户到全国各地邮局订阅，若有漏订者可直接与《光谱学与光谱分析》期刊社联系。

联系地址：北京市海淀区学院南路 76 号（南院），

《光谱学与光谱分析》期刊社

邮政编码：100081

联系电话：010-62181070, 62182998

电子信箱：chnghpxygpfx@vip.sina.com

修改稿专用邮箱：gp2008@vip.sina.com

网 址：<http://www.gpxygpfx.com>

