Capítulo 7

Explorando a Fotopletismografia por Imagem: Uma Abordagem Prática para Aplicações Biomédicas

Vitor Kauã Oliveira de Souza (UENP), Alan Floriano (IFPR), Teodiano Freire Bastos-Filho (UFES)

Abstract

This chapter addresses the extraction of cardiac signals through video cameras using imaging photoplethysmography (iPPG). The main objective is to provide a comprehensive analysis of the topic, which has gained significant relevance in recent years. The text explores the theoretical foundations of iPPG signals, the algorithmic principles underlying its main methodologies, the proposed models, and the practical implementation of a model for vital sign extraction. Additionally, case studies and real-world applications of iPPG methods are presented, followed by final considerations on technical challenges and future applications.

Resumo

Este capítulo aborda a extração de sinais cardíacos por meio de câmeras de vídeo utilizando a fotopletismografia por imagem, conhecida em inglês como Imaging Photoplethysmography (iPPG). O principal objetivo é oferecer uma análise abrangente do tema, que ganhou relevância significativa nos últimos anos. O texto explora os fundamentos teóricos dos sinais de iPPG, os princípios algorítmicos que sustentam suas principais metodologias, os modelos propostos e a implementação prática de um modelo para extração de sinais vitais. Além disso, são apresentados estudos de caso e aplicações reais dos métodos de iPPG, seguidos de considerações finais sobre os desafios técnicos e as aplicações futuras.

7.1. Introdução ao PPG e ao iPPG

A fotopletismografia, conhecida em inglês como *Photoplethysmography* (PPG), mede as variações no volume sanguíneo em camadas vasculares da pele [Allen 2007]. Para isso, utiliza-se uma fonte de luz que incide sobre o tecido, sendo parcialmente espalhada e

 ©2025 SBC – Soc. Bras. de Computação

absorvida à medida que atravessa diferentes camadas. A luz atenuada, ao ser transmitida ou refletida de volta à superfície do tecido, é captada por um sensor óptico. Esse sensor converte a intensidade da luz detectada em um sinal elétrico, o qual pode ser analisado para a extração de informações fisiológicas relevantes [Kyriacou and Allen 2021].

Esse sinal apresenta dois tipos de variações: uma de alta frequência (componente AC), que reflete as mudanças no volume arterial a cada batimento cardíaco, e uma de variação lenta, quase estática (DC). A componente DC contém informações sobre respiração, fluxo venoso, atividades do sistema nervoso simpático e termorregulação. Essa tecnologia é essencial para monitoramento cardíaco e diversas aplicações médicas, pois permite avaliar de forma simples e não invasiva a circulação sanguínea [Allen 2007].

Os primeiros relatos sobre fotopletismografia datam da década de 1930, mas foi apenas no início da década de 1970 que a técnica passou a ser aplicada em ambientes clínicos, com o surgimento dos primeiros oxímetros de pulso [Kyriacou and Allen 2021]. Desde então, a fotopletismografia tem sido amplamente explorada em dispositivos médicos, como oxímetros e monitores de frequência cardíaca, além de ser incorporada em tecnologias vestíveis voltadas para o monitoramento da saúde.

A fotopletismografia pode operar em duas configurações principais: modo de transmissão e modo de reflexão, determinadas pelo posicionamento da fonte de luz e do fotodetector [Sun and Thakor 2016]. No modo de transmissão, o tecido é colocado entre a fonte e o detector. Dessa forma, esse método é restrito a locais com pouca espessura, como dedos e lóbulos das orelhas, mas é sensível a variações ambientais e pode interferir em atividades cotidianas [Sun and Thakor 2016]. No método de PPG por transmissão, a luz emitida por um LED atravessa o tecido e é detectada por um fotodiodo posicionado do lado oposto. A quantidade de luz transmitida varia conforme o volume sanguíneo nos vasos, que oscila conforme a pulsação cardíaca.

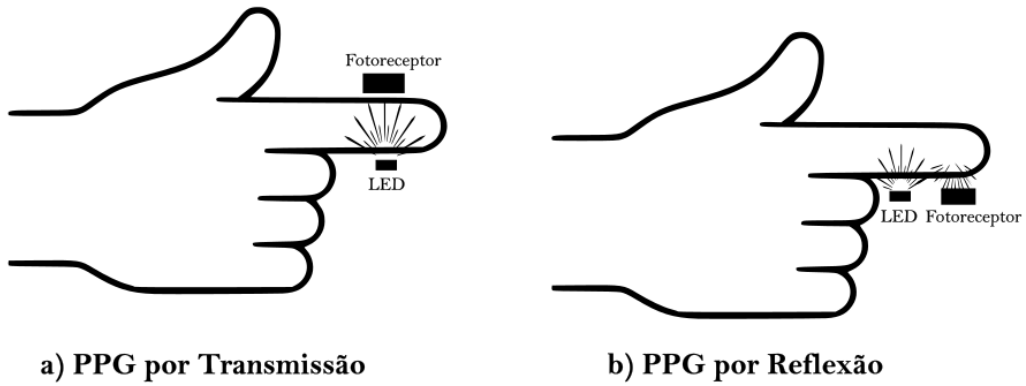

Figura 7.1. (a) PPG no modo transmissão, (b) PPG no modo reflexão.

Já no modo de reflexão, o fotodetector mede a luz que é refletida de volta, permitindo medições em praticamente qualquer área da pele. Esse método é mais utilizado em dispositivos vestíveis, como relógios inteligentes, que monitoram sinais vitais como a frequência cardíaca e a variabilidade do pulso [Sun and Thakor 2016]. Para evitar interferência direta da fonte de luz, é utilizado um escudo opaco entre os componentes

 ©2025 SBC – Soc. Bras. de Computação

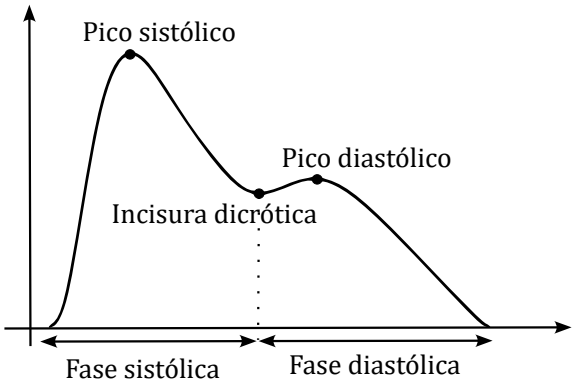

Figura 7.2. Forma de onda do sinal de PPG do dedo.

ópticos. Apesar da maior flexibilidade de aplicação, esse método exige uma boa fixação do sensor em superfícies cutâneas planas para garantir a precisão das medições [Sun and Thakor 2016].

A fotopletismografia possui ampla aplicação em dispositivos de monitoramento da saúde, sendo especialmente utilizada em oxímetros de pulso. Esses dispositivos ópticos empregam duas fontes de luz, tipicamente nas faixas do vermelho (660 nm) e do infravermelho (940 nm), para realizar a medição contínua da oxigenação arterial do sangue [Kyriacou and Allen 2021]. A partir dessa técnica, é possível estimar com precisão a saturação de oxigênio (SpO_2), a frequência cardíaca e outros parâmetros fisiológicos relevantes.

A Figura 7.2 ilustra uma forma de onda típica do sinal de PPG obtido na região do dedo. A partir dessa forma de onda, é possível extrair diversos parâmetros fisiológicos relevantes. Entre os principais, destacam-se a frequência cardíaca, obtida pela contagem dos picos sistólicos ao longo do tempo, e a variabilidade da frequência cardíaca, que reflete o equilíbrio do sistema nervoso autônomo [Allen 2007].

Além do dedo indicador, o sinal de PPG pode ser captado em diversas regiões do corpo, como o rosto, o lóbulo da orelha e o pulso, cada uma apresentando características distintas de perfusão sanguínea e resposta óptica. As variações na morfologia do sinal refletem diferenças anatômicas e vasculares específicas de cada local. A Figura 7.3 apresenta um gráfico com sinais de PPG registrados em diferentes regiões corporais, evidenciando essas variações fisiológicas.

Por se tratar de uma técnica não invasiva, a fotopletismografia oferece maior conforto aos usuários, além de ser uma tecnologia de baixo custo, o que amplia seu potencial de acessibilidade no monitoramento de sinais vitais. Nesse cenário, dispositivos vestíveis baseados em PPG, como relógios e pulseiras inteligentes, têm ganhado ampla popularidade, permitindo o acompanhamento contínuo da saúde em situações do dia a dia. Esses dispositivos são capazes de monitorar parâmetros fisiológicos em tempo real e detectar alterações significativas, como arritmias cardíacas, taquicardias e bradicardias, contribuindo para a identificação precoce de possíveis disfunções cardiovasculares.

No entanto, esse método ainda apresenta algumas limitações, como a sensibilidade ao movimento, que pode comprometer a precisão das medições, e a interferência da luz

 ©2025 SBC – Soc. Bras. de Computação

ambiente, especialmente relevante nos métodos baseados em reflexão [Kyriacou and Allen 2021].

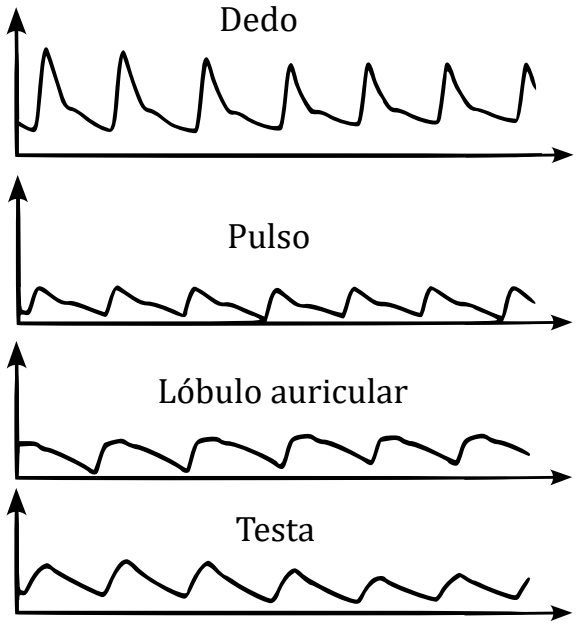

Figura 7.3. Gráfico representando os sinais de PPG em diferentes regiões do corpo, mostrando as variações na perfusão e características vasculares do corpo humano.

Nos últimos anos, a fotopletismografia por imagem, conhecida em inglês como *Imaging Photoplethysmography* (iPPG), emergiu como uma evolução dessa tecnologia [Wang et al. 2017a]. Diferente da técnica tradicional de PPG, que utiliza sensores de contato com a pele, a fotopletismografia por imagem usa câmeras de vídeo para capturar variações no volume sanguíneo [Xiao et al. 2024]. A técnica envolve a captura de sequências de imagens da pele e a análise das variações nas propriedades ópticas da pele, que são causadas pelas flutuações no volume de sangue nos vasos sanguíneos. O processamento das imagens permite a medição de parâmetros fisiológicos como a frequência cardíaca e a saturação de oxigênio, sem necessidade de contato direto com o paciente [Wang et al. 2024].

Entre os sensores ópticos utilizados, as câmeras digitais convencionais ganham destaque por sua ampla disponibilidade, baixo custo e boa sensibilidade às variações cromáticas provocadas pelo pulso sanguíneo. Essas características fazem com que sejam amplamente aplicadas em soluções práticas de iPPG, especialmente em contextos clínicos e dispositivos móveis. Por outro lado, câmeras infravermelhas (IR) vêm sendo exploradas como alternativa em ambientes com baixa iluminação ou onde há forte interferência da luz ambiente, apresentando maior robustez, embora exijam técnicas específicas de processamento.

Apesar das vantagens do monitoramento remoto por imagem, a técnica de iPPG ainda enfrenta desafios importantes. A qualidade das medições pode ser comprometida por variações na iluminação ambiente, exigindo condições controladas para garantir maior precisão. Além disso, o movimento do indivíduo ou da câmera pode introduzir ruídos nos sinais capturados, dificultando a extração confiável das informações fisiológicas

 ©2025 SBC – Soc. Bras. de Computação

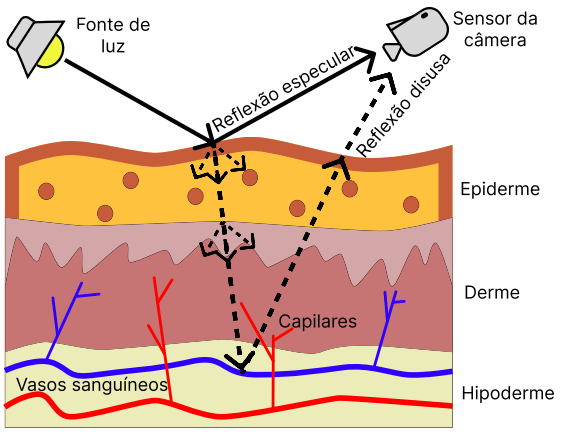

Figura 7.4. Modelo de Reflexão da Pele [Wang et al. 2017a].

[Sun and Thakor 2016].

Ainda assim, a fotopletismografia, especialmente em sua vertente baseada em imagem, representa um avanço significativo no monitoramento de saúde não invasivo. Com o contínuo progresso tecnológico, essa técnica oferece um caminho promissor para a coleta remota e em tempo real de sinais vitais, ampliando as possibilidades de cuidados médicos mais acessíveis, confortáveis e eficientes.

7.2. Fundamentação teórica do processamento de sinais de iPPG

A fundação teórica do processamento de sinais de iPPG se baseia no Modelo de Reflexão da Pele, ou do inglês, *Skin Reflection Model*. Esse modelo emprega representações matemáticas com matrizes e vetores para descrever os canais de cor, vetores unitários e outros elementos ópticos relevantes. Para compreender adequadamente a extração de sinais cardíacos por meio de métodos de iPPG, é essencial partir de uma formulação básica que considere as propriedades ópticas e fisiológicas envolvidas na reflexão da luz pela pele. Esse modelo fornece uma base sólida para a análise dos desafios encontrados e permite entender como diferentes abordagens de iPPG abordam e solucionam esses problemas [Wang et al. 2017a].

Para o modelo, ilustrado na Figura 7.4, considera-se uma fonte de luz que ilumina uma área de tecido humano que contém fluxo sanguíneo, enquanto uma câmera de vídeo registra essa região. Assume-se que a fonte de luz possui uma composição espectral constante, embora sua intensidade possa variar. A intensidade da luz captada pela câmera depende da distância entre a fonte de luz, o tecido da pele e o sensor da câmera. A pele observada pela câmera apresenta uma coloração que varia ao longo do tempo, devido a movimentos que causam mudanças na reflexão especular e na intensidade da luz, bem como às variações no fluxo sanguíneo, que alteram a cor da pele. Essas variações temporais são proporcionais à intensidade da luz refletida.

Assim, com base no modelo descrito, a reflexão de cada pixel correspondente à pele em uma sequência de imagens pode ser representada como uma função do tempo nos canais RGB, como:

 ©2025 SBC – Soc. Bras. de Computação

$$C_k(t) = I(t) \cdot (v_s(t) + v_d(t) + v_n(t)), \tag{1}$$

onde $C_k(t)$ representa os canais RGB do k-ésimo pixel da pele, e $I(t)$ indica o nível de intensidade da iluminação, o qual incorpora variações de intensidade provocadas tanto pela fonte de luz quanto pelas mudanças na distância entre a fonte, o tecido da pele e a câmera. Essa intensidade $I(t)$ é modulada por dois componentes no modelo: a reflexão especular $v_s(t)$ e a reflexão difusa $v_d(t)$. A dependência temporal desses termos se deve tanto aos movimentos corporais quanto às variações induzidas pelo fluxo sanguíneo. Por fim, o componente $v_n(t)$ representa o ruído introduzido pela quantização do sensor da câmera.

A reflexão especular age como um espelho no tecido da pele, refletindo a luz que ilumina a superfície da pele. Essa luz que é refletida não possui nenhuma informação do pulso cardíaco do indivíduo. Dessa forma, a composição espectral é equivalente à fonte de luz. Ela é dependente do tempo no sentido que o movimento do corpo influencia a estrutura geométrica entre a fonte de luz, a superfície da pele e a câmera. A reflexão especular $v_s(t)$ é denotada por:

$$v_s(t) = u_s \cdot (s_0 + s(t)) \tag{2}$$

onde u_s denota o vetor unitário correspondente à cor da luz no espectro. As variáveis s_0 e $s(t)$ representam as componentes da reflexão especular: s_0 é a parte estacionária, enquanto $s(t)$ é a parte variável, induzida pelo movimento.

Já a reflexão difusa está associada à absorção e dispersão da luz no tecido cutâneo. A presença de hemoglobina e melanina nesse tecido confere uma cromaticidade característica à componente difusa v_d. Essa componente varia ao longo do tempo devido às alterações no volume sanguíneo, sendo, portanto, uma função do tempo. A reflexão difusa $v_d(t)$ é representada por:

$$v_d(t) = u_d \cdot d_0 + u_p \cdot p(t), \tag{3}$$

onde u_d denota o vetor unitário correspondente à cor do tecido da pele, d_0 representa a intensidade da componente difusa estacionária, u_p indica a contribuição pulsátil relativa nos canais RGB e $p(t)$ representa o sinal de pulso ao longo do tempo. Substituindo esses valores na equação do modelo, obtém-se:

$$C_k = I(t) \cdot (u_s \cdot (s_0 + s(t)) + u_d \cdot d_0 + u_p \cdot p(t)) + v_n(t). \tag{4}$$

As componentes estacionárias das reflexões difusa e especular podem ser combinadas em um único termo, representando a reflexão estacionária da pele da seguinte forma:

$$u_c \cdot c_0 = u_s \cdot s_0 + u_d \cdot d_0, \tag{5}$$

 ©2025 SBC – Soc. Bras. de Computação

onde u_c denota o vetor unitário correspondente à cor da reflexão da pele e c_0 representa a intensidade dessa reflexão. Substituindo esses termos na fórmula do modelo, tem-se:

$$C_k(t) = I(t) \cdot (u_c \cdot c_0 + u_s \cdot s(t) + u_p.p(t)) + v_n(t). \quad (6)$$

A intensidade da luz capturada, $I(t)$, também pode ser expressa como a combinação de uma componente estacionária I_0 e uma componente variante no tempo, modelada como $I_0 \cdot i(t)$ — por exemplo, variações de intensidade induzidas por movimento, observadas pela câmera, sendo proporcionais ao nível de intensidade. Os sinais $i(t)$, $s(t)$ e $p(t)$ são considerados com média zero. Substituindo a modelagem na fórmula do modelo, tem-se:

$$C_k(t) = I_0 \cdot (1 + i(t)) \cdot (u_c \cdot c_0 + u_s \cdot s(t) + u_p.p(t)) + v_n(t). \quad (7)$$

Observa-se que a reflexão especular tende a ser o componente dominante na equação, frequentemente ofuscando as demais contribuições. Assim, assume-se a existência de mecanismos capazes de rejeitar as regiões em que a reflexão especular é predominante. Dessa forma, consideram-se apenas os pixels k nos quais u_d contribui de maneira significativa para a reflexão difusa, ou seja, sua influência não é desprezível.

De modo geral, o objetivo dos métodos iPPG é extrair o sinal $p(t)$ a partir de $C_k(t)$, filtrando as contribuições da reflexão especular para isolar a informação pulsátil da reflexão difusa.

7.3. Métodos para a extração do sinal de iPPG

Os métodos convencionais de extração de sinal de iPPG baseiam-se em modelos matemáticos cujo objetivo é eliminar artefatos de ruído e movimento presentes nas imagens capturadas pela câmera. Esses métodos geralmente consistem em calcular a média dos valores das componentes RGB dentro de regiões de interesse (ROIs) em cada quadro do vídeo, construindo, assim, sinais temporais RGB que são posteriormente processados para a extração do sinal de pulso.

A etapa de média dos pixels contribui para a redução do erro introduzido pela câmera. Com base na equação completa do modelo, assume-se que uma quantidade suficiente de pixels esteja focada em regiões de pele com propriedades ópticas comparáveis. Dessa forma, a média $C_k(t)$ sobre os pixels de pele pode ser aproximada como [Wang et al. 2017a]:

$$C_k(t) \approx I_0 \cdot (1 + i(t)) \cdot (u_c \cdot c_0 + u_s \cdot s(t) + u_p \cdot p(t)) \quad (8)$$

A representação fornece um sinal $C_k(t)$ em que a quantização de ruído $v_n(t)$ se torna insignificante, desde que haja um número suficientemente grande de pixels de pele na média. No entanto, observa-se que, quando essa média é calculada em uma área pequena, contendo poucos pixels de pele, o erro de quantização introduzido pela câmera permanece elevado e, portanto, não pode ser desprezado.

 ©2025 SBC – Soc. Bras. de Computação

Além disso, observa-se que os vetores de cor envolvidos na equação não dependem diretamente da posição dos pixels de pele na imagem. O sinal $C(t)$ resultante é essencialmente uma trajetória no espaço RGB ao longo do tempo t, a qual pode ser expandida e simplificada da seguinte forma:

$$\begin{aligned} C(t) =& u_c \cdot I_0 \cdot c_0 + u_s \cdot I_0 \cdot s(t) + u_p \cdot I_0 \cdot p(t) + \\ & u_c \cdot I_0 \cdot c_0 \cdot i(t) + u_s \cdot I_0 \cdot s(t) \cdot i(t) + \\ & u_p \cdot I_0 \cdot p(t) \cdot i(t) \\ & \approx u_c \cdot I_0 \cdot c_0 + u_c \cdot I_0 \cdot c_0 \cdot i(t) + u_s \cdot I_0 \cdot s(t) + \\ & u_p \cdot I_0 \cdot p(t) \end{aligned} \tag{9}$$

A aproximação se mantém por conta de que todos os termos da modulação AC são muito menores em ordem de magnitude do que o termo DC e portanto os termos de modulação do produto da equação podem ser ignorados. A aproximação apresentada no formato de equação mostra que a observação $C(t)$ é uma mistura linear de três fontes de sinais $i(t)$, $s(t)$ e $p(t)$. Isso implica que usando de uma projeção linear é possível separar essas três fontes de sinais. Dito isso, o objetivo de extrair o sinal de pulso dos sinais RGB observados pode ser traduzido como a definição de um sistema de projeção para decompor $C(t)$.

Dessa forma, diversas abordagens têm sido propostas para realizar a decomposição do sinal observado $C(t)$ e extrair o sinal de pulso. Entre as abordagens mais simples, destacam-se os métodos que utilizam diretamente os canais de cor da imagem, como o método **G**, que considera apenas o canal verde [Verkruysse et al. 2008], e o G-R [Hülsbusch 2008], que calcula a diferença entre os canais verde e vermelho. Esses métodos baseiam-se na observação de que o canal verde apresenta maior sensibilidade às variações do volume sanguíneo, sendo, portanto, um bom indicador do sinal de pulso.

Além desses, existem métodos mais sofisticados, como os baseados em *separação cega de fontes*, que empregam técnicas como a Análise de Componentes Independentes [Poh et al. 2010] e a Análise de Componentes Principais [Lewandowska and Nowak 2012] para isolar os sinais de interesse sem conhecimento prévio sobre as fontes. Outra categoria importante são os *métodos baseados em modelos* (*Model-Based Methods*), que utilizam descrições analíticas da interação entre a luz e o tecido biológico para estimar o sinal de pulso. Dentro dessa categoria, destacam-se os métodos CHROM [De Haan and Jeanne 2013] e POS [Wang et al. 2017a], amplamente utilizados por sua robustez e bom desempenho em condições desafiadoras de iluminação e movimento [Xiao et al. 2024].

Todas as abordagens e seus respectivos métodos convencionais têm o mesmo objetivo: extrair o sinal de pulso de $C(t)$ por meio de sua decomposição. No entanto, cada abordagem se distingue das demais pela metodologia específica empregada para alcançar esse objetivo. A seguir, serão abordados os métodos e suas metodologias matemáticas, incluindo também técnicas baseadas em aprendizado profundo utilizando redes neurais.

 ©2025 SBC – Soc. Bras. de Computação

7.3.1. Métodos convencionais de iPPG

Nesta parte do capítulo serão descritos os principais métodos convencionais de extração de sinais fisiológicos por meio da fotopletismografia por imagem. O objetivo é apresentar de forma clara os princípios algorítmicos fundamentais que nortearam o desenvolvimento dessas técnicas. Todos os métodos recebem como entrada uma sequência temporal de sinais RGB normalizados, denotada por $\mathbf{x}(t) = (x_r(t), x_g(t), x_b(t))^\top$, obtida a partir da média espacial da imagem e de filtros para remoção de tendências e ruídos. O objetivo final de cada abordagem é gerar uma sequência monovariada $y(t)$ que represente uma estimativa do sinal de volume de pulso.

7.3.1.1. Método ICA

A Análise de Componentes Independentes, conhecida em inglês como *Independent Component Analysis* (ICA), é uma técnica estatística que visa decompor uma mistura linear de sinais sob a suposição de independência estatística e não-gaussianidade [Poh et al. 2010]. No contexto da iPPG, considera-se que os sinais temporais RGB $\mathbf{x}(t)$ representam medições multivariadas resultantes da mistura de três fontes latentes $\mathbf{z}(t) = (z_1(t), z_2(t), z_3(t))^\top$. Esse processo de mistura instantânea pode ser descrito por:

$$\mathbf{x}(t) = \mathbf{A}\mathbf{z}(t), \tag{10}$$

onde $\mathbf{A} \in \mathbb{R}^{3\times 3}$ é uma matriz de mistura desconhecida. O objetivo da ICA é estimar uma matriz de separação $\mathbf{W}$ tal que:

$$\hat{\mathbf{z}}(t) = \mathbf{W}\mathbf{x}(t) \approx \mathbf{z}(t). \tag{11}$$

Esse problema é conhecido como separação cega de fontes, e diferentes abordagens podem ser utilizadas para estimar $\mathbf{W}$ com base na não-gaussianidade das componentes. No entanto, as soluções obtidas por ICA apresentam indeterminações típicas: as fontes podem ser recuperadas com escala, permutação e atrasos arbitrários. Apesar disso, a forma de onda original costuma ser preservada, o que é suficiente para análise no domínio tempo-frequência, como requerido na iPPG.

Uma limitação importante é que, após a separação, não é possível saber diretamente qual das três componentes contém a informação de pulso mais relevante. Para lidar com isso, adota-se uma abordagem simples: calcula-se a densidade espectral de potência (PSD) normalizada de cada componente separada, e escolhe-se aquela com o maior pico de frequência ou maior razão sinal-ruído (SNR) na faixa de 40 a 220 BPM.

 ©2025 SBC – Soc. Bras. de Computação

Entrada: $\mathbf{x}(t) = [x_r(t), x_g(t), x_b(t)]^\top$ com N amostras

Aplicar ICA: $\hat{\mathbf{z}}(t) \leftarrow \text{ICA}(\mathbf{x}(t))$

Para (cada componente $\hat{z}_i(t)$)

Calcular PSD e identificar pico de frequência

Fim-Para

Saída: $y(t) = \hat{z}_i(t)$ com maior pico na faixa de 40–220 BPM

7.3.1.2. Método PCA

A Análise de Componentes Principais, conhecida em inglês como *Principal Component Analysis* (PCA), é uma técnica amplamente utilizada em estatística multivariada, que busca maximizar a variância, minimizar as covariâncias e reduzir a dimensionalidade dos dados. Ao aplicar PCA ao vetor $\mathbf{x}(t)$, obtêm-se componentes ortogonais que explicam a variabilidade do sinal. A saída $y(t)$ é escolhida como a componente cuja energia espectral está concentrada na faixa cardíaca [Lewandowska and Nowak 2012].

No contexto da iPPG, assim como no método ICA, é preciso realizar a escolha da componente que melhor representa o sinal de pulso. Para resolver isso, calcula-se a densidade espectral de potência normalizada de cada componente principal e seleciona-se a componente com maior pico de frequência ou maior razão sinal-ruído dentro da faixa de 40 a 220 BPM.

Entrada: Sinal RGB $\mathbf{x}(t) = [x_r(t), x_g(t), x_b(t)]^\top$ com N amostras

Aplicar PCA: $\hat{\mathbf{z}}(t) \leftarrow \text{PCA}(\mathbf{x}(t))$

Para (cada componente $\hat{z}_i(t)$)

Calcular PSD($\hat{z}_i(t)$)

Fim-Para

Saída: $y(t) = \hat{z}_i(t)$ com maior pico na faixa de 40–220 BPM

7.3.1.3. Método G

Este método simples considera que o canal verde da imagem apresenta a maior quantidade de informação relacionada ao variação do volume do sangue, devido à absorção da luz verde pela hemoglobina [Verkruysse et al. 2008]. Assim, o sinal extraído é diretamente:

$$y(t) = x_g(t). \tag{12}$$

 ©2025 SBC – Soc. Bras. de Computação

Entrada: Sinal RGB temporal $\mathbf{x}(t) = [x_r(t), x_g(t), x_b(t)]^\top$ com N amostras

Saída: Sinal de pulso $y(t)$

Para ($t = 1$ até N)

$y(t) = x_g(t)$

Fim-Para

Retornar $y(t)$

7.3.1.4. Método G-R

O método G-R é uma técnica simples para extração do sinal de pulso a partir do vídeo, baseada na diferença entre os canais verde e vermelho do sinal RGB [Hülsbusch 2008]. A justificativa para essa abordagem está no fato de que o canal verde contém uma maior informação pulsátil, pois a hemoglobina absorve mais luz nesta faixa espectral, enquanto o canal vermelho apresenta menor conteúdo de pulsatilidade e pode ser usado para atenuar ruídos comuns ao sinal.

Assim, o sinal de pulso é obtido pela subtração do canal vermelho do canal verde:

$$y(t) = x_g(t) - x_r(t). \tag{13}$$

Devido a sua simplicidade, o método G-R pode apresentar limitações em ambientes com variações de iluminação intensa ou movimentos.

Algorithm 1 Método G-R

Entrada: Sinal RGB temporal $\mathbf{x}(t) = [x_r(t), x_g(t), x_b(t)]^\top$ com N amostras

Saída: Sinal de pulso $y(t)$

Para ($t = 1$ até N)

$y(t) = x_g(t) - x_r(t)$

Fim-Para

Retornar $y(t)$

Em aplicações práticas, recomenda-se complementar o método G-R com técnicas de pré-processamento, como suavização temporal e remoção de artefatos, para melhorar a qualidade do sinal extraído. Ainda assim, por sua simplicidade computacional, é uma alternativa eficiente para sistemas com recursos limitados.

7.3.1.5. Método CHROM

O método CHROM elimina o componente de reflexão especular da pele usando sinais de crominância derivados de combinações lineares dos canais RGB [De Haan and Jeanne 2013].

 ©2025 SBC – Soc. Bras. de Computação

De forma simplificada, a luz refletida da pele é composta por dois componentes, segundo o modelo dicromático de reflexão: um componente de reflexão difusa, cujas variações estão relacionadas ao ciclo cardíaco, e um componente de reflexão especular, que apresenta a cor da fonte luminosa, mas não contém sinal de pulso. A contribuição relativa desses dois componentes varia ao longo do tempo devido ao movimento da pessoa e à geometria entre câmera, pele e fonte de luz, causando dificuldades para os algoritmos de iPPG que não eliminam o componente especular aditivo. O método CHROM elimina o componente especular utilizando diferenças de cor, ou seja, sinais de crominância.

Dado o sinal RGB $\mathbf{x}(t)$, o método CHROM realiza inicialmente uma normalização por desvio padrão zero e projeta os valores normalizados em dois vetores de crominância ortogonais, definidos por:

$$X_{\text{CHROM}}(t) = 3x_r(t) - 2x_g(t), \tag{14}$$

$$Y_{\text{CHROM}}(t) = 1.5x_r(t) + x_g(t) - 1.5x_b(t). \tag{15}$$

O sinal iPPG final é então calculado por:

$$y(t) = X_{\text{CHROM}}(t) - \alpha\, Y_{\text{CHROM}}(t), \tag{16}$$

onde

$$\alpha = \frac{\sigma(X_{\text{CHROM}}(t))}{\sigma(Y_{\text{CHROM}}(t))}, \tag{17}$$

e $\sigma(\cdot)$ representa o desvio padrão.

Para lidar com variações temporais e manter a consistência do sinal ao longo do tempo, o sinal $y(t)$ é processado em janelas temporais com sobreposição (*overlap*). O resultado de cada janela é centralizado (remoção da média) e somado ao sinal final com sobreposição, técnica conhecida como *overlap-adding*. Para reduzir artefatos na junção entre janelas, pode se aplicar janelas suavizantes como a janela de Hamming no processo de soma com sobreposição. Isso permite atenuar descontinuidades entre janelas e preservar melhor o conteúdo espectral do sinal cardíaco.

 ©2025 SBC – Soc. Bras. de Computação

Entrada: Vetor RGB temporal $x(t) = [x_r(t), x_g(t), x_b(t)]^\top$ com N amostras

Saída: Sinal de pulso H

Inicializar $H = \text{zeros}(1, N)$, $l = 48$ (para câmera de 30 fps)

Para ($n = 1$ até N):

Se ($m = n - l + 1 > 0$):

$R = x_r(m:n),\ G = x_g(m:n),\ B = x_b(m:n)$

$R = \frac{R}{\mu(R)} - 1, \quad G = \frac{G}{\mu(G)} - 1, \quad B = \frac{B}{\mu(B)} - 1$ {Normalizar temporalmente}

$X = 3 \cdot R - 2 \cdot G$

$Y = 1.5 \cdot R + G - 1.5 \cdot B$

$\alpha = \frac{\sigma(X)}{\sigma(Y)}$

$h = X - \alpha \cdot Y$ {Sinal bruto}

$H(m:n) = H(m:n) + (h - \mu(h))$ {Soma com sobreposição}

Fim-Se

Fim-Para

Retornar: H

7.3.1.6. Método POS

O método *Plane-Orthogonal-to-Skin* (POS) tem como objetivo, assim como o método CHROM, eliminar os reflexos especulares na superfície da pele. Para isso, ele define um plano perpendicular ao tom de pele dominante dentro do espaço RGB normalizado temporalmente.

A Figura 7.5 busca ilustrar a distribuição da força pulsátil dentro do plano ortogonal à tonalidade da pele como uma função de um vetor de projeção **z**. São apresentados exemplos de vetores $\mathbf{z}_1$ e $\mathbf{z}_2$, que geram sinais pulsáteis de polaridades opostas (anti-fásicos), e um vetor $\mathbf{z}_3$ que resulta em um sinal com baixo conteúdo pulsátil, ou seja, dominado por ruído.

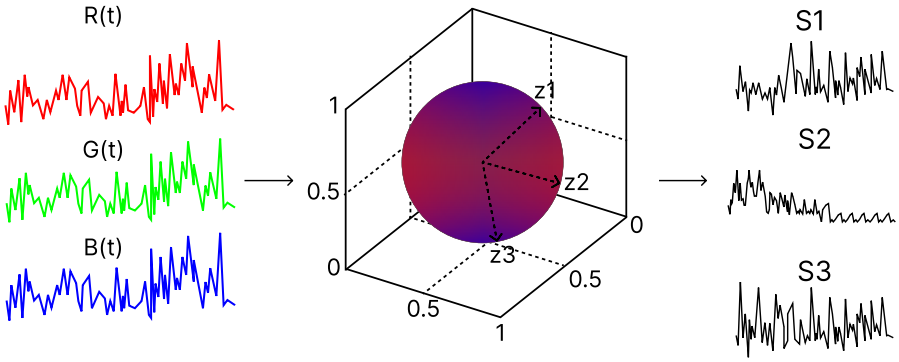

Figura 7.5. Plano ortogonal da pele [Wang et al. 2017a].

 ©2025 SBC – Soc. Bras. de Computação

Especificamente, dado o vetor RGB temporal $\mathbf{x}(t)$, o método POS é composto por três etapas principais. A primeira etapa consiste em uma normalização temporal dos sinais de cor, que remove variações de iluminação constante. Em seguida, os sinais são projetados sobre um plano ortogonal à tonalidade da pele, utilizando combinações lineares dos canais RGB:

$$X_{\text{POS}}(t) = x_g(t) - x_b(t), \tag{18}$$

$$Y_{\text{POS}}(t) = x_g(t) + x_b(t) - 2x_r(t). \tag{19}$$

Por fim, realiza-se um ajuste da direção de projeção dentro da região delimitada pelas componentes anteriores, resultando no sinal iPPG:

$$y(t) = X_{\text{POS}}(t) + \alpha, Y_{\text{POS}}(t), \tag{20}$$

onde o fator de ponderação α é dado por:

$$\alpha = \frac{\sigma(X_{\text{POS}}(t))}{\sigma(Y_{\text{POS}}(t))}, \tag{21}$$

e $\sigma(\cdot)$ representa o desvio padrão, como no método CHROM.

O método POS apresenta uma diferença fundamental em relação ao CHROM: enquanto este utiliza sinais projetados em antifase, o POS define diretamente dois eixos de projeção que produzem sinais em fase, o que tende a melhorar a consistência da pulsação extraída.

Além disso, para aumentar a relação sinal-ruído (SNR), o sinal é extraído em janelas temporais menores da sequência de vídeo. Cada segmento é processado separadamente e, ao final, os sinais parciais são recombinados utilizando a técnica de soma com sobreposição (*overlap-adding*), o que preserva a continuidade do sinal final e melhora a qualidade espectral.

 ©2025 SBC – Soc. Bras. de Computação

Entrada: Vetor RGB temporal $x(t) = [x_r(t), x_g(t), x_b(t)]^\top$ com N amostras

Saída: Sinal de pulso H

Inicializar $H = \text{zeros}(1, N)$, $l = 48$ (para câmera de 30 fps)

Para ($n = 1$ até N):

Se ($m = n - l + 1 > 0$):

$R = x_r(m:n)$, $G = x_g(m:n)$, $B = x_b(m:n)$

$R = \frac{R}{\mu(R)} - 1, \quad G = \frac{G}{\mu(G)} - 1, \quad B = \frac{B}{\mu(B)} - 1$ {Normalizar temporalmente}

$X = G - B$

$Y = G + B - 2 \cdot R$

$\alpha = \frac{\sigma(X)}{\sigma(Y)}$

$h = X + \alpha \cdot Y$ {Sinal bruto}

$H(m:n) = H(m:n) + (h - \mu(h))$ {Soma com sobreposição}

Fim-Se

Fim-Para

Retornar: H

7.3.2. Métodos de extração de sinal de iPPG com *Deep Learning*

O Aprendizado Profundo, ou do inglês, *Deep Learning*, é uma subárea da Inteligência Artificial que utiliza arquiteturas de redes neurais profundas, compostas por múltiplas camadas não lineares, para modelar padrões complexos em grandes volumes de dados [Xiao et al. 2024]. Diferentemente das abordagens tradicionais de aprendizado de máquina, que dependem de engenharia manual de atributos, os métodos de *Deep Learning* aprendem representações hierárquicas diretamente a partir de dados brutos, como imagens e vídeos [Xiao et al. 2024]. Essa capacidade os torna particularmente adequados para tarefas como a extração de sinais de iPPG.

No contexto do *Deep Learning*, destacam-se as abordagens supervisionadas para estimativa da frequência cardíaca, nas quais os modelos são treinados com dados rotulados. Isso exige um volume considerável de exemplos anotados para garantir um treinamento eficaz [Xiao et al. 2024].

Considerando a ampla variedade de técnicas atualmente disponíveis para a extração de sinais de iPPG por meio de métodos de *Deep Learning* [Xiao et al. 2024], verifica-se que esta é uma área de pesquisa em contínua expansão, com elevado potencial de inovação e aplicação. Em virtude da extensa quantidade de abordagens propostas na literatura, este capítulo se concentra especificamente nos principais métodos supervisionados, com ênfase naqueles que empregam redes neurais convolucionais, conforme descrito nos estudos de referência [Chen and McDuff 2018, Lin et al. 2019, Yu et al. 2019, Zhan et al. 2020, Lampier et al. 2022].

 ©2025 SBC – Soc. Bras. de Computação

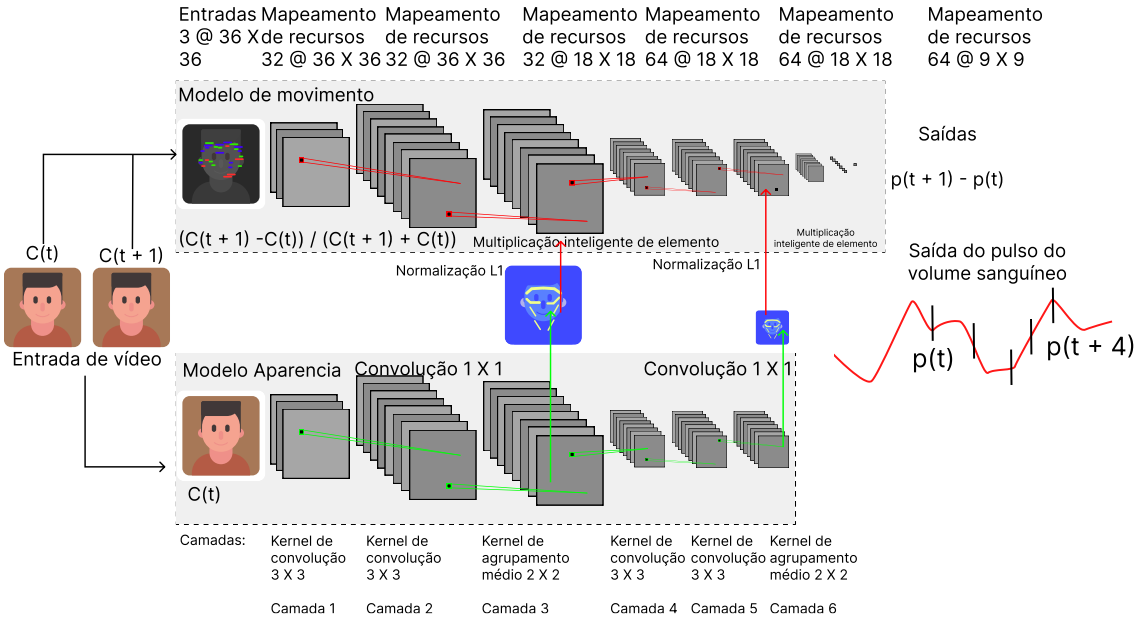

Figura 7.6. Arquitetura do modelo de convolução DeepPhys [Chen and McDuff 2018].

7.3.2.1. DeepPhys

O método DeepPhys é um método de extração iPPG com aprendizado supervisionado que utiliza de redes neurais convolucionais bidimensionais (2D CNN) para a extração do sinal. A Figura 7.6 apresenta a arquitetura do modelo DeepPhys, ilustrando uma das inovações do modelo que foi a criação de dois módulos distintos: um modelo de aparência e um modelo de movimento [Chen and McDuff 2018]. O modelo de aparência orienta o modelo de movimento a aprender representações de movimento por meio de um mecanismo de atenção. O modelo de movimento, por sua vez, utiliza a diferença normalizada entre quadros consecutivos como entrada para representar movimentos e mudanças de cor, o que melhora a robustez do DeepPhys frente a movimentos [Chen and McDuff 2018].

Além disso, o DeepPhys incorpora um mecanismo de atenção que gera máscaras de atenção suaves a partir dos quadros brutos do vídeo, atribuindo pesos maiores às regiões da pele que apresentam sinais fisiológicos mais intensos. Esse mecanismo também permite visualizar a distribuição espaço-temporal desses sinais no rosto. Embora o DeepPhys apresente um bom desempenho, em situações com variações de iluminação e artefatos de movimento, ele apresenta uma limitação importante: não consegue capturar adequadamente informações temporais dos sinais de iPPG devido à natureza estática das redes 2D CNN [Chen and McDuff 2018].

7.3.3. PhysNet

O método PhysNet é um método proposto que utiliza de redes neurais convolucionais tridimensionais (3D CNN) [Yu et al. 2019]. O modelo PhysNet foi considerado um novo marco entre os modelos 3D CNN.

Isso por que apesar de ignorar o pré-processamento da imagem, trabalhando diretamente com quadros RGB brutos como entrado, a sua arquitetura é capaz de apren-

 ©2025 SBC – Soc. Bras. de Computação

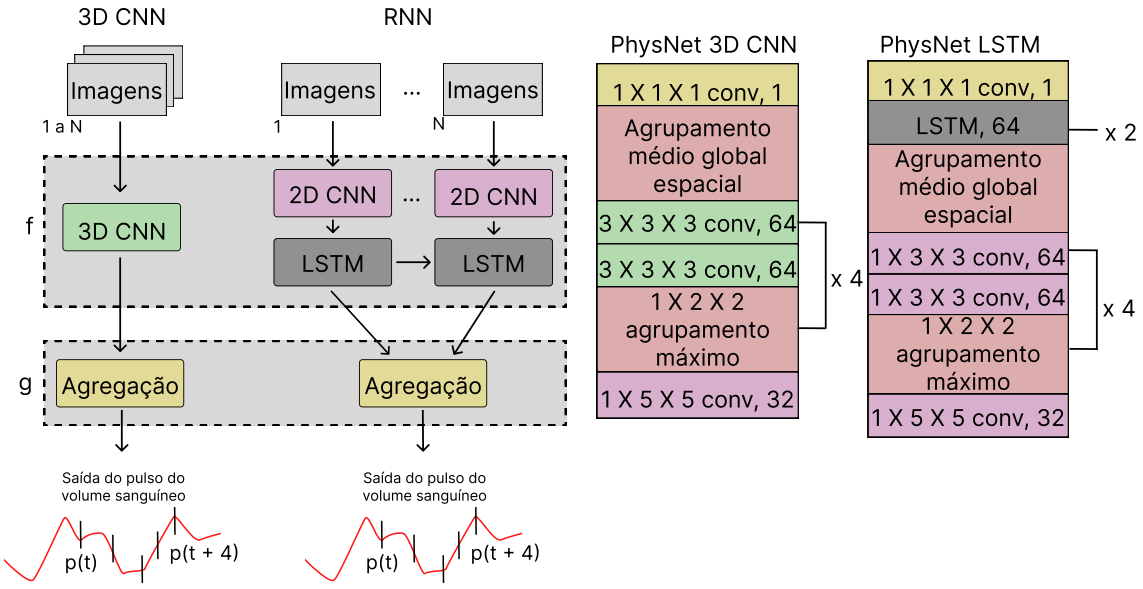

Figura 7.7. Arquitetura do modelo de convolução PhysNet [Yu et al. 2019].

der de forma eficiente a característica espaço-temporais presentes nas sequências faciais, produzindo diretamente o sinal de iPPG, sem a necessidade do pós-processamento [Yu et al. 2019]. Na Figura 7.7 é ilustrada os modelos de 3D CNN, RNN e as suas respectivas arquiteturas.

7.3.3.1. TS-CAN

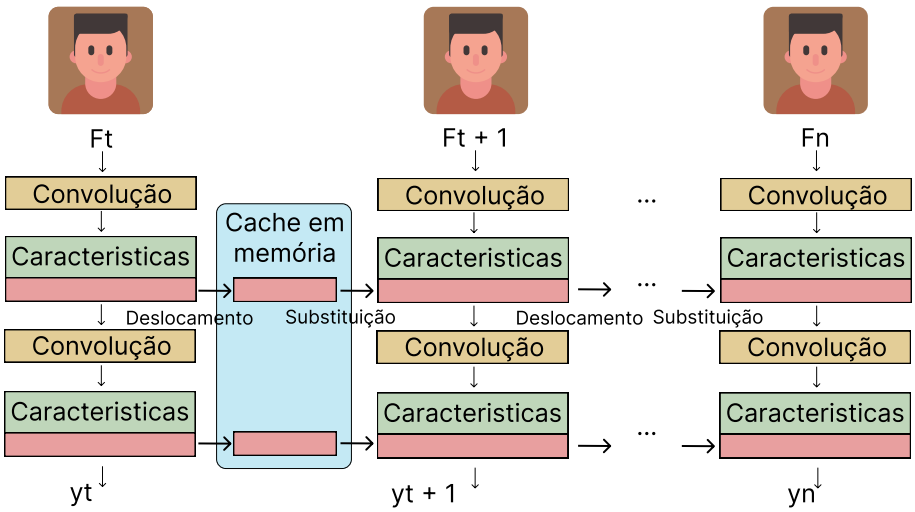

Figura 7.8. Arquitetura do modelo TSM unidirecional [Lin et al. 2019].

O método TS-CAN é uma rede neural convolucional bidimensional (2D CNN) proposta para mitigar a limitação do DeepPhys, que não consegue capturar informações temporais de forma adequada. Também conhecido como MTTS-CAN, esse método é construído com base no DeepPhys, incorporando o *Temporal Shift Module* (TSM) para

 ©2025 SBC – Soc. Bras. de Computação

recuperar a informação temporal que era antes negligenciada. Na Figura 7.8 é ilustrada a arquitetura do modelo TSM unidirecional e as propriedades de recuperação de informação temporal em cache. O princípio do TSM permite a troca de informações entre quadros adjacentes sem o uso de operações convolucionais complexas, apenas movendo blocos nos tensores ao longo do eixo temporal. Isso permite ao modelo capturar, mesmo que parcialmente, a dinâmica temporal dos sinais fisiológicos.

Diferentemente do modelo original DeepPhys, a entrada da rede de aparência no MTTS-CAN não é composta por quadros brutos do vídeo capturado, mas por quadros gerados a partir da média de múltiplos quadros adjacentes. Essa estratégia favorece a extração de informações temporais e melhora a robustez do modelo em relação a variações dinâmicas [Lin et al. 2019].

7.3.4. Modelo de CNN baseado na diferença de quadros normalizada

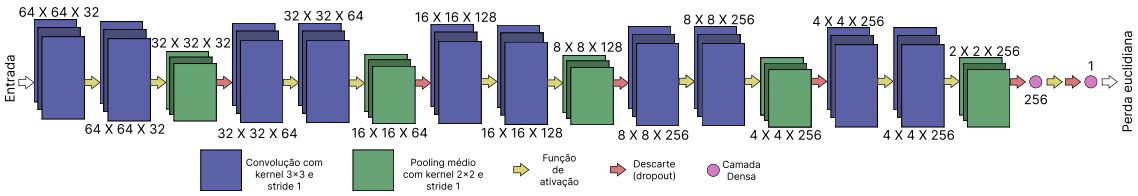

Figura 7.9. Modelo de CNN baseado na diferença de quadros normalizada [Zhan et al. 2020].

A arquitetura da rede neural convolucional (CNN) utilizada neste trabalho é ilustrada na Figura 7.9. Foi desenvolvido um modelo baseado na diferença de quadros normalizada, ou do inglês, *normalized frame difference model*, semelhante ao proposto em [Chen and McDuff 2018], com o objetivo de aprender a relação entre variações temporais na imagem obtidas por meio da normalização de diferenças entre quadros e os sinais de referência utilizados como rótulos durante o treinamento.

Diferentemente do modelo [Chen and McDuff 2018], esta arquitetura proposta opta por não empregar o módulo de atenção. Para garantir que a rede aprenda variações de cor associadas aos pixels da pele, a região facial dos sujeitos foi selecionada antes e utilizada como entrada da rede. A arquitetura CNN projetada contém dez camadas convolucionais, todas com kernels de tamanho 3×3. A cada duas camadas convolucionais, é inserida uma camada de *average pooling* com kernel 2×2, seguida por uma camada de *dropout*, com o intuito de mitigar o risco de sobreajuste. A função de ativação adotada após cada convolução é a tangente hiperbólica (*tanh*).

Em consonância com tarefas de regressão utilizando CNNs, a função de perda adotada é a distância euclidiana, que mede a discrepância entre a saída da camada totalmente conectada e o rótulo de treinamento, que corresponde ao sinal PPG diferenciado.

7.3.4.1. U-Net RGB-to-PPG

A arquitetura, denominada RGB-to-PPG [Lampier et al. 2022], foi inspirada na U-Net, uma rede convolucional originalmente desenvolvida para segmentação de imagens biológicas [Ronneberger et al. 2015]. No entanto, nesta abordagem, uma camada LSTM é

 ©2025 SBC – Soc. Bras. de Computação

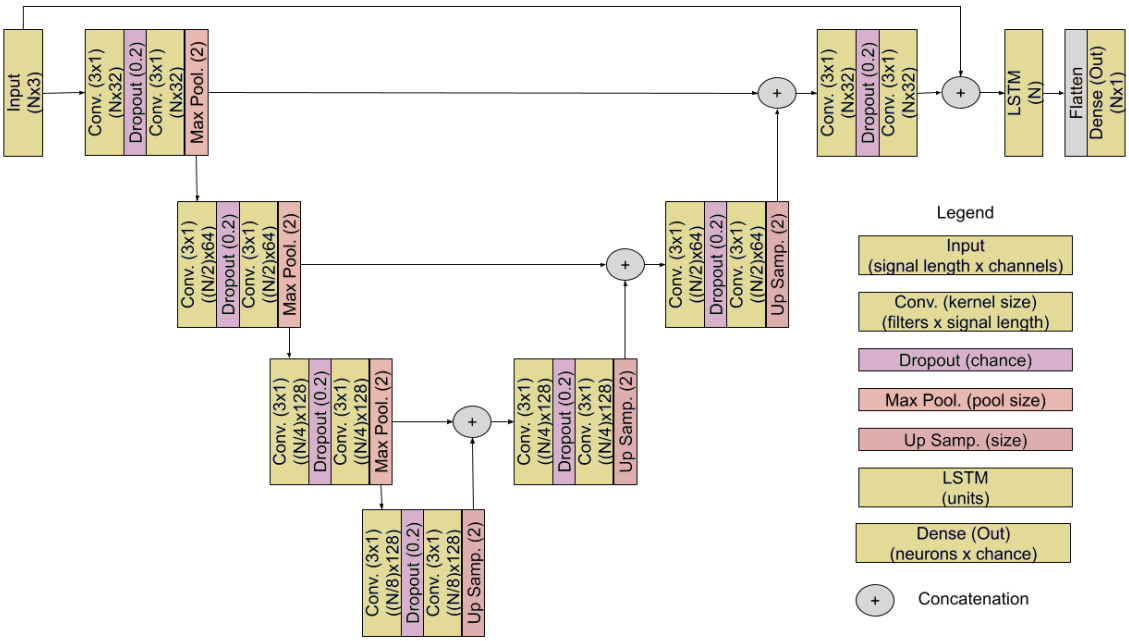

Figura 7.10. Arquitetura U-Net RGB-to-PPG [Lampier et al. 2022].

adicionada ao final da U-Net, com o objetivo de aprimorar o desempenho na tarefa de regressão. O modelo proposto recebe como entrada sinais RGB e os transforma em um sinal de pulso, a partir do qual é estimada a frequência de pulso.

A Figura 7.10 ilustra a arquitetura completa da rede proposta. As camadas convolucionais utilizam a função de ativação ReLU, enquanto as camadas LSTM empregam a tangente hiperbólica. A camada densa de saída, por sua vez, utiliza ativação linear. Os círculos representados na figura indicam operações de concatenação, responsáveis por empilhar duas entradas.

A entrada da rede consiste em uma série temporal RGB de dimensão $N \times 3$, em que N representa o número de amostras do sinal de entrada. Esses valores são previamente normalizados para o intervalo de 0 a 1, por meio da divisão por 255, correspondente à intensidade máxima em imagens com resolução de cor de 8 bits.

Para a estimativa da frequência de pulso, identificam-se os picos no sinal de iPPG gerado pelo modelo. A partir desses picos, calcula-se o inverso do intervalo entre ocorrências consecutivas. A mediana dos valores obtidos é então multiplicada por 60, convertendo a frequência de Hertz para batimentos por minuto.

7.4. Conjunto de Dados

Os conjuntos de dados, ou *datasets*, desempenham um papel essencial no desenvolvimento de métodos. Eles são utilizados para treinar, validar e testar os modelos, além de permitir a comparação entre diferentes abordagens. Neste capítulo, serão apresentados os *datasets* mais populares da área [Pirzada et al. 2024, Xiao et al. 2024].

Esses *datasets* são amplamente utilizados por fornecerem vídeos faciais acompa-

 ©2025 SBC – Soc. Bras. de Computação

nhados de sinais fisiológicos sincronizados, como os obtidos por oxímetros, registrados em diferentes condições de iluminação, movimento e qualidade de imagem. A variedade de cenários e de participantes contida nesses conjuntos é essencial para avaliar a robustez e a capacidade de generalização dos modelos de aprendizado profundo. Dessa forma, esses *datasets* se consolidaram como referência na pesquisa e no desenvolvimento de técnicas de extração de sinais de iPPG.

Tabela 7.1. Conjuntos de dados para estudos de iPPG.

Dataset	Indivíduos	Vídeos	Resolução	Padrão-ouro	Público
DEAP	32	874	720 × 576 @ 56 fps	ECG	Sim
MAHNOB-HC	27	527	1040 × 1392 @ 24 fps	ECG	Sim
UBFC-rPPG	50	50	640 × 480 @ 30 fps	PPG, HR	Sim
PURE	10	60	640 × 480 @ 30 fps	PPG, SpO2	Sim
SCAMPS	2800	2800	320 × 240 @ 30 fps	PPG, PR, RR	Sim
MMPD	22	55	1280 × 720 @ 30 fps	PPG, HR	Sim
BP4D+	140	1400	1040 × 1392 @ 25 fps	PPG, HR, BP, RR	Sim
UBFC-Phys	56	168	1024 × 1024 @ 30 fps	PPG, HR	Sim
COHFACE	40	160	640 × 480 @ 20 fps	PPG	Sim
ECG-Fitness	17	204	1920 × 1080 @ 30 fps	PPG, ECG	Sim
VIPL-HR	107	3130	960 × 720, 1920 × 1080, 640 × 480 @ 60, 30 fps	PPG, HR, SpO2	Sim
MR-NIRP	19	190	640 × 640 @ 60 fps	PPG	Sim
VicarPPG-2	50	50	1280 × 720 @ 30 fps	PPG, HR	Sim
V4V	179	1358	1720 × 720 @ 25 fps	PPG, HR, BP	Sim

7.4.1. DEAP

O conjunto de dados *DEAP* [Koelstra et al. 2011] foi inicialmente desenvolvido para a análise de emoções, mas também pode ser utilizado na avaliação de métodos de *rPPG*, uma vez que inclui sinais *PPG* autênticos. O conjunto contém dados de 32 participantes, totalizando 874 vídeos gravados com resolução de 720 × 576 e taxa de 50 quadros por segundo. Cada participante assistiu a um videoclipe musical de 1 minuto, com o objetivo de induzir diferentes estados emocionais, o que resulta em variações na frequência cardíaca. O *DEAP* coletou sinais *PPG* reais, a partir dos quais é possível calcular valores verdadeiros de frequência cardíaca.

 ©2025 SBC – Soc. Bras. de Computação

7.4.2. MAHNOB-HCI

O *MAHNOB-HCI* [Soleymani et al. 2011] é um banco de dados multimodal composto por 27 participantes, sendo que cada um gravou 20 vídeos, totalizando 527 gravações. Os vídeos foram capturados com resolução de 780 $\times$ 580 e taxa de 61 quadros por segundo. Embora seu propósito original tenha sido o reconhecimento de emoções e pesquisa em marcação implícita, o *MAHNOB-HCI* também é adequado para avaliação de métodos de medição remota da frequência cardíaca baseados em *rPPG*, devido à inclusão de sinais fisiológicos reais, como o eletrocardiograma (ECG). Todos os participantes realizaram experimentos de indução emocional e marcação implícita, durante os quais a frequência cardíaca variou em resposta às emoções. Além disso, foram utilizadas seis câmeras para capturar diferentes ângulos dos participantes (visão frontal, perfil, grande angular e close-up), tornando este conjunto adequado para testar o desempenho de métodos frente a variações de pose e ângulo.

7.4.3. UBFC-rPPG

O conjunto de dados UBFC-rPPG foi desenvolvido especificamente para a avaliação de métodos de extração de sinais de iPPG (também conhecido como *remote PPG* ou rPPG). Ele contém 50 vídeos, cada um com um indivíduo diferente, com resolução de 640 x 480 e taxa de quadros de 30 quadros por segundo. As gravações consideram variações de iluminação, tanto solar quanto interna. O UBFC-rPPG é composto por dois subconjuntos: o primeiro é uma versão reduzida, com 8 vídeos, em que os indivíduos foram instruídos a permanecer imóveis durante a captura; o segundo é um subconjunto maior, com 42 vídeos, onde os indivíduos participaram de um jogo matemático com limite de tempo, com o objetivo de aumentar sua frequência cardíaca.

O UBFC-rPPG é amplamente utilizado por pesquisadores da área devido à sua alta qualidade de vídeo e à presença de dados reais, como sinais de frequência cardíaca e PPG. Embora o conjunto de dados contenha ambos os subconjuntos, a maioria dos pesquisadores opta por utilizar o subconjunto maior devido à sua maior quantidade de dados e à qualidade superior dos vídeos gravados [Bobbia et al. 2019].

7.4.4. PURE

O conjunto de dados PURE é composto por gravações de 10 indivíduos, sendo 8 homens e 2 mulheres, com cada participante gravando 6 vídeos, totalizando 60 vídeos. As gravações foram realizadas com resolução de 640 × 480 pixels, taxa de 30 quadros por segundo e duração de um minuto por vídeo. Durante os experimentos, os participantes realizaram seis diferentes atividades, com o objetivo de gerar variações nos movimentos da cabeça. As tarefas incluíram: permanecer sentado e imóvel, conversar, movimentar a cabeça lentamente, movimentar a cabeça rapidamente, girar a cabeça em um ângulo de 20 graus e girar a cabeça em um ângulo de 35 graus. Esses movimentos foram projetados para avaliar a robustez dos métodos frente a diferentes níveis de movimentação.

Além disso, o conjunto levou em consideração variações nas condições de iluminação, utilizando luz natural proveniente de uma grande janela, sujeita à interferência de nuvens, para introduzir mudanças realistas na iluminação durante a gravação dos vídeos. Os sinais fisiológicos de referência foram obtidos com um oxímetro de dedo, com taxa de

 ©2025 SBC – Soc. Bras. de Computação

amostragem de 60 Hz, garantindo medições precisas da frequência cardíaca real. Cabe destacar que todas as imagens do conjunto PURE estão armazenadas no formato PNG sem perdas, o que assegura a fidelidade visual necessária para uma estimativa precisa dos sinais de iPPG [Stricker et al. 2014].

7.4.5. SCAMPS

O conjunto de dados SCAMPS é uma coleção de dados fisiológicos sintéticos de larga escala, composta por 2800 vídeos com resolução de 320 × 240 pixels e taxa de 30 quadros por segundo [McDuff et al. 2022]. Ele fornece rótulos de referência no nível de quadro, incluindo sinais de PPG, intervalos de pulso, formas de onda respiratórias, intervalos respiratórios e 10 ações faciais distintas. Além disso, o SCAMPS também oferece rótulos no nível de vídeo, abrangendo diversos indicadores fisiológicos.

Os parâmetros fornecidos são utilizados para gerar sinais de PPG com duração de 20 segundos e frequência de 300 Hz, juntamente com as intensidades das unidades de ação facial. Cada vídeo é sintetizado com base nesses sinais, combinados com intensidades de ação facial e atributos de aparência escolhidos aleatoriamente, como textura da pele, cor do cabelo, vestimentas, condições de iluminação e cenários de fundo.

O grande volume e a diversidade dos dados sintéticos presentes no SCAMPS são extremamente úteis para a pesquisa na área, especialmente em tarefas onde a obtenção de dados reais com tal riqueza de informações seria complexa e onerosa. No entanto, vale destacar que o SCAMPS é principalmente utilizado para fins de treinamento de modelos, e não tanto para validação ou testes finais [McDuff et al. 2022].

7.4.6. MMPD

O MMPD é o primeiro conjunto de dados desenvolvido inteiramente a partir de gravações feitas com câmeras de celulares ou *smartphones* [Tang et al. 2023]. A base é composta por 33 indivíduos e um total de 660 vídeos, com duração de um minuto cada, gravados originalmente em resolução 1280 × 720 e 30 quadros por segundo. Para facilitar o compartilhamento dos dados coletados, os vídeos foram comprimidos para uma resolução de 320 × 240.

O conjunto foi cuidadosamente projetado para abranger uma diversidade de tons de pele, com quatro categorias diferentes, e diversas condições de iluminação, incluindo LED com alta e baixa intensidade, luz incandescente e luz natural. Além disso, ele inclui diferentes atividades, como repouso, rotação de cabeça, conversação e caminhada, criando um cenário variado que permite aos pesquisadores avaliar a robustez de seus métodos em diferentes contextos ambientais.

Adicionalmente, o MMPD contém quatro experimentos focados no impacto de movimentos mais intensos e bruscos. Nesses testes, os participantes realizaram atividades físicas vigorosas, como elevações de joelhos, para aumentar sua frequência cardíaca antes da gravação dos vídeos. Após cada sessão de exercício, os indivíduos recebiam um período adequado de descanso para que a frequência cardíaca se estabilizasse antes de iniciar o próximo experimento.

O MMPD também disponibiliza rótulos reais, incluindo valores de frequência car-

 ©2025 SBC – Soc. Bras. de Computação

díaca e sinais de PPG de referência, que são recursos valiosos para estudos de estimação de sinais fisiológicos [Tang et al. 2023].

7.4.7. BP4D+

O conjunto de dados BP4D+ é uma base multimodal projetada para análise de emoções espontâneas, com foco na estimativa remota de sinais fisiológicos, como frequência cardíaca e pressão arterial. Ele contém gravações de vídeos de 140 indivíduos, cada um participando de 10 tarefas emocionais, resultando em 1400 vídeos no total. As gravações são feitas a uma taxa de 25 quadros por segundo e incluem tanto vídeos RGB quanto térmicos, ambos capturados na mesma frequência de quadros [Zhang et al. 2016].

Além dos vídeos, o BP4D+ fornece dados fisiológicos, como medições de pressão arterial (sistólica, diastólica e média), frequência cardíaca (batimentos por minuto), taxa de respiração e atividade galvânica da pele (EDA), que são coletados simultaneamente aos vídeos. Essas medições são acompanhadas por modelos 3D dinâmicos da face dos participantes, o que permite uma análise detalhada das expressões faciais durante a execução das tarefas [Zhang et al. 2016].

As tarefas emocionais foram projetadas para induzir respostas emocionais específicas nos participantes, proporcionando um cenário controlado para a análise de emoções e suas correlações com os sinais fisiológicos. O BP4D+ tem sido amplamente utilizado em pesquisas que envolvem a estimativa de sinais fisiológicos a partir de vídeos faciais, como a fotopletismografia por imagem, análise de expressões faciais para reconhecimento de emoções e diagnóstico de condições psicológicas, além do desenvolvimento de interfaces afetivas que respondem às emoções dos usuários. Sua riqueza multimodal e a diversidade de dados tornam o BP4D+ um recurso valioso para o avanço da pesquisa nessa área [Zhang et al. 2016].

7.4.8. UBFC-Phys

O conjunto de dados UBFC-Phys foi desenvolvido inicialmente para o reconhecimento de emoções e é composto por 56 indivíduos, sendo 46 do sexo feminino e 10 do sexo masculino. Cada participante foi instruído a realizar três tarefas distintas: descansar, conversar e resolver problemas matemáticos, resultando em um total de 168 gravações de vídeo. As gravações foram realizadas com resolução de 1024×1024 pixels e taxa de 35 quadros por segundo [Meziatisabour et al. 2021].

Além dos vídeos, o conjunto UBFC-Phys utiliza um dispositivo de pulseira inteligente para coletar sinais PPG, considerados como referência padrão-ouro para medições de frequência cardíaca. Para complementar as gravações, os participantes preencheram questionários antes e após os experimentos, com o objetivo de registrar dados relacionados ao nível de ansiedade, fornecendo informações adicionais sobre o estado emocional durante as tarefas [Meziatisabour et al. 2021].

7.4.9. COHFACE

O *dataset* COHFACE [Heusch et al. 2017] é um conjunto de dados público criado pelo com o intuito de permitir que pesquisadores avaliem seus métodos de RPPG/ IPPG de maneira padronizada e justa. A base contém dados de 40 participantes, sendo 28 homens

 ©2025 SBC – Soc. Bras. de Computação

e 12 mulheres, e cada indivíduo contribuiu com quatro gravações de vídeo, totalizando 160 vídeos. As gravações foram feitas com resolução de 640 × 480 pixels e uma taxa de 20 quadros por segundo. Para obter os sinais fisiológicos reais, todos os participantes utilizaram sensores de PPG de contato durante as filmagens.

Durante a coleta dos dados, foram consideradas duas condições distintas de iluminação. Em uma delas, os vídeos foram gravados com iluminação artificial de estúdio, onde as janelas foram mantidas fechadas para bloquear a luz natural e garantir uma iluminação estável com luzes artificiais. Na outra condição, os vídeos foram gravados com iluminação natural, mantendo as janelas abertas e desligando todas as luzes artificiais.

Apesar de sua importância e ampla utilização, a principal limitação do COHFACE é o fato de que os vídeos foram fortemente comprimidos, o que introduz uma quantidade significativa de ruído. Esse ruído pode afetar negativamente a extração precisa dos sinais rPPG, comprometendo a acurácia dos métodos avaliados com essa base de dados [Heusch et al. 2017].

7.4.10. ECG-Fitness

ECG-Fitness [Spetlik et al. 2018] é um *dataset* público composto por registros de 17 participantes, sendo 14 do sexo masculino e 3 do sexo feminino, que realizaram quatro tipos distintos de atividades: fala, remo, exercícios em bicicleta ergométrica e em aparelho elíptico. As gravações foram feitas com o uso de duas webcams Logitech C920 e uma câmera térmica FLIR, sob três condições de iluminação diferentes: luz natural proveniente de janelas próximas, iluminação com lâmpadas halógenas de 400 W e luzes LED de 30 W. Para cada participante, foram gerados 12 vídeos, cobrindo todas as combinações entre os três tipos de iluminação e os quatro estados de atividade, totalizando assim 204 vídeos no conjunto. As gravações foram feitas em resolução Full HD(1920 × 1080 pixels) a uma taxa de 30 quadros por segundo, com duração de um minuto cada. Um dos aspectos mais notáveis do ECG-Fitness é que ele é o único *dataset* conhecido que inclui a atividade de remo entre os registros [Spetlik et al. 2018].

7.4.11. VIPL-HR

O VIPL-HR [Niu et al. 2018] é um *dataset* público multimodal de grande escala e alta complexidade, composto por gravações de vídeos de 107 indivíduos. Ele inclui três tipos distintos de vídeos: vídeos RGB, vídeos no espectro infravermelho próximo (NIR) e vídeos gravados por câmeras de celulares smartphones. Essas gravações foram realizadas utilizando câmeras RGB, câmeras RGB-D e câmeras de smartphones. No total, o *dataset* reúne 3.130 vídeos faciais em luz visível. Os vídeos RGB foram obtidos tanto por câmeras RGB quanto por câmeras RGB-D, com resoluções de 960 × 720 pixels a 25 fps e 1920 × 1080 pixels a 30 fps, respectivamente. Já os vídeos NIR foram capturados com câmeras RGB-D, com resolução de 640 × 480 pixels a 30 fps. Os vídeos de smartphone, por sua vez, foram registrados em 1920 × 1080 pixels a 30 fps [Niu et al. 2018].

O uso desses diferentes tipos de dispositivos busca permitir a avaliação da robustez dos métodos propostos dada as diferentes modalidades de vídeo. Além disso, o VIPL-HR introduz dois fatores que influenciam a coleta dos dados: movimento da cabeça (estável, movimento intenso, falando) e condições de iluminação (ambiente de laboratório, escuro

 ©2025 SBC – Soc. Bras. de Computação

e claro), fornecendo um cenário mais realista para os testes. O conjunto também disponibiliza rótulos fisiológicos reais, como frequência cardíaca (HR), oxigenação do sangue (SPO2) e volume de pulso sanguíneo (BVP) [Niu et al. 2018].

7.4.12. MR-NIRP

O MR-NIRP [Koelstra et al. 2011] é o primeiro *dataset* de vídeos de dados fisiológicos que inclui cenários de direção, oferecendo uma abordagem mais realista em comparação com os tradicionais ambientes controlados. Ele é composto por 190 vídeos de 19 indivíduos, capturados tanto durante a condução de um veículo quanto dentro de um carro estacionado. Durante as gravações, os sujeitos também realizaram atividades como falar e mover a cabeça aleatoriamente, simulando situações comuns ao dirigir. Os vídeos foram capturados em uma resolução de 640 × 640 pixels com uma taxa de 60 quadros por segundo.

O objetivo principal do MR-NIRP é permitir a avaliação do desempenho de diferentes métodos de RPPG em contextos de direção, ampliando os testes para além das condições controladas da realização da coleta de dados. Os vídeos são sincronizados com sinais reais de PPG, obtidos por meio de um oxímetro de pulso colocado no dedo dos participantes. As gravações incluem simultaneamente dados em RGB e infravermelho próximo (NIR), embora muitos estudos optem por utilizar os dados NIR para treinar e testar os modelos de deep learning. No entanto, o *dataset* apresenta algumas limitações, como a presença de valores nulos (zeros) nos sinais de PPG, o que pode dificultar a avaliação precisa dos métodos de rPPG aplicados nesses dados [Koelstra et al. 2011].

7.4.13. VicarPPG-2

O VicarPPG-2 [Gudi et al. 2020] é um conjunto de dados público composto por gravações de vídeos de 10 voluntários, com idade média de 29 anos. Foram registrados 40 vídeos, cada um com duração de 5 minutos, gravados em resolução de 1280 × 720 pixels e a uma taxa de 60 quadros por segundo. Cada participante realizou a gravação de vídeos distintos. No primeiro, os indivíduos permaneciam em estado estático. No segundo, realizavam cinco movimentos planejados de corpo e cabeça: inclinação lateral da cabeça , movimento vertical da cabeça , combinação dos dois , movimentação dos olhos com a cabeça imóvel, e movimentos naturais da cabeça enquanto ouviam música. No terceiro vídeo, os participantes eram expostos a um jogo projetado para gerar estresse, enquanto no quarto, apareciam em estado de relaxamento, logo após terem passado por exercícios físicos que induzem à fadiga. Para obter os sinais fisiológicos reais, o *dataset* utilizou oxímetros de pulso CMS50E, conectados aos dedos dos participantes, registrando sinais PPG autênticos [Gudi et al. 2020].

7.4.14. V4V

O conjunto de dados V4V [Revanur et al. 2021] compreende uma coleção de gravações de vídeos e dados fisiológicos e conta com vídeos de 179 indivíduos, abrangendo diferentes etnias, como afro-americanos, caucasianos e asiáticos. Cada voluntário participou de até 10 tarefas experimentais, planejadas para provocar emoções específicas, totalizando 1.358 vídeos. As gravações têm duração variável, entre 5 e 206 segundos, e foram realizadas com resolução de 1280 × 720 pixels e 25 quadros por segundo. Para a coleta

 ©2025 SBC – Soc. Bras. de Computação

de dados fisiológicos reais, o V4V utiliza o sistema de aquisição BIOPAC MP150, que registra sinais como PPG, frequência cardíaca, pressão arterial e outras medidas vitais. Apesar de seu grande volume de dados e da variedade de desafios emocionais propostos, o conjunto mantém condições de iluminação consistentes em todos os vídeos, garantindo uniformidade na qualidade das imagens [Revanur et al. 2021].

7.5. Métricas de Avaliação

As métricas de avaliação em fotopletismografia por imagem são fundamentais para quantificar a precisão e a qualidade dos modelos aplicados na estimativa de parâmetros fisiológicos, como a frequência cardíaca. Como as técnicas de iPPG estimam sinais contínuos ao longo do tempo, é necessário utilizar métricas de regressão para avaliar o desempenho dos métodos. A seguir, são apresentadas as principais métricas utilizadas na avaliação de modelos baseados em iPPG.

7.5.1. Relação Sinal-Ruído

A Relação Sinal-Ruído, conhecida do inglês como *Signal-to-Noise Ratio* (SNR) é uma métrica amplamente empregada para avaliar a qualidade dos sinais de iPPG. Sua função principal é quantificar o quanto da informação contida no sinal está efetivamente associada à atividade cardíaca, em comparação com os componentes considerados ruído [De Haan and Jeanne 2013].

Para o cálculo da SNR, equação 22, adota-se a razão entre a energia espectral concentrada na vizinhança da frequência fundamental do pulso e a energia remanescente no intervalo de 40 a 240 batimentos por minuto (bpm). A frequência fundamental é determinada com precisão por meio de um sinal de ECG registrado simultaneamente, servindo como referência confiável.

$$\mathrm{SNR} = 10 \cdot \log_{10} \left(\frac{\sum_{f=40}^{240} \left(U_t(f) \cdot \hat{S}(f) \right)^2}{\sum_{f=40}^{240} \left((1 - U_t(f)) \cdot \hat{S}(f) \right)^2} \right) \tag{22}$$

Onde:

- $\hat{S}(f)$ é o espectro do sinal de pulso (obtido por iPPG);
- f representa a frequência em batimentos por minuto (bpm), no intervalo de 30 a 240 bpm;
- $U_t(f)$ é uma janela binária (template), que assume valor 1 nas regiões em torno da frequência fundamental do pulso e sua primeira harmônica, e 0 fora dessas regiões.

Dado que a frequência cardíaca varia ao longo do tempo, especialmente durante atividades físicas, a análise é feita em janelas temporais curtas, utilizando uma abordagem de janela deslizante. A SNR final é obtida pela média dos valores calculados em cada janela, permitindo uma avaliação dinâmica e realista da qualidade dos sinais de iPPG ao longo do tempo.

 ©2025 SBC – Soc. Bras. de Computação

7.5.2. Erro Absoluto Médio

O Erro Absoluto Médio, do inglês, *Mean Absolute Error* (MAE) é uma das métricas tradicionais utilizadas na avaliação de modelos de regressão. Ele calcula a média das diferenças absolutas entre os valores preditos e os valores reais. O MAE fornece uma visão clara do erro médio cometido pelo modelo em suas previsões, sem considerar a direção do erro (se positivo ou negativo).

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{23}$$

Onde:

- y_i são os valores reais,
- $\hat{y}_i$ são os valores preditos,
- n é o número total de observações.

7.5.3. Erro Quadrático Médio

O Erro Quadrático Médio, do inglês *Mean Squared Error* (MSE) é uma métrica que calcula a média dos quadrados das diferenças entre os valores reais e preditos. O MSE é útil quando se deseja enfatizar grandes desvios entre as previsões e os valores reais.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{24}$$

7.5.4. Raiz do Erro Quadrático Médio

A Raiz do Erro Quadrático Médio, do inglês *Root Mean Squared Error* (RMSE) é a raiz quadrada do MSE. Essa métrica expressa o erro na mesma unidade dos dados originais, o que facilita a interpretação dos resultados e permite uma análise direta da magnitude dos desvios entre os valores estimados e os reais. A métrica RMSE é especialmente útil para comparar diferentes modelos de previsão, já que permite uma interpretação direta em termos da magnitude do erro.

$$\text{RMSE} = \sqrt{\text{MSE}} \tag{25}$$

7.5.5. Coeficiente de Determinação

O Coeficiente de Determinação, também conhecido como R^2, mede a proporção da variabilidade dos dados explicada pelo modelo. Ele varia entre 0 e 1, onde um valor de 1 indica que o modelo explica toda a variabilidade dos dados, e um valor de 0 indica que o modelo não explica nada da variabilidade. Um R^2 mais alto indica um modelo melhor ajustado aos dados.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{26}$$

 ©2025 SBC – Soc. Bras. de Computação

Onde $\bar{y}$ é a média dos valores reais.

7.6. Gráficos de Avaliação

Além das métricas numéricas, a análise por meio de gráficos é uma etapa complementar indispensável para verificar a consistência das estimativas em relação aos sinais de referência. A seguir, são apresentadas as principais representações gráficas utilizadas na avaliação de modelos baseados em iPPG.

7.6.1. Espectrograma

O espectrograma é uma representação tempo-frequência que permite observar como o conteúdo espectral de um sinal varia ao longo do tempo. Na análise de sinais de iPPG, ele desempenha um papel importante ao revelar a frequência cardíaca predominante e seus harmônicos, bem como sua estabilidade e variações dinâmicas. Um exemplo ilustrativo é apresentado na Figura 7.11.

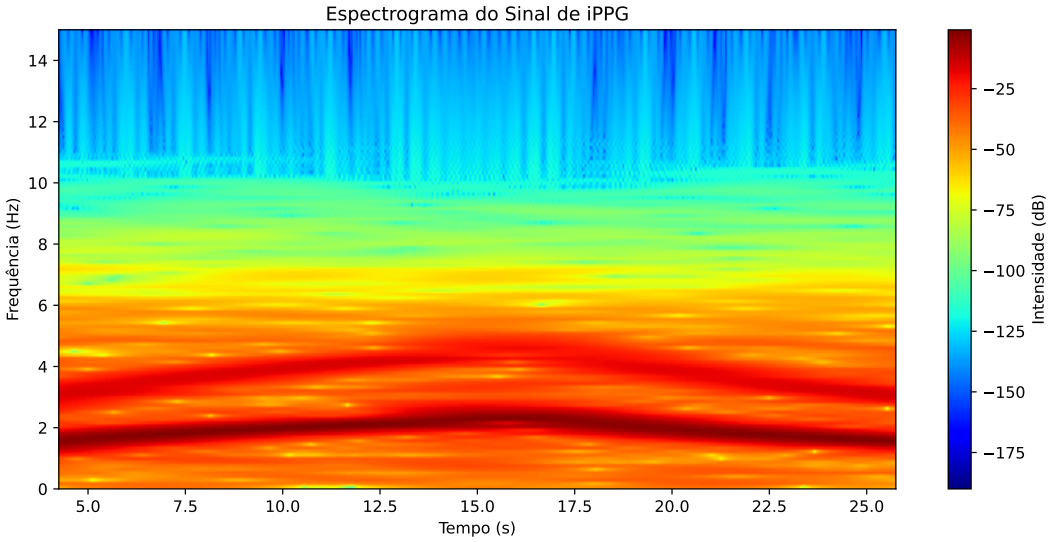

Figura 7.11. Exemplo de um gráfico de espectrograma.

Essa representação é especialmente relevante na avaliação comparativa entre diferentes métodos de extração de iPPG. Ao aplicar espectrogramas aos sinais obtidos por distintas abordagens, é possível visualizar não apenas a presença da frequência cardíaca estimada, mas também o nível de ruído, a ocorrência de variações abruptas ou artefatos e a fidelidade espectral em relação ao comportamento fisiológico esperado.

Por exemplo, um método robusto tende a produzir espectrogramas com uma faixa espectral bem definida e contínua ao longo do tempo, enquanto métodos mais suscetíveis a interferências podem exibir múltiplas bandas incoerentes, interrupções ou espalhamento espectral. Além disso, a presença clara de harmônicos pode indicar que a morfologia do pulso foi preservada, reforçando a qualidade da estimativa.

 ©2025 SBC – Soc. Bras. de Computação

7.6.2. Gráfico de Dispersão

O gráfico de dispersão, também conhecido como Scatter Plot, é uma ferramenta fundamental para comparar visualmente os valores reais com os valores preditos por um modelo. Esse gráfico é especialmente útil na avaliação de modelos de regressão, pois permite observar se há uma correlação entre os dois conjuntos de dados, facilitando a identificação de padrões e a avaliação da precisão do modelo.

No gráfico de dispersão, os valores reais (y_i) são representados no eixo vertical (eixo y), enquanto os valores preditos ($\hat{y}_i$) são plotados no eixo horizontal (eixo x). Para um modelo de regressão ideal, espera-se que os pontos do gráfico estejam distribuídos ao longo da linha $y = x$, o que indica que as previsões do modelo estão próximas dos valores reais. Quanto mais próximo o conjunto de pontos estiver da linha de identidade (linha reta onde $y = x$), melhor será o desempenho do modelo.

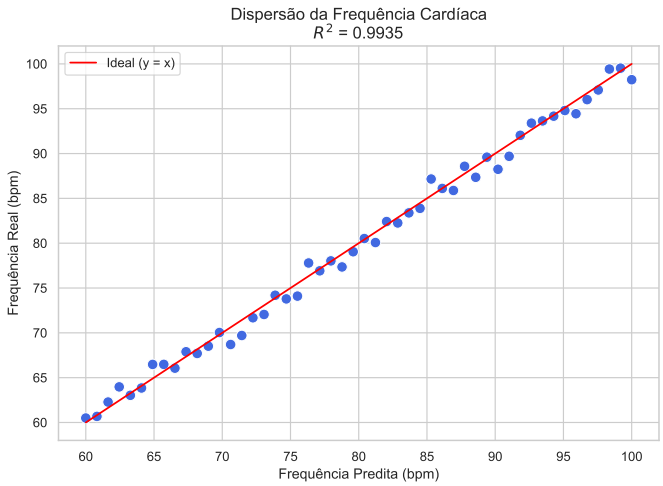

Figura 7.12. Exemplo de um gráfico de dispersão.

A análise do gráfico de dispersão oferece diversas informações sobre a qualidade do modelo, como:

- Correlações: Se os pontos formam uma linha reta ou uma curva, isso indica a presença de uma correlação entre os valores reais e preditos. No caso de uma correlação linear, os pontos estarão próximos da linha $y = x$. Em casos de correlação não linear, os pontos podem formar uma curva.
- Erros sistemáticos: A presença de desvios sistemáticos, como agrupamentos de pontos em certas áreas ou uma distribuição não uniforme ao longo da linha $y = x$, pode indicar que o modelo está cometendo erros em certas faixas de valores. Por exemplo, se os pontos tendem a se concentrar mais em uma parte do gráfico, pode ser um sinal de que o modelo tem dificuldades em prever valores em outra parte do intervalo.
- Homocedasticidade: A dispersão dos pontos ao longo do gráfico pode revelar informações sobre a variância dos erros. Se os pontos estão igualmente distribuídos

 ©2025 SBC – Soc. Bras. de Computação

ao longo de toda a linha $y = x$, isso sugere que o modelo tem uma variância constante nos erros, ou seja, homocedasticidade. No entanto, se a dispersão dos pontos for maior em algumas áreas e menor em outras, pode indicar heterocedasticidade (variância não constante).

7.6.3. Histograma dos erros

O histograma é uma ferramenta visual fundamental para avaliar a qualidade de modelos preditivos, especialmente na estimativa de sinais fisiológicos por iPPG. Quando aplicado aos erros do modelo, ele exibe a frequência de ocorrência de diferentes valores de erro, permitindo observar a forma da distribuição. No eixo horizontal são representados os valores dos erros (diferença entre os valores preditos e os reais), enquanto o eixo vertical indica a frequência com que esses erros ocorrem.

Uma distribuição simétrica e centrada em torno de zero sugere que o modelo não possui viés sistemático e que os erros são majoritariamente aleatórios, o que é desejável. Por outro lado, uma concentração de erros positivos ou negativos pode indicar viés, enquanto a presença de caudas longas ou picos incomuns pode apontar para *outliers* ou falhas na suposição de normalidade dos resíduos. Essas observações são essenciais para verificar a robustez do modelo e orientar ajustes ou melhorias. A Figura 7.13 apresenta um exemplo desse tipo de análise.

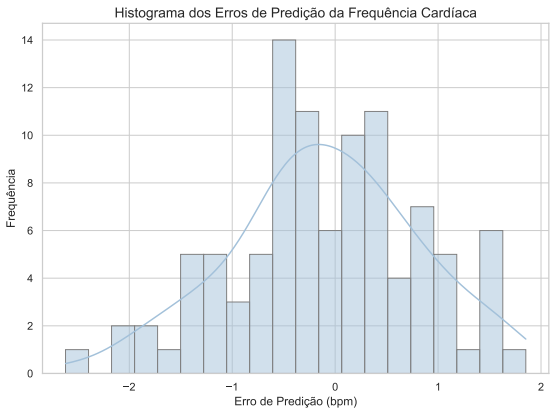

Figura 7.13. Exemplo de histograma dos erros.

7.6.4. Boxplot dos erros

O boxplot é uma representação gráfica compacta que resume a distribuição dos erros de um modelo, permitindo visualizar rapidamente sua dispersão, simetria e presença de *outliers*. A Figura 7.14 ilustra um exemplo desse tipo de gráfico.

No gráfico, a caixa representa o intervalo interquartil (IQR), entre o primeiro (Q1) e o terceiro quartil (Q3), com a linha interna indicando a mediana. As extremidades dos "bigodes" mostram os limites inferiores e superiores (até 1,5 vezes o IQR), enquanto pontos fora desses limites são considerados *outliers*.

Na análise dos erros, o boxplot permite verificar:

 ©2025 SBC – Soc. Bras. de Computação

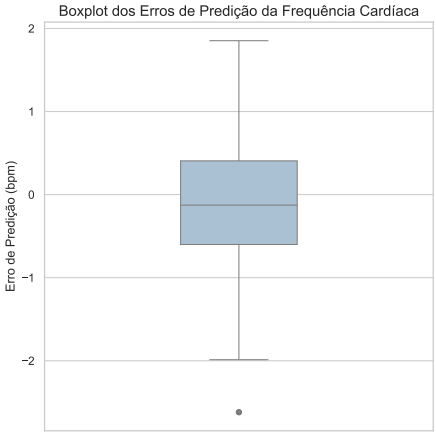

Figura 7.14. Exemplo de boxplot dos erros.

- **Centralidade**: A mediana próxima de zero sugere ausência de viés sistemático.
- **Dispersão**: Uma caixa estreita indica baixa variabilidade nos erros; uma caixa larga, maior incerteza nas previsões.
- **Outliers**: Erros extremos podem sinalizar limitações do modelo em certos casos ou a presença de dados atípicos.

Na avaliação de modelos de extração de sinais de iPPG, onde a precisão nas estimativas de parâmetros fisiológicos é fundamental, o boxplot dos erros é uma ferramenta eficaz para avaliar se o modelo é consistente e se há pontos que exigem ajustes ou maior atenção.

7.6.5. Gráfico de Bland-Altman

O gráfico de Bland-Altman é uma ferramenta visual amplamente utilizada para avaliar a concordância entre valores reais (y_i) e preditos ($\hat{y}_i$) [Giavarina 2015], especialmente em sinais de iPPG. Ele representa a diferença entre os valores preditos e reais em função da média entre eles, permitindo identificar possíveis desvios sistemáticos ou erros concentrados em determinadas faixas de valor.

$$\text{Erro} = \hat{y}_i - y_i \tag{27}$$

Esse tipo de análise é útil para verificar a presença de viés (quando os erros não se distribuem em torno de zero) e para detectar variações na precisão do modelo ao longo do intervalo de estimativas. Uma distribuição uniforme e simétrica dos erros ao redor de zero indica um bom desempenho, enquanto padrões sistemáticos sugerem limitações do modelo em certas condições.

A Figura 7.15 mostra um exemplo típico desse gráfico, utilizado para avaliar o desempenho de estimativas de frequência cardíaca a partir de sinais de iPPG. Este gráfico complementa outras métricas ao oferecer uma perspectiva visual direta sobre a consistência das previsões em diferentes regiões do sinal analisado.

 ©2025 SBC – Soc. Bras. de Computação

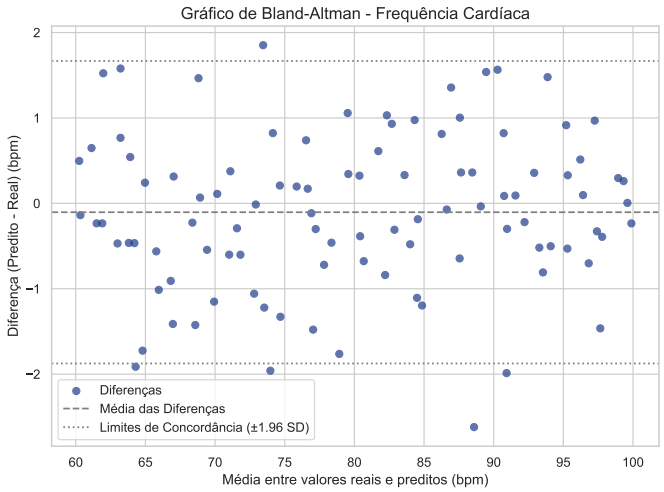

Figura 7.15. Exemplo de gráfico de Bland-Altman.

7.7. Implementação prática de um modelo de iPPG

A implementação prática de um modelo de extração de sinais de iPPG envolve uma série de etapas, que vão desde a configuração da câmera e iluminação até a obtenção do sinal do pulso extraído. Os métodos de iPPG são variados, com diferenças significativas entre eles. O modelo que será descrito a seguir utiliza o método tradicional de extração baseado no método POS, amplamente utilizado para a extração de sinais cardíacos a partir de variações de cor na pele, especialmente na região facial.

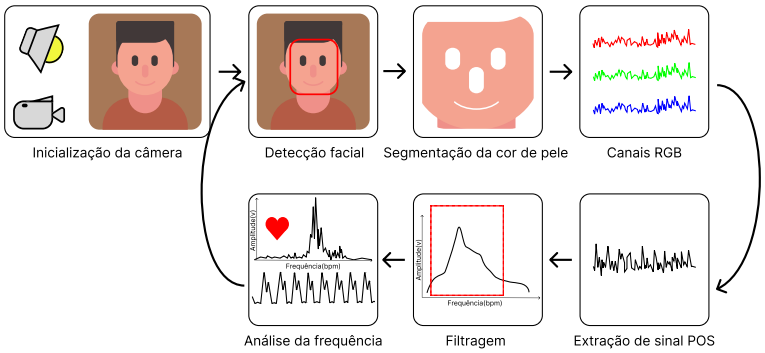

Figura 7.16. Visão geral da captura e processamento do sinal de iPPG.

7.7.1. Aquisição dos dados

A primeira etapa no processo de implementação de um modelo de iPPG é a captura de imagens da região facial do participante. A qualidade da aquisição do sinal depende fortemente das condições do captura de vídeo. É recomendado:

- **Câmera:** Utilize uma câmera RGB com uma taxa de quadros adequada (por exemplo de 30 fps) e resolução suficiente para capturar detalhes faciais (por exemplo,

 ©2025 SBC – Soc. Bras. de Computação

640x480 pixels). Webcams, câmeras DSLR, câmeras de smartphones ou sensores como o Kinect podem ser utilizados. Além disso, para garantir uma captura estável e precisa, recomenda-se desativar o foco automático e o equilíbrio de branco automático da câmera. O foco automático pode causar variações no sinal devido ao ajuste constante da lente, enquanto o equilíbrio de branco automático pode alterar as cores do sinal, comprometendo a precisão da extração. Ajuste manualmente o foco e o equilíbrio de branco para obter uma imagem estável e consistente ao longo da captura.

- **Iluminação:** É essencial garantir uma fonte de luz estável e difusa, com variação temporal mínima. Flutuações na iluminação ambiente podem introduzir ruídos consideráveis, prejudicando a qualidade do sinal extraído. A iluminação homogênea melhora a precisão da detecção dos sinais fisiológicos.

- **Posicionamento do Participante:** O indivíduo deve estar posicionado de frente para a câmera, com o rosto claramente visível e evitando movimentos excessivos durante a captura. Movimentos significativos da cabeça ou do corpo podem comprometer a estabilidade e a qualidade do sinal. Para resultados ótimos, recomenda-se uma distância entre 0,9 m e 1,2 m entre o participante e a câmera [Han et al. 2015].

7.7.2. Detecção da face e seleção da região de interesse

A detecção da face é realizada por meio de algoritmos especializados, disponíveis em bibliotecas de processamento de imagens como OpenCV [Bradski and Kaehler 2008] e Mediapipe [Lugaresi et al. 2019], que apresentam alto desempenho, especialmente em tempo real. Após a identificação da face, é determinada a Região de Interesse (ROI) para a extração do sinal. As áreas mais comumente selecionadas para essa finalidade na literatura incluem a testa, as bochechas ou toda a face, dependendo do contexto e da precisão desejada. A ROI pode ser ajustada automaticamente a cada novo quadro utilizando algoritmos de rastreamento, como o Kanade-Lucas-Tomasi [Wu et al. 2013], que seguem pontos característicos do rosto ao longo do tempo, reduzindo os efeitos de pequenos movimentos e garantindo a precisão da extração do sinal.

7.7.3. Segmentação da pele

A segmentação da pele é uma etapa essencial para isolar a área relevante e minimizar a influência de outras que não são relacionadas à pele. A segmentação pode ser feita utilizando algoritmos baseados em limiar de cor ou em técnicas mais elaboradas de aprendizado de máquina.

Um método simples envolve a utilização do espaço de cores YCbCr para identificar os pixels da pele [Saxen and Al-Hamadi 2014]. Para isso, são aplicados limiares nas componentes de cor, como mostrado abaixo:

$$\text{MáscaraPele}(x,y) = \begin{cases} 1, & \text{se } Cb_{\min} \leq Cb(x,y) \leq Cb_{\max} \text{ e } Cr_{\min} \leq Cr(x,y) \leq Cr_{\max} \\ 0, & \text{caso contrário} \end{cases} \tag{28}$$

 ©2025 SBC – Soc. Bras. de Computação

Os limiares $Cb_{\min}, Cb_{\max}, Cr_{\min}, Cr_{\max}$ são definidos com base em estudos empíricos sobre a distribuição dos valores da cor da pele humana [Saxen and Al-Hamadi 2014]. Alternativamente, a segmentação pode ser feita utilizando algoritmos de agrupamento tradicionais ou, mais recentemente com redes neurais convolucionais (CNN).

7.7.4. Construção dos sinais de cores

Uma vez capturadas as regiões de interesse das imagens, realiza-se a extração dos sinais brutos de cor (canais R, G e B). Para cada quadro, os valores médios dos pixels da ROI são calculados para formar três séries temporais, representadas como:

$$s_R(t) = \frac{1}{N}\sum_{i=1}^{N} I_R(x_i, y_i, t) \tag{29}$$

$$s_G(t) = \frac{1}{N}\sum_{i=1}^{N} I_G(x_i, y_i, t) \tag{30}$$

$$s_B(t) = \frac{1}{N}\sum_{i=1}^{N} I_B(x_i, y_i, t) \tag{31}$$

onde $s_R(t), s_G(t)$ e $s_B(t)$ são as séries temporais para os canais vermelho, verde e azul, respectivamente, $I_R(x_i, y_i, t)$, $I_G(x_i, y_i, t)$ e $I_B(x_i, y_i, t)$ são os valores de intensidade dos pixels da ROI em cada quadro t, e N é o número de pixels na ROI. Esse processo é importante para atenuar o ruído presente em pixels individuais.

7.7.5. Extração do sinal com o método POS

Os sinais extraídos de cada canal RGB são normalizados com base na média de uma janela temporal de análise escolhida. A normalização é realizada dividindo-se o valor de cada canal pela média da série dentro da janela. Os sinais normalizados são então organizados em um vetor tridimensional, representando os canais de cor normalizados. Depois, o método transforma os sinais RGB normalizados em um novo espaço de sinais pulsáteis utilizando uma projeção ortogonal. Após a projeção, o sinal obtido é então ajustado com base em um parâmetro de ponderação (α), e o sinal de pulso final é gerado por meio da soma com sobreposição das estimativas anteriores, como visto anteriormente na explicação do método.

7.7.6. Filtragem dos sinais

O sinal extraído por meio do método POS é, posteriormente, submetido a um processo de filtragem utilizando um filtro passa-faixa, cuja principal finalidade consiste em isolar as componentes espectrais associadas à frequência cardíaca, atenuando simultaneamente ruídos de baixa e alta frequência que não apresentam relevância fisiológica. A faixa espectral de interesse para a frequência cardíaca humana situa-se, em geral, entre 0,75 Hz e 4,0 Hz, correspondendo a um intervalo aproximado de 45 a 240 batimentos por minuto (bpm). Entre os filtros passa-faixa comumente adotados nesse contexto, destacam-se os modelos de Butterworth, Chebyshev e Elíptico, sendo a escolha do tipo de filtro deter-

 ©2025 SBC – Soc. Bras. de Computação

minada por critérios como a taxa de atenuação fora da banda passante, a linearidade da resposta em fase e a complexidade computacional, conforme as exigências da aplicação.

7.7.7. Extração da frequência cardíaca

Concluída a etapa de filtragem, procede-se à conversão do sinal para o domínio da frequência, com o emprego da Transformada Rápida de Fourier (FFT). Essa técnica permite decompor o sinal temporal em suas componentes harmônicas, fornecendo uma representação espectral detalhada. A partir dessa análise espectral, identifica-se o pico de maior amplitude dentro da faixa de interesse, o qual corresponde à frequência cardíaca dominante, expressa em Hertz (Hz). Para a conversão dessa medida para batimentos por minuto, utiliza-se a relação direta $\text{bpm} = \text{Hz} \times 60$, valor padronizado na quantificação da frequência cardíaca.

7.8. Aplicações

Os avanços nas pesquisas relacionadas à área de iPPG têm possibilitado uma expansão significativa no campo de aplicações dessas abordagens. A seguir, são discutidas algumas das possíveis aplicações emergentes dessas abordagens, incluindo aquelas que já estão sendo investigadas por estudos recentes.

7.8.1. Medição de múltiplos sinais vitais

A técnica de iPPG surgiu inicialmente com o propósito de monitorar remotamente a frequência cardíaca. Com o amadurecimento das pesquisas na área, essa tecnologia expandiu suas possibilidades de aplicação, permitindo a estimativa de diversos sinais fisiológicos de forma não intrusiva.

Atualmente, a técnica de iPPG tem sido adaptada para a medição de parâmetros como pressão arterial [Zeng et al. 2025], frequência respiratória, variabilidade da frequência cardíaca e saturação de oxigênio [Lampier et al. 2023]. No caso da pressão arterial, seu monitoramento remoto representa uma alternativa promissora aos métodos convencionais, sobretudo para a detecção de quadros de hipertensão sem a necessidade de dispositivos de contato direto com a pele.

A saturação de oxigênio no sangue, por sua vez, é um indicador crítico da capacidade do organismo de transportar oxigênio adequadamente. Níveis reduzidos podem sugerir hipóxia e demandam atenção clínica imediata. Embora a técnica de iPPG já venha sendo utilizada para estimar esse parâmetro, os resultados ainda são moderadamente precisos, sendo necessário o aprimoramento de técnicas para maior confiabilidade diagnóstica [Lewandowska and Nowak 2012].

7.8.2. Monitoramento em hospitais

O monitoramento de sinais vitais sem contato apresenta grande potencial em ambientes hospitalares, principalmente por eliminar a necessidade de sensores tradicionais, como os utilizados em eletrocardiogramas e na fotopletismografia convencional, que exigem fixação direta à pele do paciente. Essa característica é especialmente vantajosa para indivíduos com pele sensível ou comprometida, como pacientes internados em Unidades de Terapia Intensiva (UTI), vítimas de queimaduras ou recém-nascidos em Unidades de

 ©2025 SBC – Soc. Bras. de Computação

Terapia Intensiva Neonatal (UTIN) [Wang et al. 2023, Zeng et al. 2024].

Além do uso em ambientes clínicos tradicionais, a técnica de iPPG pode ser aplicada em triagens hospitalares, permitindo o monitoramento simultâneo de múltiplos pacientes, e também em contextos cirúrgicos, onde pode ser empregada para avaliar a perfusão sanguínea em tempo real.

7.8.3. Monitoramento de treino físico

O monitoramento da frequência cardíaca é fundamental em exercícios físicos, especialmente para indivíduos que buscam controlar sua frequência cardíaca dentro da zona alvo para treinamento cardiovascular ou queima de gordura. Até o momento, os métodos tradicionais de monitoramento de frequência cardíaca baseados em contato, como sensores de ECG em faixas torácicas e sensores de PPG em pulseiras de pulso, têm sido amplamente utilizados em aplicações de treino físico para consumidores. No entanto, essas medições baseadas em contato são geralmente desconfortáveis, inconvenientes, e difíceis de visualizar durante o treino, o que limita sua eficácia informativa.

Como alternativa, um sistema de monitoramento baseado em câmeras, incorporando a técnica de iPPG e um monitor, pode ser integrado diretamente em equipamentos de academia, como esteiras ou bicicletas ergométricas [Wang et al. 2017b]. Esse sistema permite a medição, visualização e análise da frequência cardíaca do usuário de forma contínua e automática, sem a necessidade de qualquer dispositivo corporal. Esse método sem contato oferece uma solução mais eficaz para o acompanhamento da frequência cardíaca ao longo do exercício, otimizando o treino de forma mais prática e conveniente. Além disso, o sistema pode avaliar a recuperação da frequência cardíaca imediatamente após o exercício, fornecendo informações sobre a saúde cardiovascular do indivíduo, como a presença de arritmias, ou até mesmo prever sua mortalidade.

7.8.4. Monitoramento de saúde em casa

O método de iPPG pode ser aplicado no ambiente doméstico para acompanhar e avaliar a saúde de uma pessoa (mesmo que não seja paciente) no cotidiano, com o intuito de promover melhorias em seu estilo de vida, como no caso de atividades físicas para o coração, hábitos alimentares, controle de estresse mental e qualidade do sono. Um exemplo disso é o monitoramento da frequência cardíaca: é possível integrar uma câmera RGB comum a um espelho, criando um sistema inteligente de monitoramento de saúde. Ao se posicionar em frente ao espelho, o indivíduo tem sua frequência cardíaca medida automaticamente, com o valor sendo exibido no espelho.

Além disso, esse sistema pode ser instalado acima da cama para monitorar continuamente a frequência cardíaca (e sua variabilidade) durante o sono [Vogels et al. 2018]. Com base nessas medições, é possível analisar os diferentes estágios do sono (como sono leve, sono profundo e movimento rápido dos olhos (REM)), ajudando a melhorar a qualidade do sono, por exemplo, ao configurar um alarme inteligente para o momento ideal de despertar. A tecnologia também pode ser integrada a sistemas de monitoramento em vídeo em residências, usados para monitorar idosos ou bebês, com funcionalidades como avaliação de marcha, detecção de quedas, monitoramento da saída da cama ou identificação de outras emergências.

 ©2025 SBC – Soc. Bras. de Computação

7.8.5. Computação afetiva

Os métodos de iPPG têm se destacado na computação afetiva, sobretudo pela capacidade de extrair sinais fisiológicos de forma não invasiva a partir de vídeos. Estudos iniciais demonstraram que a variação da frequência cardíaca extraída medida pela técnica de iPPG pode ser usada para estimar níveis de estresse com acurácia de até 85% [Meziati Sabour et al. 2021].

Desde então, diversas pesquisas vêm expandindo sua aplicação, incluindo a criação de bases de dados como a UBFC-Phys, voltada à análise de estresse e emoções [Meziati Sabour et al. 2021]. O reconhecimento de emoções, por sua vez, tem se beneficiado do uso de iPPG, inicialmente em propostas para detectar microexpressões faciais. Estudos mais recentes também exploram o uso de grafos e redes neurais para combinar informações fisiológicas e visuais na detecção de estados emocionais [Liu et al. 2024]. Além disso, há investigações em andamento sobre o uso de iPPG para reconhecimento de dor. Espera-se que, em um futuro próximo, essas abordagens avancem em aplicações como interfaces humano-máquina e avaliações fisiológicas automatizadas [Li et al. 2014].

7.8.6. Vigilância por vídeo

Um grande número de câmeras foi instalado em cidades, como em estações centrais, ruas e bares, com o objetivo de monitorar pedestres para fins de segurança. Essas câmeras são geralmente conectadas a sistemas inteligentes que utilizam algoritmos de visão computacional para analisar o fluxo de pessoas, entender comportamentos dos pedestres, detectar ações agressivas ou violentas, e até prever outras emergências. O princípio é utilizar padrões físicos (como características espaciais ou de movimento temporal) dos pedestres para reconhecer comportamentos ou eventos perigosos.

De forma semelhante, a técnica de iPPG pode ser integrada a esses sistemas de câmeras para medir e analisar os padrões fisiológicos dos pedestres à distância, facilitando a identificação e a classificação de comportamentos. Como a vigilância por vídeo geralmente ocorre em ambientes externos, vários desafios adicionais precisam ser considerados, como as condições de iluminação, a resolução facial e os movimentos corporais, em comparação com as aplicações internas abordadas anteriormente neste capítulo.

7.8.7. Entretenimento

A técnica de iPPG oferece um grande potencial para inovar sistemas de entretenimento, como televisores inteligentes, realidade virtual, realidade aumentada e dispositivos de jogos cinéticos. Ao monitorar variáveis fisiológicas do usuário, como a frequência cardíaca e a variabilidade da frequência cardíaca, a técnica de iPPG pode adaptar a experiência de interação de maneira mais personalizada e responsiva.

Em contextos como jogos ou filmes, por exemplo, a tecnologia de iPPG pode analisar as respostas emocionais do usuário e ajustar a intensidade, o ritmo ou o conteúdo da experiência com base nas reações detectadas. Isso resulta em uma interação mais imersiva, ajustando automaticamente os estímulos para maximizar o prazer e o engajamento. Além disso, o monitoramento das emoções pode ser integrado a plataformas de mídia para compreender melhor as preferências e reações dos usuários, ajudando a prever gostos e aprimorar a personalização do conteúdo. Com isso, cria-se uma experiência

 ©2025 SBC – Soc. Bras. de Computação

mais dinâmica e alinhada ao estado emocional do usuário, transformando a forma como interagimos com a tecnologia no universo do entretenimento.

7.8.8. Monitoramento de motoristas

A prevenção de acidentes de trânsito causados por fatores comportamentais e cognitivos do motorista tem motivado o desenvolvimento de tecnologias baseadas em visão computacional. Sistemas modernos embarcados em veículos utilizam câmeras para monitorar a postura da cabeça, piscadas de olhos, direção do olhar, movimentação da boca e expressões faciais, emitindo alertas diante de sinais de fadiga, sonolência, distração, estresse ou ansiedade [Smart Eye 2023].

Mais recentemente, técnicas de iPPG têm ganhado destaque como solução não invasiva para o monitoramento contínuo de sinais vitais, como frequência cardíaca e variabilidade da frequência cardíaca, indicadores fortemente relacionados ao estado emocional e cognitivo do condutor [Ahmed et al. 2025]. Algoritmos de aprendizado profundo têm mostrado resultados animadores na extração de sinais de iPPG em cenários reais com iluminação variável, movimento do veículo e mudanças na expressão facial do condutor. Esses sistemas também têm sido propostos para a detecção de direção sob influência de álcool, associando dados de expressões faciais e iPPG a modelos de redes neurais com bons resultados [Keshtkaran 2025]. Além do uso automotivo, o potencial da técnica de iPPG se estende ao monitoramento de pilotos em aeronaves ou operadores de máquinas pesadas, favorecendo intervenções preventivas em contextos críticos.

Apesar dos avanços, desafios persistem quanto à generalização dos modelos para diferentes perfis demográficos e à robustez em ambientes altamente dinâmicos. Por isso, o desenvolvimento de sistemas de monitoramento de motoristas baseados em iPPG exige não apenas inovações em algoritmos, mas também a construção de bases de dados diversificadas e a integração com sistemas de assistência à condução em tempo real.

7.8.9. Detecção de *Deepfake*

Deepfake, uma combinação das palavras *deep learning* (aprendizado profundo) e *fake* (falso), refere-se a algoritmos de aprendizado profundo utilizados para simular e criar conteúdos de áudio e vídeo [Rana et al. 2022]. Atualmente, *Deepfake* tornou-se um campo altamente popular, sendo amplamente aplicado em técnicas de inteligência artificial para troca de rostos, síntese de voz, geração de rostos e vídeos [Rana et al. 2022].

O surgimento dessa tecnologia tornou possível manipular e gerar conteúdos de vídeo e áudio incrivelmente realistas, com uma difícil detecção, o que torna desafiador para os consumidores desses conteúdos sintetizados discernirem o que é verdadeiro e o que é falso. Em função disso, pesquisadores têm se dedicado ao desenvolvimento de métodos para distinguir esses conteúdos. Estudos demonstraram que a medição da frequência cardíaca a partir de vídeos faciais pode ser uma ferramenta eficaz para determinar e diferenciar se um vídeo é real ou falso. Como resultado, os métodos de iPPG começaram a ser empregados na detecção de *Deepfakes*, com o surgimento de abordagens mais específicas para essa aplicação, alcançando resultados promissores e demonstrando o grande potencial dos métodos iPPG nesta área. Destaca-se que a pesquisa em reconhecimento de *Deepfake* por meio de métodos iPPG continua a ser uma das áreas mais promissoras da

 ©2025 SBC – Soc. Bras. de Computação

pesquisa científica [Xiao et al. 2024].

7.8.10. *Anti-spoofing* Facial

Com o avanço contínuo da tecnologia ao longo dos anos, os métodos de segurança e criptografia têm se desenvolvido de maneira significativa. Um dos métodos de segurança mais comuns atualmente são as chaves biométricas, como impressões digitais e características faciais. Celulares, aplicativos bancários e de mensagens já adotam a biometria como uma chave de acesso. No entanto, como mencionado anteriormente, tecnologias como o *deepfake* de vídeo utilizam inteligência artificial para realizar a troca de rostos, o que torna a biometria facial vulnerável a ataques *spoofing*. Além do *deepfake*, agentes maliciosos podem obter fotos ou vídeos contendo o rosto de um indivíduo alvo e usá-los para contornar esse tipo de segurança, acessando assim os dados da vítima.

Em razão disso, o interesse por parte dos pesquisadores em desenvolver técnicas para combater esse tipo de ataque tem crescido consideravelmente. Com o rápido avanço dos métodos de iPPG, estudiosos têm identificado o potencial dessas técnicas para aprimorar os sistemas de reconhecimento facial, oferecendo uma solução mais robusta contra ataques *spoofing* [Xiao et al. 2024, Wang et al. 2024].

7.9. Limitações, desafios e estudos futuros

As aplicações dos métodos e abordagens de fotopletismografia por imagem (iPPG) são amplas, mas, conforme discutido ao longo do trabalho, ainda existem desafios significativos em sua implementação. Esses desafios estão presentes tanto nas abordagens tradicionais quanto nas que utilizam aprendizado profundo, indicando a necessidade contínua de estudos e melhorias.

Um dos principais obstáculos enfrentados está relacionado à sensibilidade ao movimento durante a captura do vídeo. Essa limitação afeta a estabilidade do sinal extraído, pois movimentos do indivíduo podem comprometer o processo de detecção facial e, consequentemente, a área de interesse utilizada na extração do sinal. Quando a detecção facial é feita apenas no primeiro quadro, por exemplo, qualquer movimentação fora da área inicial pode resultar em quadros inutilizáveis. Mesmo quando a detecção ocorre a cada quadro, o movimento contínuo pode demandar maior processamento e causar diferenças entre o tempo de gravação e o número de quadros válidos coletados.

Outro fator importante é a iluminação durante a etapa de gravação. A presença de sombras causadas por movimentações, especialmente quando há segmentação da cor da pele, pode interferir na extração precisa do sinal. Além disso, a qualidade da luz ambiente influencia diretamente no desempenho do método, sendo observada a necessidade de uma iluminação estável e de intensidade adequada. Em experimentos, foi verificado que luzes muito intensas podem prejudicar a obtenção do sinal de iPPG, o que reforça a importância de um controle cuidadoso das condições de iluminação. Além disso, as características ópticas da pele influenciam na eficácia da extração, assim, diferenças na refletância cutânea podem dificultar a detecção das variações de cor associadas ao pulso [Wang and Shan 2020].

Em resumo, apesar do grande potencial da técnica de iPPG como ferramenta não

 ©2025 SBC – Soc. Bras. de Computação

invasiva de monitoramento fisiológico, seu uso efetivo ainda depende da superação de limitações relacionadas ao movimento, à iluminação e à adaptação às características visuais dos indivíduos. Esses pontos representam áreas importantes para pesquisas futuras, com foco na melhoria da robustez, da precisão e da aplicabilidade em diferentes cenários.

Referências

[Ahmed et al. 2025] Ahmed, S. G., Verbert, K., Zaki, N., Khalil, A., Aljassmi, H., and Alnajjar, F. (2025). Ai innovations in rppg systems for driver monitoring: Comprehensive systematic review and future prospects. *IEEE Access*, 13:22893–22913.

[Allen 2007] Allen, J. (2007). Photoplethysmography and its application in clinical physiological measurement. *Physiological measurement*, 28(3):R1.

[Bobbia et al. 2019] Bobbia, S., Macwan, R., Benezeth, Y., Mansouri, A., and Dubois, J. (2019). Unsupervised skin tissue segmentation for remote photoplethysmography. *Pattern Recognition Letters*, 124:82–90. Award Winning Papers from the 23rd International Conference on Pattern Recognition (ICPR).

[Bradski and Kaehler 2008] Bradski, G. and Kaehler, A. (2008). *Learning OpenCV: Computer vision with the OpenCV library*. "O'Reilly Media, Inc.".

[Chen and McDuff 2018] Chen, W. and McDuff, D. (2018). Deepphys: Video-based physiological measurement using convolutional attention networks. pages 356–373.

[De Haan and Jeanne 2013] De Haan, G. and Jeanne, V. (2013). Robust pulse rate from chrominance-based rppg. *IEEE transactions on biomedical engineering*, 60(10):2878–2886.

[Giavarina 2015] Giavarina, D. (2015). Understanding bland altman analysis. *Biochemia medica*, 25(2):141–151.

[Gudi et al. 2020] Gudi, A., Bittner, M., and van Gemert, J. (2020). Real-time webcam heart-rate and variability estimation with clean ground truth for evaluation. *CoRR*, abs/2012.15846.

[Han et al. 2015] Han, B., Ivanov, K., Wang, L., and Yan, Y. (2015). Exploration of the optimal skin-camera distance for facial photoplethysmographic imaging measurement using cameras of different types. In *Proceedings of the 5th EAI International Conference on Wireless Mobile Communication and Healthcare*, pages 186–189.

[Heusch et al. 2017] Heusch, G., Anjos, A., and Marcel, S. (2017). A reproducible study on remote heart rate measurement. *CoRR*, abs/1709.00962.

[Hülsbusch 2008] Hülsbusch, M. (2008). *An Image-Based Functional Method for Opto-Electronic Detection of Skin-Perfusion*. Phd thesis, RWTH Aachen University.

[Keshtkaran 2025] Keshtkaran, E. (2025). *Automated Methods for Estimating Blood Alcohol Concentration Level from Facial Cues*. Ph.d. thesis, Edith Cowan University.

 ©2025 SBC – Soc. Bras. de Computação

[Koelstra et al. 2011] Koelstra, S., Muhl, C., Soleymani, M., Lee, J.-S., Yazdani, A., Ebrahimi, T., Pun, T., Nijholt, A., and Patras, I. (2011). Deap: A database for emotion analysis; using physiological signals. *IEEE transactions on affective computing*, 3(1):18–31.

[Kyriacou and Allen 2021] Kyriacou, P. A. and Allen, J. (2021). *Photoplethysmography: technology, signal analysis and applications*. Academic Press.

[Lampier et al. 2023] Lampier, L. C., Floriano, A., Valadão, C. T., Silva, L. A., Caldeira, E. M. D. O., and Bastos-Filho, T. F. (2023). A deep learning approach for estimating spo 2 using a smartphone camera. *IEEE Transactions on Instrumentation and Measurement*, 72:1–8.

[Lampier et al. 2022] Lampier, L. C., Valadão, C. T., Silva, L. A., Delisle-Rodríguez, D., Caldeira, E. M. d. O., and Bastos-Filho, T. F. (2022). A deep learning approach to estimate pulse rate by remote photoplethysmography. *Physiological Measurement*, 43(7):075012.

[Lewandowska and Nowak 2012] Lewandowska, M. and Nowak, J. (2012). Measuring pulse rate with a webcam. *Journal of Medical Imaging and Health Informatics*, 2(1):87–92.

[Li et al. 2014] Li, X., Chen, J., Zhao, G., and Pietikainen, M. (2014). Remote heart rate measurement from face videos under realistic situations. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4264–4271. IEEE.

[Lin et al. 2019] Lin, J., Gan, C., and Han, S. (2019). Tsm: Temporal shift module for efficient video understanding. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 7083–7093.

[Liu et al. 2024] Liu, I., Liu, F., Zhong, Q., Ma, F., and Ni, S. (2024). Your blush gives you away: detecting hidden mental states with remote photoplethysmography and thermal imaging. *PeerJ Computer Science*, 10:e1912.

[Lugaresi et al. 2019] Lugaresi, C., Tang, J., Nash, H., McClanahan, C., Uboweja, E., Hays, M., Zhang, F., Chang, C.-L., Yong, M., Lee, J., et al. (2019). Mediapipe: A framework for perceiving and processing reality. In *Third workshop on computer vision for AR/VR at IEEE computer vision and pattern recognition (CVPR)*, volume 2019.

[McDuff et al. 2022] McDuff, D., Wander, M., Liu, X., Hill, B. L., Hernandez, J., Lester, J., and Baltrusaitis, T. (2022). Scamps: Synthetics for camera measurement of physiological signals.

[Meziati Sabour et al. 2021] Meziati Sabour, R., Benezeth, Y., De Oliveira, P., Chappé, J., and Yang, F. (2021). Ubfc-phys: A multimodal database for psychophysiological studies of social stress. *IEEE Transactions on Affective Computing*.

 ©2025 SBC – Soc. Bras. de Computação

[Meziatisabour et al. 2021] Meziatisabour, R., Benezeth, Y., Oliveira, P., Chappé, J., and Yang, F. (2021). Ubfc-phys: A multimodal database for psychophysiological studies of social stress. *IEEE Transactions on Affective Computing*, PP:1–1.

[Niu et al. 2018] Niu, X., Han, H., Shan, S., and Chen, X. (2018). VIPL-HR: A multi-modal database for pulse estimation from less-constrained face video. *CoRR*, abs/1810.04927.

[Pirzada et al. 2024] Pirzada, P., Wilde, A., and Harris-Birtill, D. (2024). Remote photoplethysmography for heart rate and blood oxygenation measurement: A review. *IEEE Sensors Journal*, 24(15):23436–23453.

[Poh et al. 2010] Poh, M.-Z., McDuff, D. J., and Picard, R. W. (2010). Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. *Opt. Express*, 18(10):10762–10774.

[Rana et al. 2022] Rana, M. S., Nobi, M. N., Murali, B., and Sung, A. H. (2022). Deepfake detection: A systematic literature review. *IEEE access*, 10:25494–25513.

[Revanur et al. 2021] Revanur, A., Li, Z., Ciftci, U. A., Yin, L., and Jeni, L. A. (2021). The first vision for vitals (V4V) challenge for non-contact video-based physiological estimation. *CoRR*, abs/2109.10471.

[Ronneberger et al. 2015] Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, pages 234–241. Springer.

[Saxen and Al-Hamadi 2014] Saxen, F. and Al-Hamadi, A. (2014). Color-based skin segmentation: an evaluation of the state of the art. In *2014 IEEE International Conference on Image Processing (ICIP)*, pages 4467–4471. IEEE.

[Smart Eye 2023] Smart Eye (2023). Driver monitoring system (dms). https://www.smarteye.se/solutions/automotive/driver-monitoring-system/.

[Soleymani et al. 2011] Soleymani, M., Lichtenauer, J., Pun, T., and Pantic, M. (2011). A multimodal database for affect recognition and implicit tagging. *IEEE transactions on affective computing*, 3(1):42–55.

[Spetlik et al. 2018] Spetlik, R., Cech, J., Franc, V., and Matas, J. (2018). Visual heart rate estimation with convolutional neural network.

[Stricker et al. 2014] Stricker, R., Müller, S., and Gross, H.-M. (2014). Non-contact video-based pulse rate measurement on a mobile service robot. *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication*, 2014:1056–1062.

[Sun and Thakor 2016] Sun, Y. and Thakor, N. (2016). Photoplethysmography revisited: From contact to noncontact, from point to imaging. *IEEE Transactions on Biomedical Engineering*, 63(3):463–477.

 ©2025 SBC – Soc. Bras. de Computação

[Tang et al. 2023] Tang, J., Chen, K., Wang, Y., Shi, Y., Patel, S., McDuff, D., and Liu, X. (2023). Mmpd: Multi-domain mobile video physiology dataset.

[Verkruysse et al. 2008] Verkruysse, W., Svaasand, L. O., and Nelson, J. S. (2008). Remote plethysmographic imaging using ambient light. In *Optics Express*, volume 16, pages 21434–21445. Optical Society of America.

[Vogels et al. 2018] Vogels, T., Van Gastel, M., Wang, W., and De Haan, G. (2018). Fully-automatic camera-based pulse-oximetry during sleep. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 1349–1357.

[Wang et al. 2023] Wang, H., Huang, J., Wang, G., Lu, H., and Wang, W. (2023). Contactless patient care using hospital iot: Cctv-camera-based physiological monitoring in icu. *IEEE Internet of Things Journal*, 11(4):5781–5797.

[Wang et al. 2024] Wang, J., Shan, C., Liu, L., and Hou, Z. (2024). Camera-based physiological measurement: Recent advances and future prospects. *Neurocomputing*, 575.

[Wang et al. 2017a] Wang, W., den Brinker, A. C., Stuijk, S., and de Haan, G. (2017a). Algorithmic principles of remote-ppg. In *IEEE Transactions on Biomedical Engineering*, volume 64, pages 1479–1491. Institute of Electrical and Electronics Engineers (IEEE).

[Wang et al. 2017b] Wang, W., den Brinker, A. C., Stuijk, S., and de Haan, G. (2017b). Robust heart rate from fitness videos. *Physiological measurement*, 38(6):1023.

[Wang and Shan 2020] Wang, W. and Shan, C. (2020). Impact of makeup on remote-ppg monitoring. *Biomedical Physics & Engineering Express*, 6(3):035004.

[Wu et al. 2013] Wu, Y., Lim, J., and Yang, M.-H. (2013). Online object tracking: A benchmark. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2411–2418.

[Xiao et al. 2024] Xiao, H., Liu, T., Sun, Y., Li, Y., Zhao, S., and Avolio, A. (2024). Remote photoplethysmography for heart rate measurement: A review. *Biomedical Signal Processing and Control*, 88.

[Yu et al. 2019] Yu, Z., Li, X., and Zhao, G. (2019). Recovering remote photoplethysmograph signal from facial videos using spatio-temporal convolutional networks. *CoRR*, abs/1905.02419.

[Zeng et al. 2024] Zeng, Y., Yu, D., Song, X., Wang, Q., Pan, L., Lu, H., and Wang, W. (2024). Camera-based cardiorespiratory monitoring of preterm infants in nicu. *IEEE Transactions on Instrumentation and Measurement*.

[Zeng et al. 2025] Zeng, Y., Zhu, Y., Song, X., Wang, Q., Yang, J., and Wang, W. (2025). Camera-based neonatal blood pressure estimation from multisite and multiwavelength pulse transit time-a proof of concept in nicu. *IEEE Internet of Things Journal*.

 ©2025 SBC – Soc. Bras. de Computação

[Zhan et al. 2020] Zhan, Q., Wang, W., and De Haan, G. (2020). Analysis of cnn-based remote-ppg to understand limitations and sensitivities. *Biomedical optics express*, 11(3):1268–1283.

[Zhang et al. 2016] Zhang, Z., Girard, J., Wu, Y., Zhang, X., Liu, P., Ciftci, U., Canavan, S., Reale, M., Horowitz, A., Yang, H., Cohn, J., Ji, Q., and Yin, L. (2016). Multimodal spontaneous emotion corpus for human behavior analysis. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016-December.

 ©2025 SBC – Soc. Bras. de Computação