



高低频通道特征交叉融合的低光人脸检测算法

许皓, 钱宇华, 王克琪, 刘畅, 李俊霞

引用本文:

许皓, 钱宇华, 王克琪, 刘畅, 李俊霞. 高低频通道特征交叉融合的低光人脸检测算法[J]. *智能系统学报*, 2024, 19(2): 472–481.

XU Hao, QIAN Yuhua, WANG Keqi, et al. Low-light face detection method based on the cross fusion of high-and low-frequency channel features[J]. *CAAI Transactions on Intelligent Systems*, 2024, 19(2): 472–481.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202208034>

您可能感兴趣的其他文章

跨年龄人脸验证技术研究

Research on age invariant face verification technology

智能系统学报. 2021, 16(2): 247–253 <https://dx.doi.org/10.11992/tis.202011029>

神经网络多层特征信息融合的人脸识别方法

Face recognition method based on neural network multi-layer feature information fusion

智能系统学报. 2021, 16(2): 279–285 <https://dx.doi.org/10.11992/tis.201907053>

利用场景光照识别优化的双目活体检测方法

Binocular camera based face liveness detection with optimized scene illumination recognition

智能系统学报. 2020, 15(1): 160–165 <https://dx.doi.org/10.11992/tis.201912026>

一种新融合算法的维吾尔族人脸识别

A new fusion algorithm for uyghur face recognition

智能系统学报. 2018, 13(3): 431–436 <https://dx.doi.org/10.11992/tis.201710014>

一种多层特征融合的人脸检测方法

Face detection method fusing multi-layer features

智能系统学报. 2018, 13(1): 138–146 <https://dx.doi.org/10.11992/tis.201707018>

人脸识别背后的数据清理问题研究

Research on the data cleansing problem for face recognition technology

智能系统学报. 2017, 12(5): 616–623 <https://dx.doi.org/10.11992/tis.201706025>

DOI: 10.11992/tis.202208034

网络出版地址: <https://link.cnki.net/urlid/23.1538.TP.20231204.1535.004>

高低频通道特征交叉融合的低光人脸检测算法

许皓^{1,2}, 钱宇华^{1,2,3}, 王克琪¹, 刘畅^{1,2}, 李俊霞^{1,2}

(1. 山西大学 大数据科学与产业研究院, 山西 太原 030006; 2. 山西大学 计算机与信息技术学院, 山西 太原 030006; 3. 山西大学 计算智能与中文信息处理教育部重点实验室, 山西 太原 030006)

摘要: 低光条件下捕获的人脸图像存在着噪声严重、对比度低等不足, 极大影响了现有人脸检测器的准确性, 另外, 现有的低光图像检测算法欠缺小区域人脸信息的提取能力。为此, 提出一种基于深度学习的两阶段人脸检测算法, 即利用现有的低光图像增强算法对人脸图像进行增强后再进行检测。为提升检测算法对人脸信息的提取能力, 设计一种新型的高低频通道特征交叉融合方法, 该方法首先使用高低频通道特征可分离模块分离出不同尺度特征的高低频信息, 然后对上述信息进行交叉融合, 提升网络提取高频细节信息和低频色域信息的能力, 进而提高检测网络的性能。对比试验和消融试验验证了该研究方法的有效性, 试验结果表明该方法优于基准方法 4.0% mAP。

关键词: 低光人脸检测; 高低频通道特征; 低光增强; 多尺度特征融合; 计算机视觉; 图像处理; 深度学习; 频率域分析

中图分类号: TP183 文献标志码: A 文章编号: 1673-4785(2024)02-0472-10

中文引用格式: 许皓, 钱宇华, 王克琪, 等. 高低频通道特征交叉融合的低光人脸检测算法 [J]. 智能系统学报, 2024, 19(2): 472-481.

英文引用格式: XU Hao, QIAN Yuhua, WANG Keqi, et al. Low-light face detection method based on the cross fusion of high-and low-frequency channel features[J]. CAAI transactions on intelligent systems, 2024, 19(2): 472-481.

Low-light face detection method based on the cross fusion of high-and low-frequency channel features

XU Hao^{1,2}, QIAN Yuhua^{1,2,3}, WANG Keqi¹, LIU Chang^{1,2}, LI Junxia^{1,2}

(1. Institute of Big Data Science and Industry, Shanxi University, Taiyuan 030006, China; 2. School of Computer and Information Technology, Shanxi University, Taiyuan 030006, China; 3. Key Laboratory of Computational Intelligence and Chinese Information Processing, Shanxi University, Taiyuan 030006, China)

Abstract: Face images captured under low-light conditions suffer from significant noise and low contrast, which negatively impact the accuracy of existing face detection systems. In addition, existing low-light image detection algorithms struggle to extract information from small facial areas. To tackle these issues, this paper proposes a two-stage face detection algorithm based on deep learning. This algorithm enhances low-light images before initiating the detection process using an established low-light image enhancement method. The objective is to enhance the ability of the detection method to extract facial information. Thus, a new cross-fusion method of high- and low-frequency channel features is designed. The first step involves using a separable module for high- and low-frequency channel features, enabling the separation of different scale features. These separated features are then cross-fused to improve the ability of the network to extract high-frequency details and low-frequency color information. This, in turn, improves the performance of the detection network. The comparative and ablation experiments validate the effectiveness of the proposed method. The experimental results demonstrate that our method surpasses the baseline method by 4.0% mAP.

Keywords: low-light face detection; features of high-and low-frequency channels; low-light enhancement; multiscale feature fusion; computer vision; image processing; deep learning; frequency domain analysis

收稿日期: 2022-08-22. 网络出版日期: 2023-12-05.

基金项目: 科技创新 2030-“新一代人工智能”重大项目 (2021-ZD0112400); 国家自然科学基金项目 (62136005); 山西省揭榜挂帅项目 (202201020101006).

通信作者: 钱宇华. E-mail: jinchengqyh@126.com.

人脸检测在安防监控、行人重定位与人机交互等诸多领域具有广泛的应用。近年来, 正常光照条件下的人脸检测技术已取得长足的发展, 并

被实际应用在人们生活的方方面面。然而在低照度条件下,现有的人脸检测方法常常难以产生令人满意的效果。

在光照不足情况下获得的图像,其质量会遭受一定程度的退化,如能见度降低、信号细节损失等。这些退化不仅影响人眼的视觉感知,还降低了现有基于可见光视觉任务的性能。更重要的是,这种退化带来了人脸结构信息不可逆转的破坏以及噪声的干扰,并且夜间人脸具有微小、模糊和遮挡等问题,很大程度上增加了可见光视觉检测任务在夜间检测的难度。

为促进该领域的发展,研究人员构建了 DARK FACE^[1]数据集。该数据集包含 6000 张夜间拍摄的人脸图像,为系统评价夜间人脸检测效果做出了贡献,但是,基于现有的人脸检测器如 Center-Face^[2]在该数据集上几乎检测不到任何人脸,这与其在 WIDER FACE^[3]数据集上表现的性能形成了鲜明的对比。因此,基于可见光的视觉检测方法并不适用于低照度条件。

为使人脸检测方法在低照度条件下具有更好的泛化性和鲁棒能力,研究者使用图像增强方法^[4-13]对其增强,提高图像能见度。Gamma 校正^[6]技术的出现在一定程度上提高了图像质量,增强了人类检测器的性能,但是不能很好地解决图像特征的退化问题。图像特征的退化使网络不能充分提取有效特征,而无效特征会极大地干扰检测网络的定位过程,产生较大的偏差。

因此,本研究提出一种基于深度学习的两阶段人脸检测算法,先利用现有的低光图像增强方法(zero-reference deep curve estimation, Zero-DCE^[4])对低光人脸图像进行增强,恢复图像的色域信息,降低低光图像的噪声影响。然后,针对特征退化问题,提出了一种新型的适应于夜间人脸检测的特征融合方法——高低频信息交叉融合方法(cross feature pyramid networks, Cross-FPN)。该方法通过高低频通道可分离模块(high-low frequency channel separable modules, HLC)将特征图不同频率通道的信息进行分离,帮助检测网络更好地提取不同频率的有效特征。然后将不同尺度间的频率信息交叉融合,提升网络对夜间图像更深层次信息的表征能力,从而有效提升其夜间小人脸检测精度。

在试验部分,利用 DARK FACE 数据集^[1]在夜间人脸检测任务上进行了多方法对比和消融试验分析。试验结果表明,本方法优于现有的基准检测器,显著提升了检测器的精度,证明了方法

的有效性。同时通过消融试验分析了高低频通道特征交叉融合方法对夜间人脸检测的重要性及合理性。

1 相关工作

1.1 低光图像增强方法

低光图像增强方法旨在改善或消除照明不足图像存在的低亮度、低对比度、噪声和伪影等问题,提升其感知质量。传统的图像增强方法仅能增强动态范围偏小的图像对比度,通常依赖于局部统计和强度映射等理论,例如,直方图均衡化^[5]和 Gamma 校正^[6]。受 Retinex 理论^[7]的启发,Guo^[8]提出了(low-light image enhancement, LIME)图像增强算法,通过加入一种结构先验来细化初始光照图,最后根据该理论合成增强图。Rahman 等^[9]将低光图像的光照和反射率分离,解决了光照和反射率之间的模糊性。大量的工作^[10-13]也侧重于通过分别估计照明和反射率来对图像进行增强。最近,基于深度模型的图像增强方法取得了长足的进步。文献[14-17]基于 Retinex 理论^[7]构造了端到端学习深度网络模型,进一步提升了图像低光增强的效果。但夜间人脸图像不易获取光照充足的监督信息(Ground Truth),并且传统的增强方法导致增强后的图像具有更多的噪声、伪影等问题,因此,需要使用无监督学习方法进行增强。Zero-DCE^[4]作为近两年领域关注度较高的无监督图像增强方法,为两阶段夜间人脸检测方法提供了新思路。该方法将增强任务转换为光照曲线估计问题,对输入图像进行动态范围内的像素级调整,从而获得增强图像。

1.2 小人脸检测

小人脸检测一直都是检测领域的研究热点,也是待解决的难点之一。传统的解决方案是在数据层面对小人脸数据进行过采样(Oversampling)和数据增强(Augmentation)。随着深度学习的发展,许多研究证明了特征金字塔网络(feature pyramid networks, FPN)有助于增强模型对于小人脸的检测能力。受该网络结构的启发,文献[18]引入融合因子,控制深层传递到浅层的信息,使 FPN 更适用于小人脸检测。Mei 等^[19]提出了一种基于图像恢复金字塔注意力模块的图像修复模型,能够从多尺度特征金字塔中提取到长距离与短距离的特征关系,进一步加强特征间的联系。大量工作^[20-25]对 FPN 结构进行灵活设计,旨在充分挖掘不同分辨率特征之间的潜在特征,利用各个大型感受野的判别信息来集成多路径特征,提

升目标或人脸检测器的检测性能。

虽然上述工作证明了 FPN 对于小人脸检测的有效性, 但是对夜间场景下的检测方法的研究却很少。受低光图像存在的一些退化问题影响, FPN 方法在特征融合过程中会产生信息学习偏差。因此, 提出一种高低频通道特征融合方法, 使其更适应于夜间人脸检测。

1.3 夜间人脸或目标检测

随着大规模检测数据集的出现, 以及深度学习技术的进步, 光照充足环境中的人脸检测取得了显著的进步。但是对于低光条件或约束环境中的人脸检测研究仍不充分, 其中, 域自适应方法取得了一定的突破。Sasagawa 等^[26] 提出 Yolo-in-the-Dark 方法, 使用 glue 层和生成模型合并了不同域中的预训练模型, 旨在缩小两个域之间的差距。Wang 等^[27] 充分利用现有的正常光数据, 提出一种高低级联合适应模式。Liang 等^[28] 以低光图像增强方法为基础, 设计了一种循环曝光生成 (recurrent exposure generation, REG) 模块, 并将其与多重曝光检测 (multiple exposure detection, MED) 模块无缝耦合, 有效抑制了不均匀的照明

和噪声, 显著提高人脸检测性能。

上述域自适应方法具有数据利用率低、鲁棒性差和不易训练等缺点, 导致网络不能很好地耦合夜间人脸特征, 而低光图像增强方法也只是在提高图像的感知质量, 并没有帮助检测网络更好地学习潜在特征。为了更好地挖掘特征的潜在信息, 提出了一种特征高低频通道信息交叉融合低光人脸检测方法, 在高低频率通道上对特征进行有序、高效地融合, 使得检测网络更关注不同尺度的高频细节信息和低频语义信息, 增强了人脸区域定位能力。

2 Cross-FPN 低光人脸检测方法

本节将介绍高低频通道信息交叉融合的低光人脸检测方法的整体架构, 并对各个模块进行详细说明。由图 1 可知, 本方法提出的网络架构主要由低光增强模块和 Cross-FPN 网络组成。Cross-FPN 网络包含高低频通道交叉融合网络 (high-low frequency channel cross fusion network, HLC-NET) 和跳跃连接操作 (skip layer, SL)。

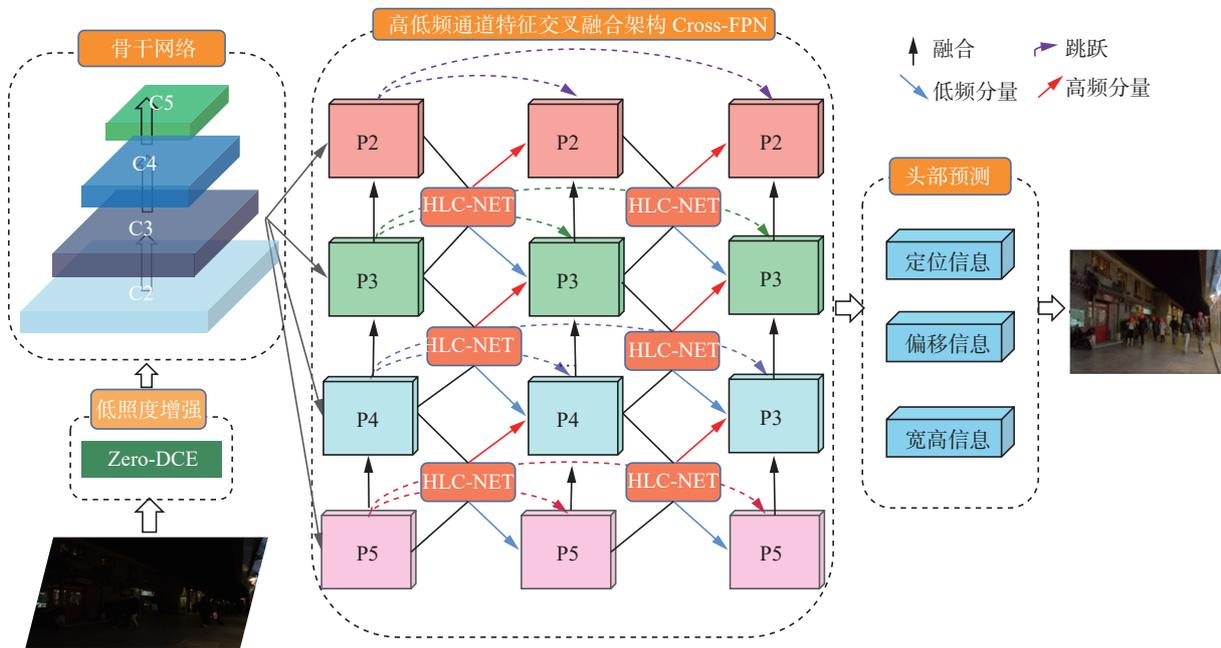


图 1 基于 Cross-FPN 的低光人脸检测器模型
Fig. 1 Low light face detector model based on Cross-FPN

2.1 低光图像增强模块

低光人脸图像普遍存在受光不均匀的现象, 即仅有一侧人脸受光, 所以增强的图像会出现过曝或曝光不足等问题。本研究希望在去除噪声和保证色彩增强的情况下, 自适应调整低光人脸图像的光照。

为解决上述问题, 基于文献 [4], 本方法的低光图像增强部分使用非线性映射方法。该方法由多次迭代的二次曲线组成, 利用 7 层卷积神经网络对下式的参数进行学习。

$$L(x; a) = x + ax(1 - x) \tag{1}$$

$$L_n = L(L_{n-1}; \alpha_n) \tag{2}$$

式中: $L_0(\cdot)$ 表示输入图像; $L_n(\cdot)$ 表示迭代 n 次的二次曲线; a_n 表示通过网络预测的 RGB 图像。ZeroDCE 法的优势在于利用无监督的数据信息进行亮度提升和色彩增强, 与其他基于 Retinex 理论^[7]的深度学习方法相比, 不会引入冗余的信息和伪影。

2.2 Cross-FPN 网络

Corss-FPN 网络旨在设计一个高效通用的夜间人脸检测任务的特征交叉融合网络, 其内部为多层高低频通道特征分离模块(HLC), 用于提取特征高低频通道信息, 并通过多条交叉融合路径和横向跳跃连接操作来丰富模型的表征能力。

2.2.1 跳跃连接模块 (SL)

横向跳跃连接利用尺度空间中的多向语义信息流来丰富特征表达, 并促进网络跨越不同尺度进行复杂层次的特征学习。

根据文献 [29], 跳跃连接模块在每一层中设计稠密连接 (Dense-link), 以提高特征的复用率。考虑到稠密连接的计算复杂度和信息的复用效率, 将稠密连接改进为单体连接 (S-link), 图 2 给出了两者结构的对比。具体地, 单体连接在每个级别 k 中, l 层仅接收第一层的特征信息。即

$$\mathbf{X}_k^l = \text{Skip}(\text{Add}(\mathbf{X}_k^l, \mathbf{X}_k^l)) \quad (3)$$

式中: \mathbf{X}_k^l 表示第 k 层的第 l 个特征图; $\text{Skip}(\cdot)$ 表示横向跳跃连接。为了减少计算量并防止通道特征信息在交互时的丢失, 在特征图之间使用 $\text{Add}(\cdot)$ 连接。

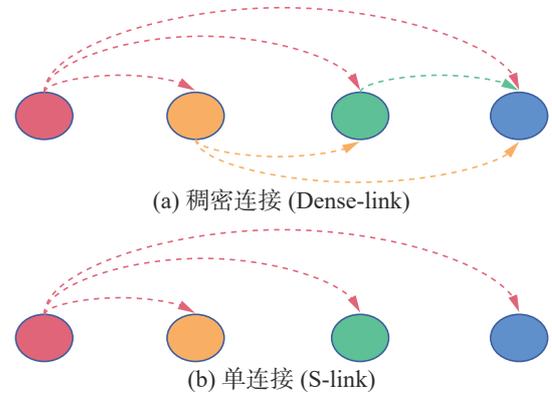


图 2 Dense-link 与 S-link 对比
Fig. 2 Contrast between Dense-link and S-link

2.2.2 高低频通道特征交叉融合网络

高低频通道交叉融合网络 (HLC-NET) 对不同尺度特征的频率通道信息进行融合。由图 3 可知, 主要由高低频率通道可分离模块 (HLC) 堆叠构成。

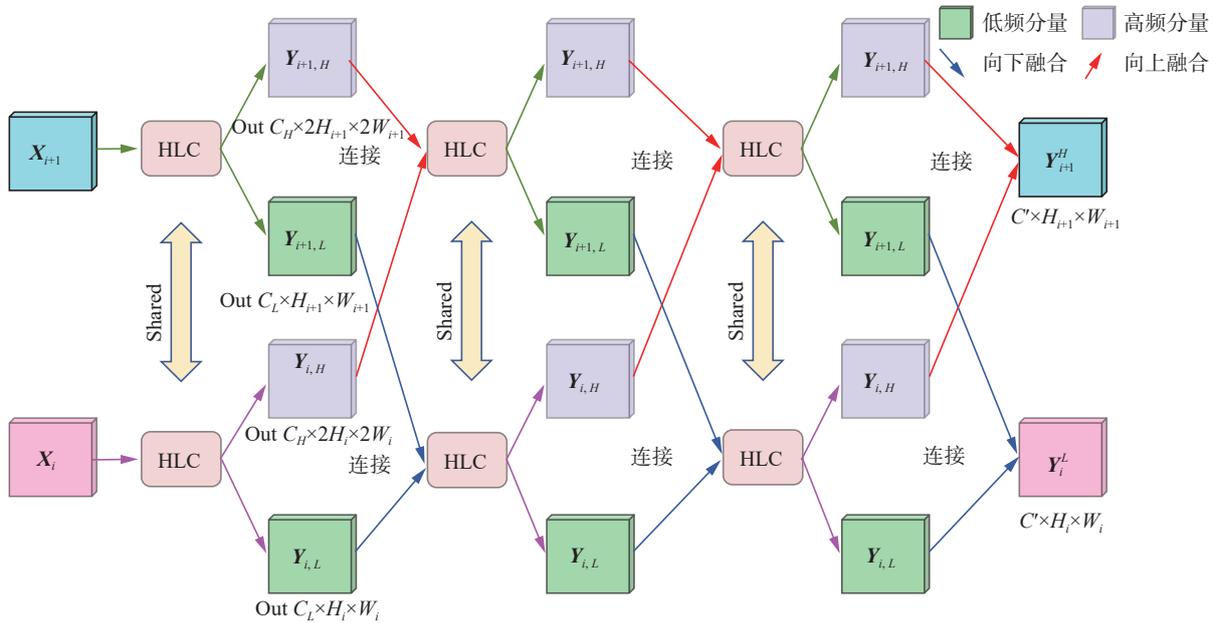


图 3 HLC-NET 模型架构
Fig. 3 HLC-NET model architecture

对于给定相邻层之间的特征图 $\mathbf{X}_i \in \mathbf{R}^{C \times H \times W}$ 和 $\mathbf{X}_{i+1} \in \mathbf{R}^{C \times 2H \times 2W}$ 。 $i+1$ 层输出浅层特征, 具有较小的感受野和丰富的细节信息, 更加适合检测小人脸。 i 层输出深层特征, 具有较大的感受野和较好的语义信息, 更适合检测大人脸。

为结合两者优势, HLC-NET 将相邻两层特征通过 HLC 模块通道分离, 其中同一层 HLC 模块

的参数共享, 根据下式分别产生 4 种频率通道特征。

$$\mathbf{Y} = \text{HLC}(\mathbf{X}_{i+1}, \mathbf{X}_i), \mathbf{Y} \in \{\mathbf{Y}_{i,H}, \mathbf{Y}_{i,L}, \mathbf{Y}_{i+1,H}, \mathbf{Y}_{i+1,L}\} \quad (4)$$

对于上层模块输出的 2 种不同尺度、4 类不同频率的通道特征, HLC-NET 将这些特征分别在同一频率通道上进行拼接融合, 并通过多次交叉迭代得到高频通道特征 \mathbf{Y}_{i+1}^H 和低频通道特征 \mathbf{Y}_i^L 。上述过程如下式所示

$$\begin{aligned} \text{HLC}_n &= \text{HLC}(\text{HLC}_{n-1}) \\ \mathbf{Y}_i^H &= \text{HLC}_n(\text{Concat}(\mathbf{Y}_{i,H}, \text{Down}(\mathbf{Y}_{i+1,H}))) \\ \mathbf{Y}_i^L &= \text{HLC}_n(\text{Concat}(\mathbf{Y}_{i,L}, \text{Down}(\mathbf{Y}_{i+1,L}))) \end{aligned} \quad (5)$$

式中: $\text{HLC}_n(\cdot)$ 表示 HLC 模块经过 n 次迭代后输出的结果; $\text{Down}(\cdot)$ 表示特征图的下采样操作; $\text{Concat}(\cdot)$ 表示特征图的拼接操作。如下算法叙述了 HLC-NET 对特征进行交叉融合的过程。

算法 1 高低频通道交叉融合算法过程

输入 \mathbf{X}_i 为输入低频特征矩阵; \mathbf{X}_{i+1} 为输入低频特征矩阵; a_{in} 为高频通道数; n 为迭代次数; \mathbf{W} 为卷积参数;

输出 \mathbf{Y} 为包含不同尺度的高低频特征分量

- 1) while $i \leq n$ do;
- 2) $\mathbf{X}_i^H \leftarrow \mathbf{X}_i * a_{in}, \mathbf{X}_{i+1}^H \leftarrow \mathbf{X}_{i+1} * a_{in}, a_{in} \in [0, 1]$;
- 3) $\mathbf{X}_i^L \leftarrow \mathbf{X}_i * (1 - a_{in}), \mathbf{X}_{i+1}^L \leftarrow \mathbf{X}_{i+1} * (1 - a_{in}), a_{in} \in [0, 1]$;
- 4) $(\mathbf{W}^{H \rightarrow L}, \mathbf{W}^{H \rightarrow H}, \mathbf{W}^{L \rightarrow L}, \mathbf{W}^{L \rightarrow H}) \leftarrow \text{Split}(\mathbf{W})$;
- 5) $\mathbf{X}_{i+1} \leftarrow \text{Up}(f(\mathbf{X}_L, \mathbf{W}^{L \rightarrow H})) + f(\mathbf{X}_H, \mathbf{W}^{H \rightarrow H})$;
- 6) $\mathbf{X}_i \leftarrow \text{Up}(f(\mathbf{X}_L, \mathbf{W}^{L \rightarrow L})) + f(\mathbf{X}_H, \mathbf{W}^{H \rightarrow L})$;
- 7) end while;
- 8) return $\mathbf{Y} = [\mathbf{Y}_i^H, \mathbf{Y}_i^L, \mathbf{Y}_{i+1}^H, \mathbf{Y}_{i+1}^L]$.

2.3 高低频通道特征分离模块 (HLC)

在夜间图像中, 信息往往以不同频率传输, 高频信号即图像变化剧烈的地方, 一般体现为图像的边缘信息。低频信号即图像变化平缓的地方, 一般体现为图像的色域信息。卷积操作作为

一种特征映射也可以看作是不同频率信息的融合。

融合低光条件下人脸的高频信息和低频信息会增强检测网络对人脸的定位能力。HLC 模块把卷积的输出映射到不同频率的通道上, 作为不同频率信息的融合。对于普通的卷积, 输入和输出特征都具有相同分辨率的映射, 无法有效提取特征的高频与低频部分。基于尺度空间理论^[30], 更高分辨率的特征承载着更多的细节信息。所以, HLC 模块将高频通道特征的分辨率放大为原来的 2 倍。

由图 4 可知, 根据文献 [31], 设 $\mathbf{X} \in \mathbf{R}^{C \times H \times W}$ 为 HLC 模块的输入特征, \mathbf{Y} 为 HLC 模块的输出特征, 其中 C 表示输入特征图的通道数量, H 和 W 表示为特征的空间维度。HLC 模块利用向量变换将 \mathbf{X} 分离为高尺度特征 \mathbf{X}_H 和低尺度特征 \mathbf{X}_L 。其中 $\mathbf{X}_H \in \mathbf{R}^{I_H \times 2H \times 2W}$, $\mathbf{X}_L \in \mathbf{R}^{I_L \times H \times W}$, I_H 是输入通道根据通道分离比率 a_{in} 设置的高频通道数, 即 $I_H = a_{in} \times C$, I_L 则取 $1 - a_{in}$; 同理, 可计算获得输出通道的高低通道数 O_H 与 O_L , 其中, $a_{in}, a_{out} \in [0, 1]$ 表示原通道变换为高或尺度特征通道的比率。 \mathbf{Y} 由 \mathbf{X}_H 和 \mathbf{X}_L 经过变换形成高频向量 $\mathbf{Y}_H \in \mathbf{R}^{O_H \times 2H \times 2W}$ 和低频向量 $\mathbf{Y}_L \in \mathbf{R}^{O_L \times H \times W}$ 。其中 $\mathbf{Y}_H = \mathbf{X}^{H \rightarrow H} + \mathbf{X}^{L \rightarrow H}$, $\mathbf{Y}_L = \mathbf{X}^{H \rightarrow L} + \mathbf{X}^{L \rightarrow L}$, $\mathbf{X}^{H \rightarrow *}$ 为高频分量向其他频率分量的映射, $\mathbf{X}^{L \rightarrow *}$ 为低频分量向其他频率分量的映射。

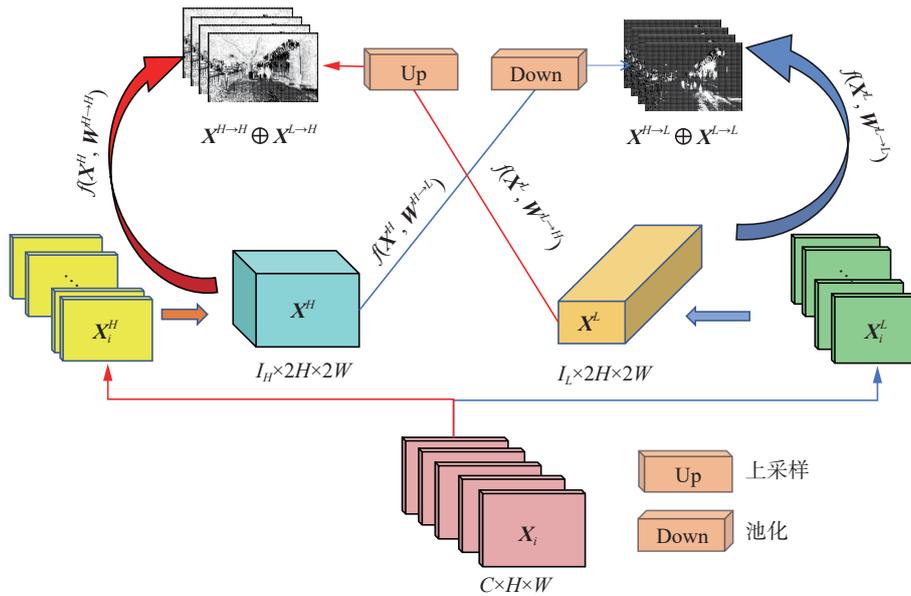


图 4 HLC 模型架构

Fig. 4 HLC model architecture

为了得到高低频通道分离的 \mathbf{X} , HLC 将卷积参数 \mathbf{W} 根据 $a \in \{a_{in}, a_{out}\}$ 比率分为 $\mathbf{W} = (\mathbf{W}_H, \mathbf{W}_L)$, \mathbf{W}_H

进一步分为 $\mathbf{W}_H = (\mathbf{W}^{H \rightarrow H}, \mathbf{W}^{L \rightarrow H})$, 同理, $\mathbf{W}_L = (\mathbf{W}^{H \rightarrow L}, \mathbf{W}^{L \rightarrow L})$, 具体如下式所示

$$\begin{aligned}
\mathbf{W}^{H \rightarrow H} &= \text{Split}(\mathbf{W}, (O_H, I_H)) \\
\mathbf{W}^{H \rightarrow L} &= \text{Split}(\mathbf{W}, (O_L, I_H)) \\
\mathbf{W}^{L \rightarrow H} &= \text{Split}(\mathbf{W}, (O_H, I_L)) \\
\mathbf{W}^{L \rightarrow L} &= \text{Split}(\mathbf{W}, (O_L, I_L))
\end{aligned} \quad (6)$$

式中: $\text{Split}(\cdot)$ 表示对 \mathbf{W} 参数的输入通道和输出通道的索引分别按照 I_* 和 O_* 切片分组, 其中, $*$ $\in [H, W]$ 。然后分离得到的 X_H 和 X_L 特征通过上述产生的卷积参数运算, 最终获得输出特征 \mathbf{Y} 的高低频通道特征, 具体生成过程如下

$$\begin{aligned}
\mathbf{Y}_H &= \text{Up}(f(X_L, \mathbf{W}^{L \rightarrow H})) + f(X_H, \mathbf{W}^{H \rightarrow H}) \\
\mathbf{Y}_L &= \text{Down}(f(X_H, \mathbf{W}^{H \rightarrow L})) + f(X_L, \mathbf{W}^{L \rightarrow L})
\end{aligned} \quad (7)$$

式中: $f(X^*, \mathbf{W}^{* \rightarrow *})$ 表示为 X^* 对应的卷积参数 $\mathbf{W}^{* \rightarrow *}$ 上计算的结果; $\text{Up}(\cdot)$ 表示上采样操作; $\text{Down}(\cdot)$ 表示下采样平均池化操作。

由图 4 可知, HLC 模块输出结果是该方法提取到的中间特征图。HLC 模块将输入特征映射到不同频率的通道上, 高频通道特征关注更多边缘信息, 低频通道特征关注更多色域信息。从不同尺度的特征图提取高低频通道信息并交叉融合, 可为检测网络提供更多不同尺度的通道特征信息, 有助于检测网络分辨有效特征。此外, 低频分量和高频分量的分辨率是不同的, 符合 FPN 网络中的尺度特征融合原则。

3 试验结果

为了测试提出方法的性能, 利用公开的夜间人脸检测数据集 DAKE FACE^[1], 以通用的网络架构和现有的精准人脸检测器为基准, 进行了大量的试验。

3.1 数据集介绍

本研究采用公开数据集 DARK FACE^[1] 作为实验数据集。该数据集提供了 6 000 张真实夜间环境下拍摄的低光人脸图像, 其标签为低光场景下的人脸标注, 共计 43 849 个手动标注的人脸边界框, 且标注的人脸有较大的尺度方差, 范围为 1 像素 \times 2 像素 \sim 335 像素 \times 296 像素。每张图像通常有 1~20 个带标注的人脸信息。数据集包含教学楼、街道、桥梁、立交桥和公园等场景。图像分辨率为 1 080 \times 720。因数据集在夜间低光环境下拍摄, 图像具有受光不均、人脸微小、遮挡和抖动等问题。测试阶段随机挑选 1 000 张图像作为测试集。由于原始测试未公开, 试验参考 COCO 测试指标^[32], 通过平均精度 (mAP) 作为衡量标准, 平均精度 (mAP) 计算方式为精确率 (Precision) 和召回率 (Recall) 曲线下的面积。

3.2 网络架构

试验采用 Center-Face^[2] 人脸检测器为基准检测器, 其作为一阶段 (One-stage) 检测器, 能够实时高效、高精度地预测人脸边框和特征点位置。该检测器以人脸边框中心为中心点, 使用向外发散的高斯分布热图 (HeatMap)、中心点偏移损失 (Offset) 和边界框宽高损失 (WH) 作为监督损失。其机制更符合人脸检测, 即以点为面。根据文献 [2], 采用 MobileNet_V2 网络作为主干网络, 取 P_2 、 P_3 、 P_5 、 P_7 为网络输出的 4 层特征图作为 Cross-FPN 的输入, 输入通道分别为 24、32、96、320。此外, 针对人脸检测器的参数还进行了微调, 将文献 [2] 中的损失比例参数 (1, 0.1, 0.1) 置为 (1, 0.1, 0.5)。

对于 Cross-FPN 结构, 采用 4 层 HLC-NET 迭代结构, 取最后一轮的迭代结果作为检测头的输入特征。其中, HLC-NET 通过 3 次 HLC 模块堆叠将分离的高低频通道特征交叉融合。3 次高低频通道划分参数 α_{in} 和 α_{out} 分别取 (0.5, 0.75)、(0.75, 0.5)、(0.5, 0.5)。

3.3 试验细节

数据预处理阶段, 使用先进的数据增强策略处理训练样本, 包括随机反转、随机亮度、随机裁剪、颜色抖动和几种过滤方法。训练集图像大小设置为 640 \times 640。

训练阶段, 将批量大小 (batch size) 设置为 4, 使用容量为 32 GB 的单 GPU 训练该网络模型, 15 h 即到达收敛状态。为了达到更快的收敛速度, 本研究采用 Adam 优化器^[33] 训练, 初始学习率设置为 0.01, 学习率动态调整为指数衰减, Gamma 指数为 0.99。将具有真实人脸边框信息且 IoU > 0.35 的区域视为正样本, 正样本和负样本的采样率设置为 1:3。

推理阶段, 首先将图像大小恢复至初始分辨率 1 080 \times 720。然后根据文献 [2] 的策略, 采用 3 \times 3 卷积核进行最大值池化操作, 替代非极大值抑制, 重叠面积比率设置为 0.35, 并保留前 300 个预测边界框。

3.4 试验结果及分析

3.4.1 传统检测器结果及分析

表 1 给出了最先进的人脸检测器和普通的目标检测器的测试结果。普通的目标检测器 Fast R-CNN^[34] 性能最差, mAP 为 3.6%。这表明夜间图像质量不佳或传统目标检测算法缺乏提取人脸特征或检测人脸边界框的独特设计。但现有先进的人脸检测器 single stage headless face detector

(SSH^[35])、RetinaFace^[36]、PyramidBox^[37]、deconvolutional single shot detector (DASD)^[38] 和 Center-Face^[2] 性能表现也不突出, mAP 多为 20% 以下的检测精度。这表明在光照不足条件下的图像, 会给人脸检测等高级任务造成极大的阻碍。

3.4.2 基于低光增强的结果及分析

本节进行了多种先进的低光图像增强技术的对比试验, 试图寻求更加精确的检测方案。试验中检测阶段均使用 Center-Face 人脸检测器。试验结果如图 5 所示, 基于传统算法^[6, 8] 的低光图像增强方法取得了较好的精度, mAP 分别为 26.9% 和 28.1%。其他增强方法^[4, 14-16] 比原始检测效果有一定提升, 但由于增强后的图像存在噪声、伪影等问题, 只能检测到特征较为明显的人脸, 并且噪声、伪影等问题并没有得到抑制, 甚至进行了放大, 网络对人脸位置的感知能力减弱。Cross-FPN 方法可以有序交叉融合高低频通道信息, 使检测网络更加关注有效特征, 因此尤其是小人脸的检测效果是突出的。

表 1 不同检测方法的精度对比

Table 1 Accuracy comparison of different detection methods %

检测类型	方法	mAP
人脸/目标检测	Fast R-CNN ^[34]	3.6
	SSH ^[35]	8.1
	RetinaFace ^[36]	10.7
	PyramidBox ^[37]	15.8
	Center-Face ^[2]	17.4
	DSFD ^[38]	20.1
低光增强 (Center-Face)	Gamma ^[6]	26.9
	RetinexNet ^[14]	25.7
	Kind ^[15]	26.3
	Kind ++ ^[16]	28.4
	LIME ^[8]	28.1
	Zero-DCE ^[4]	28.8
特征金字塔 (Center-Face) (Zero-DCE ^[4])	FPN ^[39]	40.6
	DilateEncoder ^[40]	43.8
	本研究方法 (Cross-FPN)	44.6



图 5 不同增强方法结果比较

Fig. 5 Comparison of Results of Different Enhancement Methods

3.4.3 基于 FPN 方法结果及分析

基于可见光的人脸检测器虽然在低光图像增强技术的辅助下性能有所提升, 但是微小人脸的检测效果依旧不容乐观。在 DAKE FACE 数据集中, 小人脸的比例非常高。引入 FPN 结构, 帮助神经网络学习更多尺度的人脸特征。由表 1 可知, 在低光图像增强的基础上, 对比了一系列基于 FPN 的检测方法和提出的 Cross-FPN 方法。其中, Cross-FPN 方法比 FPN 方法^[39] 的精度提升了 4.0 个百分点; 比文献 [40] 提出的 DilateEncoder 方法^[40], 精度提升了 0.8 个百分点。结果表明, 在 FPN 方法的基础上, Cross-FPN 方法使得多尺度

特征的高低频信息更加突出, 预测出的锚框 (anchor) 更加密集, 提高了小人脸的检测精度。

3.5 消融试验及分析

本节进行了消融试验, 证明提出的 Cross-FPN 方法的有效性和合理性。

试验对比了有无 Cross-FPN 方法的检测网络的性能 (表 1), 引入提出的方法, 检测性能比引入 FPN 后的精度提升了 4.0 个百分点。

3.5.1 有无预训练的影响

由表 2 可知, 试验比较了有无预训练的网络性能, 试验结果表明, 在 ImageNet 数据集上进行过预训练的 MobileNet_V2 网络表现并不好, 这是

由于光照充足域的特征较为复杂, 并且与夜间域的特征相差较大, 使得网络对夜间域的特征能力减弱。

表 2 有无预训练的指标对比
Table 2 Comparison of pre-trained Indicators %

预训练	mAP
是	42.9
否	44.6

3.5.2 不同 α_{in} 和 α_{out} 影响

由表 3 可知, 试验比较了使用不同输入通道 α_{in} 比率和输出通道比率 α_{out} 的情况下, 基于 Cross-FPN 的低光人脸检测的效果和精度。试验表明, 输出高频通道比率越大, Cross-FPN 的效果相对较好, 说明神经网络更加关注于高频特征信息。以高频细节信息作为补充, 神经网络可以提取到更多有效特征, 有助于检测头更好地区分正负样本。

表 3 不同 α_{in} 与 α_{out} 精度对比
Table 3 Comparison of accuracy of α_{in} and α_{out} %

$(\alpha_{in}, \alpha_{out})$	mAP
(0.5, 0.25)(0.25, 0.75)(0.75, 0.5)	43.9
(0.5, 0.5)(0.5, 0.5)(0.5, 0.5)	44.3
(0.5, 0.75)(0.75, 0.5)(0.5, 0.5)	44.6

3.5.3 融合方式的影响

由表 4 可知, 试验比较了高低频通道不同的融合方式。试验表明, 在不同特征的高频通道特征之间相互融合的效果略好于高低频相互融合的效果。人脸检测高级视觉任务并不是提取特征的感受野越大越好, 它更需要关注特征的位置信息。这种位置信息对于夜间人脸检测更加重要。将特征的高频和低频通道信息有序融合, 突出不同频率下的特征, 提高了检测定位的能力, 证明了 Cross-FPN 方法可行。

表 4 通道融合方式的消融试验
Table 4 Ablation studies of channel fusion %

方法/融合方式	mAP
高频与低频	44.2
高频与高频	44.6

4 结束语

提出方法在低光图像增强和 FPN 结构的基础上, 针对夜间人脸特征, 提出了一种新型的高低频通道分离的特征融合方式(HLC), 构建了多尺度高低频通道特征交叉融合方法(Cross-FPN)。大量的试验验证了高低频率通道的特征信息有序

结合对夜间人脸检测的有效性。由于方法旨在结合多尺度特征、深浅层特征和锚框尺寸优化等有利于夜间小人脸检测任务的理论方法, 使得其检测结果更为精准。因此, 在一定程度上会牺牲计算时间以获得较高准确度。在今后的工作中, 将致力于平衡不同光照强度下图像的特征频域信息, 寻求利用不同源域下特征的频率通道信息高效交互方式, 可以在较少运行时间下提高夜间人脸检测精度。

参考文献:

- [1] YUAN Ye, YANG Wenhan, REN Wenqi, et al. UG2+Track 2: A collective benchmark effort for evaluating and advancing image understanding in poor visibility environments[EB/OL]. (2019-04-09)[2022-01-01]. <https://doi.org/10.48550/arXiv.1904.04474>.
- [2] XU Yuanyuan, YAN Wan, YANG Genke, et al. Center-Face: joint face detection and alignment using face as point[J]. Scientific programming, 2020, 2020: 1-8.
- [3] YANG Shuo, LUO Ping, LOY C C, et al. WIDER FACE: a face detection benchmark[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 5525-5533.
- [4] GUO Chunle, LI Chongyi, GUO Jichang, et al. Zero-reference deep curve estimation for low-light image enhancement[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020: 1777-1786.
- [5] PIZER S M, AMBURN E P, AUSTIN J D, et al. Adaptive histogram equalization and its variations[J]. *Computer vision, graphics, and image processing*, 1987, 39(3): 355-368.
- [6] FARID H. Blind inverse Gamma correction[J]. *IEEE transactions on image processing*, 2001, 10(10): 1428-1433.
- [7] LAND E H, MCCANN J J. Lightness and retinex theory[J]. *Journal of the optical society of America*, 1971, 61(1): 1-11.
- [8] GUO Xiaojie. LIME: a method for low-light image enhancement[C]//Proceedings of the 24th ACM International Conference on Multimedia. New York: ACM, 2016: 87-91.
- [9] RAHMAN Z U, JOBSON D J, WOODDELL G A. Retinex processing for automatic image enhancement[J]. *Journal of electronic imaging*, 2004, 13(1): 100-110.
- [10] FU Xueyang, ZENG Delu, HUANG Yue, et al. A fusion-

- based enhancing method for weakly illuminated images[J]. *Signal processing*, 2016, 129: 82–96.
- [11] FU Xueyang, ZENG Delu, HUANG Yue, et al. A weighted variational model for simultaneous reflectance and illumination estimation[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 2782–2790.
- [12] WANG Shuhang, ZHENG Jin, HU Haimiao, et al. Naturalness preserved enhancement algorithm for non-uniform illumination images[J]. *IEEE transactions on image processing: a publication of the IEEE signal processing society*, 2013, 22(9): 3538–3548.
- [13] LI Mading, LIU Jiaying, YANG Wenhan, et al. Structure-revealing low-light image enhancement via robust retinex model[J]. *IEEE transactions on image processing*, 2018, 27(6): 2828–2841.
- [14] WEI Chen, WANG Wenjing, YANG Wenhan, et al. Deep retinex decomposition for low-light enhancement[EB/OL]. (2018–08–14)[2022–01–01]. <https://arxiv.org/abs/1808.04560.pdf>
- [15] ZHANG Yonghua, ZHANG Jiawan, GUO Xiaojie. Kindling the darkness: a practical low-light image enhancer[C]//Proceedings of the 27th ACM International Conference on Multimedia. New York: ACM, 2019: 1632–1640.
- [16] ZHANG Yonghua, GUO Xiaojie, MA Jiayi, et al. Beyond brightening low-light images[J]. *International journal of computer vision*, 2021, 129(4): 1013–1037.
- [17] 王克琪, 钱宇华, 梁吉业, 等. 局部-全局关系耦合的低照度图像增强 [J]. *中国科学:信息科学*, 2022, 52(3): 443–460.
- WANG Keqi, QIAN Yuhua, LIANG Jiye, et al. Local-global coupling relationship based low-light image enhancement[J]. *Scientia sinica (informationis)*, 2022, 52(3): 443–460.
- [18] GONG Yuqi, YU Xuehui, DING Yao, et al. Effective fusion factor in FPN for tiny object detection[C]//2021 IEEE Winter Conference on Applications of Computer Vision. Waikoloa: IEEE, 2021: 1159–1167.
- [19] MEI Y, FAN Y, ZHANG Y, et al. Pyramid attention network for image restoration[J]. *International journal of computer vision*, 2023, 131(12): 3207–3225.
- [20] LIM J S, ASTRID M, YOON H J, et al. Small object detection using context and attention[C]//2021 International Conference on Artificial Intelligence in Information and Communication. Jeju Island: IEEE, 2021: 181–186.
- [21] XU A, YAO A, LI A, et al. Auto-FPN: automatic network architecture adaptation for object detection beyond classification[C]//2019 IEEE/CVF International Conference on Computer Vision. Seoul: IEEE, 2020: 6648–6657.
- [22] LUO Yihao, CAO Xiang, ZHANG Juntao, et al. CE-FPN: enhancing channel information for object detection[J]. *Multimedia tools and applications*, 2022, 81(21): 30685–30704.
- [23] ZHANG Tianwen, ZHANG Xiaoling, KE Xiao. Quad-FPN: a novel quad feature pyramid network for SAR ship detection[J]. *Remote sensing*, 2021, 13(14): 2771.
- [24] JIANG Yiqi, TAN Zhiyu, WANG Junyan, et al. Giraffe-Det: a heavy-neck paradigm for object detection[EB/OL]. (2022–02–09)[2022–04–04]. <https://doi.org/10.48550/arXiv.2202.04256>.
- [25] CAO Junxu, CHEN Qi, GUO Jun, et al. Attention-guided context feature pyramid network for object detection[EB/OL]. (2020–05–23)[2022–04–04]. <https://arxiv.org/abs/2005.11475>.
- [26] SASAGAWA Y, NAGAHARA H. YOLO in the dark - domain adaptation method for merging multiple models[C]//Proceedings of the European Conference on Computer Vision. [S.l.]: Springer, 2020: 345–359.
- [27] WANG Wenjing, YANG Wenhan, LIU Jiaying. HLA-face: joint high-low adaptation for low light face detection[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021: 16190–16199.
- [28] LIANG Jinxiu, WANG Jingwen, QUAN Yuhui, et al. Recurrent exposure generation for low-light face detection[J]. *IEEE transactions on multimedia*, 2022, 24: 1609–1621.
- [29] HUANG Gao, LIU Zhuang, VAN DER MAATEN L, et al. Densely connected convolutional networks[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 2261–2269.
- [30] LINDEBERG T. Scale-space theory in computer vision[M]. [S. l.]: Springer Science & Business Media, 2013.
- [31] CHEN Yunpeng, FAN Haoqi, XU Bing, et al. Drop an octave: reducing spatial redundancy in convolutional neural networks with octave convolution[C]//2019 IEEE/CVF International Conference on Computer Vision. Seoul: IEEE, 2020: 3434–3443.
- [32] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft COCO: common objects in context[C]//European Confer-

- ence on Computer Vision. Cham: Springer, 2014: 740–755.
- [33] DA K. A method for stochastic optimization[EB/OL]. (2014–12–22)[2022–01–01]. <https://doi.org/10.48550/arXiv.1412.6980>.
- [34] GIRSHICK R. Fast R-CNN[C]//2015 IEEE International Conference on Computer Vision. Santiago: IEEE, 2016: 1440–1448.
- [35] NAJIBI M, SAMANGOUEI P, CHELLAPPA R, et al. SSH: single stage headless face detector[C]//2017 IEEE International Conference on Computer Vision. Venice: IEEE, 2017: 4885–4894.
- [36] DENG Jiankang, GUO Jia, VERVERAS E, et al. Retina-Face: single-shot multi-level face localisation in the wild[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020: 5202–5211.
- [37] TANG Xu, DU D K, HE Zeqiang, et al. PyramidBox: A context-assisted single shot face detector[C]//European Conference on Computer Vision. Cham: Springer, 2018: 812–828.
- [38] LI Jian, WANG Yabiao, WANG Changan, et al. DSFD: dual shot face detector[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2020: 5055–5064.
- [39] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 936–944.
- [40] CHEN Qiang, WANG Yingming, YANG Tong, et al. You only look one-level feature[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021: 13034–13043.

作者简介:



许皓, 硕士研究生, 主要研究方向为夜间人脸检测、目标检测、低光增强。E-mail: 552420610@qq.com。



钱宇华, 教授、博士生导师, 山西大学大数据科学与产业研究院院长, 主要研究方向为人工智能、大数据、AI for Science。先后主持国家自然科学基金重点项目、国家重点研发计划项目、国防重点项目等国家/省部级项目 20 余项。曾获山西省科学技术奖(自然科学类)一等奖, 入选全球高被引科学家。发表学术论文 200 余篇。E-mail: jinchengqyh@126.com。



王克琪, 博士研究生, 主要研究方向为深度学习、计算机视觉、低照度图像增强。E-mail: wkq_2021@163.com。