

随着计算机技术的不断发展,方便有效的心率检测技术成为人们的迫切需求。作为一种非接触式的心率检测技术,rPPG^[1]成为了国内外众多学者的研究热点,rPPG 是指通过相机等传感器捕捉由心脏跳动造成的皮肤颜色周期性变化的无创光学技术。在每一次心跳中,血液的流动过程会引起所测组织区域内部血管中血液容积发生变化,这些变化反映在皮肤表面会呈现出像素的明暗变化,利用光电技术采集皮肤表面像素变化的程度从而获得 PPG 信号,PPG 信号经过滤波、去噪等分离出心率、呼吸等生命体征信号。另外,PPG 信号中包含丰富的生命体征信息,其中可以表示心肺功能的信息(心率、心率变异性等)对心血管疾病的诊断和治疗具有重要意义。

5.1 背景介绍

心率(Heart Rate,HR)/心率变异性/脉搏都是由心脏跳动引起的信号,三者在一一般情况下可以统称为心率。实际上,三者仍存在一些不同。其中,心率指普通人在安静状态下每分钟的心跳次数,成年人在安静状态下每分钟 60~100 次,会因为年龄、性别等个人生理状态差异而不同,是人最为基本的生命体征之一。心率变异性是指逐次心跳差异的变化情况,即相邻两次心跳之间的微小差异,包含了心血管系统的神经调节信息,可用作心肺神经功能的评估;心率变异性以静态仰卧状态下 5 分钟心电图的低频功率/高频功率比值作为基准,正常状态下的范围为 1.5~2.0。脉搏又指脉率,是指在人体的体表能够触及到的动脉搏动,一般以每分钟脉搏的次数来计量。心脏的收缩舒张直接影响着动脉的扩张和回缩,因动脉随心脏收缩和舒张而出现的周期性的起伏活动包含了大量的生理信息,可以用于心血管疾病的诊断与预测。

目前人们根据是否需要接触人体将检测心率的技术路线分为两种:一种是接触式的检测技术,另一种是非接触式的检测技术。

(1) 接触式检测技术是指传感器需要接触人们的皮肤或者身体其他相关部位才能获取相应生命体征信号的方法。例如,心电图(Electro Cardio Gram, ECG)^[2]通过在胸部贴附电极进行心脏监测,一些基于心冲击图(Ballisto Cardio Graphy, BCG)^[3]或者 rPPG 的智能手表或者手环需贴身佩戴才能获取受试者的脉动信号。受到应用场景与装置安装的限制,接触式检测方法存在不可避免的缺陷,例如,脉搏血氧仪等监测设备因其不方便携带,在实

际场景中无法做到实时有效检测,并且有些设备(如心电监测仪)需要专业的医护人员进行操作、维护。目前市场上存在一些具有心率、呼吸等监测功能的智能手表、智能手环等可穿戴式产品,因其携带方便备受人们推崇,但对于一些皮肤敏感者、皮肤烧伤患者或者重症术后患者不能适用。

(2) 针对上述接触式检测方法或者设备存在的问题,非接触式的检测技术或相关设备表现出独特的优势。非接触式的心率检测方法是指在不接触人们体表的条件下获取生命体征信号,进而评估身体状态。

本实验基于 rPPG 技术提取目标心率,主要是利用深度学习获取受试者的心率值。将 rPPG 技术与深度学习技术相结合,能够有效减少人为主观因素的干扰,提取到鲁棒的特征表示,并且在不断的训练过程中可以有效减少不均匀光照、面部表情等外部噪声的影响。利用深度学习检测心率的流程包括视频/图像数据的预处理或人脸检测,将检测到的人脸数据送到神经网络进行训练,对于训练得到的原始 PPG 信号进行降噪、平滑,最后分离出目标心率。基于深度学习利用 rPPG 进行心率检测的流程如图 5.1 所示。

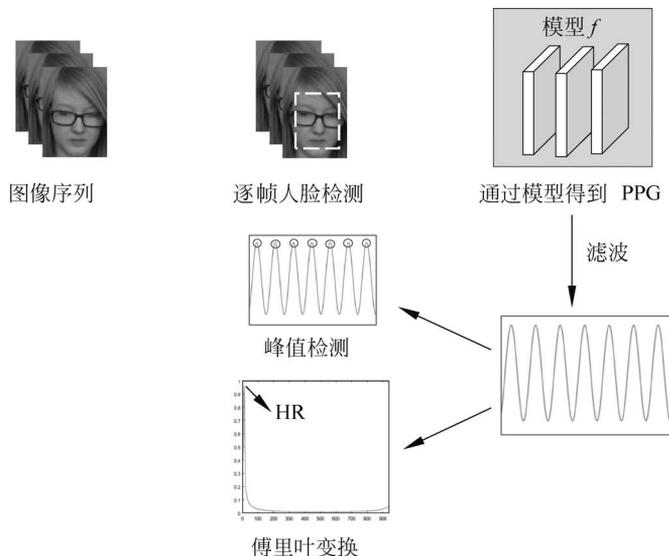


图 5.1 基于深度学习利用 rPPG 进行心率检测的流程

5.2 算法原理

皮肤反射光的细微变化可用于提取人体的生理信号。基于 rPPG 技术进行心率检测,其重点在于对面部光照反射信息的提取与处理,面部反射信息包含着丰富的特征信息,如纹理、边缘等低级信息,也有类似语义信息等高级信息,信息提取方法对心率检测结果是否准确起着重要的作用。本实验所实现的基于深度学习的心率检测算法模型为 DeepPhys^[4],可以有效提取目标的心率。

5.2.1 基于皮肤反射模型的帧差表示

现有的非接触式心率检测技术通常使用相机采集人脸视频,为了建立起光照、相机拍摄

以及人脸皮肤生理状态之间的数学模型,早期研究使用了 Lambert-Beer law (LBL)^[5] 或 Shafer^[6] 的二向反射模型。为了更好地对面部表观颜色变化与帧差表示之间的关系建模,本实验使用由 Chen 所提出的皮肤反射模型,该模型具体如下。

首先,假设光源具有恒定不变的光谱,但强度不同,用以模拟实际场景下的自然光照,然后定义图像中第 k 个皮肤像素的表达式:

$$\mathbf{C}_k(t) = I(t)(\mathbf{v}_s(t) + \mathbf{v}_d(t)) + \mathbf{v}_n(t) \quad (5-1)$$

其中, $\mathbf{C}_k(t)$ 表示单一 RGB 像素值的向量; $I(t)$ 表示光照强度,光照强度随光源以及光源、皮肤组织和相机的距离而变化; $\mathbf{v}_s(t)$ 表示来自皮肤的镜面反射; $\mathbf{v}_d(t)$ 表示来自皮肤组织吸收和散射所形成的漫反射; $\mathbf{v}_n(t)$ 表示相机传感器量化的噪声。光照强度 $I(t)$ 、镜面反射 $\mathbf{v}_s(t)$ 、漫反射 $\mathbf{v}_d(t)$ 都可通过线性变换分解为静态定量和随时间而变的动态变量之和的形式。

漫反射 $\mathbf{v}_d(t)$ 表示为

$$\mathbf{v}_d(t) = \mathbf{u}_d d_0 + \mathbf{u}_p p(t) \quad (5-2)$$

其中, \mathbf{u}_d 表示皮肤的单位颜色向量; d_0 表示静态反射强度; \mathbf{u}_p 表示由血红蛋白和黑色素吸收所引起的相对脉动强度; $p(t)$ 表示血液脉冲信号。

镜面反射 $\mathbf{v}_s(t)$ 可表示为

$$\mathbf{v}_s(t) = \mathbf{u}_s (s_0 + \Phi(m(t), p(t))) \quad (5-3)$$

其中, \mathbf{u}_s 表示光源光谱的单位颜色向量; s_0 与 $\Phi(m(t), p(t))$ 分别表示镜面反射中的固定部分以及发生变化的部分; $m(t)$ 表示所有的非生理变化,如光源闪烁、头部旋转、表情变化等。

光照强度 $I(t)$ 表示为

$$I(t) = I_0 + (1 + \varphi(m(t), p(t))) \quad (5-4)$$

其中, I_0 表示亮度强度的固定部分; $I_0 \varphi(m(t), p(t))$ 表示由相机观察到的强度变化,由于姿态变化和生理状态之间形成的复杂关系, $\varphi(m(t), p(t))$ 通常是复杂的非线性函数。

为了简化上述表达式,由皮肤镜面反射和漫反射所构成的静态分量可表示为

$$\mathbf{u}_c c_0 = \mathbf{u}_s s_0 + \mathbf{u}_d d_0 \quad (5-5)$$

其中, \mathbf{u}_c 表示皮肤反射的单位颜色向量; c_0 表示反射强度。将式(5-2)~式(5-5)代入式(5-1)表示的皮肤反射模型,得到更加清晰的皮肤反射模型表达式:

$$\mathbf{C}_k(t) = I_0 (1 + \phi(m(t), p(t))) \times (\mathbf{u}_c c_0 + \mathbf{u}_s \varphi(m(t), p(t)) + \mathbf{u}_p p(t)) + \mathbf{u}_n(t) \quad (5-6)$$

由于时变分量要比固定分量在数量级上小得多,因此可以忽略变化项, $\mathbf{C}_k(t)$ 可近似为

$$\begin{aligned} \mathbf{C}_k(t) \approx & \mathbf{u}_c I_0 c_0 + \mathbf{u}_c I_0 c_0 \phi(m(t), p(t)) \\ & + \mathbf{u}_s I_0 \varphi(m(t), p(t)) + \mathbf{u}_p I_0 p(t) + \mathbf{u}_n(t) \end{aligned} \quad (5-7)$$

对于任何基于视频的生理信号测量,式(5-7)的目的都是从 $\mathbf{C}_k(t)$ 中获取 $p(t)$,然而很多方法都忽略了生理信号与其他因素的联系 $\varphi(m(t), p(t))$ 与 $\Phi(m(t), p(t))$,并假设 $\mathbf{C}_k(t)$ 与 $p(t)$ 是单一的线性关系,选取的 $m(t)$ 即皮肤 ROI 区域小且稳定,不受外界光照影响。实际上这种情况并不适用,因此检测信号的精度并不高,需要建立一种新的模型来表示 $\mathbf{C}_k(t)$ 和 $p(t)$ 之间的复杂关系。

Chen 等对皮肤反射模型开展了进一步的研究,提出将一种新的标准化帧差作为输入,

这种方法避免了光流法对于亮度恒定这一约束条件的缺点,更加符合实际场景中生理信号 $p(t)$ 动态变化这一事实。根据以往的研究结果,利用空间上的像素平均可以有效减少由相机带来的噪声 $\mathbf{u}_n(t)$ 。实验根据前期的相关研究工作^[7],使用双三次插值将每帧图像下采样到 $L \times L$ 像素,并且选取 $L=36$ 。因此每个皮肤像素可以表示为

$$C_l(t) \approx u_c I_0 c_0 + u_c I_0 c_0 \phi(m(t), p(t)) + u_s I_0 \phi(m(t), p(t)) + u_p I_0 p(t) \quad (5-8)$$

其中, $C_l(t)$ 表示经过下采样的皮肤像素, $l=1, 2, \dots, L^2$ 表示每帧图像的像素标号。

为了减少受试者本身肤色以及光源的影响,应尽可能较少固定分量的影响,并且考虑到处理的图像帧来自短时间窗口,因此可以通过对时间维度 t 进行求导并化简得到:

$$C'_l(t) \approx u_c I_0 c_0 \left(\frac{\partial \phi}{\partial m} m'(t) + \frac{\partial \phi}{\partial p} p'(t) \right) + u_s I_0 \left(\frac{\partial \phi}{\partial m} m'(t) + \frac{\partial \phi}{\partial p} p'(t) \right) + u_p I_0 p'(t) \quad (5-9)$$

对于实际情况,由于光源到图像空间各个像素点的距离不同,导致亮度强度 I_0 在空间上是不均匀分布的,因此可利用亮度在空间上的时间平均化减少亮度强度的影响,经过空间的像素平均得到表达式为

$$\frac{C'_l(t)}{C_l(t)} \approx \mathbf{1} \left(\frac{\partial \phi}{\partial m} m'(t) + \frac{\partial \phi}{\partial p} p'(t) \right) + \text{diag}^{-1}(u_c) u_p \frac{1}{c_0} p'(t) + \text{diag}^{-1}(u_c) u_s \frac{1}{c_0} \left(\frac{\partial \phi}{\partial m} m'(t) + \frac{\partial \phi}{\partial p} p'(t) \right) \quad (5-10)$$

其中, $\mathbf{1}=[111]^T$ 。在式(5-10)中, $\overline{C_l(t)}$ 需要在短时间窗口上进行计算以防止由非生理变化 $m(t)$ 等引起的问题,在 Δt 足够小的条件下皮肤像素值仅与生理变化 $p(t)$ 相关。式(5-10)可进一步表示为

$$D_l(t) = \frac{C'_l(t)}{\overline{C_l(t)}} \sim \frac{C_l(t + \Delta t) - C_l(t)}{\overline{C_l(t + \Delta t) + C_l(t)}} \quad (5-11)$$

其中, $D_l(t)$ 为最终的归一化帧差表示,作为模型的输入,其对应的 PPG 信号真值表示为

$$p'(t) = p(t + \Delta t) - p(t)$$

上述即为完整的基于皮肤反射模型的帧差表示建立过程。

5.2.2 基于注意力机制的外观表示

注意力机制是人类观察事物的一种机制。当人类观察物体时,会自动关注目标的某些信息,在后续的观察中会增加相应的关注度,在有限的条件下获取更多的关键信息。在深度学习中,进行图像分类或检测等任务时也会有需要重点关注的区域,这与人脑的注意力机制很相似,因此在计算机视觉领域,注意力机制也可以发挥良好的作用。

在神经网络中,注意力机制可以理解为放大某些特定区域的信息,即使用掩码对原特征层赋予不同的权重信息,强调局部区域,并通过反复的训练回传使得局部区域的信息加强,形成整个网络的注意力。注意力机制通常是端到端的,并与整个网络共同训练,能够利用梯度下降等方法进行优化。假设神经网络的某一层输出为

$$\mathbf{o} = \{\mathbf{o}_1, \mathbf{o}_2, \dots, \mathbf{o}_n\} \quad (5-12)$$

其中, \boldsymbol{o}_i 表示第 i 个维度的输出向量, 经过注意力机制模型赋予不同权重以后表示为

$$\boldsymbol{A}\boldsymbol{o}_i = f_{A\boldsymbol{o}}(\boldsymbol{w}_i, \boldsymbol{o}_i) \quad (5-13)$$

其中, $f_{A\boldsymbol{o}}$ 表示经过注意力机制加权以后的输出模型, \boldsymbol{w}_i 为第 i 个向量的权重, $\boldsymbol{A}\boldsymbol{o}_i$ 为 \boldsymbol{o}_i 经过注意力模型后的输出。

注意力分为强注意力(hard attention)和软注意力(soft attention)两种。如果选取的掩码是固定的模板, 即为强注意力, 这种注意力可以使网络关注到预先设定的区域, 但不够灵活, 且无法与神经网络一同训练。另一种注意力可以随网络一同训练, 称为软注意力, 它在不同层之间增加了可以微分的权重矩阵, 更加灵活高效。目前注意力机制多数都为软注意力。根据注意力机制使用的位置不同又可分为空间注意力与通道注意力机制。本实验使用的为空间注意力机制, 空间注意力机制结构如图 5.2 所示。

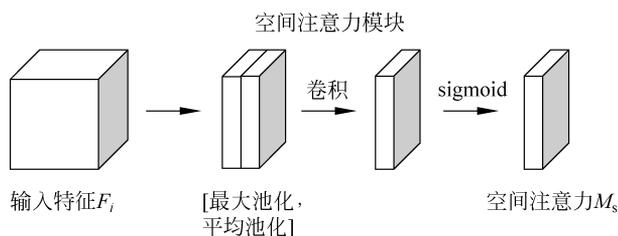


图 5.2 空间注意力机制结构

如图 5.2 所示, 将上一网络层的输出特征图 F_i 输入空间注意力模块, 经过池化、卷积等操作, 使用 sigmoid 对权重进行激活, 可得到空间注意力掩码 M_s 。将掩码与原输入特征按照空间上的对应关系进行相乘获取加权后的输出, 使用了注意力机制的神经网络如图 5.3 所示。

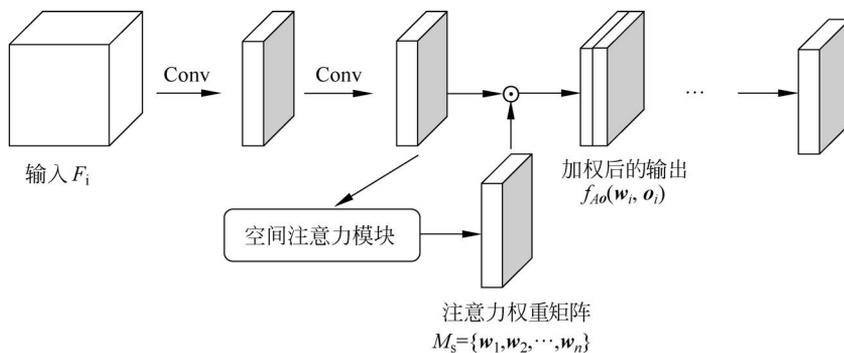


图 5.3 注意力机制的神经网络

指导信息 M_s 的计算表示为

$$M_s = \frac{H_j W_j \sigma(f_j(x_a^j))}{2 \|\sigma(f_j(x_a^j))\|} \quad (5-14)$$

即输出特征使用激活函数 sigmoid 对 $f_j(x_a^j)$ 进行激活后, 利用 L_1 正则化生成掩码。

5.2.3 网络模型

图像的空间信息是人们最常关注, 也是最容易获取的信息, 在图像的检测、分类等各种任务中占据着重要的地位。在 DeepPhys 中主要使用的是以 VGG-16 为主体的基本网络结

构(帧差表示分支与外观指导分支双分支结构),具体的网络模型结构如图 5.4 所示。所获得的 PPG 信号经过带通滤波、信号平滑去噪、傅里叶变换等操作得到目标心率信号。

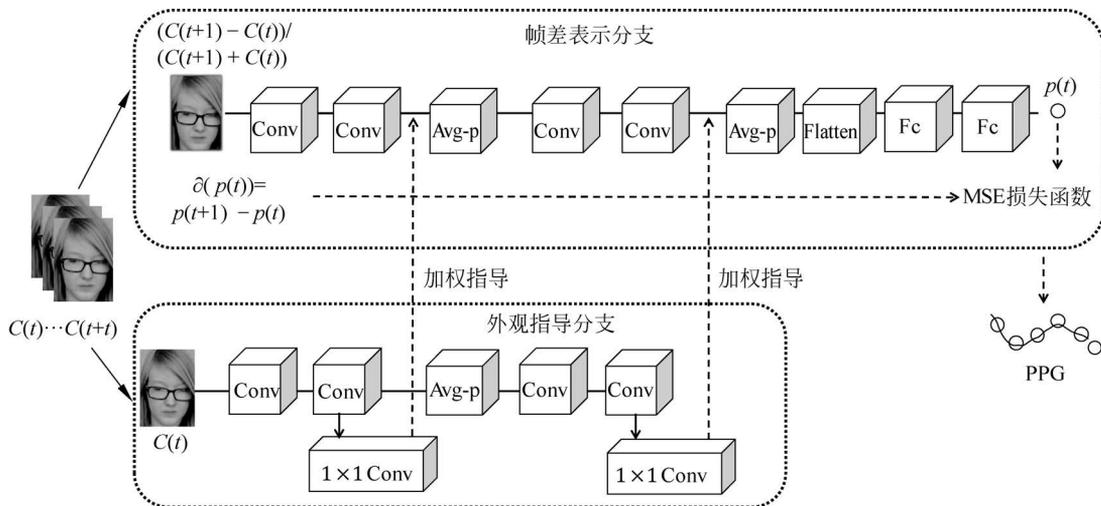


图 5.4 网络模型结构

5.3 实验操作

5.3.1 代码介绍

本实验所需要的环境配置如表 5.1 所示。

表 5.1 实验环境配置

操作系统	Ubuntu 18.04 LTS
开发语言	Python 3.8.8
深度学习框架	PyTorch 1.4.0
相关库	Numpy 1.15.4
	h5py 3.1.0
	Opencv 3.3.0

实验代码下载地址可扫描书中提供的二维码获得。工程文件中包含的文件较多,对于本实验而言,仅需关注示例模型 DeepPhys 相关代码文件,因此在下列代码文件目录结构中,仅展示了相关代码文件,其余文件不予展示。



```

----- 工程根目录
|--- log.py ----- 运行日志文件
|--- loss.py ----- 损失函数选择文件
|--- main.py ----- 主函数
|--- meta_params.json ----- 参数选择文件
|--- models.py ----- 模型选择文件
|--- optim.py ----- 优化器选择文件
|--- parallel.py ----- 数据并行处理文件
|--- dataset
|--- DeepPhysDataset.py ----- DeepPhys 数据加载文件
    
```

```

├── dataset_loader.py ----- 数据加载文件
├── nets
│   ├── blocks
│   │   ├── attentionBlocks.py ----- 注意力机制代码文件
│   │   └── motionBlock.py ----- 帧差表示代码文件
│   ├── funcs
│   │   └── complexFunctions.py ----- 数学计算公式文件
│   ├── layers
│   │   └── complexLayers.py ----- 部分网络层文件
│   └── models
│       └── DeepPhys.py ----- DeepPhys 模型
├── unused
│   └── Facedetect.py ----- 作者构建的人脸检测文件
├── utils
│   ├── dataset_preprocess.py ----- 数据处理调用文件
│   ├── image_preprocess.py ----- 视频/图像数据库处理文件
│   ├── text_preprocess.py ----- 真值数据处理文件
│   └── funcs.py ----- 滤波器、波形图绘制等文件
└── README.md ----- 说明文件

```

5.3.2 数据集介绍

本实验使用的数据集为 Bobbia 等^[8]制作的 UBFC-RPPG 数据集,UBFC-RPPG 中共 49 个 ID,每个 ID 包含一段长约 1 分钟的视频,该视频数据通过低成本摄像头(Logitech C920 HD Pro)进行录制,帧率为 30fps,分辨率为 $640 \times 480\text{px}$,保存为未经过压缩的 AVI 格式。受试者同时与指夹脉搏血氧仪传感器(Contec Medical CMS50E)相连采集对应的实际 PPG 信号。该数据集目前网上公开的数据 ID 共 43 个,具体的数据集下载地址可扫描书中提供的二维码获得。



将数据集下载并解压后得到相应的视频数据和真值文件,解压后的文件夹具体结构如下:

```

├── UBFC ----- 数据集名字
│   ├── subject1 ----- 目标 1 文件夹
│   │   ├── vid.avi ----- 目标 1 视频文件
│   │   └── ground_truth.txt ----- 目标 1 真值文件
│   ├── subject3 ----- 目标 3 文件夹
│   │   ├── vid.avi ----- 目标 3 视频文件
│   │   └── ground_truth.txt ----- 目标 3 真值文件
│   ...
│   └── subject49 ----- 目标 49 文件夹
│       ├── vid.avi ----- 目标 49 视频文件
│       └── ground_truth.txt ----- 目标 49 真值文件

```

其中,vid. avi 表示对应目标 ID 下的视频文件,时长在 1 分钟左右,ground_truth. txt 表示对应的 PPG 信号真值文件,根据以往的研究工作,使用其中的第二列作为对应视频帧的 PPG 信号真值。数据集中某 4 个 ID 的示例图像如图 5.5 所示。



图 5.5 UBFC 数据集中某 4 个 ID 的示例图像

5.3.3 实验操作与结果

1. 训练过程

打开文件夹,在 `meta_params.json` 选择相应的参数,具体包括模型、数据集、优化器种类、学习率、训练轮次等。待所需要相关参数设置完毕,可运行 `main.py` 文件,选择训练过程,获取训练结果,可采用以下两种方式运行 `main.py` 文件:

(1) 使用 Pycharm 集成环境打开相关代码文件,在 `main.py` 文件中单击“运行”按钮运行代码。

(2) 打开终端,cd 到当前项目目录,在终端中运行如下命令操作:

```
python ./main.py
```

训练时,将在终端显示网络损失以及训练轮次等相关参数的变化情况。训练结束后,在 `meta_params.json` 中设置的模型保存路径下将会发现以 `.pth` 为扩展名的训练模型文件,该模型保存了网络训练中产生的权重信息。

2. 测试过程

在 `main.py` 文件中,将训练过程修改为测试过程,并导入测试模型,选定参数,运行 `main.py` 获取测试结果,具体运行方式在训练过程中已经说明,此处不再阐述。测试结果包括两方面:一是各种生理指标在终端上的直接显示,例如,测试目标当前的心率值,预测 PPG 信号与实际 PPG 信号的误差、相关系数等;另一种是波形图的绘制,如图 5.6 所示,该图展示了测试 ID5 前 300 帧图像的 PPG 波形预测结果。利用傅里叶变换将预测的 PPG 信号转化为频域信号,经过带通滤波、去噪、平滑以后计算能量谱,能量谱中最大的幅度值对应的横坐标即为当前一段时间内的心率,经过时间上的转换,就可以得到目标一分钟的平均心

率值,例如该目标经过上述流程处理后可得到当前目标心率为 102 次/分钟。

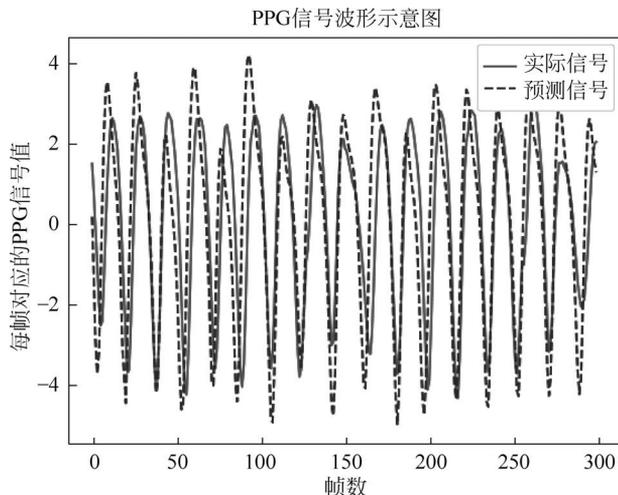


图 5.6 目标 5 前 300 帧图像预测波形图

5.4 总结与展望

随着人工智能技术的发展与进步,智能产品和智能技术已经大量应用到人们日常生活的各个领域,例如,人脸识别、目标检测、自动驾驶、生物医学等。如今,利用人工智能技术辅助日常生活已经成为了一种常态,近年来,随着智慧医疗的概念进入大众的视线中,利用 AI 技术辅助医疗诊治、检测生命指标成为了人们的研究热点。本次实验基于皮肤反射模型的帧差表示实现了目标心率值的提取,此外,还通过引入注意力机制,基于人脸外表空间信息生成加权指导掩码,提升了心率检测的准确性。

心率检测作为其中最基本也最为重要的研究问题之一,一直备受研究学者青睐。本次使用的远程光电容积脉搏波描记法利用视频中人体面部区域的微小颜色变化来提取心率,并对心率提取任务中如何减少光照、面部姿态等外界噪声,提取更具代表性的特征进行了重点讨论,与传统的心率检测方法相比,该技术具有更高的精度和实用性。

参考文献

- [1] Verkruyse W, Svaasand L O, Nelson J S. Remote plethysmographic imaging using ambient light[J]. *Optics Express*, 2008, 16(26): 21434-21445.
- [2] Li C, Zheng , Tai . Detection of eeg characteristic points using wavelet transforms [J]. *IEEE Transactions on Biomedical Engineering*, 1995, 42(1): 21-28.
- [3] Inan O T, Migeotte P E, Park K S, et al. Ballistocardiography and seismocardiography: a review of recent advances[J]. *IEEE Journal of Biomedical and Health Informatics*, 2014, 19(4): 1414-1427.
- [4] Chen W, McDuff D. Deepphys: video-based physiological measurement using convolutional attention networks[C]//*Proceedings of European Conference on Computer Vision*, 2018.
- [5] Xu S, Sun L, Rohde G K. Robust efficient estimation of heart rate pulse from video[J]. *Biomedical Optics Express*, 2014, 5(4): 1124-1135.

- [6] Wang W, Den Brinker A C, Stuijk S, et al. Algorithmic principles of remote PPG [J]. IEEE Transactions on Biomedical Engineering, 2016, 64(7): 1479-1491.
- [7] Wang W, Stuijk S, De Haan G. Exploiting spatial redundancy of image sensor for motion robust rPPG [J]. IEEE Transactions on Biomedical Engineering, 2014, 62(2): 415-425.
- [8] Bobbia S, Macwan R, Benezeth Y, et al. Unsupervised skin tissue segmentation for remote photoplethysmography [J]. Pattern Recognition Letters, 2019, 124: 82-90.