

InGaAs 近红外人脸图像检测超轻量算法研究

苏晏园^{1,2,3}, 范广宇^{1,2*}, 龚海梅^{1,2*}, 李 雪^{1,2}, 陈永平^{1,2}

- (1. 中国科学院上海技术物理研究所 传感技术国家重点实验室, 上海 200083;
2. 中国科学院红外成像材料与器件重点实验室, 上海 200083;
3. 上海科技大学, 上海 201210)

摘要: InGaAs 近红外探测器广泛应用于航天航空、军事与民生领域。为了实现 InGaAs 探测器智能化, 结合人脸检测应用, 提出了可部署于低功耗移动智能设备的超轻量 InGaAs 近红外人脸检测算法。主要针对近红外人脸样本较少与低功耗设备部署问题展开研究, 采用迁移学习与二值量化方案训练网络。算法首先通过大规模可见光人脸数据集实现了基于 SSD 的预训练人脸检测网络。然后使用二值量化方案大幅压缩网络参数空间大小与计算量, 但同时造成网络准确度下降。为进一步提升网络二值量化效果, 为二值量化过程引入了特征均值信息, 并以对抗卷积形式弥补了准确度损失。最后, 算法通过小规模近红外人脸数据对预训练二值网络进行微调, 实现最终网络。所实现的二值量化人脸检测网络在采集的近红外人脸验证集中可以获得 71.18% 平均准确度。

关键词: 二值化; 近红外人脸检测; SSD; 网络压缩; InGaAs 探测器

中图分类号: TP391.4 **文献标志码:** A **DOI:** 10.3788/IRLA20220078

Research of ultra-light InGaAs NIR face detection algorithm

Su Yanyuan^{1,2,3}, Fan Guangyu^{1,2*}, Gong Haimei^{1,2*}, Li Xue^{1,2}, Chen Yongping^{1,2}

- (1. State Key Laboratories of Transducer Technology, Shanghai Institute of Technical Physics, Chinese Academy of Sciences, Shanghai 200083, China;
2. Key Laboratory of Infrared Imaging Materials and Detectors, Shanghai Institute of Technical Physics, Chinese Academy of Sciences, Shanghai 200083, China;
3. ShanghaiTech University, Shanghai 201210, China)

Abstract: InGaAs NIR detectors are widely used in aerospace, military and civilian fields. In order to realize the intelligence of InGaAs detectors, combined with face detection applications, an ultra-lightweight InGaAs NIR face detection algorithm that can be deployed in low-power mobile smart devices is proposed. This paper mainly studies the problems of few NIR face samples and low-power device deployment, and uses transfer learning and binary quantization to train the network. The algorithm first realizes a pre-trained face detection network based on SSD through a large-scale visible light face dataset. Then, the binary quantization scheme is used to greatly compress the network parameter space size and calculation amount, but the network accuracy is reduced at the same time. In order to further improve the effect of network binary quantization, this paper introduces feature mean information for the binary quantization process and makes up for the loss of accuracy in the form of adversarial convolution. Finally, the algorithm fine-tunes the pre-trained binary network through small-scale NIR

收稿日期:2022-01-29; 修订日期:2022-03-09

作者简介:苏晏园,男,硕士生,主要从事深度学习、红外探测、目标检测方面的研究。

导师简介:陈永平,男,研究员,博士生导师,博士,主要从事高性能航天遥感 CMOS 探测器、CMOS 电路与红外 MEMS 集成、高灵敏硅基 APD 等方面的研究。

face data to achieve the final network. The binarization face detection network implemented in this paper can achieve an average accuracy of 71.18% in the collected NIR face verification set.

Key words: binarization; NIR face detection; SSD; model compression; InGaAs detector

0 引言

深度神经网络由于在特征表达与特征学习方面的优势,逐步取代传统算法。当前计算机视觉任务主要应用于可见光领域不同计算机视觉任务,包括图像分类^[1]、语义分割^[2]、目标检测^[3]等。但是,可见光相机在夜视,雾霾等场景中无法获取可靠数据,从而限制了应用范围。InGaAs 探测器响应波段为近红外波段,具有高灵敏度、高探测率、室温工作、低功耗、高均匀性等优异的技术特性。因此在夜视、航空航天等军事与民用领域广泛应用,弥补了可见光相机应用中的不足。

文中的目标是将深度学习应用于基于 InGaAs 探测器的智能探测设备,并结合人脸检测^[4]应用,实现适用于低功耗探测设备的近红外人脸检测智能算法。人脸检测是重要的计算机视觉任务,是人脸识别^[5]、情绪分析^[6]、安防报警^[7]等任务的基础。近红外人脸检测可以采用主动成像,确保了人脸成像质量,并且能够屏蔽部分打印人脸及电子显示设备中人脸的干扰。

文中对算法设计中两个主要技术问题进行研究:(1) InGaAs 近红外人脸小样本学习。由于能够自主采集近红外人脸图像较少,同时网络中相关红外波段人脸数据不足,文中无法建立完整近红外人脸数据集用于训练与测试,需通过少量样本实现网络设计。(2) 模型轻量化设计。DNNs 使用包含大量参数的复杂神经网络,以此获得丰富的图像特征,因而需要大量算力。对于低功耗探测器而言,硬件算力限制了算法设计,算法部署难度大。

目前,已有许多网络轻量化方法被提出,例如模型剪枝^[8]、模型量化^[9]等。网络量化将浮点数值表示的网络参数映射为固定的离散数值。以此将复杂的浮点计算替换为高效的固定比特计算或位运算,从而降低网络所需计算量。得益于 Nvidia GPU 等硬件支持,8 bit 量化成为主流量化应用方案。但是,8 bit 量化压缩率有限,为达到更高压缩率,二值量化成为重要研究方向。

另一方面,深度学习网络一般使用 GPU 以浮点数完成卷积所需的乘加运算,对于移动设备的算力与带宽提出了较高要求。而压缩后的二值网络可以部署于 FPGA 中,使用同或门(XNOR)与位计数代替浮点数乘法与累加操作,降低计算量的,实现最高 128 倍的理论性能提升。

二值量化神经网络研究起始于 BNN^[10],作者提出了二值网络梯度优化的基础方案,实现了 DNNs 中卷积层及激活层权重二值量化工作。但是,在 ImageNet^[11] top-1 识别任务中,BNN 仅能实现 42.2% 识别准确度,相比于单精度网络 72.4% 的准确度存在较大差距。2020 年,ReActNet^[12] 优化二值网络中的 Sign 函数及 PReLU 函数,最终识别准确度可达 69.4%。将二值网络与 1 网络准确度差距缩小至 3%。

深度学习需要大数据作为支撑,但为不同的任务单独建立数据集成本过高,包含大量重复工作。因此小样本学习成为重要的研究方向。目前,数据增强,元学习,迁移学习等方案被用于解决小样本问题。其中,迁移学习通过已有知识迁移新知识的方法,使新网络快速应用于新任务。

文中将二值量化、迁移学习用于网络设计,提出了名为 ReactFace 的超轻量红外人脸检测网络方案。它实现了以下目标:(1) 基于 Single Shot MultiBox Detector (SSD) 的二值化人脸检测基础网络设计;(2) 通过特征校准的 Sign 函数和对抗卷积层减少二值量化带来的准确率损失;(3) 实现可见光人脸检测算法向近红外人脸检测算法迁移。最终网络能够部署于 FPGA 低功耗设备中,并且对于 Fddb 人脸测试集,ReActNet 与单精度网络准确度仅相差 5%,在采集的红外人脸测试集中达到 71.18% 平均检测准确率。

1 相关工作

1.1 SSD 目标检测网络

SSD 网络^[13]是单阶段、端到端的目标检测网络。SSD 采用全卷积网络设计方式,解决了图像尺寸变化产生的影响。同时,由于单阶段网络避免重复计算特征,实现了较快网络计算速度。采用结合 anchor

先验框等算法, SSD 实现了接近双阶段网络^[14]的检测效果。

为了达到更高压缩率, 文中采用 MobileNet v1^[15]作为 SSD 的特征提取卷积网络。MobileNet v1 采用卷积分离方法, 减少了卷积网络参数的同时, 保持着良好的特征提取能力, 适合于各类移动设备应用。

1.2 ReActNet 二值量化

2020 年, ReActNet^[12] 探究了 ReLU 和 Sign 函数对二值卷积网络的影响, 证明了 ReLU 函数对网络特征学习能力产生许多影响。并据此提出了 ReActNet 网络应用于 ImageNet 分类任务中。ReActNet 网络为 Sign 和 ReLU 函数增加偏置, 并命名新函数为 RSign 及 RPRReLU, 如公式 (1) 与公式 (2)。

$$x_i^b = \text{RSign}(x_i^r) = \begin{cases} 1, & x_i^r > \alpha_i \\ -1, & x_i^r \leq \alpha_i \end{cases} \quad (1)$$

式中: x_i^r 为第 i 层 Sign 函数输入特征; α_i 为学习变量, 经过训练的网络能够通过 α_i 调整 Sign 函数阈值。经过阈值调整后的 Sign 函数将实现更精确的二值化效果。

公式 (2) 为 RPRReLU 函数计算公式, 式中: x_i 为输入特征; $\gamma_i, \beta_i, \zeta_i$ 为学习变量。与 PReLU 相比, RPRReLU 增加了 γ_i, ζ_i 变量, 使更多有效特征得以保留。

$$\text{RPRReLU}(x_i) = \begin{cases} x_i - \gamma_i + \zeta_i, & \text{if } x_i > \gamma_i \\ \beta_i(x_i - \gamma_i) + \zeta_i, & \text{if } x_i \leq \gamma_i \end{cases} \quad (2)$$

ReActNet 同时修改了 MobileNet v1 中的可分离卷积模块, 如图 1 所示。与原卷积分离模块不同, ReActNet 模块增加了 RSign 与 RPeLU 函数, 并通过 shortcut 方式加强了网络的训练效果。与 XNOR-Net 相比, ReActNet 对于 ImageNet top1 识别准确度提升了 18.2%。

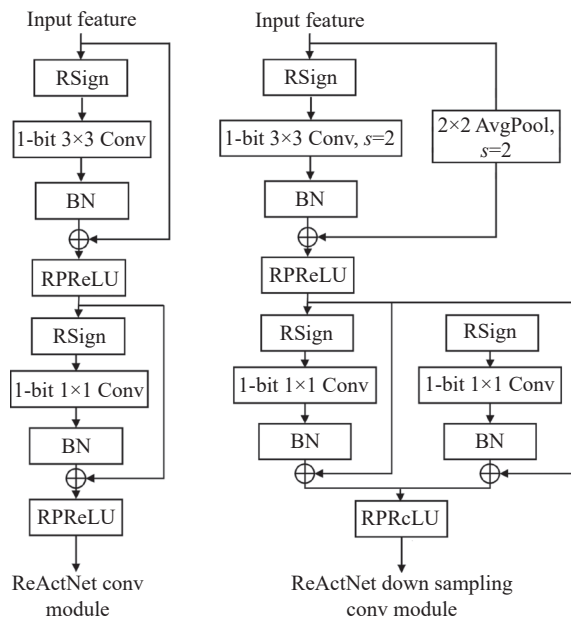
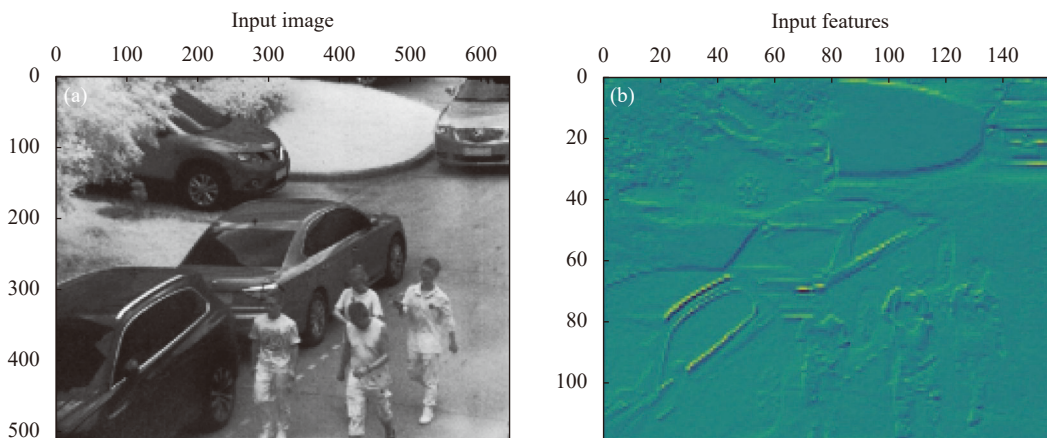


图 1 ReActNet 模块框图

Fig.1 Module of ReActNet

目前, ReActNet 网络仅在图像分类任务中取得了良好的效果。文中将其应用于目标检测任务中发现, 与识别任务不同, 使用 ReActNet 为特征提取网络的模型检测准确度出现了大幅下降。

图 2 展示了卷积网络特征传递情况。(a) 为输入图片; (b) 为 ReActNet 模块输入特征图, 即图 1 中 F1 特征; (c) 为经过 RSign 函数量化后的特征图片, 即 F2 特征, 其中黄色特征代表 1, 紫色特征代表 -1; (d) 为二值卷积输出, 即 F3 特征。从图中可以发现, 因为 RSign 将特征映射为二值数据, 无法保留特征重要程度信息, 造成卷积网络特征获取能力下降, (d) 中特征丢失大量信息。为了解决 Sign 函数带来的问题, 文中引进较少的参数, 改进网络结构并优化了 Sign 函数。



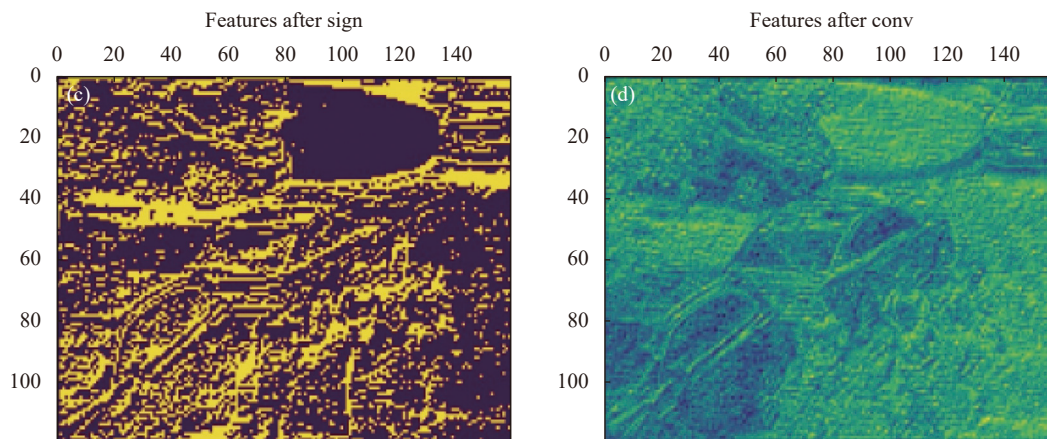


图 2 (a) 输入图片; (b) 卷积层输入特征; (c) 二值化特征; (d) 卷积重建后特征

Fig.2 (a) Input image; (b) Input features of convolution layer; (c) Binarized features; (d) Reconstructed features after convolution

1.3 Representative Batch Normalization

2021 年,GAO 等人^[16]对卷积神经网络中 BatchNorm (BN) 层进行了探究。其发现 BN 层的批次依赖性能够实现稳定的训练和更好的网络表示,但不可避免地忽略了实例之间的表示差异。为进一步优化 BN 层工作,GAO 首先证明了输入特征的均值包含部分推理信息,但 BN 推理将均值信息删除。为利用特征均值信息,GAO 提出了 Representative Batch Normalization (RBN) 引进了均值用于特征中心校准,使归一化中心随着输入图像变化而偏移。最后,作者还引进缩放校准使特征稳定分布。

传统 BN 计算如公式 (3)~(5) 所示,式中: X 为待校准特征; $E(*)$ 为期望计算; $Var(*)$ 为方差计算; ε 为微小常量, γ, β 为学习变量。

$$\text{中心化: } X_m = X - E(X) \quad (3)$$

$$\text{缩放: } X_s = \frac{X_m}{\sqrt{Var(X) + \varepsilon}} \quad (4)$$

$$\text{仿射变换: } Y = X_s \gamma + \beta \quad (5)$$

RBN 为 BN 增加了中心校准以及缩放校准,如公式 (6)~(9) 所示。公式 (6) 中 X 为输入特征, w_m 为学习变量, K_m 为 $(k_0, k_1, \dots, k_n)^T$, k_i 为各通道期望,即 $(\mu_0, \mu_1, \dots, \mu_c)$ 。公式 (9) 中, w_s, w_b 为学习变量, K_s 与 K_m 相似,此处为各通道方差的集合。

$$\text{中心校准: } X_{cm(n,c,h,w)} = X_{(n,c,h,w)} + w_{m(n,c)} \odot K_{m(n,c)} \quad (6)$$

$$\text{中心化: } X_m = X_{cm} - E(X_{cm}) \quad (7)$$

$$\text{缩放: } X_s = \frac{X_m}{\sqrt{Var(X_{cm}) + \varepsilon}} \quad (8)$$

$$\text{缩放校准: } X_{cs} = X_s \cdot R(w_v \odot K_s + w_b) \quad (9)$$

1.4 通过模型微调实现迁移学习

2020 年, Zhuang^[17] 等人对迁移学习进行详细的总结。迁移学习将已有知识域称为源域,待学习知识域称为目标域,可用于解决目标域样本不足的问题。2014 年, Bengio 等人将 ImageNet 分为类别不重合的两个数据集,并使用预训练与微调方案,实现 DNNs 在图像分类任务中迁移学习任务。Zhuang^[17] 证明了对于相关性较强的两个任务,通过源域预训练网络,小样本目标域数据微调训练方式,可以将源域学习的通用特征迁移至目标域。同时,迁移后网络特征提取能力随预训练网络提取能力增强而增强。

2 模型设计

文中研究目标为轻量化 InGaAs 近红外人脸检测网络。因为自主采集图像较少,同时互联网中相关波段开源人脸数据不足,文中无法建立完整近红外人脸数据集用于训练与测试。得益于近红外人脸特征与可见光下人脸特征相近,文中将采用迁移学习方案实现近红外人脸检测。

最终, ReActFace 网络设计分为两个步骤: 1) 通过网络中大规模可见光人脸数据训练由二值化特征提取网络和单精度预测网络组成的可见光二值化预训练网络; 2) 通过小样本近红外人脸数据对网络进行迁移学习,实现近红外二值化人脸检测网络。

2.1 可见光二值化预训练网络设计

基于 SSD 目标检测模型与 ReActNet 二值化工

作,文中首先实现了适应于可见光人脸检测的二值化卷积神经网络,并以此用作该工作的基准网络。为了解决基准网络中存在的特征学习能力不足的问题,文中接着提出了对抗卷积模块、带特征校准的 Sign 函

数等方案加强了二值网络的学习能力。

ReActFace 的总体设计基于 SSD 网络,如图 3(a)所示,网络包括二值特征提取网络、提升网络感受野的附加卷积层以及卷积预测网络。网络以图像作为

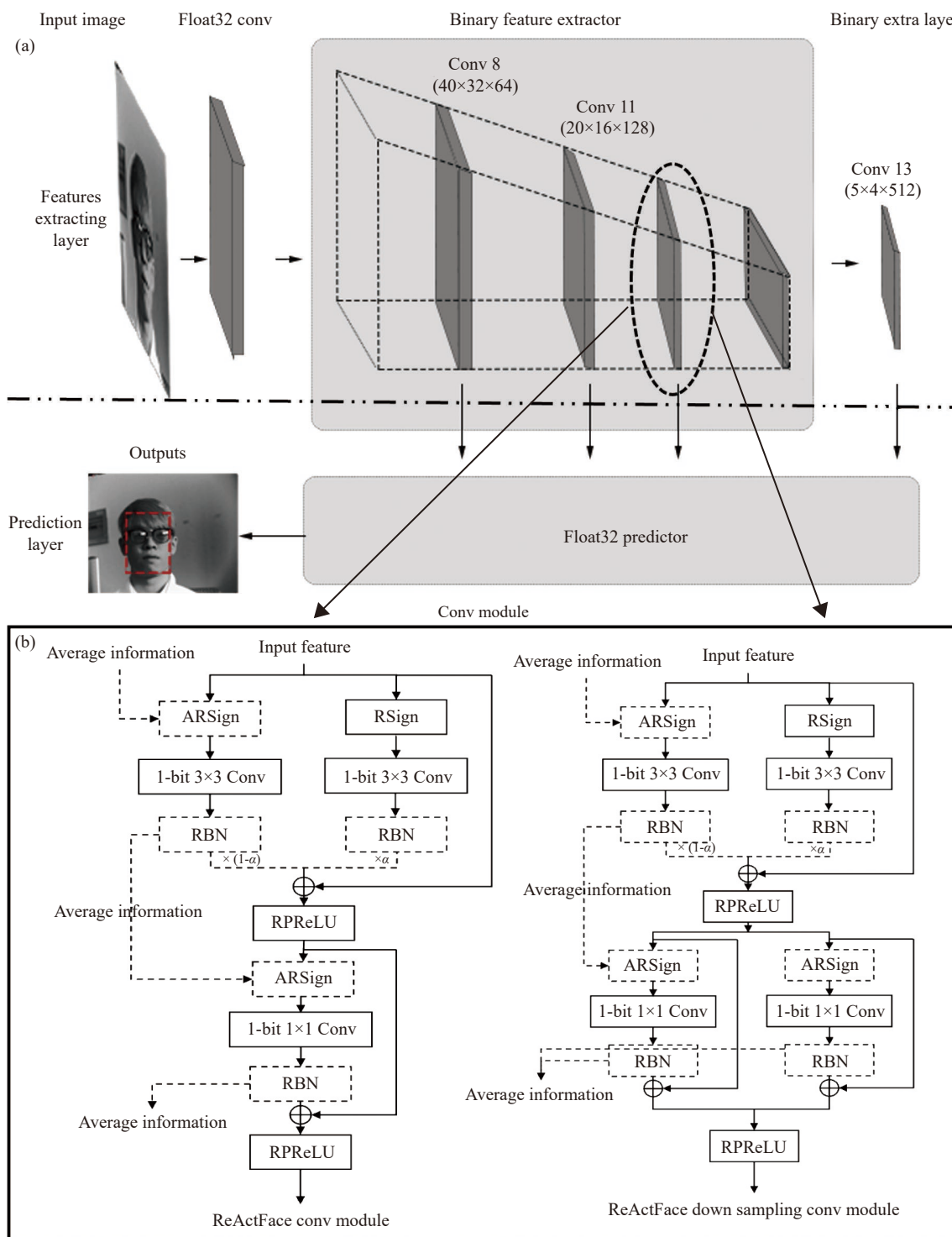


图 3 (a) 网络框图; (b) 优化后的卷积模块框图

Fig.3 (a) Block diagram of network; (b) Block diagram of optimized convolution module

输入, 返回人脸标记选框的尺寸和位置。其中, 二值特征提取网络由图 3(b) 所示。

2.1.1 通过特征校准的 Sign 函数提升二值量化效果

对于 Sign 函数而言, 公式 (1) 中 RSign 函数为 Sign 函数增加了偏置 α_i 。 α_i 是通过训练集数据获得的固定常量。与 BN 相似, RSign 函数关注了数据的整体分布, 但忽视了实例之间的差异。因为 Sign 函数在 0 处出现了突变, 输入数据轻微波动将引入大量噪

声, 网络更关注实例间的差异。文中为 RSign 函数引入了输入特征均值作为实例信息, 优化 Sign 函数判断。优化后的函数如公式 (10) 所示, 文中命其为 ARSign:

$$X_s = \begin{cases} +1 & X + w_s \odot \mu_m \geq 0 \\ -1 & X + w_s \odot \mu_m < 0 \end{cases} \quad (10)$$

式中: X 为输入特征; w_s 为待学习参数。 μ_m 经过卷积层后的特征均值, 即图 3(b) 中的 Average Information。

图 4 展示了 ARSign 工作效果。(1) 为经过尺寸

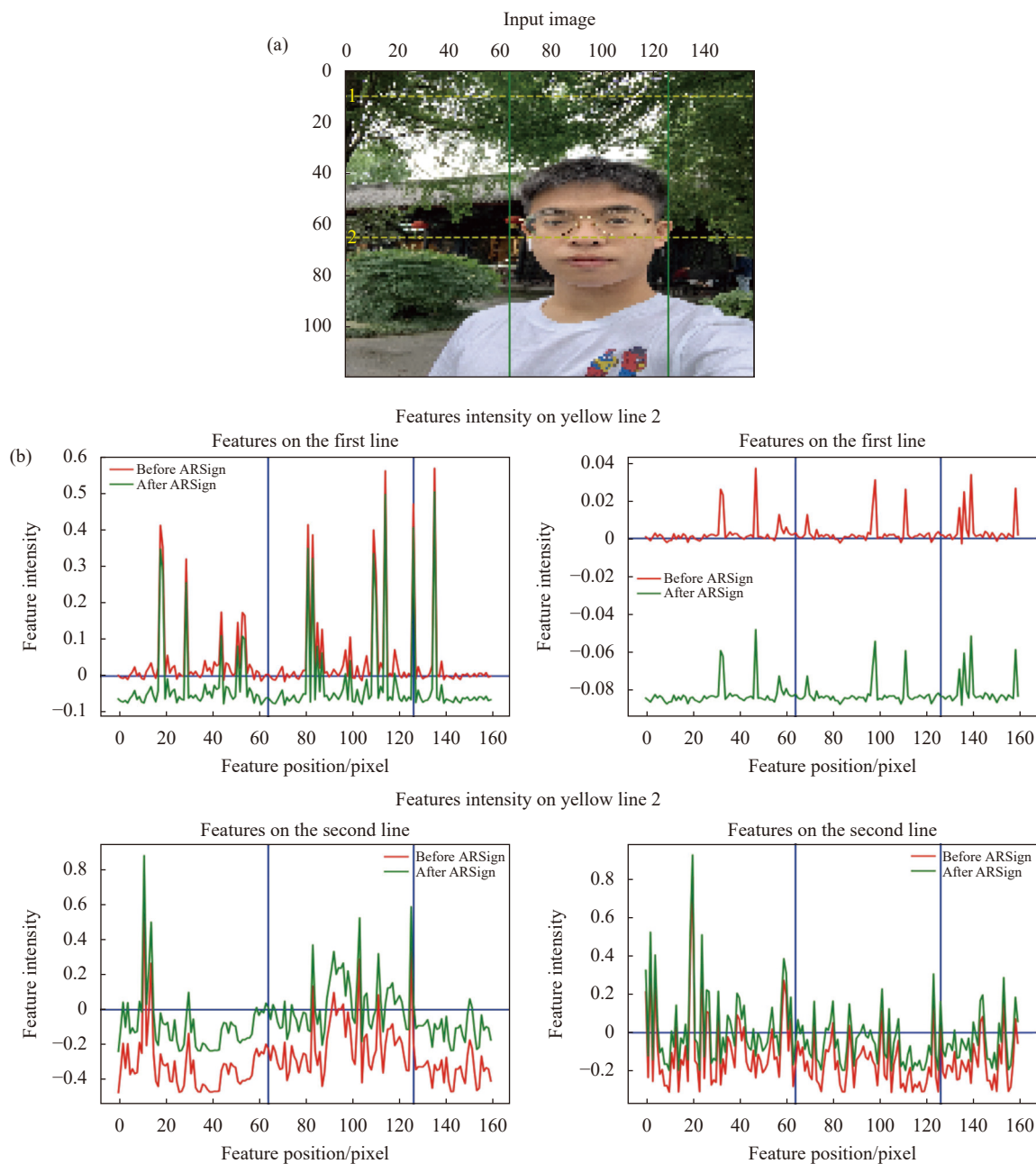


图 4 (a) 输入图像, (b) 黄线处特征强度

Fig.4 (a) Input image; (b) Feature intensity at the yellow line

变换的输入图像,图中存在两条黄色虚线,分别为黄线 1 和黄线 2。可以发现,黄线 1 没有包含人脸特征,而黄线 2 在竖向蓝线选中区域存在人脸特征。(2) 为黄线 1 和黄线 2 处特征强度,其中,红线为特征经过 RSign 的输出,绿线为经过 ARSign 输出。左右两图分别为两个不同特征层示例。图表通过蓝线分割了人脸区域与非人脸区域、强度大于 0 的区域以及小于 0 的区域。通过 ARSign 函数,黄线 1 中非目标特征得到衰弱,使强度低于 0,经过 Sign 函数后将被正确判断为不存在特征。而对于黄线 2 中的特征,ARSign 函数进行增强,使更多人脸特征强度高于 0,但同时也引进非目标特征噪声。总体而言,ARSign 增强了特征信噪比,提升了检测效果。

2.1.2 通过对抗卷积模块提升二值量化效果

文中通过对抗卷积模块进一步增强了模型,增

强后的网络结构如图 3(b) 中所示。相比于图 1 中 ReActNet 普通卷积模块与下采样模块,文中做了以下改进:

(1) 在 ReActNet 的基础上,引进了 RBN 替换了传统 BN 层,优化了网络标准化。

(2) 网络增加了 ARSign->3×3 Conv->RBN 的强化通路,用以学习困难对象特征。

(3) 引进 α 参数,使两条通路实现对抗优化。进一步放大通路间的差异性。

图 5 展示了输入图片后,对抗卷积模块中两条通路经过二值化后的特征图。(a)~(e) 分别代表输入图片、输入卷积层的特征、特征经过 RSign 输出、经过 ARSign 输出及对抗卷积模块输出。从图中可以发现,RSign 和 ARSign 提取了两组不同的特征。最终,对抗卷积层输出了更细致的纹理特征。

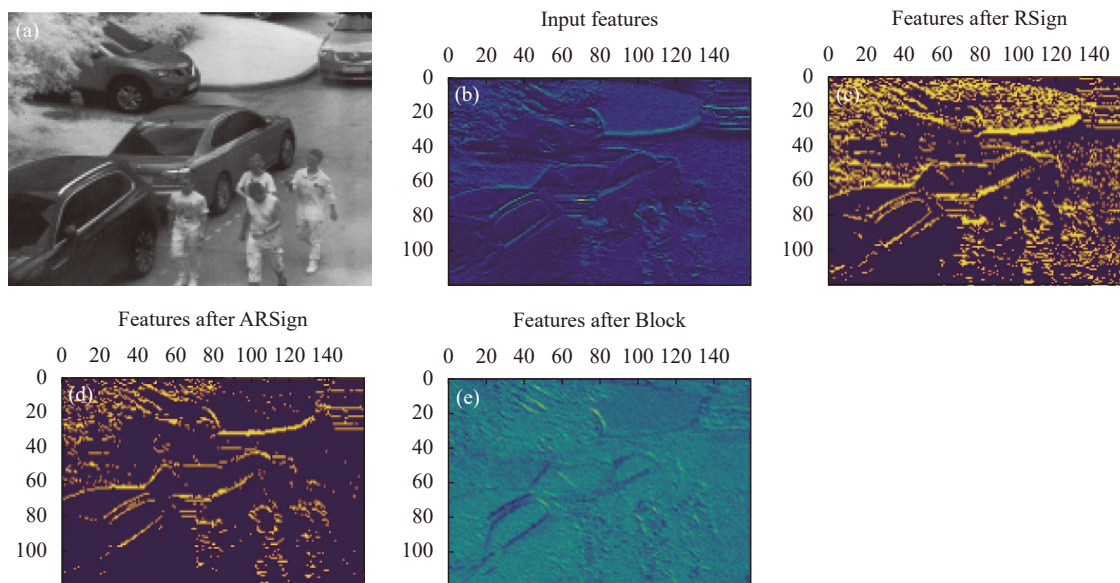


图 5 (a) 输入图片; (b) 卷积层输入特征; (c) ARSign 通路二值化特征; (d) RSign 通路二值化特征; (e) 卷积重建后特征

Fig.5 (a) Input image; (b) Input features of convolution layer; (c) Binarized features of ARSign pathway; (d) Binarized features of RSign pathway; (e) Reconstructed features after convolution

2.2 近红外二值化网络微调训练

经过可见光人脸图像预训练后的二值化人脸检测网络,将保留大量通用的人脸特征,以及少量可见光领域的偏移特征。由于用于微调的近红外人脸数据较少,为了更好的保留网络中通用的人脸特征,文中对网络微调学习率以及参与微调的网络层进行探究。

实验中首先固定了学习率,将预测层、附加网络和特征提取网络依次加入微调,最终发现仅固定特征提取网络获得最好的检测效果。固定更多或更少的网络将导致过拟合或欠拟合的问题,使检测准确度下降。对于学习率,通过探究设定为 0.0002,并使微调训练循环次数设定在 100。

3 实验测试

3.1 数据集介绍

文中使用 WiderFace Dataset^[18] 训练集进行可见光人脸检测模型预训练。数据集包含 32 203 张图像并且标记了 393 703 张人脸, 分类成为 61 种 event。数据集中人脸在比例、姿势和遮挡方面具有高度可变性, 并且将每一种 event 随机分成 40%/10%/50% 用作训练, 验证以及测试集。WiderFace 数据集以 PASCAL VOC 2012 方式标注以及储存。

WiderFace 数据集依照图片中人脸的尺寸, 角度等依据, 将人脸数据集分为 Easy, Medium 与 Hard 三类别。由于 Test 数据集需要提交 WiderFace 作者进行测试, 本研究以 WiderFace 验证集测试模型训练效果。

实现模型训练及验证后, 网络使用 Face Detection Data Set and Benchmark(FDDB)^[19] 数据集进行测试。FDDB 数据集包含了取自 Labeled Faces in the Wild (LFW)^[20] 的一组 2845 张图片中的 5 171 张面孔。

最后, 文中使用 InGaAs 近红外摄像机采集了 313 张含 612 个人脸的近红外人脸图片, 并将其 1:1 分为迁移训练集与验证集进行训练以及测试。

3.2 训练方案

实验使用一张 RTX Titan 24 G 显存的 GPU 进行训练, 使用 Pytorch 实现网络。训练相关参数如下:

(1) 输入图片尺寸: 320×240

(2) 网络训练采用随机梯度下降优化器 (SGD) 学习率为 0.03, 并在 80, 100 训练 batch 时降为 0.003 和 0.0003。

(3) 批量大小: 36;

(4) 训练循环次数: 200;

(5) 损失函数: CrossEntropyLoss;

(6) 微调近红外网络时, 批量大小为 8, 学习率为 0.0002;

为提升网络鲁棒性, 训练数据通过随机旋转、随机截取等方式模拟更多使用场景, 实现数据增强^[21]。

3.3 实验结果对比

实验依次测试了 ReactFace 网络人脸检测准确率与网络压缩率。为了进一步测试各优化模块的实现效果, 进行了模块消融实验。

实验结果中, Easy, Medium 以及 Hard 表示 WiderFace 验证集中三种检测难度人脸的识别准确率。实验认为, 预测框与真实选框 IOU 大于 0.5 被认为是判断正确, 反之为错误。测试输入图像尺寸为 320×240。FDDB 数据集以 ROC 曲线展示实现效果。

3.3.1 可见光预训练网络实现效果

本节展示了 ReActFace 网络检测准确度结果, 表 1 为 WiderFace 验证集测试结果, 图 6 为 FDDB 测试集结果。其中, 基准网络指将 SSD 网络通过 ReActNet 直接二值量化的网络, 单精度 SSD 网络代表同结构的单精度人脸识别网络。

结合表 1 与图 6, 可以发现, ReActFace 网络相较于基准网络检测准确度发生明显提升。对于 WiderFace 验证集, 在 Medium 与 Hard 检测难度中, ReActNet 与单精度网络仅有 2.30% 与 3.48% 的差距。对于 FDDB 人脸测试集, ReActFace 网络仍保持了良好的鲁棒性, 检测效果与单精度网络相近。

表 1 ReActFace 在 WiderFace 验证集中测试结果

Tab.1 ReActFace results in the WiderFace validation

| set | | | |
|-------------|----------|------------|----------|
| Net | Easy mAP | Medium mAP | Hard mAP |
| Baseline | 66.38% | 54.02% | 23.45% |
| ReActFace | 71.19% | 65.72% | 37.29% |
| Float32 SSD | 76.48% | 67.02% | 40.77% |

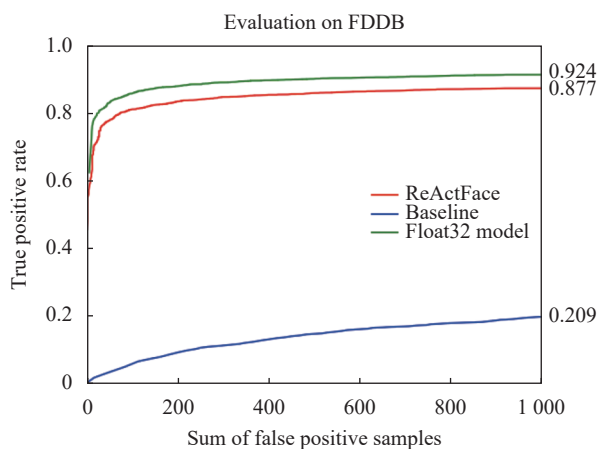


图 6 ReActFace 在 FDDB 测试集中测试结果 (Baseline 指基准网络, Float32 Model 指单精度 SSD 网络)

Fig.6 ReActFace test results in FDDB test set (Baseline refers to the benchmark network, Float32 Model refers to the full precision SSD network)

3.3.2 预训练网络模块消融实验

依次将 ARSign 层替换为 RSign, RBN 替换为 BN, 并以单通路替换对抗卷积模块进行模块消融实验。表 2 展示了实验在 WiderFace 验证集中测试结果, 图 7 展示了 FDDB 测试集测试结果。

表 2 模块消融实验在 WiderFace 验证集测试结果

Tab.2 Module ablation results in the WiderFace validation set test

| Net | Easy mAP | Medium mAP | Hard mAP |
|-------------------------|----------|------------|----------|
| ReActFace | 71.19% | 65.72% | 37.29% |
| ReActFace-ARSign | 71.00% | 65.30% | 37.00% |
| ReActFace-RBN | 70.90% | 65.20% | 36.62% |
| ReActFace-Contrast Conv | 65.50% | 60.25% | 33.72% |
| Baseline | 66.38% | 54.02% | 23.45% |

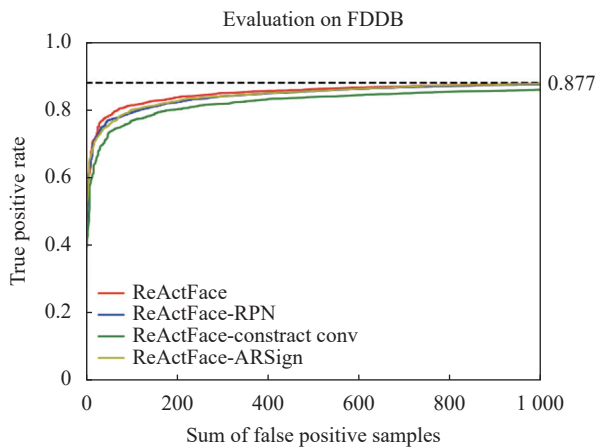


图 7 消融实验在 FDDB 中测试结果

Fig.7 Module ablation results in FDDB validation set test

从表 2 中可以发现各卷积模块对网络均产生了影响, 其中, 构建旁路的卷积对抗网络对提升网络准确率贡献最大。ARSign, RBN 增强了网络对于中、困难样本学习。从图 7 中 FDDB 测试结果可以发现, 在相同的误检样本的条件下, ReActFace 能够保持更好的准确率。实验证明了 ARSign、RBN 函数与对抗卷积模块通过特征校准、构建对抗 Sign 通路等方式提升了二值网络在检测任务中特征提取能力。

3.3.3 近红外人脸检测网络实现效果

表 3 展示了通过特征迁移学习后实现的近红外人脸检测网络测试效果。可以发现, ReActFace 网络相比于基准网络检测准确度提升达 7.46%。但是, 单

精度人脸检测网络经过迁移训练后, 检测效果却不如 ReActFace, 文中推测复杂的卷积网络易于出现过拟合现象。单精度人脸检测网络相比于 ReActFace 学习了更多可见光下的特征而非通用人脸特征, 导致迁移效果不如 ReActFace。图 8 展示了 ReActFace 网络近红外人脸图片的检测效果。

表 3 近红外人脸数据集测试结果

Tab.3 Result in NIR-Face data set

| Net | mAP |
|-----------------------|--------|
| Baseline | 63.72% |
| ReActFace | 71.18% |
| Float32 Face Detector | 53.37% |



图 8 近红外人脸检测效果

Fig.8 Result on NIR Images

3.3.4 网络二值量化效果

表 4 展示了网络二值量化后, 特征提取网络与网络整体参数量变化。可以发现, 二值量化后的特征提取网络实现了二倍以上压缩率。网络整体因为单精度预测层存在, 网络参数下降了 40%。特征提取网络中仍存在单精度保存的参数, 可以通过 8 位整型量化, 实现进一步压缩。

表 4 近红外人脸数据集测试结果

Tab.4 Result in NIR-Face data set

| Net | Size of feature extractor/kb | Size of Net/kb |
|-----------------------|------------------------------|----------------|
| ReActFace | 264.8 | 627.1 |
| Float32 Face Detector | 676.2 | 1 040.8 |

4 结束语

文中对二值量化网络在近红外人脸检测领域应用进行了探究, 依次完成单精度可见光人脸检测预训练网络、单精度网络二值量化与优化工作, 以及近红

外人脸微调工作。提出了特征校准 ARSign 二值函数, 并通过对抗卷积模块提升了二值特征提取效果, 最终网络被命名为 ReActFace。ReActFace 在 WiderFace 验证集和 FDDB 测试集中验证了预训练模型效果, 在采集的近红外人脸数据集中获得 71.18% 平均检测准确度。未来, 该研究将继续实现单精度网络二值量化工作, 并使用 FPGA 进行硬件实现。

参考文献:

- [1] Xie Xingxing, Cheng Gong, Wang Jiabao, et al. Oriented r-cnn for object detection[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 3520-3529.
- [2] Su Yunzheng, Hao Qun, Cao Jie, et al. Point cloud semantic segmentation method based on segmented blocks merging [J]. *Infrared and Laser Engineering*, 2021, 50(10): 20200482. (in Chinese)
- [3] Li Weipeng, Yang Xiaogang, Li Chuanxiang, et al. An improved semi-supervised transfer learning method for infrared object detection neural network [J]. *Infrared and Laser Engineering*, 2021, 50(3): 20200511. (in Chinese)
- [4] Zhang Shifeng, Chi Cheng, Lei Zhen, et al. Refineface: Refinement neural network for high performance face detection [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, 43(11): 4008-4020.
- [5] Meng Qiang, Zhao Shichao, Huang Zhida, et al. Magface: A universal representation for face recognition and quality assessment[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 14225-14234.
- [6] Abdullah S M A, Ameen S Y A, Sadeeq M A, et al. Multimodal emotion recognition using deep learning [J]. *Journal of Applied Science and Technology Trends*, 2021, 2(2): 52-58.
- [7] Wang Yunfeng, Li Zuopeng. Application research of object detection algorithm in edge environment [J]. *Computer Engineering and Applications*, 2021, 57(16): 220-228. (in Chinese)
- [8] Li Hao, Kadav Asim, Durdanovic Igor, et al. Pruning filters for efficient convnets [EB/OL]. (2016-08-31) [2022-03-09]. <https://arxiv.org/abs/1608.08710>.
- [9] Li Zhisheng, Wang Lei, Guo Shasha, et al. Laius: An 8-bit fixed-point CNN hardware inference engine[C]//2017 IEEE International Symposium on Parallel and Distributed Processing with Applications and 2017 IEEE International Conference on Ubiquitous Computing and Communications (ISPA/IUCC), 2017: 143-150.
- [10] Courbariaux Matthieu, Hubara Itay, Soudry Daniel, et al. Binarized neural networks: Training deep neural networks with weights and activations constrained to +1 or -1 [EB/OL]. (2016-02-09)[2022-03-09]. <https://arxiv.org/abs/1602.02830>.
- [11] Krizhevsky Alex, Sutskever Ilya, Hinton Geoffrey E. Imagenet classification with deep convolutional neural networks [J]. *Advances in Neural Information Processing Systems*, 2012, 25: 1097-1105.
- [12] Liu Zechun, Shen Zhiqiang, Savvides Marios, et al. Reactnet: Towards precise binary neural network with generalized activation functions[C]//European Conference on Computer Vision, 2020: 143-159.
- [13] Liu Wei, Anguelov Dragomir, Erhan Dumitru, et al. Ssd: Single shot multibox detector[C]//European Conference on Computer Vision, 2016. SPRINGER: 21-37.
- [14] Ren Shaoqing, He Kaiming, Girshick Ross, et al. Faster r-cnn: Towards real-time object detection with region proposal networks [J]. *Advances in Neural Information Processing Systems*, 2015, 28: 91-99.
- [15] Howard Andrew G, Zhu Menglong, Chen Bo, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications [EB/OL]. (2017-04-17)[2022-03-09]. <https://arxiv.org/abs/1704.04861v1>.
- [16] Gao Shang-Hua, Han Qi, Li Duo, et al. Representative batch normalization with feature calibration[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 8669-8679.
- [17] Zhuang Fuzhen, Qi Zhiyuan, Duan Keyu, et al. A comprehensive survey on transfer learning [C]//Proceedings of the IEEE, 2020, 109(1): 43-76.
- [18] Yang Shuo, Luo Ping, Loy Chen-Change, et al. Wider face: A face detection benchmark[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 5525-5533.
- [19] Jain Vidit, Learned-Miller Erik. FDDB: A benchmark for face detection in unconstrained settings [R]. UMass Amherst Technical Report, 2010.
- [20] Huang Gary B, Mattar Marwan, Berg Tamara, et al. Labeled faces in the wild: A database for studying face recognition in unconstrained environments[C]//Workshop on Faces in 'Real-Life' Images: Detection, Alignment, and Recognition, 2008.
- [21] Shorten Connor, Khoshgoftaar Taghi M. A survey on image data augmentation for deep learning [J]. *Journal of Big Data*, 2019, 6(1): 1-48.