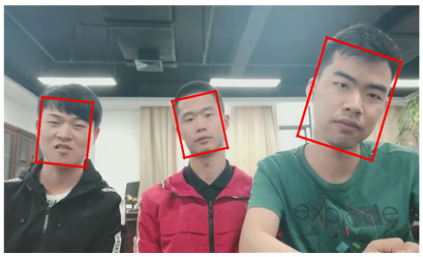
Opto-Electronic Engineering
光 电 工 程

Article

2020 年，第 47 卷，第 1 期

DOI: 10.12086/oee.2020.190135

级联网络和金字塔光流的旋转不变人脸检测



孙　锐[1,2]，阚俊松[1,2*]，吴柳玮[1,2]，王　鹏[3]

[1]合肥工业大学计算机与信息学院，安徽 合肥 230009；
[2]工业安全与应急技术安徽省重点实验室，安徽 合肥 230009；
[3]合肥进毅智能技术有限公司，安徽 合肥 230088

摘要：在无约束的开放空间中，由于面部姿态变化、背景环境复杂、运动模糊等，人脸检测仍是一个具有挑战性的任务。本文针对视频流中人脸检测存在的平面内旋转问题，将人脸关键点与金字塔光流相结合，提出了基于级联网络和金字塔光流的旋转不变人脸检测算法。首先利用级联渐进卷积神经网络对视频流中前一帧进行人脸位置和关键点的定位；其次为获取关键点与人脸候选框间光流映射，使用独立的关键点检测网络对当前帧进行再次定位；之后计算前后两帧之间关键点光流位移；最后通过关键点光流位移与人脸候选框的映射关系，对视频中检测到的人脸进行校正，从而完成平面内旋转人脸不变性检测。实验经 FDDB 公开数据集上测试，证明该方法精确度较高。并且，在 Boston 面部跟踪数据集上进行动态测试，证明该人脸检测算法能有效解决平面内旋转人脸检测问题。对比其它检测算法，该算法检测速度有较大优势，同时视频中窗口抖动问题得到了很好解决。

关键词：旋转不变性；关键点检测；级联渐进网络；金字塔光流；人脸检测

中图分类号：TP391.41；TP183　　**文献标志码：**A

引用格式：孙锐，阚俊松，吴柳玮，等. 级联网络和金字塔光流的旋转不变人脸检测[J]. 光电工程，2020，**47**(1): 190135

Rotating invariant face detection via cascaded networks and pyramidal optical flows

Sun Rui[1,2], Kan Junsong[1,2*], Wu Liuwei[1,2], Wang Peng[3]

[1]School of Computer and Information, Hefei University of Technology, Hefei, Anhui 230009, China;
[2]Anhui Province Key Laboratory of Industry Safety and Emergency Technology, Hefei, Anhui 230009, China;
[3]Hefei Jinyi Science and Technology, Hefei, Anhui 230088, China

Abstract: In the unconstrained open-space, face detection is still a challenging task due to the facial posture changes, complex background environment, and motion blur. The rotation-invariant algorithm based on cascaded network and pyramid optical flow is proposed. Firstly, the cascading progressive convolutional neural network is adopted to locate the face position and facial landmark of the previous frame in the video stream. Secondly, the independent facial landmark detection network is used to reposition the current frame, and the optical flow mapping displacement of the facial landmark between the two frames is calculated afterwards. Finally, the detected face is corrected by the mapping relationship between the optical flow displacement of the facial landmark and the bounding

收稿日期：2019-03-25；　收到修改稿日期：2019-05-14
基金项目：国家自然科学基金资助项目(61471154)；中央高校基本科研业务费专项资金资助项目(JZ2018YYPY0287)
作者简介：孙锐(1976-)，男，博士，教授，主要从事计算机视觉的研究。E-mail：sunrui@hfut.edu.cn
通信作者：阚俊松(1995-)，男，硕士研究生，主要从事计算机视觉的研究。E-mail：2931338359@qq.com

box, thereby completing the rotation-invariant face detection. The experiment was tested on the FDDB public datasets, which proved that the method is more accurate. Moreover, the dynamic test on the Boston head tracking database proves that the face detection algorithm can effectively solve the problem of rotation-invariant face detection. Compared with other detection algorithms, the detection speed of the proposed algorithm has a great advantage, and the window jitter problem in the video is well solved.

Keywords: rotation-invariant; facial landmark; calibration networks; pyramid optical flow; face detection

Citation: Sun R, Kan J S, Wu L W, *et al.* Rotating invariant face detection via cascaded networks and pyramidal optical flows[J]. *Opto-Electronic Engineering*, 2020, **47**(1): 190135

1 引 言

人脸识别是计算机视觉领域比较成功的应用之一。随着视频监控在城市空间的快速普及，公安部门采集了海量无约束开放环境下的视频，视频流中的人脸检测存在尺度变化、局部遮挡、运动模糊以及光照变化等复杂问题，特别是人脸旋转会影响整个人脸识别系统性能和效率。本文针对平面内人脸旋转问题展开研究并提出一种新颖的人脸检测方法。

传统的人脸检测方法主要在 Viola 和 Jones[1]的工作基础上进行改进，该框架也被扩展到处理旋转人脸检测[2]，通过手工设计不同特征[3-4]在一定程度上改进人脸检测性能，但是这些特征增加了计算复杂度，对环境适应能力差。近年来，随着卷积神经网络的不断发展进步，传统手工设计特征方法逐步演化到深度卷积神经网络学习特征[5-6]。人脸检测中如关键点定位[7]、面部空间结构学习[8]、克服严重遮挡[9]等，通过采用深度卷积神经网络[13-14]，大大提高了人脸检测的精度。

现阶段，经典的检测算法都针对通用目标提出。将 Faster R-CNN[10]、SSD[11]、YOLO[12]等系列通用网络框用到人脸检测(如 Face R-CNN、S^3FD 等)。虽然对比传统方法在准确性上有较高的优势，但并没有很好地针对人脸旋转问题，常见的旋转不变性人脸检测还是以传统方法为主[15-16]。实际应用中，虽然通用检测网络的检测效果得到提高，但是单纯运用较深的网络会使处理速度较慢，同时缺乏时间一致性。一些最新文献中以 cascade CNN[17]为基本模型设计人脸检测，如文献[18-20]，但 cascade CNN 在视频流中进行人脸检测缺少一种方法来保存面部信息，不仅可能出现人脸位置的跳变，而且如果一个面部短暂扭曲或发生遮挡，会导致检测突然失败。深度学习也推动了目标跟踪的发展，如文献[21-22]等将光流引入卷积神经网络，以实现目标跟踪。本文将光流引入人脸检测，提出了一种扩展级联卷积神经网络的方法，以实现一种适应随时间推移的平面内旋转不变检测。

综上所述，本文主要贡献如下：

1) 优化级联卷积神经网络

由于关键点的位置对人脸候选框的校正有直接影响。为了让校正结果更加准确，一方面将前一帧人脸关键定位任务进行分离并去除了部分网络的最大池化以提高人脸关键点定位精度；另一方面，为了加速当前帧中人脸的检测效率，设计了独立的关键点检测网络，并对部分关键点进行校验，排除检测异常的关键点。

2) 加入金字塔光流映射

通用网络框架由于没有考虑到时间信息，产生的人脸边界框不稳定。尽管脸部不发生移动，边界框的大小和位置也会有变化，有时会因此丢失一些中间帧的跟踪。为了克服这些问题，我们使用一个光流映射来保存人脸以及先前计算的信息，从而减少因面部短暂扭曲或遮挡时而产生的检测失败。

2 整体框架

解决人脸旋转不变性的常见方式有：1) 通过高度复杂网络学习旋转不变特征；2) 对样本进行划分，训练多个模型；3) 在检测到人脸之后，对人脸进行校正。前两种方式会大大降低检测效率。本文根据特征点位置进行候选框的调整，实现人脸的不变性检测。视频流中人脸的旋转不变性检测整体流程图框架如图 1。

本文的思路是初始化过程加载级联渐进网络，其中级联渐进网络分为三个阶段，阶段一、阶段二为区分人脸与非人脸，阶段三区分人脸和非人脸和人脸关键点定位。视频流中前一帧经过三层级联渐进卷积神经网络进行人脸识别和关键点定位后，再用独立的关键点检测网络对当前帧人脸区域进行关键点再次定位，计算前一帧与当前帧之间光流映射。通过关键点光流位移与人脸候选框的映射关系，对视频中检测到的人脸进行校正，完成视频流中旋转不变性人脸检测。

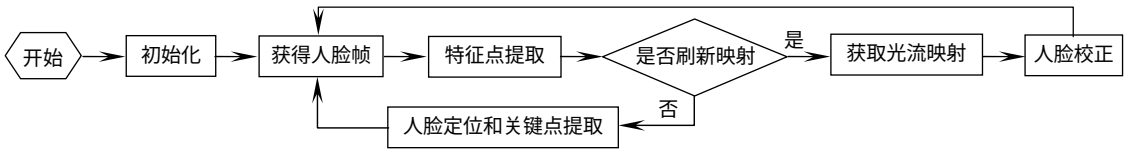


图 1 整体流程图

Fig. 1 Flow chart of developed algorithm

3 旋转不变人脸检测

本文先对输入图像进行双线性插值，调整到不同比例，构建图像金字塔，再通过级联卷积神经网络获得回归边界框和关键点坐标，最后通过光流校正获得回归边界框。该算法先构造图像金字塔，用于两部分：

1) 检测人脸中利用到了图像金字塔。检测人脸过程中这些不同比例的图像作为三个阶段的输入进行训练，目的是可以检测不同尺寸的人脸。

2) 计算光流，用于人脸校正。两帧之间利用金字塔光流信息差获得光流映射，进行人脸定位与校正。

构造过程分为两步，先对图像进行高斯平滑，再借助亚采样获得上一幅图像的缩略图。设当前帧 J^0 分辨率大小为 $M\times N$，金字塔第 l 层是由 $l-1$ 层图像 J^{l-1} 经高斯窗口函数 W 卷积和向下采样得到，获取金字塔：

$$J^l(i,j)=\sum_{m=-s}^{s}\sum_{n=-s}^{s}W(m,n)J^{l-1}(2i+m,2j+n)\quad,\qquad(1)$$

其中：$0\le i<M/2^l$，$0\le j<N/2^l$，$0\le l\le t$，t为分解层数，W 大小为 $(2s+1)\times(2s+1)$。

在文献[23]中，使用多个 CNN 来进行人脸检测。但是本文方法针对旋转不变性的人脸检测，需要更精准的关键点定位和图像边缘信息的提取。为此，将面部检测任务与关键点定位任务进行分离，设计了三阶段级联渐进卷积神经网络，结构示意图如图 2 所示。

如图 2(a)所示，第一阶段基本的构造是一个全连接网络。上一步构建完成的图像金字塔通过一个全卷积网络进行初步特征提取与边框标定，获取候选窗口以及其边界框回归向量，使用边界框回归的方法校准候选边框，经非极大值抑制合并高度重叠的候选框。

如图 2(b)(即第二阶段)为一个卷积神经网络，相比图 2(a)(即第一阶段)增加了全连接层，对输入数据的筛选更加严格。其中所有候选边框来源于第一阶段，对候选框重新选择，拒绝大量假候选框，经边界框回归校准，用非极大值抑制对候选框合并。图 2(c)(第三阶段)一个卷积神经网络，相比图 2(a)，增加了卷积层与全连接层，该阶段不仅需要进行人脸非人脸的分类和边界框的回归还需要进行人脸关键点定位。

3.1 三阶段级联渐进卷积神经网络

在第一阶段中获取候选窗口以及其边界框回归向量，校准候选边框向量，经非极大值抑制合并候选框。第二阶段中所有备候选边框来源于第一阶段，对备选框进行重新选择，拒绝大量假候选框。

人脸分类：

$$L_i^{\det}=-(y_i^{\det}\log(p_i)+(1-y_i^{\det})(1-\log(p_i)))\quad,\qquad(2)$$

其中：p_i 是人脸的概率，$y_i^{\det}$ 为背景的真实标签，$y_i^{\det}\in\{0,1\}$，式(2)为人脸分类的交叉熵损失函数。

边界框回归：

$$L_i^{\text{box}}=\left\|\hat{y}_i^{\text{box}}-y_i^{\text{box}}\right\|_2^2\quad,\qquad(3)$$

其中 $y_i^{\det}\in R^4$。阶段一、阶段二输入源的训练：

$$\min\sum_{i=1}^{N}\sum_{j\in\{\det,\text{box}\}}\alpha_j\beta_i^jL_i^j,\ \beta_i^j\in\{0,1\}\quad,\qquad(4)$$

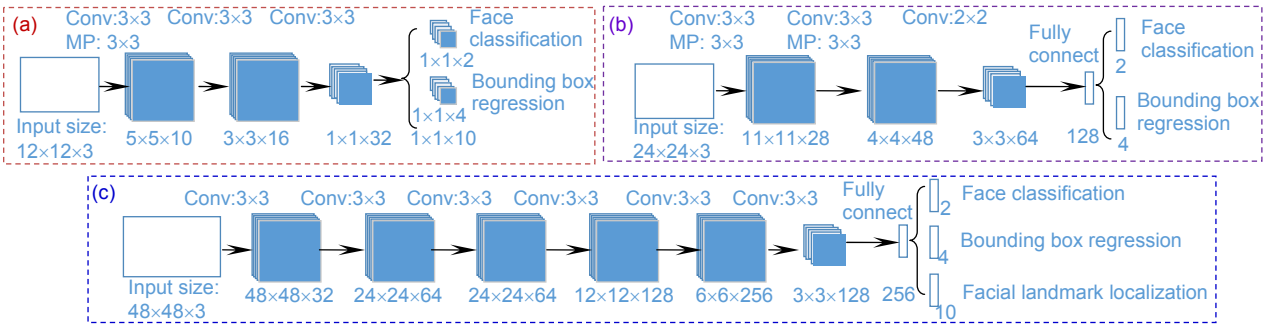


图 2 三阶段级联渐进卷积神经网络。(a) 第一阶段；(b) 第二阶段；(c) 第三阶段

Fig. 2 Three-stage cascade progressive convolutional neural network. (a) First stage; (b) Second stage; (c) Third stage

其中：α_j 为任务重要性，β_i^j 为样本标签，L_i^j 为上面的损失函数，$\hat{y}_i^{\text{box}}$ 为通过网络预测得到，y_i^{box} 为实际的真实标签。级联渐进网络 1 与级联渐进网络 2 中 α_{det} 设置为 1，α_{box} 为 0.5。整个训练过程就是上面函数最小化的过程。

由于人脸特征点占图像比例较小，虽然最大池化操作能减小输入大小，使神经网络能专注于重要元素，但是容易导致人脸特征点的信息损失。本论文取消了最大池化，同时网络将输出五个面部特征位置。对比阶段一、二，第三阶段中不仅需要进行人脸非人脸的分类和边界框的回归还需要进行人脸关键点的定位：

$$L_i^{\text{landmark}}=\left\|\hat{y}_i^{\text{landmark}}-y_i^{\text{landmark}}\right\|_2^2,\ y_i^{\text{landmark}}\in R^{10} \quad 。(5)$$

式(5)为通过欧氏距离计算的回归损失。$\boldsymbol{y}$ 是左上角坐标(x、y)、长、宽四个元素(参数)所组成。计算网络预测的地标位置和真实地标的欧氏距离，并最小化。

第三阶段输入源的训练：

$$\min\sum_{i=1}^{N}\sum_{j\in\{\text{det,box,landmark}\}}\alpha_j\beta_i^jL_i^j,\ \beta_i^j\in\{0,1\} \quad , \quad (6)$$

级联渐进网络 3 的 α_{det} 设置为 1，α_{box} 为 0.5，α_{landmark} 为 1。

级联渐进卷积神经网络在检测图像序列中人脸时，每一帧都是单独处理的。这意味着，通过在整个图像上移动不同大小的窗口并对其进行评估。由于没有考虑到时间信息，产生的人脸边界框也不稳定。

3.2 金字塔 LK 光流映射

光流是相邻帧之间描述物体运动信息的一种方法，常运用于视频中对物体的跟踪。本算法引入光流的概念，通过光流场模式分类对运动人脸的检测。同时解决了在人脸跟踪过程中新出现了人脸。而无法检测到的情况。基本的光流方程：

$$I(x+\Delta x,y+\Delta y,t+\Delta t)=I(x,y,t) \quad , \quad (7)$$

其中：$I(x,y,t)$ 为像素点 (x,y) 在 t 时刻的照度值，将式(7)左边泰勒级数展开：

$$I(x,y,t)+\Delta x\cdot\frac{\partial I}{\partial x}+\Delta y\cdot\frac{\partial I}{\partial y}+\Delta t\cdot\frac{\partial I}{\partial t}+e=I(x,y,t) \quad , (8)$$

令 $u=\Delta x/\Delta t$，$v=\Delta y/\Delta t$，则 (u,v) 表示光流场，u 和 v 分别代表该点光流的 x 轴的分量和 y 轴分量。设 $I_x=\partial I/\partial x$，$I_y=\partial I/\partial y$，$I_t=\partial I/\partial t$，将 u，v，I_x，I_y，I_t 代入式(7)得到光流约束方程：

$$I_x\cdot u+I_y\cdot v+I_t=0 \quad 。 \quad (9)$$

算法使用光流映射防止人脸的丢失，同时防止跟踪的脸部区域发生不规律变化。假设人脸图像上一个小的空间邻域内具有相同的光流速度，且存在有限点，根据光流约束方程，可得：

$$I_x(x_i)u+I_y(x_i)v+I_t(x_i)=0 \quad , \quad (10)$$

式中：$x_i\in\Omega,(i=0,1,\cdots,m-1)$，$m$ 为局部小范围内存在点个数，求解结果即是光流映射值，其过程如下：

在图像一个小空间邻域Ω内，光流估计误差为

$$E_{\text{LK}}=\sum_{x_i\in\Omega}(I_x(x_i)u+I_y(x_i)v+I_t(x_i))^2 \quad 。 \quad (11)$$

设 $\boldsymbol{v}=(u,v)^{\text{T}}$，$\nabla I(x)=(I_x,I_y)^{\text{T}}$，最小化 E_{LK} 得到光流计算公式：

$$\boldsymbol{A}^{\text{T}}\boldsymbol{A}\boldsymbol{v}=\boldsymbol{A}^{\text{T}}\boldsymbol{b} \quad , \quad (12)$$

式中：$\boldsymbol{A}=[\nabla I(x_1),\nabla I(x_2),\cdots,\nabla I(x_m)]$ 为系数矩阵，$\boldsymbol{b}=-[I_t(x_1),I_t(x_2),\cdots,I_t(x_m)]^{\text{T}}$ 为常数项，解上式得：

$$\boldsymbol{v}=(\boldsymbol{A}^{\text{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\text{T}}\boldsymbol{b} \quad , \quad (13)$$

式中：$\boldsymbol{A}^{\text{T}}\boldsymbol{A}$ 是一个 2×2 的矩阵，

$$\boldsymbol{A}^{\text{T}}\boldsymbol{A}=\begin{bmatrix}\sum\limits_{x_i\in\Omega}I_x^2(x_i) & \sum\limits_{x_i\in\Omega}I_x(x_i)I_y(x_i)\\ \sum\limits_{x_i\in\Omega}I_y(x_i)I_x(x_i) & \sum\limits_{x_i\in\Omega}I_y^2(x_i)\end{bmatrix} 。$$

当矩阵 $\boldsymbol{A}^{\text{T}}\boldsymbol{A}$ 是非奇异矩阵时，就可以得到光流 v 的解析解。通过矩阵 $\boldsymbol{A}^{\text{T}}\boldsymbol{A}$ 的特征值判断 v 的可靠性，设 $\boldsymbol{A}^{\text{T}}\boldsymbol{A}$ 的特征值为 λ_1 和 λ_2，且 $\lambda_1\geq\lambda_2$。若 λ_1 和 λ_2 均大于某个阈值 γ，则由式(11)计算光流值，若不能满足 λ_1 和 λ_2 均大于某一个阈值 γ 的条件，则不进行光流计算。

在前面的步骤中已经获得了初始人脸位置和人脸关键点，对视频中人脸关键点做相应标记。为了加速视频帧处理，本文设计了单独的人脸关键点检测网络，在前一帧经级联渐进卷积神经网络所产生的人脸候选框和关键点位置的基础上，对当前帧人脸区域进行人脸关键点检测。特征点提取网络结构如图 3 所示，该网络基本构造是包含全连接层的神经网络，完成对关键点定位任务。

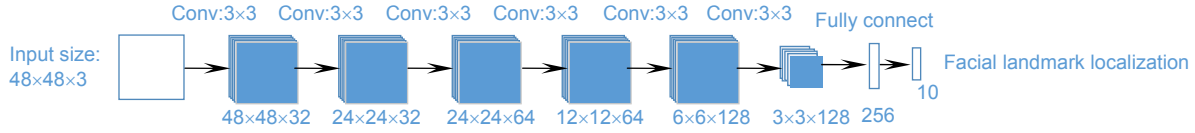


图 3 特征点提取网络

Fig. 3 Facial landmark extraction network

由于人脸特征点的特征窗口要求较高，如果窗口选择太小，易出现矩阵 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ 为奇异矩阵的情况，无法得到光流的解析解，求出的光流无法描述物体的运动；如果人脸特征点窗口选择太大，则又不能满足图像灰度的一致性假设。为此，构建了金字塔分层光流。

构建图像高斯金字塔，图像做高斯平滑和向下采样。构建步骤与人脸检测过程中的金字塔构造方式相同，人脸金字塔光流示意图如图 4 所示。图中 I 为前一帧图像的灰度，J 为当前帧图像的灰度，前一帧图像金字塔的第 $0,1,\cdots,n-1$ 层分别用 $I^0,I^1,\cdots,I^{n-1}$ 来表示，其中 I^0 表示原始图像，位于金字塔的第 0 层。金字塔自下而上每一层的像素数会不断减少，通过对上层的小尺度、低分辨率的图像的分析所得到的信息指导下层大尺度、高分辨率图像的分析。

图像金字塔算法处理图像，将在图像开始时金字塔顶层，即较小的空间尺度上进行关键点的预测跟踪，再通过金字塔迭代向下直到金字塔的底层的处理来修正初始运动向量的假定。

获得图像金字塔后，通过计算关键点的偏移映射来校正人脸。为了防止错误的校正，对于检测到的特征点本文采用正反向误差检测判断是否有效。利用人脸关键点结合检测到的人脸中心点，将这些点作为有效的特征点。通过 LK 稀疏光流法对下一帧进行特征点和候选框的预测，同时采用金字塔搜索，保证了在独立关键点检测网络中下一帧检测的速度和准确度。

3.3 人脸校正

经过这样由粗到细的光流估计，对于每个关键点赋予速度矢量，形成视频中人脸图像的运动场，使视频流中运动的人脸关键点获得准确的定位。

每当调用金字塔光流映射时，它会生成一个包含人脸位置与人脸关键点的映射面，通过获取的特征点与面部的映射关系校正人脸边界框。为了降低人脸不变性检测器对计算能力的要求，视频流中级联渐进卷积神经网络每间隔 n 帧执行一次，每帧都进行误差检测，判断是否更新映射图。光流映射图如图 5 所示。

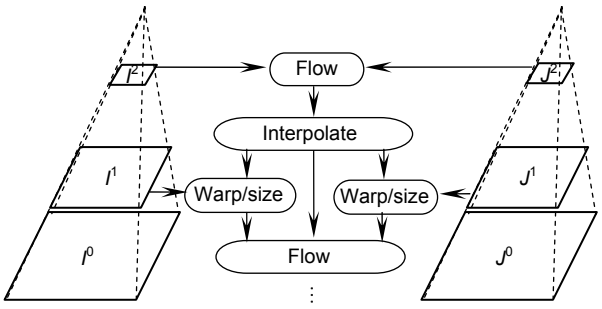


图 4 人脸金字塔光流示意图

Fig. 4 Face pyramid optical flow

算法遵循图 1 所示的流程图。在初始化阶段，程序打开视频输入原始人脸图像，如图 5(a)所示，加载级联卷积神经网络，并通过改进后的级联卷积神经网络构建初始映射图，如图 5(b)所示。在每帧的算法框架内进行处理，并进行光流计算。文中光流只是针对一定区域而非整幅图像的计算，大大降低运算量，同时图像金字塔不需要重新生成，光流计算中利用了检测过程中所生成的金字塔。两帧之间利用光流信息差获得光流映射(如图 5(c)所示)，从而进行人脸的定位与校正(校正结果如图 5(d)所示)。

3.4 误差检测

对于预测得到特征点，通过正反向误差算法判断误差是否有效。视频中的人脸是连续变化的，它具有正反向的连续性，即当前帧无论是按照时间上的正向预测还是反向预测，所产生的轨迹是相似的。所以可以借此来判断关键点是否有效。给定的人脸视频序列 $S=(I_t,I_{t+1},\cdots,I_{t+k})$，正向获取 k 帧，得到 $T_{\mathrm{f}}^k=(x_t,x_{t+1},\cdots,x_{t+k})$，反向获取 k 帧，得到 $T_{\mathrm{b}}^k=(\hat{x}_t,\hat{x}_{t+1},\cdots,\hat{x}_{t+k})$，这里 I_t 为 t 时刻的视频帧，x_t 为 t 时刻目标点在人脸映射图的位置，T_{f}^k 是正向预测所有 $t+k$ 时间里关键点的位置集合，T_{b}^k 是反向预测所有 $t+k$ 时间关键点位置集合。算法误差的定义：

$$E_{\mathrm{FB}}(T_{\mathrm{f}}^k\,|\,S)=dis(T_{\mathrm{f}}^k,T_{\mathrm{b}}^k)=\left\|x_t-\hat{x}_t\right\| \text{ 。} \tag{14}$$

从时间 t 的初始位置 $x(t)$ 开始预测，直到时间 $t+p$ 时的位置 $x'(t+p)$，再从预测位置 $x'(t+p)$ 反向预测时间 t 的预测位置 $x'(t)$，则对于时间 $t+p$ 的预测结果为

$$x(t+p)=\begin{cases}x'(t+p), & \delta<\tau\\ x(t+p-1), & \delta\geq\tau\end{cases}, \tag{15}$$

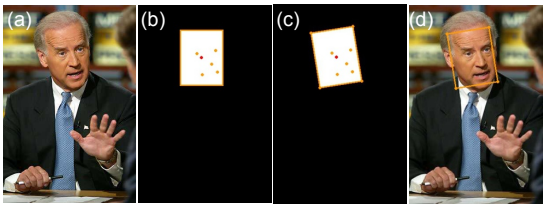


图 5 光流映射。(a) 原始人脸图像；(b) 关键点及候选框检测图；(c) 关键点映射；(d) 校正结果

Fig. 5 Optical flow mapping. (a) Original face image; (b) Facial landmark and bounding box detection; (c) Facial landmark mapping; (d) Calibration result

其中：$\delta=\left|x'(t)-x(t)\right|$，$\tau$ 为误差参数的阈值，当预测的结果大于所设的阈值则将上一帧的人脸映射赋予当前帧。

4 实验与分析

4.1 实验设置

性能测试实验采用 Intel(R) Core(TM) i7-8700 CPU、NVIDIA GeForce GTX 1080 Ti GPU 和 16 G 内存的机器配置，使用 python 语言在 Tensorflow1.10.0 框架中实现。

本文在 Wider Face[24]、CelebA[25]、LFPW[26]和 BiolD[27]数据集上训练网络，在 FDDB[28]和 Boston 人脸跟踪数据集[29]上进行评估和效率测试。Wider Face 数据集包含了 32203 张图片并标记了 393703 个边界框。CelebA 包括 202599 张人脸图像和相应的 5 个关键点，每张图片有 40 个二进制属性注释。LFPW 包含来自网络的 1432 张人脸图像，它分为 1132 张训练图像和 300 张测试图像。该数据集包含一定程度的遮挡人脸，在姿态、光照和表情方面有很大变化，用于在无约束条件下测试人脸关键点检测。BiolD 数据集包含在各种光照和复杂背景下的 1521 张面部图像，其中眼睛位置被手工标注。FDDB 数据集包含了在 2845 张图片中标注了 5171 张面部。

训练集和验证集在文本 trainImage 和 testImage 中定义。这些文本文件的每一行都以图像名开始，接着是人脸边界框的边界位置，然后是五个人脸关键点的位置。对数据进行数据标注，再送入网络进行训练。训练数据组成为 3:1:1:2=负样本：正样本：部分面部样本：特征面部样本。其中，负样本设置为特征数据的交并比小于 0.3 的样本；正样本设置为特征数据的交并比高于 0.65 的样本；部分面部样本设置为特征数据的交并比介于 0.4 和 0.65 之间；特征面部样本为面部标上 5 个特征点。负样本和正样本用于面部分类任务，负样本和部分面部样本用于边界框回归，特征面部样本用于面部特征点定位。

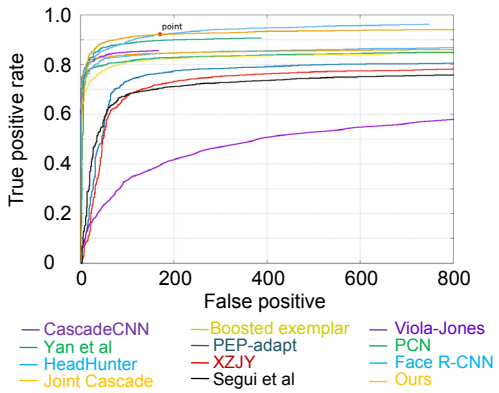


图 6 人脸检测方法比较

Fig. 6 Comparison of face detection methods

4.2 静态检测结果

为进行人脸检测算法之间的优势对比，利用 FDDB 上的 ROC 曲线对人脸检测方法进行性能评估。曲线如图 6 所示，其纵坐标为真阳性率，即将所有阈值之上的检测框的检测结果累加起来除以样本总数。横坐标没有采用假阳性率，而是直接采用假阳性，即为所有检测框中负样本数。所有这些人脸检测得到的假阳性数都随着真实阳性率的增加而迅速增加，除了 Viola-Jones，其它方法都是近几年提出的新方法。从图 6 中实验对比可以发现，当假阳性数量小于 170 时，如图中点 point 所示，本文方法性能优于其他方法。当假阳性数量大于 170 时，本文方法性能与 Face R-CNN[30]较接近。虽然以 Faster R-CNN 为基础模型的 Face R-CNN 获得了最好的性能效果，但由于模型较大，应用场景受到限制。本文方法在逼近 Face R-CNN 性能的同时，也适合快速处理视频流。

随机从 FDDB 数据中选取几张图片，使用本文中算法进行人脸检测，检测效果如图 7 所示。

4.3 动态检测结果

视频人脸检测为了权衡准确性和速度，算法每 10 帧使用级联渐进卷积神经网络检测一次。通过实验发现 10 左右为较好刷新率。图 8 为 Boston 头部跟踪数据集上进行人脸检测的结果。所选取的示例帧依次为视频 user_01_video_04 的第 39、44、49、54、58、63、74、90 帧和 user_01_video_03 的第 44、65、103、130 帧，其中 user_01_video_04 为旋转人脸，

图 7 FDDB 数据集上检测结果

Fig. 7 Results on the FDDB

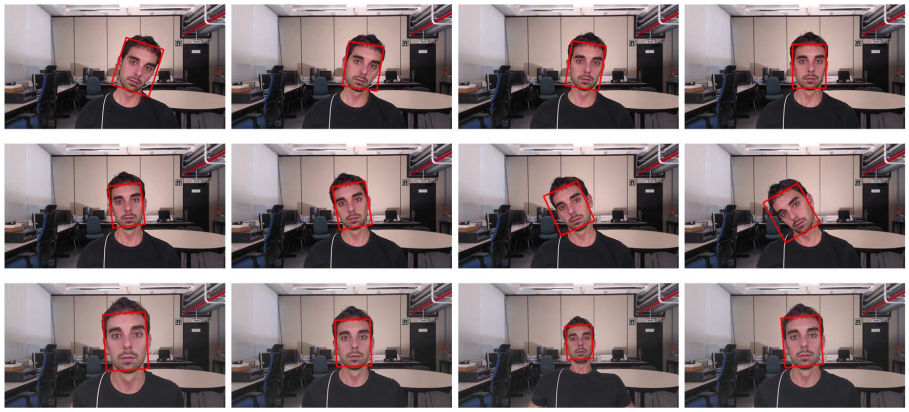


图 8 Boston 头部跟踪数据集检测结果

Fig. 8 Test results on Boston head tracking database

user_01_video_03 为尺度变化人脸。实验说明了对于视频流中单张人脸的人脸检测，本文方法能较好地解决平面内旋转问题和人脸尺度变化问题。

Boston 头部跟踪数据集中只记录单用户头部运动。在实际应用场景下，视频流人脸数目常常不定，我们拍摄了包含多张人脸的视频，并对该视频进行人脸检测。多用户人脸视频的检测结果如图 9 所示，实验结果表明本文方法也适用于检测多用户头部运动，有较好的适应性。

将本文的算法与其它旋转不变的人脸检测器在“标准.mp4”视频上进行速度比较，这些图像的最小人脸尺寸为 100×100。实验视频统一长度为 10 s，帧率为 30 f/s，画面大小为 640×480。表 1 给出常见人脸检测算法效率比较。从表中可以看出，本文算法比 Faster R-CNN(VGG16)、SSD500(VGG16)、R-FCN(ResNet-50) 等速度更快，相比于 Cascade CNN 在速度上也有一定提升。且模型尺寸只有 3.7 M，远小于 Faster R-CNN 与 SSD500 等通用网络架构。与其它模型尺寸对比，本文方法模型尺寸较小，适用于移动端设备。相比通过高度复杂网络学习旋转不变特征和对样本进行划分训练多个模型的方法大大减少了时间成本。

图 9 多用户人脸视频检测结果

Fig. 9 Test results on Multi-face video

5 结 论

视频流中人脸检测面临众多挑战，人脸的旋转角度变化和尺度变化影响了实际应用环境下人脸识别的精度。针对这个问题，本文提出了一种使用级联网络与金字塔光流相结合的算法，通过关键点与人脸候选框之间的映射，解决人脸平面内旋转。该方法通过对 FDDB 和 Boston 头部跟踪数据集的评估以及效率的测试，实验表明该算法能有效解决平面内旋转人脸检测

表 1 视频流中人脸检测算法效率及相应模型大小

Table 1 Speed and model size between different methods

Method	CPU/fps	GPU/fps	Model size/M
Faster R-CNN(VGGM)	1	21	350
Faster R-CNN(VGG16)	0.5	11	547
Cascade CNN	31	68	4.2
Cascade CNN+STN	15	31	4.7
Divide-and-Conquer	14	21	2.2
SSD500	1	21	95
R-FCN(ResNet-50)	0.8	16	123
Our	34	64	3.7

问题。摄像点在室内、室外公共区域的布置环境，使得采集的视频中人脸不可避免会发生旋转，该算法对视频监控领域有着重要应用前景。

参考文献

[1] Viola P, Jones M. Rapid object detection using a boosted cascade of simple features[C]//*Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2001.

[2] Li B, Yang A M, Yang J. Rotated face detection using AdaBoost[C]//*Proceedings of 2010 2nd International Conference on Information Engineering and Computer Science*, 2010: 1–4.

[3] Froba B, Ernst A. Face detection with the modified census transform[C]//*Proceedings of the Sixth IEEE International Conference on Automatic Face and Gesture Recognition*, 2004: 91–96.

[4] Jin H L, Liu Q S, Lu H Q, *et al*. Face detection using improved LBP under Bayesian framework[C]//*Proceedings of the Third International Conference on Image and Graphics*, 2004: 306–309.

[5] Farfade S S, Saberian M J, Li L J. Multi-view face detection using deep convolutional neural networks[C]//*Proceedings of the 5th ACM on International Conference on Multimedia Retrieval*, 2015: 643–650.

[6] Ranjan R, Patel V M, Chellappa R. A deep pyramid deformable part model for face detection[C]//*Proceedings of 2015 IEEE 7th International Conference on Biometrics Theory, Applications and Systems*, 2015.

[7] Yang S, Luo P, Loy C C, *et al*. From facial parts responses to face detection: a deep learning approach[C]//*Proceedings of 2015 IEEE International Conference on Computer Vision*, 2015.

[8] Bas A, Huber P, Smith W A P, *et al*. 3D morphable models as spatial transformer networks[C]//*Proceedings of 2017 IEEE International Conference on Computer Vision Workshops*, 2017.

[9] Li X X, Liang R H. A review for face recognition with occlusion: from subspace regression to deep learning[J]. *Chinese Journal of Computers*, 2018, **41**(1): 177–207.
李小薪, 梁荣华. 有遮挡人脸识别综述: 从子空间回归到深度学习[J]. 计算机学报, 2018, **41**(1): 177–207.

[10] Ren S Q, He K M, Girshick R, *et al*. Faster R-CNN: towards real-time object detection with region proposal networks[C]//*Proceedings of the 28th International Conference on Neural Information Processing Systems*, 2015: 91–99.

[11] Liu W, Anguelov D, Erhan D, *et al*. Single shot MultiBox detector[C]//*Proceedings of the 14th European Conference on Compute Vision* (*ECCV*), 2016: 21–37.

[12] Redmon J, Divvala S, Girshick R, *et al*. You only look once: unified, real-time object detection[C]//*Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*, 2016.

[13] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[C]//*Proceedings of the 3rd International Conference on Learning Representations*, 2015.

[14] Li H X, Lin Z, Shen X H, *et al*. A convolutional neural network cascade for face detection[C]//*Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition*, 2015: 5325–5334.

[15] Pan R, Wei H Q. Research on human face detection and recognition based on rotation invariance[J]. *Computer Engineering and Design*, 2009, **30**(8): 1941–1943, 1997.
潘榕, 魏慧琴. 基于旋转不变性的人脸定位识别研究[J]. 计算机工程与设计, 2009, **30**(8): 1941–1943, 1997.

[16] Wang W Q, Zhang X Y, Gao C Q, *et al*. Scale invariant face recognition from single sample[J]. *Journal of Image and Graphics*, 2012, **17**(3): 380–386.
王炜强, 张晓阳, 曹春芹, 等. 尺度不变单样本人脸识别方法[J]. 中国图象图形学报, 2012, **17**(3): 380–386.

[17] Bao X A, Hu L L, Zhang N, *et al*. Fast face detection algorithm based on cascade network[J]. *Journal of Zhejiang Sci-Tech University*, 2019, **41**(3): 347–353.
包晓安, 胡玲玲, 张娜, 等. 基于级联网络的快速人脸检测算法[J]. 浙江理工大学学报, 2019, **41**(3): 347–353.

[18] Liu W Q. Research on face detection algorithm based on cascaded convolutional neural networks[D]. Xiamen: Xiamen University, 2017.
刘伟强. 基于级联卷积神经网络的人脸检测算法的研究[D]. 厦门: 厦门大学, 2017.

[19] Sun K, Li Q M, Li D Q. Face detection algorithm based on cascaded convolutional neural network[J]. *Journal of Nanjing University of Science and Technology*, 2018, **42**(1): 40–47.
孙康, 李千目, 李德强. 基于级联卷积神经网络的人脸检测算法[J]. 南京理工大学学报, 2018, **42**(1): 40–47.

[20] Lin L Y. A visual object tracking method via CNN and optical flow with online learning[D]. Guangzhou: Guangdong University of Technology, 2018.
林露樾. 融合卷积神经网络以及光流法的目标跟踪方法[D]. 广州: 广东工业大学, 2018.

[21] Wang Z L, Huang M, Zhu Q B, *et al*. The optical flow detection method of moving target using deep convolution neural network[J]. *Opto-Electronic Engineering*, 2018, **45**(8): 38–47.
王正来, 黄敏, 朱启兵, 等. 基于深度卷积神经网络的运动目标光流检测方法[J]. 光电工程, 2018, **45**(8): 38–47.

[22] Zhang K P, Zhang Z P, Li Z F, *et al*. Joint face detection and alignment using multitask cascaded convolutional networks[J]. *IEEE Signal Processing Letters*, 2016, **23**(10): 1499–1503.

[23] Yang S, Luo P, Loy C C, *et al*. WIDER FACE: a face detection benchmark[C]//*Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*, 2016.

[24] Liu Z W, Luo P, Wang X G, *et al*. Deep learning face attributes in the wild[C]//*Proceedings of 2015 IEEE International Conference on Computer Vision*, 2015: 3730–3738.

[25] Sun Y, Wang X G, Tang X O. Deep convolutional network cascade for facial point detection[C]//*Proceedings of 2013 IEEE Conference on Computer Vision and Pattern Recognition*, 2013: 3476–3483.

[26] Köstinger M, Wohlhart P, Roth P M, *et al*. Annotated facial landmarks in the wild: a large-scale, real-world database for facial landmark localization[C]//*Proceedings of 2011 IEEE International Conference on Computer Vision Workshops*, 2011: 2144–2151.

[27] Jain V, Learned-Miller E G. FDDB: A benchmark for face detection in unconstrained settings[R]. *UMass Amherst Technical Report*, 2010.

[28] Cascia M L, Sclaroff S. Fast, reliable head tracking under varying illumination[C]// *Proceedings of 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1999: 604–610.

[29] Wang H, Li Z F, Ji X, *et al*. Face R-CNN[C]//*2017 IEEE Conference on Computer Vision and Pattern Recognition*, 2017.

Rotating invariant face detection via cascaded networks and pyramidal optical flows

Sun Rui[1,2], Kan Junsong[1,2*], Wu Liuwei[1,2], Wang Peng[3]

[1]School of Computer and Information, Hefei University of Technology, Hefei, Anhui 230009, China;
[2]Anhui Province Key Laboratory of Industry Safety and Emergency Technology, Hefei, Anhui 230009, China;
[3]Hefei Jinyi Science and Technology, Hefei, Anhui 230088, China

Effect picture

Overview: In recent years, with the rapid deployment of video surveillance in urban space, the public security department has collected video in a massive unconstrained open environment. There are complex problems such as scale change, partial occlusion, motion blur and illumination change in the face detection of video stream. In particular, face rotation affects the performance and efficiency of the entire face recognition system. In this paper, the in-plane rotation problem of face detection in video stream is combined with the pyramid optical flow, and a rotating invariant face detection algorithm based on cascaded network and pyramid optical flow is proposed. Firstly, the cascading progressive convolutional neural network is used to locate the face position and facial landmark of the previous frame in the video stream. Secondly, the optical flow mapping between the facial landmark and the bounding box is obtained, and the independent facial landmark network is used to detect the current frame. After that, the optical flow displacement of the key points between the two frames is calculated. Finally, the detected face of the video is corrected by the mapping relationship between the optical flow displacement of the key point and the face candidate frame, thereby completing the rotation-invariant face detection. The experiments were tested on the FDDB public datasets. The ROC curve on the FDDB evaluates the performance of the face detection method. When the number of false positives is less than 160, the performance of our method is better than other methods. When the number of false positives is more than 160, the face detection result is close to Face R-CNN, which proves that the method has higher accuracy. Moreover, the dynamic test on the Boston head tracking database proves that the face detection algorithm can effectively solve the problem of rotation and scale change of the target area in the plane. The speed of this algorithm with other rotationally invariant face detectors on standard .mp4 video is compared. The minimum face size of these images is 100×100. The experimental video has a uniform length of 10 s, a frame rate of 30 frames/s, and a picture size of 640×480. Experiments show that the algorithm detection speed has a great advantage, and the window jitter problem in the video is well solved. The average detection rate of the algorithm in this paper is higher than the general video frame rate, and the model size is small, which is suitable for mobile devices. Time costs are greatly reduced compared to the methods of learning rotational invariant features and segmenting samples by highly complex networks.

Citation: Sun R, Kan J S, Wu L W, *et al*. Rotating invariant face detection via cascaded networks and pyramidal optical flows[J]. *Opto-Electronic Engineering*, 2020, **47**(1): 190135

Supported by National Natural Science Foundation of China (61471154) and Fundamental Research Funds for Central Universities (JZ2018YYPY0287)

* E-mail: 2931338359@qq.com