

# 一种场景自适应的双分支牛脸高效识别算法

焦杰<sup>1</sup>, 齐咏生<sup>1,2,3\*</sup>, 刘利强<sup>1,2,3</sup>, 李永亭<sup>1,2,3</sup>, 王朝霞<sup>1,2,3</sup>

(1. 内蒙古工业大学电力学院, 内蒙古呼和浩特 010051; 2. 大规模储能技术教育部工程研究中心, 内蒙古呼和浩特 010080;  
3. 内蒙古自治区高等学校智慧能源技术与装备工程研究中心, 内蒙古呼和浩特 010080)

**摘要:** 随着智慧牧业的高速发展, 牛脸识别已成为牧场智能化养殖的关键, 但牧场养殖环境复杂且动物的自主能动性差, 导致牛脸数据采集与识别过程会受到模糊、遮挡和光照等环境因素的严重干扰。针对此问题, 提出一种复杂场景自适应选择双分支牛脸高效识别算法。该算法首先设计了基于像素融合的数据增强策略, 通过 Beta 分布计算融合系数, 将牛的左右脸图像按融合系数进行像素级整合, 在丰富样本特征信息同时, 增强网络学习模糊和遮挡下的牛脸特征, 提升网络对复杂场景的泛化能力; 其次, 在主干特征提取网络中引入一种新型注意力机制 CDAA (Composite Dual-branch Adaptive Attention), 可随着场景信息变换, 自适应加强通道与空间注意力分支的权重, 提高网络在复杂场景下的特征筛选能力; 之后, 设计 FaceNet 与 U-LBP (Uniform Local Binary Patterns) 结合的双分支特征提取结构, 并将提取的特征向量实现自适应加权融合, 增加网络在过亮或过暗环境下的鲁棒性; 最后, 在损失函数中加入改进交叉熵损失 (Focal Loss), 根据场景信息复杂度动态调控权重系数, 实现对难易分类样本自主控制。为检测算法的有效性和实时性, 在特定数据集上进行消融试验, 与多种典型识别算法进行对比。实验结果表明, 提出的算法能很好满足实时性要求, 在开集测试集上准确率达到 87.53%, 识别速度达到 108 帧/s, 且在复杂场景下, 识别效果均优于对比网络。

**关键词:** 复杂场景; 图像融合; 双分支结构; 牛脸识别; 场景自适应

**基金项目:** 国家自然科学基金 (No. 62363029, No. 62241309); 内蒙古科技计划项目 (No. 2020GG0283, No. 2021GG164); 内蒙古自然科学基金 (No. 2022MS06018, No. 2021MS06018)

**中图分类号:** TP391.4

**文献标识码:** A

**文章编号:** 0372-2112(2024)09-3251-11

**电子学报 URL:** <http://www.ejournal.org.cn>

**DOI:** 10.12263/DZXB.20230829

## A Scene-Adaptive Dual-Branch Efficient Cattle Facial Recognition Algorithm

JIAO Jie<sup>1</sup>, QI Yong-sheng<sup>1,2,3\*</sup>, LIU Li-qiang<sup>1,2,3</sup>, LI Yong-ting<sup>1,2,3</sup>, WANG Zhao-xia<sup>1,2,3</sup>

(1. School of Electric Power, Inner Mongolia University of Technology, Hohhot, Inner Mongolia 010051, China;

2. Engineering Research Center of Large Energy Storage Technology, Ministry of Education, Hohhot, Inner Mongolia 010080, China;

3. Intelligent Energy Technology and Equipment Engineering Research Centre of Colleges and Universities in the Inner Mongolia Autonomous Region, Hohhot, Inner Mongolia 010080, China)

**Abstract:** With the rapid development of intelligent animal husbandry, cattle facial recognition has become a key aspect of intelligent farming in cattle ranches. However, due to the complexity of the ranching environment and the limited autonomy of animals, the process of collecting and identifying cattle facial data is severely affected by environmental factors such as blurriness, occlusion, and lighting. To address this issue, a complex scene-adaptive dual-branch efficient cattle facial recognition algorithm is proposed. This algorithm first designs a data augmentation strategy based on pixel fusion. By calculating fusion coefficients using the beta distribution, the left and right facial images of cattle are integrated at the pixel level, enriching the sample's feature information. Simultaneously, the algorithm enhances the network's ability to learn cattle facial features under blurriness and occlusion, improving its generalization ability to complex scenes. Furthermore, a novel attention mechanism called composite dual-branch adaptive attention (CDAA) is introduced into the main feature extraction network. This mechanism adaptively strengthens the weights of the channel and spatial attention branches as scene information changes, enhancing the network's feature selection ability in complex scenarios. Next, a dual-branch feature extraction structure combining

FaceNet and U-LBP (Uniform Local Binary Patterns) is designed. The extracted feature vectors are adaptively weighted and fused to increase the network's robustness in overly bright or dark environments. Finally, an improved cross-entropy loss (Focal Loss) is incorporated into the loss function. Weight coefficients are dynamically adjusted based on the complexity of the scene information to autonomously control the classification of difficult and easy samples. To evaluate the effectiveness and real-time performance of the algorithm, ablation experiments are conducted on a specific dataset, comparing it with various typical recognition algorithms. The experimental results indicate that the proposed algorithm effectively meets real-time requirements, achieving an accuracy of 87.53% on the open test set with a recognition speed of 108 frames per second. Moreover, in complex scenarios, the recognition performance of the proposed algorithm surpasses that of the comparative networks.

**Key words:** complex scenarios; image fusion; dual-branch structure; cattle face recognition; scene adaption

**Foundation Item(s):** National Natural Science Foundation of China (No.62363029, No.62241309); Inner Mongolia Science and Technology Project (No.2020GG0283, No.2021GG164); Inner Mongolia Natural Science Foundation (No.2022MS06018, No.2021MS06018)

## 1 引言

近年来,随着畜牧业的发展,自动化和信息化被应用于牛场养殖,通过精准化养殖<sup>[1,2]</sup>模式对牛只进行监测管理<sup>[3]</sup>.牛个体身份识别是实现精准化养殖的关键,传统的识别方法主要是通过佩戴基于射频识别技术(Radio Frequency Identification, RFID)<sup>[4]</sup>的电子耳标进行识别.但该方法存在易丢失和易伪造风险,且费时费力、识别的人力成本较高.随着生物识别<sup>[5]</sup>技术的不断发展,研究人员开始利用传统的特征提取算子<sup>[6]</sup>和数字图像处理技术<sup>[7,8]</sup>识别牛脸.Kumar等<sup>[9]</sup>提出一种多分辨率算法,使用加速稳健特征(Speeded Up Robust Features, SURF)<sup>[10]</sup>与局部二值模式(Local Binary Pattern, LBP)<sup>[11]</sup>从不同尺度的高斯金字塔级别提取特征,该方法对于数据集的要求较高,难以广泛应用.

目前,随着深度学习算法在计算机视觉<sup>[12,13]</sup>领域的广泛应用,基于深度学习的人脸识别<sup>[14]</sup>技术成为身份认证的研究热点.基于此,牛脸识别<sup>[15]</sup>相关课题也已得到深入研究,其成果具有广泛的应用前景.其中,Meng等<sup>[16]</sup>提出一种基于双通道卷积神经网络(Two-branch Convolutional Neural Network, TB-CNN)的牛脸识别模型.将采集到的两张不同角度的牛面部图像输入到不同通道的卷积神经网络中进行特征提取,对两通道的特征进行特征融合,并将全局平均池化层与分类器结合,对牛脸进行识别.Weng等<sup>[17]</sup>指出光照问题是影响奶牛面部识别准确性的最重要因素之一,光照不均匀将大大提高识别错误率,而且,姿态的不稳定性也会影响识别的准确性.基于此问题,论文利用残差网络思想,提出一种基于奶牛面部图像的改进卷积神经网络(Res\_5\_2Net).Gong等<sup>[18]</sup>基于ResNet-50<sup>[19]</sup>,并使用不同数量的sk\_bottleneck,通过整合多个感受野的信息,集成多个信息接收域,在多个尺度上提取面部特征.

此外,研究人员发现注意力机制可以帮助识别网

络关注图像的一些显著区域,依据重要性程度去提取对当前任务有用的特征,并抑制冗余特征.为此,Li等<sup>[20]</sup>基于通道注意力机制,提出一种新型注意力机制来提高遮挡场景下牛脸识别准确率,该算法以MobileNetV1<sup>[21]</sup>为主干网络,插入组成通道注意力的压缩模块和激励模块获得识别增益,即使中间部位的很多特征信息被遮挡,也能做到牛脸高效识别.杨蜀秦等<sup>[22]</sup>提出一种融合坐标信息的奶牛面部识别模型,在YOLOV4<sup>[23]</sup>网络的特征提取层和检测头部分别引入坐标注意力机制和包含坐标通道的坐标卷积模块,以增强模型对目标位置的敏感性.

以上基于深度学习的牛脸识别算法均在自建的数据集上取得了较好的识别效果.然而,上述研究均是在正常环境或单一场景的条件下实现,针对复杂场景下的牛脸识别研究较少,例如模糊、逆光、遮挡这复杂环境干扰都会对牛脸识别的准确率造成极大影响,而实际牛场还存在多种干扰并存的场景.

针对上述现实问题,本文提出一种场景自适应的双分支牛脸高效识别算法,该算法为应对复杂场景的干扰将场景信息数字化体现,并在各个模块充分融合场景信息.随着场景信息改变,算法进行自适应调整来应对干扰.算法首先在特征提取网络输入端将牛的左右脸图像进行像素融合,使用融合系数反映遮挡和模糊场景复杂度,以此模拟实际场景,增加训练样本的多样性,丰富样本目标特征信息.在网络特征提取方面,设计FaceNet与U-LBP(Uniform Pattern LBP)结合的双分支特征提取结构,将两个分支提取的特征向量依据场景逆光度自适应加权融合,增加网络在过亮或过暗环境下的鲁棒性.此外,在FaceNet主干网络中引入一种随场景信息变换,提供自适应权重系数的复合双分支注意力机制,做到不同复杂场景下使用动态加权实现注意力机制不同需求的自动调整,提升网络对复杂场景的适应能力.由于数据集经过数据增强后增加了大量复杂场景下的牛脸图像,导致难分类样本增多.为

此,在原三元组损失(Triplet Loss)的基础上加入改进交叉熵损失 Focal Loss,并依据前述场景信息复杂度指标调控权重系数,提高难分类样本对 Loss 的贡献度,实现对难易分类样本的自动控制.最后,为验证算法的可移植性,将算法部署到嵌入式平台 Jeston Xavier NX 进行了测试和验证.

## 2 牛脸识别算法

现有的牛脸识别算法在测试结果中均有较高的识别准确率,但在实际应用过程中由于牛棚内环境比较复杂,牛脸图像通常受到模糊、遮挡和光照变化的严重干扰,导致复杂场景下牛脸识别达不到满意的效果.其中,模糊场景多数由运动和悬浮沉积物引起,会弱化图像部分特征信息;遮挡场景是由栏杆和同类遮挡引起,会导致遮挡目标的某些细节或特征被隐藏;光照影响是由于目标所处位置光线差以及拍摄角度问题引起.为解决上述问题,本文在轻量化网络的基础上,提出一种复杂场景自适应选择的双分支牛脸高效识别算法,在保证算法轻量化的前提下,提升牛脸识别准确率.算法总体结构框图如图 1 所示.

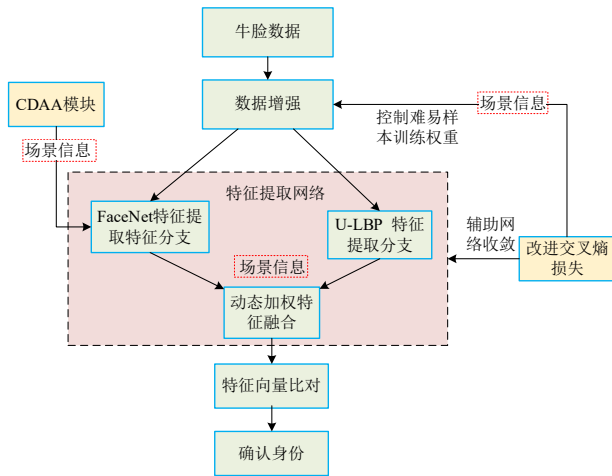


图 1 算法结构框图

由图 1 的结构框图可以看出,在网络结构方面,CDAA(Composite Dual-branch Adaptive Attention)模块融合场景信息提升网络对复杂场景中目标特征提取能力;双分支特征提取网络融合场景信息,并进行动态加权特征融合,抑制光照干扰.在损失函数方面,改进交叉熵损失融合场景信息,自适应调整权重系数,控制难易样本训练权重.场景信息作为复杂场景的重要评价指标,应用在 CDAA 注意力机制、双分支特征提取网络以及损失函数上,随着场景复杂度的变换,通过动态加权的方式,做到算法自适应调整,从而提高算法在实际复杂场景下的适应性和鲁棒性.

### 2.1 数据增强

牛脸数据在采集和识别过程中,常常由于动物不易控制和拍摄角度等原因,导致截取到的牛脸图像存在模糊和遮挡现象.实际数据集中多数为正常环境拍摄的牛脸,典型复杂场景下的牛脸图像样本偏少.为此,本文在识别任务中,设计图像融合的数据增强方法获取典型特征图像,并取得较好效果.类似图像融合的数据增强方法已被用来模拟水下目标的重叠、模糊和遮挡现象,以此提高水下目标检测的能力<sup>[24,25]</sup>.本文设计了一种像素融合的数据增强方法,将牛脸图像的左右脸通过 Beta 分布计算融合系数的方式进行像素整合,随着融合系数的改变,生成的融合图像所体现的场景复杂度也随之改变,这样能更好地模拟现实场景中目标由于遮挡和模糊造成的特征信息弱化和丢失,扩展训练样本的多样性.通过像素融合方法进行数据增强,一方面可在网络训练过程中同时计算两张图像数据,减少内存消耗;另一方面,有助于网络学习模糊和遮挡下的牛脸特征,提升网络对复杂场景的泛化能力.

为保证图像融合的有效性,首先统一牛脸图像大小,将牛的左右脸图像进行不失真的图像放缩,转化为 160×160 大小.然后,使用 Beta 分布计算融合系数<sup>[26]</sup>的方法进行图像融合,融合后图像大小与图像不失真放缩后的大小一致.最后,考虑到现实环境中存在复杂场景和正常场景两种情况,若训练样本全部使用左右脸融合后的图像,会导致网络无法学习正常场景中牛的左右脸特征,导致识别效果不佳.为此,设计阈值控制融合样本的数量,随机函数生成纯小数,阈值设为  $\tau$ ,当大于等于  $\tau$  时生成融合训练样本. Beta 分布计算融合系数的方法计算过程如下:

$$\Gamma(a) = \int_0^1 t^{a-1} \cdot e^{-t} dt$$

$$\Gamma(b) = \int_0^1 t^{b-1} \cdot e^{-t} dt \quad (1)$$

$$\lambda = \text{Beta}(a, b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$$

$$z_i = \lambda x_i + (1 - \lambda)y_i$$

其中,  $x_i, y_i$  为同一头牛的不同左脸训练样本和右脸训练样本;  $z_i$  为融合训练样本;  $\lambda$  是由参数为  $a, b$  的 Beta 分布计算出的融合系数,  $a, b$  的值由实验测试结果确定; Beta 函数的计算用于归一化概率密度函数,通常依赖于伽玛函数  $\Gamma$ , 通过结合  $\Gamma(a)$  和  $\Gamma(b)$  两个伽玛函数的结果来计算  $\text{Beta}(a, b)$ , 得到融合系数.

### 2.2 特征提取模块

FaceNet 主干网络使用 Inception-ResNet<sup>[27]</sup>起到了提取特征的作用,为方便算法后期的移植,本文将主干特征提取网络替换为 MobileNetV2<sup>[28]</sup>.传统的 SE(Squeeze-and-Excitation)通道注意力机制<sup>[29]</sup>通过“挤

压”获得每个通道的全局统计信息,通过“激励”学习每个通道的权重,最后将每个通道的特征与对应的注意力权重相乘,得到加权后的特征表示.空间注意力机制可以使模型自动选择性地聚焦在输入图像的不同位置上,从而捕获重要的空间位置特征.单独的通道注意力机制,更多关注的是通道信息,却忽略了空间信息,对于一些逆光场景鲁棒性较差,增加空间注意力机制后,可以结合二者的优势在一定程度上抑制环境因素的干扰.但仅仅将通道注意力机制和空间注意力机制简单地进行组合,会导致注意力机制在特征提取过程中无法区分通道和空间注意力的重要性.尤其是处于不同场景下的图像包含的信息更为复杂,网络需要关注的侧重点也是不同的.

为此,引入一种新型注意力机制 CDAA,设计通道和空间注意力并行结构,并通过构造评价指标设计自适应权重系数,有效加强相应注意力分支的权重.相较于传统注意力机制,CDAA 注意力机制能有效融合图像所处复杂场景信息,对通道和空间注意力作用进行动态调整,提高网络在复杂场景下的特征提取能力.如表 1 所示为主干特征提取网络结构.

CDAA 注意力模块以 ECA(Efficient Channel Attention)注意力机制为基础,引入空间位置信息,形成通道和空间并行的双分支注意力机制.同时,使用动态加权方法,为通道注意力机制和空间注意力机制两个分支赋予权重系数  $\alpha$  和  $\beta$ ,该权重系数代表两个注意力分支的加强程度,由逆光度和融合系数指标确定.其中,融合系数指标按照 2.1 节式(1)计算给出,而逆光度指标的计算公式描述如下:

表 1 主干特征提取模块结构

Input	Operator	$t$	$c$	$n$	$s$
$160^2 \times 3$	Conv2d	—	32	1	2
$160^2 \times 3$	CDAA	—	32	1	—
$80^2 \times 32$	Bottleneck	1	16	1	1
$80^2 \times 16$	Bottleneck	6	24	2	2
$40^2 \times 24$	Bottleneck	6	32	3	2
$20^2 \times 32$	Bottleneck	6	64	4	2
$10^2 \times 64$	Bottleneck	6	96	3	1
$10^2 \times 96$	Bottleneck	6	160	3	2
$5^2 \times 160$	Bottleneck	6	320	1	1
$5^2 \times 620$	Conv2d $1 \times 1$	—	1 280	1	1
$5^2 \times 1 280$	Avgpool $7 \times 7$	—	—	1	—
$1 \times 1 \times 1 280$	Conv2d $1 \times 1$	—	$k$	—	—

$$\eta = (Q_{\text{dark}} + Q_{\text{bright}}) / Q_{\text{total}} \quad (2)$$

逆光拍摄的图像会呈现出一种高对比度和高亮度的状态,根据图像高对比度和高亮度的属性,可以把图像转换为灰度图.通过对大量逆光场景的灰度直方图分布结果统计可知,灰度值处于  $0 \sim 30$  像素的个数,表示极暗像素;灰度值处于  $220 \sim 255$  的像素个数,表示极亮像素.本文经过统计的方法得到极暗像素与极亮像素个数,最后再与图像的总像素个数做比值就可以得到一个逆光度的评价指标.极亮像素和极暗像素的个数越多,代表图像的逆光度越大.式(2)中  $Q_{\text{dark}}$  为灰度值处于  $0 \sim 30$  的像素个数,表示极暗像素; $Q_{\text{bright}}$  为灰度值处于  $220 \sim 255$  的像素个数,表示极亮像素; $Q_{\text{total}}$  为图像总像素个数, $\eta$  为图像逆光度.CDAA 模块结构如图 2 所示,具体描述见后文.

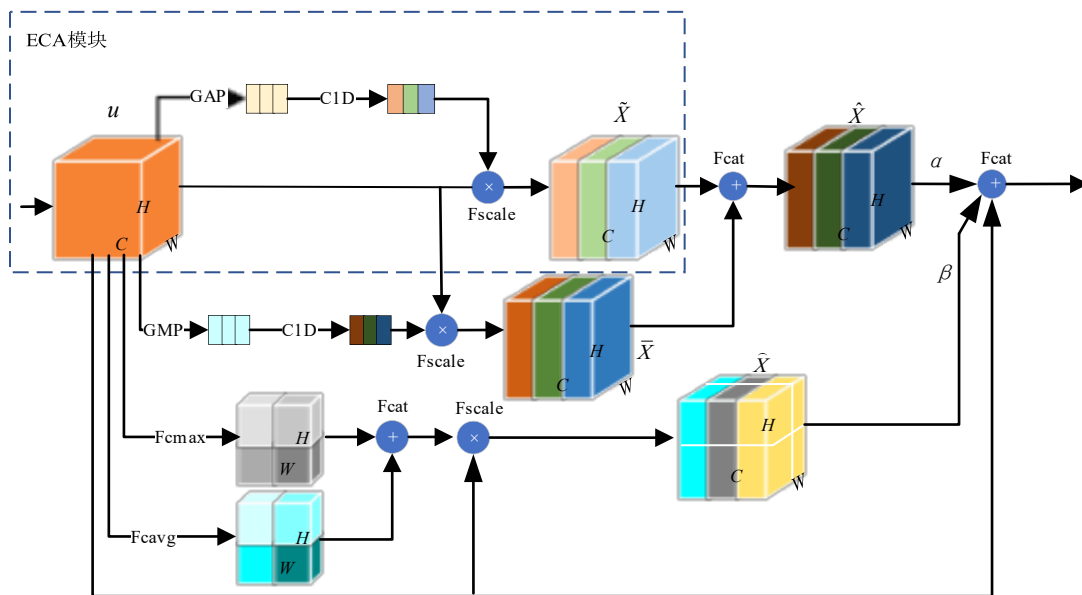


图 2 CDAA 注意力模块结构图

### 2.2.1 通道注意力分支

该分支通过融合全局平均池化和全局最大池化特征实现通道注意力分支,使网络可以用全局信息来选择性地强调信息特征,并抑制冗余特征,对目标边缘特征具有选择性地突出,对处理模糊和遮挡场景下的图像有较好的效果.全局平均池化特征通道具体计算公式如下:

$$z_c = \text{GAP}(u_c) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W u_c(i, j) \quad (3)$$

$$\omega_z = \sigma(\text{C1D}_k(z)) \quad (4)$$

$$\tilde{X}_c = \text{F}_{\text{scale}}(u_c \cdot \omega_z) \quad (5)$$

式(3)中  $z \in \mathbb{R}^c$  为不降维的聚合特征,由压缩输出  $u$  的空间维度  $H \times W$  生成,  $\text{GAP}$  (Global Average Pooling) 为全局平均池化;式(4)中  $\text{C1D}$  (Conv1D) 表示一维卷积,  $k$  为卷积核大小,  $\sigma$  为 Sigmoid 激活函数.  $\text{ECA}$  模块使用全局平均池化后,确定自适应卷积核大小  $k$  并进行一维卷积,然后 Sigmoid 函数学习各个通道的权值得到归一化后的标量  $\omega_z$ ,经过式(5)中  $\text{F}_{\text{scale}}$  将归一化的权值乘到各个通道得到特征  $\tilde{X}_c$ ,其中  $\tilde{X}_c \in \tilde{X} = \{\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_c\}$ . 全局最大池化特征通道具体计算公式如下:

$$s_c = \text{GMP}(u_c) = \max(u_c) \quad (6)$$

$$\omega_s = \sigma(\text{C1D}_k(s)) \quad (7)$$

$$\bar{X}_c = \text{F}_{\text{scale}}(u_c \cdot \omega_s) \quad (8)$$

$$\hat{X} = \text{F}_{\text{cat}}(\tilde{X}, \bar{X}) \quad (9)$$

$\text{GMP}$  (Global Max Pooling) 为全局最大池化,与全局平均池化特征通道的计算过程同理,  $\bar{X}_c$  为该模块提取的特征,其中  $\bar{X}_c \in \bar{X} = \{\bar{X}_1, \bar{X}_2, \dots, \bar{X}_c\}$ ,之后与  $\text{ECA}$  模块进行特征融合构成完整的通道注意力分支.式(9)中  $\text{F}_{\text{cat}}$  代表特征融合操作.

### 2.2.2 通道注意力分支

空间注意力分支可用来赋予特征信息不同位置的重要性,突出重要的空间位置,抑制空间信息冗余,该分支对逆光或强光背景下的图像特征提取敏感,可有效抑制强光背景或逆光因素.具体计算公式如下:

$$p_{(i,j)} = \text{F}_{\text{cmax}}(u_{(i,j)}) = \max(u_{(i,j)}) \quad (10)$$

$$q_{(i,j)} = \text{F}_{\text{cavg}}(u_{(i,j)}) = \frac{1}{c} \sum_{m=1}^c u_{(i,j)} \quad (11)$$

$$\hat{X}_c = \text{F}_{\text{scale}}(u_c, \delta(\text{F}_{\text{cat}}(p_{(i,j)}, q_{(i,j)}))) \quad (12)$$

式(10)中  $\text{F}_{\text{cmax}}$  为通道最大池化,式(11)中  $\text{F}_{\text{cavg}}$  为通道平均池化,二者都是在空间维度  $H \times W$  中沿通道方向进行运算,  $u_{(i,j)}$  由压缩空间维度  $H \times W$  中某个位置上所有通道得到,经过通道最大池化和通道平均池化分别提取特征  $p_{(i,j)}$  和  $q_{(i,j)}$ .式(12)中  $\hat{X}_c$  为模块提取特征,其中  $\hat{X}_c \in \hat{X} = \{\hat{X}_1, \hat{X}_2, \dots, \hat{X}_c\}$ .

### 2.2.3 自适应加权算法

为有选择性地区分通道注意力机制和空间注意力机制在不同复杂场景下的作用,采用自适应加权方法有效区分空间和通道信息的重要性,使网络针对不同场景实现注意力机制强化学习.利用通道和空间信息将上下文信息编码为局部特征的同时,结合当前场景的复杂性,强化相应通道和空间注意力分支的权重,使网络针对不同场景实现动态调整,以此来增加对复杂场景的表达能,解决传统注意力模块针对不同复杂场景适应性差的问题.如当图像处于模糊和遮挡场景中,则加强通道注意力分支权重,提高网络对于牛脸边缘特征以及未遮挡部分特征的提取能力;而当图像处于具有强背景逆光场景时,加强空间注意力分支权重,对牛脸特征进行聚焦,增加背景与目标特征之间的判别性.

通过式(9)和式(12)分别计算通道注意力机制和空间注意力机制两个分支对应的特征信息  $\hat{X}$  和  $\bar{X}_c$ ,在进行特征融合之前,引入自适应权重系数  $\alpha$  和  $\beta$  分别代表通道注意力分支和空间注意力分支的权重.图像所处的场景复杂度由融合系数与逆光度之和表示.而  $\alpha$  的取值由融合系数在整个场景复杂度中所占的比例确定; $\beta$  的取值由逆光度在整个场景复杂度中所占的比例确定.  $\alpha$  和  $\beta$  的计算公式如下:

$$\alpha = \frac{\lambda}{\eta + \lambda} + 1 \quad (13)$$

$$\beta = \frac{\eta}{\eta + \lambda} + 1 \quad (14)$$

式(13)和式(14)中,若  $\eta + \lambda = 0$ ,则表示不参与运算,此时  $\alpha$  和  $\beta$  同时置为 1,代表网络使用原注意力机制,不进行分支加强.注意力分支加强后进行特征融合,具体表达如下:

$$X = \text{F}_{\text{cat}}(u, (\hat{X} \cdot \alpha), (\bar{X}_c \cdot \beta)) \quad (15)$$

其中,  $X$  为注意力分支加强后的融合特征.

## 2.3 双分支结构

### 2.3.1 FaceNet 特征提取特征分支

FaceNet 是一种基于 CNN 与三元组损失 (Triplet Loss) 的高效人脸识别算法,该算法识别原理主要是利用不同人脸图像的低耦合性及相同人脸图像在不同角度下的高内聚性.通过深度卷积网络进行特征提取,获得一个长度为 128 的特征向量,经过 L2 标准化后将不同的人脸特征向量转化为同一数量级,方便比较.其结构如图 3 所示<sup>[30]</sup>.

### 2.3.2 U-LBP 特征提取分支

光照变化对牛脸识别准确率的影响多数由光照变化较大导致图像处于一种高对比度和高亮度的状态,

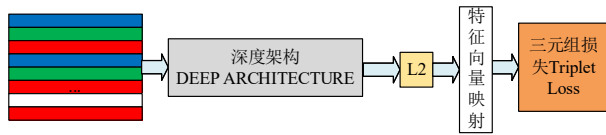


图3 FaceNet结构图

这种图像中的牛脸常常以过亮或过暗的形式输入识别网络,使得网络提取的牛脸信息差异过大,没法准确识别.为此,本算法引入U-LBP特征提取分支解决上述问题.

U-LBP是一种改进LBP(Local Binary Patterns)的图像特征提取方法,可以对原始LBP算法进行降维,有效减少数据量的同时保留较好的图像信息.当LBP对应的二进制数最多经过两次0与1之间的跳变,则标记其为等价模式类,除此外都标记为另一类.通过这种方式可以将模式数量由原来的256种减少为 $P(P-1)+2$ 种, $P$ 为邻域像素个数.对于邻域内8个像素点来说,等价模式数量为58,再加其他类构成59类.

特征提取通过将 $160 \times 160$ 的图像划分为8行8列共64块区域,并提取每个区域的直方图进行顺序排列得到大小为 $n \times 64$ 维的特征向量,其中, $n$ 代表LBP特征模式数量.特征向量提取过程如图4所示.

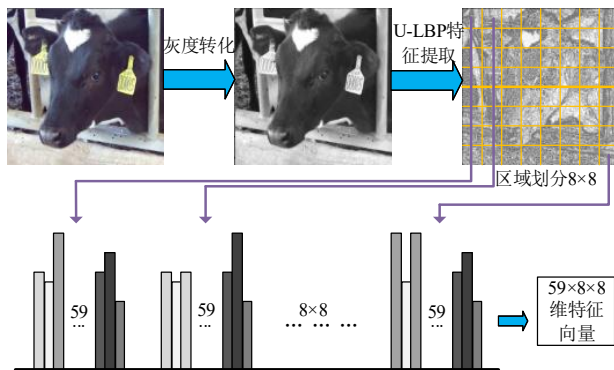


图4 U-LBP特征提取示意图

128维特征向量已经被广泛验证并证明在人脸比对和身份验证任务中表现出很高的准确性和鲁棒性.为与FaceNet特征提取特征分支得到的向量进行特征融合,并找到最具信息量的特征子空间,需要将U-LBP提取的特征向量使用主成分分析(Principal Component Analysis, PCA)进行降维,在保留原始特征判别能力的基础上,去除冗余信息.

### 2.3.3 动态加权特征融合

双分支特征提取结构在FaceNet网络结构的基础上引入U-LBP特征提取分支,将双分支提取的特征向量进行动态加权融合.由于U-LBP具有二进制编码的方式,并将LBP的输出模式限定为等价的统一模式,有效限制了输出的种类,减少了由于光照变化引起的噪

声.这些统一模式对于光照变化具有一定的不变性.同时,U-LBP特征中的统一模式具有相对均匀的分布.这意味着不同统一模式的概率分布相对平均,使得在不同光照条件下,不同区域的图像具有相似的统一模式分布,对于图像中的局部噪声也相对不敏感.因此,U-LBP特征向量对于光照强度的变化相对较为稳定.将其与FaceNet网络提取的特征加权融合,一方面增强牛脸特征信息,另一方面提升网络整体的鲁棒性,可有效应对牛脸图像由于光照变化造成的识别准确率偏低问题.双分支结构如图5所示.

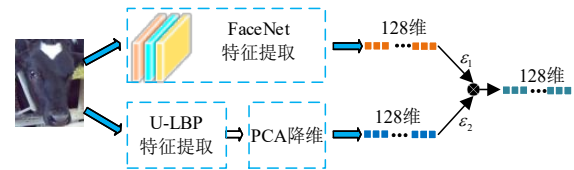


图5 双分支结构示意图

融合过程中如果将双分支的特征向量等权重融合,可能导致某一分支包含的更重要或更有区分性的信息被稀释,从而减弱网络的性能.而且双分支结构是针对不同的任务进行设计,尤其对于不同场景的切换,双分支需要根据输入图像的特性和处理任务的需求,动态调整每个分支的重要性.例如,在处理常见场景中的牛脸图像时,网络会适当降低U-LBP特征提取分支权重.这种灵活的适应性确保了网络可以更有效地处理不同的视觉任务.而等权重融合的网络将无法区分双分支结构的重要性,会使网络难以适应这种不同的需求.

为此,本算法引入动态融合系数 $\varepsilon_1$ 和 $\varepsilon_2$ 来自适应调整双分支特征向量融合时所占权重大小.动态融合系数的值由输入图像的逆光度确定,由于U-LBP特征提取分支对于光照干扰较大时,具有显著作用.当逆光度小于0.5时,牛脸图像受光照影响较小,此时随着逆光度不断增大,FaceNet特征提取分支中的CDAA注意力机制可以有效应对这种较小的干扰,因此,不断加强FaceNet特征向量融合系数,削弱U-LBP特征向量融合系数.当图像逆光度大于0.5时,代表图像极亮像素和极暗像素的个数较多,牛脸所处场景的光照与正常光照相差较大,此时,加强U-LBP特征向量融合系数.因此,设计如下计算公式进行双分支特征向量融合:

$$\varepsilon_1 = \begin{cases} 0.5 \times (1 - \eta), & \eta \geq 0.5 \\ 0.5 \times (1 + \eta), & \eta < 0.5 \end{cases} \quad (16)$$

其中, $\varepsilon_1$ 为FaceNet特征向量的加权系数.

$$\varepsilon_2 = 1 - \varepsilon_1 \quad (17)$$

其中, $\varepsilon_2$ 为U-LBP特征向量的加权系数.

$$\varphi = \varphi_1 \times \varepsilon_1 + \varphi_2 \times \varepsilon_2 \quad (18)$$

其中,  $\phi_1$  为图像经过 FaceNet 网络提取的特征向量,  $\phi_2$  为图像经过 U-LBP 算法提取的特征向量,  $\phi$  为加权融合后标准 128 维向量.

## 2.4 损失函数设计

由于实际采集的数据中易分类的样本占多数, 难分类的样本占少数, 仅依靠 FaceNet 网络模型的原损失函数 Triplet Loss 无法控制难分类与易分类样本的训练权重. 为加强网络对难分类样本的识别, 同时削弱易分类样本的权重, 本算法设计了一种改进交叉熵损失 Focal Loss 来控制难易分类样本训练权重的策略, 并结合 Triplet Loss 作为整体损失, 以此提升算法识别的准确率. 其数学表达式为:

$$FL(P_i) = -(1 - P_i)^\gamma \log(P_i) \quad (19)$$

其中,  $P_i$  为模型对样本的预测概率,  $P_i$  值越大, 代表样本越容易分类,  $P_i$  值越小, 代表样本越难分类.  $\gamma$  为调控权重系数,  $(1 - P_i)^\gamma$  用于降低易分类样本权重的因子, 当  $P_i$  值趋于 1 时, 该因子的值接近 0, 样本对 Loss 的贡献度小; 当  $P_i$  值趋于 0 时, 该因子的值接近 1, 样本对 Loss 的贡献度大, 以便网络更好地学习难分类样本. 由于调控权重系数  $\gamma$  可以控制易分类样本的权重降低速度, 本算法对其进行优化, 使用本文提出的融合系数  $\lambda$  与逆光度  $\eta$  确定  $\gamma$  的取值. 具体实现方法如下:

$$\gamma = \begin{cases} 1 + \lambda + \eta, & \text{if } \lambda \leq 0.5 \\ 1 + (1 - \lambda) + \eta, & \text{otherwise} \end{cases} \quad (20)$$

其中, 当  $\lambda$  取值小于 0.5 时, 图像右脸信息所占权重较大, 当  $\lambda$  取值大于 0.5 时, 图像左脸信息所占权重较大. 因此, 当样本图像的融合系数越接近 0.5 且逆光度较大时, 牛脸识别难度较大. 此时, 动态调控权重系数  $\gamma$  (增大), 可以有效提高难分类样本对 Loss 贡献度, 提高网络对难分类样本的学习能力.

## 3 实验与分析

为满足后期算法可移植性的要求, 算法首先在服务器进行训练. 训练环境均在 Ubuntu18.04 Linux 操作系统下完成, 硬件环境配置为: i7-11700@2.50 GHz CPU、NVIDIA GeForce RTX 3090 显卡、运行内存 24 GB. 软件环境配置为: Cuda11.0, Cudnn8.1.0 加速, 基于 pytorch 1.8.1 深度学习框架. 模型训练策略: 使用 Adam 优化器进行优化, 总迭代次数 1 100 次, 迭代动量系数 0.9, 批量大小 32, 权重衰减系数 0.000 5, 模型初始学习率设置 0.01. 之后, 将训练好的模型移植到嵌入式主板 Jetson Xavier NX 进行测试.

### 3.1 数据集制作

本文算法针对复杂场景下的牛脸识别任务, 为了充分验证本文算法的有效性和真实性, 数据集中的牛脸均采集自内蒙古欧太牧场, 使用三路摄像头采集牛

只的左、右、正三个角度的面部数据, 以此保证数据的多样性. 首先, 使用刻录机与三路摄像头采集 20 s 左右的牛脸视频, 将采集到的视频数据存储在移动硬盘中; 接着, 通过视频剪辑软件将采集的视频数据进行剪辑, 只保留视频中拍到完整牛脸的部分; 之后使用 python 算法对保留的视频进行分帧处理, 并筛选去重, 防止相似度高的图像重复出现; 最后通过 YOLOV7 算法精确地检测分帧后图像中的牛脸, 将其裁剪保存, 形成最终的牛脸识别数据集.

本次研究采集 280 头肉牛的牛脸数据进行模型训练, 其中每头牛包含左、右、正三个角度各 20 张, 总计 16 800 张图片. 为增加样本多样性, 通过翻转、放缩、色域变换、改变对比度等数据增强的方法对数据集进行扩充, 扩充后的图片数量为 22 500. 将数据集中栏杆遮挡的牛脸、牛与牛之间遮挡的牛脸、受到不同强度光照的牛脸、由于牛头摆动造成图像模糊的牛脸以及部分正常的牛脸图像选取 500 张作为测试集, 生成的相同牛脸与不同牛脸的对照组有 2 236 组. 剩余图片按照 0.9:0.1 的比例划分为训练集和验证集. 此外, 再采集不参与训练的 98 头牛共计 500 张牛脸图像作为开集测试集. 数据具体划分数量如表 2 所示.

表 2 牛脸数据集划分表

训练集	验证集	闭集测试集	开集测试集
19 800	2 200	500	500

### 3.2 性能评价标准

本算法使用识别准确率 Acc (Accuracy)、验证率 Val (Validation Rate) 和每秒帧数 FPS (Frames Per Second) 作为模型的性能评价指标. 其计算公式如下:

$$\text{Acc} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (21)$$

$$\text{Val} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (22)$$

TP、TN、FP、FN 分别代表真正例、真反例、假正例、假反例. Acc 是模型正确判断同一牛脸样本与正确判断不同牛脸样本在总样本中所占的比例. Val 是模型正确地将验证集中属于目标身份的牛脸识别为正例的比例, 即正确判断同一牛脸样本在所有同一牛脸样本中所占的比例. FPS 代表了牛脸识别的速度, 即每秒内识别图片的数量, 可以通过识别一张图片所需的时间再取倒数得到 FPS 值.

### 3.3 算法测试

#### 3.3.1 像素融合系数实验

为验证基于像素融合数据增强策略的有效性, 测试 Beta 分布中由不同参数  $a$ 、 $b$  确定的融合系数在自制牛脸数据集上的识别效果. 本文使用 5 组不同的 Beta 分布进行测试, 测试结果如表 3 所示.

表3 Beta分布测试结果

Beta分布	开集 Acc/%
(0.3,0.3)	76.32
(0.5,0.5)	77.35
(0.7,0.7)	76.96
(1,1)	75.63
(2,2)	75.28

由表3可知,Beta(0.5,0.5)确定的融合系数,识别准确率最高.此时,融合训练样本的训练误差和泛化差距达到最佳.因此,实验中Beta分布计算融合系数的参数设置为Beta(0.5,0.5).

### 3.3.2 消融实验

为验证本文算法改进的有效性,首先在自制的牛脸数据集上进行消融实验,分别使用开集和闭集测试集进行验证,实验结果如表4所示.

表4 消融实验

数据增强	注意力机制	U-LBP特征提取分支	改进损失函数	FPS/(帧/s)	Acc/%		Val/%	
					闭集	开集	闭集	开集
				126	97.61	79.63	97.12	75.22
√				126	98.92	83.61	98.32	77.35
√	√			115	98.38	84.01	99.85	79.08
√	√	√		110	99.24	85.28	98.54	78.22
√	√	√	√	108	99.32	87.53	99.46	83.12

由表4可知,使用基于像素融合的数据增强方法,将牛的左右脸图像融合,网络在训练过程中不仅可以有效学习复杂场景下牛脸特征,而且融合后图像包含了部分左右脸信息,在闭集和开集测试集上识别准确率都有明显提升.增加CDAA注意力机制后,主干网络在复杂场景下特征提取能力明显提升,开集验证集上的识别准确率也提高了.增加LBP特征提取分支后,与改进的FaceNet主干特征提取网络构成双分支结构,并将两个分支提取的有效特征向量自适应加权融合,网络在过亮或过暗环境下的鲁棒性明显改善,开集测试集上的识别准确率提升了5.65%,闭集测试集上识别准确率提升了1.63%.此外,将改进交叉熵损失Focal Loss引入原损失函数,增加网络对难分类样本的学习能力,开集测试集上识别准确率提升7.9%,闭集测试集上识别准确率提升1.71%.算法改进后的识别速度FPS值与未改进前的网络非常接近,仅牺牲了较少的识别速度,大幅提升了识别准确率.

由于闭集测试集中的牛脸图片为参加训练的牛

脸,算法改进前后识别准确率提升不高,均接近100%.开集测试集中的牛脸图片为不参与训练的牛脸,更能反映算法在实际应用中的泛化能力.为体现改进算法在复杂场景下的有效性,选取不同场景下的牛脸图像输入网络提取特征向量,并对向量间的欧氏距离进行对比,确定其是否为同一头牛.测试结果如表5所示.

由表5三组复杂场景对比测试结果可以看出,在模糊和遮挡场景下,网络增加图像融合的数据增强方法和CDAA注意力机制模块后,牛脸特征向量的欧式距离均不同程度地减小.由于第二组图像牛脸遮挡面积较大,而且两张图像的光照差异较大,导致两张图像特征向量的欧式距离较大.对于第三组有光照干扰的牛脸图像,网络增加双分支结构后,牛脸特征向量的欧式距离大幅缩小.在优化损失函数后,三组图像特征向量的欧氏距离均达到最小.综上,由消融实验结果与表5牛脸特征向量欧氏距离实验可充分验证本文算法的有效性,从另一个侧面也反映出本文算法在复杂场景下具有较好的鲁棒性.

表5 不同场景对比测试

网络结构	模糊干扰(欧式距离)		遮挡干扰(欧式距离)		光照干扰(欧式距离)	
	0	150	0	150	0	150
MobileNetV2-FaceNet	0.680	0.680	1.109	1.109	0.665	0.665
+图像融合	0.589	0.589	1.077	1.077	0.651	0.651
+注意力机制	0.443	0.443	1.060	1.060	0.590	0.590
+双分支结构	0.436	0.436	0.951	0.951	0.268	0.268
+损失函数	0.428	0.428	0.720	0.720	0.155	0.155

### 3.3.3 模型对比测试

为进一步评估模型的优劣,将本文算法与其他当前典型的识别算法进行对比,包括:FaceNet、MobileFaceNet、SK\_ResNet<sup>[18]</sup>和ECCSA-MFC<sup>[31]</sup>。其中,FaceNet为应用最为广泛的人面部识别算法,主干特征提取网络使用Inception-ResNet;MobileFaceNet为改进后的轻量化面部识别网络;SK\_ResNet为ResNet-50的改进算法,使用不同数量的SK\_Bottleneck来融合多个感知信息,并在多个尺度上提取牛的面部特征;ECCSA-MFC算法在MobileFaceNet的主干网络引入融合空间信息的高效通道注意力机制,增加特征提取范围,并使用余弦退火进行动态学习率调优。对比测试结果如表6所示。

由表6可知,本文算法在识别精度上显著高于其他识别算法,由于本文算法在改进过程中引入CDA注意力机制,并设计了双分支结构,导致模型大小和参数量略微增加,识别速度略微减小。MobileFaceNet算法和ECCSA-MFC算法的模型大小和参数量相对较小,但无法解决复杂场景的干扰,导致识别准确率偏低,无法应用于实际环境。相比模型较为复杂的FaceNet,本文算法在保证网络轻量化的同时,拥有较快的识别速度,识别准确率比FaceNet网络高出2.02个百分点。SK\_ResNet由于使用了多个SK\_Bottleneck来融合多个感知信息,参数量、计算量以及模型大小都高于本文算法,而且识别准确率与识别速度稍低于本文算法。综上,对比结果可以看出,本文算法在识别准确率上具有明显优势,能更好适应各类典型复杂场景。

表6 不同模型对比结果

算法名称	参数量	计算量 /GFLOPS	模型大小	开集 Acc/%	FPS/ (帧/s)
FaceNet	22.79 M	2.85	87.4 MB	85.51	83
MobileFaceNet	0.99 M	0.390 45	4 MB	78.63	126
SK_ResNet	5.91 M	1.65	22.56 MB	84.22	97
ECCSA-MFC	1.20 M	0.446 55	4.42 MB	81.13	112
本文算法	2.16 M	0.327 62	15 MB	87.53	108

### 3.4 算法移植测试

为了保证算法在提升检测精度的同时满足移植要求,将本文算法训练好的模型移植到嵌入式平台Jetson Xavier NX上进行测试,测试过程中使用RetinaFace面部检测算法与本文算法搭建牛脸识别平台,检测算法检测到牛脸后将该区域裁剪,输入到后端识别网络进行牛脸识别。识别过程选取处于常见复杂场景的98只牛进行牛脸识别,识别准确率Acc达到了84.7%,由于嵌入式平台的算力受限以及在识别算法的基础上加入了面部检测算法,导致识别速度FPS下降为67帧/s。选取如图6所示的多个典型复杂场景下的具体识别效果图作为具体样例来体现算法在复杂场景下应用的真实性。



图6 复杂场景下牛脸识别结果

图6中横向图片为一组,每张图片左下角的标号为牛对应的身份编号。第一组图片为模糊场景下牛脸识别结果;第二组图片为遮挡环境下牛脸识别结果,同时,该组图片选取的两头牛具有较高的相似度;第三组图片为光照干扰场景下的牛脸识别结果。由三组图片左下角识别结果与其耳标编号相验证,可知牛脸识别结果均正确,验证了本文算法模型嵌入式平台移植的有效性,在保证实时性要求的同时达到了较高的识别准确率。

## 4 结论

本文充分考虑了牛场实际环境的特点以及在该环境中牛脸识别受到的干扰因素,在满足移植性要求的前提下,提出了一种结合图像融合与双分支结构的场景自适应牛脸识别算法。

(1)针对模糊和遮挡现象,将牛脸图像的左右脸通过Beta分布计算融合系数的方式进行像素融合,增加训练样本的多样性,丰富样本目标的特征信息。同时,帮助网络学习现实场景中遮挡和模糊的现象。

(2)在主干特征提取网络引入新型注意力机制CDA模块,自适应调节通道注意力机制和空间注意力机制在不同复杂场景下的权重,提高复杂场景下牛脸特征提取能力。

(3)设计FaceNet与U-LBP双分支特征提取结构,利用自适应加权聚合的方法将两个分支提取的特征向

量有效融合,解决光照因素的影响。

(4)在三元组损失中引入改进交叉熵损失 Focal Loss,根据场景信息动态调控权重系数,提高难分类样本对 Loss 贡献度,从而控制难易分类样本权重,提高网络对难分类样本的学习能力。

#### 参考文献

- [1] MORRONE S, DIMAURO C, GAMBELLA F, et al. Industry 4.0 and precision livestock farming (PLF): An up to date overview across animal productions[J]. *Sensors*, 2022, 22(12): 4319.
- [2] MANCUSO D, CASTAGNOLO G, PORTO S M C. Cow behavioural activities in extensive farms: Challenges of adopting automatic monitoring systems[J]. *Sensors*, 2023, 23(8): 3828.
- [3] MAHMUD M S, ZAHID A, DAS A K, et al. A systematic literature review on deep learning applications for precision cattle farming[J]. *Computers and Electronics in Agriculture*, 2021, 187: 106313.
- [4] WILLIAMS L R, FOX D R, BISHOP-HURLEY G J, et al. Use of radio frequency identification (RFID) technology to record grazing beef cattle water point use[J]. *Computers and Electronics in Agriculture*, 2019, 156: 193-202.
- [5] WANG M, YIN X F, ZHU Y M, et al. Representation learning and pattern recognition in cognitive biometrics: A survey [J]. *Sensors*, 2022, 22(14): 5111.
- [6] SHU X, TANG H, HUANG S C. Face spoofing detection based on chromatic ED-LBP texture feature[J]. *Multimedia Systems*, 2021, 27(2): 161-176.
- [7] 吴靖, 叶晓晶, 黄峰, 等. 基于深度学习的单帧图像超分辨率重建综述[J]. *电子学报*, 2022, 50(9): 2265-2294.  
WU J, YE X J, HUANG F, et al. A review of single image super-resolution reconstruction based on deep learning[J]. *Acta Electronica Sinica*, 2022, 50(9): 2265-2294. (in Chinese)
- [8] BINDAL N, GHUMAAN R S, SOHI P J S, et al. A systematic review of state-of-the-art noise removal techniques in digital images[J]. *Multimedia Tools and Applications*, 2022, 81(22): 31529-31552.
- [9] KUMAR S, SINGH S K, SINGH A K. Muzzle point pattern based techniques for individual cattle identification[J]. *IET Image Processing*, 2017, 11(10): 805-814.
- [10] PAUL M, KARSH R K, AHMED TALUKDAR F. Image hashing based on shape context and speeded up robust features (SURF)[C]//2019 International Conference on Automation, Computational and Technology Management (ICACTM). Piscataway: IEEE, 2019: 464-468.
- [11] LIU Y J. Athletes' throwing action recognition method based on PCA-LBP algorithm[J]. *International Journal of Computational Intelligence Studies*, 2023, 12(1/2): 130-141.
- [12] 邵志文, 周勇, 谭鑫, 等. 基于深度学习的表情动作单元识别综述[J]. *电子学报*, 2022, 50(8): 2003-2017.  
SHAO Z W, ZHOU Y, TAN X, et al. Survey of expression action unit recognition based on deep learning[J]. *Acta Electronica Sinica*, 2022, 50(8): 2003-2017. (in Chinese)
- [13] 程旭, 宋晨, 史金钢, 等. 基于深度学习的通用目标检测研究综述[J]. *电子学报*, 2021, 49(7): 1428-1438.  
CHENG X, SONG C, SHI J G, et al. A survey of generic object detection methods based on deep learning[J]. *Acta Electronica Sinica*, 2021, 49(7): 1428-1438. (in Chinese)
- [14] 汪宇玲, 陈立, 黎明, 等. 基于迹变换和旋转增量调制特征的模糊人脸识别[J]. *电子学报*, 2021, 49(12): 2437-2448.  
WANG Y L, CHEN L, LI M, et al. Rotational delta modulation feature and its application in blurry face recognition based on trace transform[J]. *Acta Electronica Sinica*, 2021, 49(12): 2437-2448. (in Chinese)
- [15] FUENTES S, GONZALEZ VIEJO C, TONGSON E, et al. The livestock farming digital transformation: Implementation of new and emerging technologies using artificial intelligence[J]. *Animal Health Research Reviews*, 2022, 23(1): 59-71.
- [16] WENG Z, MENG F S, LIU S Q, et al. Cattle face recognition based on a Two-Branch convolutional neural network [J]. *Computers and Electronics in Agriculture*, 2022, 196: 106871.
- [17] WENG Z, FAN L Z, ZHANG Y, et al. Facial recognition of dairy cattle based on improved convolutional neural network[J]. *IEICE Transactions on Information and Systems*, 2022, E105.D(6): 1234-1238.
- [18] GONG H, PAN H H, CHEN L, et al. Facial recognition of cattle based on SK-ResNet[J]. *Scientific Programming*, 2022, 2022: 5773721.
- [19] MA X X, LI Z W, ZHANG L. An improved ResNet-50 for garbage image classification[J]. *Tehnicki Vjesnik - Technical Gazette*, 2022, 29(5): 1552-1559.
- [20] LI Z, LEI X M. Cattle face recognition under partial occlusion[J]. *Journal of Intelligent & Fuzzy Systems*, 2022, 43(1): 67-77.
- [21] CHEN T T, LIN G J, ZHANG H J, et al. Non-uniform illumination face recognition method based on improved MobileNetV1[J]. *Journal of Computers*, 2023, 34(1): 145-155.

- [22] 杨蜀秦, 刘杨启航, 王振, 等. 基于融合坐标信息的改进 YOLO V4 模型识别奶牛面部[J]. 农业工程学报, 2021, 37(15): 129-135.  
YANG S Q, LIU Y Q H, WANG Z, et al. Improved YOLO V4 model for face recognition of dairy cow by fusing coordinate information[J]. Transactions of the Chinese Society of Agricultural Engineering, 2021, 37(15): 129-135. (in Chinese)
- [23] WANG Y F, HUA C C, DING W L, et al. Real-time detection of flame and smoke using an improved YOLOv4 network[J]. Signal, Image and Video Processing, 2022, 16(4): 1109-1116.
- [24] 张艳, 李星汕, 孙叶美, 等. 基于通道注意力与特征融合的水下目标检测算法[J]. 西北工业大学学报, 2022, 40(2): 433-441.  
ZHANG Y, LI X S, SUN Y M, et al. Underwater object detection algorithm based on channel attention and feature fusion[J]. Journal of Northwestern Polytechnical University, 2022, 40(2): 433-441. (in Chinese)
- [25] LIN W H, ZHONG J X, LIU S, et al. ROIMIX: Proposal fusion among multiple images for underwater object detection[C]//2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Piscataway: IEEE, 2020: 2588-2592.
- [26] ZHANG H Y, CISSE M, DAUPHIN Y N, et al. MixUp: Beyond empirical risk minimization[EB/OL]. (2018-04-27)[2023-08-29]. <https://arxiv.org/abs/1710.09412>.
- [27] PENG C, LIU Y K, YUAN X P, et al. Research of image recognition method based on enhanced inception-ResNet-V2[J]. Multimedia Tools and Applications, 2022, 81(24): 34345-34365.
- [28] PANG Y, YU W B, ZHANG Y A, et al. Sheep face recognition and classification based on an improved Mobile-netV2 neural network[J]. International Journal of Advanced Robotic Systems, 2023, 20(1): 172988062311529.
- [29] 张红斌, 熊其鹏, 蒋子良, 等. 联合 SENet 异构层特征融合与集成学习的材质图像识别[J]. 控制与决策, 2022, 37(6): 1632-1642.  
ZHANG H B, XIONG Q P, JIANG Z L, et al. Material image recognition combining heterogeneous-layer feature fusion of SENet and ensemble learning[J]. Control and Decision, 2022, 37(6): 1632-1642. (in Chinese)
- [30] SCHROFF F, KALENICHENKO D, PHILBIN J. FaceNet: A unified embedding for face recognition and clustering[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2015:

815-823.

- [31] 张宏鸣, 周利香, 李永恒, 等. 基于改进 MobileFaceNet 的羊脸识别方法[J]. 农业机械学报, 2022, 53(5): 267-274.  
ZHANG H M, ZHOU L X, LI Y H, et al. Sheep face recognition method based on improved MobileFaceNet[J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(05): 267-274. (in Chinese)

### 作者简介



**焦杰** 男, 1998年3月出生于山西省大同市. 现为内蒙古工业大学电力学院硕士研究生. 基于计算机视觉与深度学习的目标检测与识别.  
E-mail: 1344148511@qq.com



**齐咏生** 男, 1975年12月出生于内蒙古包头市. 现为内蒙古工业大学电力学院教授、副院长. 主要研究方向为移动机器人协同控制技术.  
E-mail: qys@imut.edu.cn



**刘利强** 男, 1975年5月出生于内蒙古包头市. 现为内蒙古工业大学电力学院教授、硕士生导师. 主要研究方向为计算机视觉、新能源发电技术.  
E-mail: llqiang@imut.edu.cn



**李永亭** 女, 1975年11月出生于内蒙古乌海市. 现为内蒙古工业大学电力学院副教授、硕士生导师. 主要研究方向为复杂工业过程故障监测与诊断.  
E-mail: liyt@imut.edu.cn



**王朝霞** 男, 1985年12月出生于内蒙古呼和浩特市. 现为内蒙古工业大学电力学院讲师. 主要研究方向为为3维高度成像技术、卫星组网观测技术.  
E-mail: wangzx@imut.edu.cn